Novelty Detection in Dermatological Images

Gabriela Mariel Maletti

Technical University of Denmark
Department of Informatics and Mathematical Modelling
Ph.D. Thesis No. 116
Kgs. Lyngby 2003
Denmark

July 5, 2003

This dissertation is submitted to Informatics and Mathematical Modeling at the Technical University of Denmark in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The work has been supervised by Associate Professor Bjarne K. Ersbøll and Professor Knut Conradsen.

It is dedicated to my sons Sergio Carlos and Julio Guillermo.

Kgs. Lyngby, June 24, 2003

Gabriela Mariel Maletti

Abstract

The problem of novelty detection is considered for a set of dermatological image data. Different points of view are analyzed in detail.

First, novelty detection is treated as a contextual classification problem. Different learning phases can be detected when the sample size is increased. The detection of the emergence a new class is considered here. A model that estimates the minimal amount of information required to recognize structure in the data as a function of class separability is also proposed.

Secondly, texture alteration detection is considered a novelty detection problem. The possibility of avoiding pattern registration by transforming the data to a space invariant of registration is explored through a canonical analysis tool. Problems of expressing the data in a way they can be compared are also considered here, for instance, pattern registration. An approach for patterns alignment using an algorithm that iteratively approximates the minimal value of an assumed smooth function is proposed. Registration outputs are evaluated and analyzed with the Multi-variate Alteration Detection Transform.

Resumé

Problemet med detektion af ændringer (novelty detection) i sæt af dermatologiske billeddata betragtes i denne afhandling. Problemet anskues og analyseres fra forskellige løsningsmæssige vinkler.

Først bliver novelty detection behandlet som et kontekstuelt klassifikationsproblem. Ved gradvist at øge samplestørrelsen detekteres forskellige faser af indlæringen. Her lægges der især vægt på detektion af inklusionen af en ny klasse. En model, som estimerer den mindste nødvendige mængde information for at opdage struktur i data som funktion af seperabilitet af klasserne, bliver også foreslået.

Dernæst betragtes detektion af ændringen i tekstur som et novelty detection problem. Muligheden for at undgå geometrisk registrering af billederne undersøges ved først at transformere data så det bliver uafhængigt af registrering. Dernæst analyseres de transformerede data ved hjælp af en variant af kanonisk korrelationsanalyse. Problemer relateret til at udtrykke data på en måde, så de kan analyseres samlet, for eksempel som mønster (co-)registrering, betragtes også her. En metode til registrering af mønstre, der baserer sig på en algoritme, som iterativt approksimerer den minimale værdi af en antaget glat funktion foreslås. De registrerede billeder bliver evalueret og analyseret ved hjælp af en transformation kaldet: Multivariate Alteration Detection.

Acknowledgments

I would like to thank my supervisors Associate Professor Bjarne Ersbøll and Professor Knut Conradsen for their support and advice. I'm grateful to Prof. Conradsen, for opening the doors to this Institution for me. Particulary, thank you to Bjarne, for his moral support and encouragement during the whole period.

I would like to thank to the Department of Informatics and Mathematical Modelling for providing the research facilities for doing my work, specially for their understanding when I attended conferences with my small children.

I am also grateful to Dr. Jorge Lira for his advice during the time, that finally was my "external research", at the Laboratory of Remote Sensing at the Geophysics Institute of the National Autonomous University of Mexico. I also would like to mention and thank Dr. Boris Escalante for his comprehension, when I moved to Denmark.

I would like to thank to each one of the members of the Section for Image Analysis and Computer Graphics for their pleasant company, particularly to Helle Welling, Klaus Hilger, Michael Edberg, Henrik Aanæs, David Delgado, Mikkel Stegmann, Karsten Hartelius and Nicolas Guilbert. I would like to thank Dr. Allan Nielsen and Dr. Rasmus Larsen for their comments. It would not have been possible to do my work without the kind help of Mogens Dyrdal, Finn Kuno Christensen, Ulla Nørhave, Mette Larsen, Niels Rasmussen, Paul Erik Thamdrup and Eina Boeck.

Thanks also to The National Hospital of Denmark, to Prof. David Owens and to Torsana Diabetes Diagnostics for providing digital images of malignant melanoma and fundus, to the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions of psoriasis lesions.

To the SITE Project funded by a grant from the Danish Technical Research

Foundation (Project Number *STVF* 56-00-0123), for supporting the present work. I would like to thank the National Autonomous University of Mexico, for supporting my first year of Ph.D. studies economically. I am also grateful to the Otto Mønsted Foundation for supporting me financially during my studies.

I very much appreciate the support of my friends (in alphabetical order) Martha Alegre, Dubhe Chavira, Els Defoor, Ericka García, Andrea Porchia, Ana Rodriguez, Alicia Thomsen and Silvia Venegas during all the period. I also would like to thank to Morten Lange, for this comprehension during the time we have been at his place. I specially would like to mention and thank the kindness of "Mormormor" Grethe Larsen to all the members of my family.

I would like to thank to my teachers Alicia Raatz, Armando Castro, Verónica Stockmayer, Luisa Herrmann, Ursula Waidelich, Raúl Mereles and Nora Meza, and to Catalina and Waltraud Plocher. Finally, thanks to my Opa Karl, my parents Esther and Carlos, to my sister Graciela and my brother Roberto for their support. I would like to thank to my husband Sergio, for making the concept of maximization of the redundancy understandable to me. I wish to express my gratitude to my life priority, my sons Sergio Carlos and Julio Guillermo, for their love and energy.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

There are two main aspects that are considered in the present thesis. The first one, is novelty detection as a theoretical aspect; the second one, regards the data set used to test different approaches proposed. Novelty detection is focused on in different ways, each one presented in a chapter. In each case, the data set consisting of dermatological images is used to illustrate the schemes proposed. Basically, the present chapter is a short description of what this thesis is about.

1.1 Novelty Detection

A definition of novelty detection follows.

Definition 1.1 Novelty detection is to find dissimilarities in a set of patterns.

Following [25], a pattern is an array of entities of the same nature; a class of patterns is a set of patterns with similar characteristics. In image analysis, novelty detection implies to determine the change in the pixel features quantitatively. A pixel feature is the value of a pixel attribute, like texture, local spectral mean, etc.. Two patterns are considered similar if they satisfy the same homogeneity criteria.

There are different problems we could think of as a novelty detection problem. For instance, the case for which the patterns within a set are ordered according to some criteria. In Figure 1.1 we have a synthetical example of a set of pixel values corresponding to an image row. Starting from column 1, we could say that the first 20

pixels satisfy the same homogeneity criteria, because they are distributed randomly around the same mean value. At column 21, the pixel values change significantly: they are, in fact, distributed around a different mean. This change could be considered the first novelty. The pixels are again similar until the the appearance of a second novelty, which occurs at column 61.

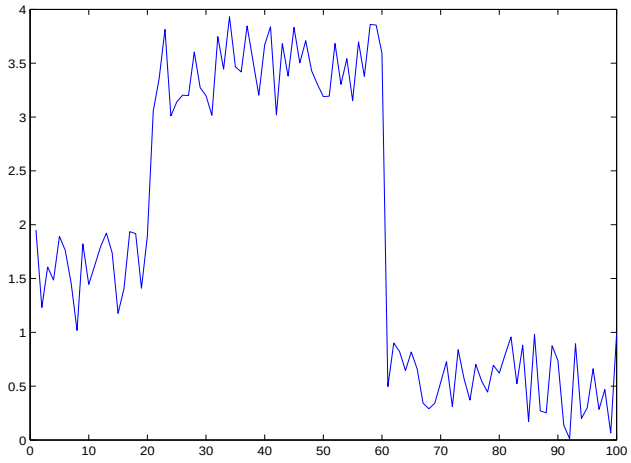


Figure 1.1: Pixel values of a synthetical image row.

Examples like this are used in Chapter 2, in which novelty detection is considered a contextual classification problem. Given a set of ordered patterns, we want to find when the first dissimilarity between those patterns occurs. However, how do we know what is similar and what is not similar? How do we define the homogeneity criteria of a class of patterns? When did we learn enough about a given class of patterns to be able to correctly say that a new pattern satisfies an homogeneity criteria or not? How does the separability between the densities of two patterns influence the detection of their dissimilarities? Review of work done about the matter and a new approach is presented. The problem is formally defined as follows. For a given prototype pixel -assumed to be surrounded by pixels belonging to the same class-, we want to find the maximal circular set centered on it for which it is true that all the pixels belong to the same class as the center one. In other words, we want to detect the emergence of a new class.

In Chapter 3, the problem is first to find the component showing maximal change in texture descriptors of pairs of segmented patterns. As an attempt to avoid the pattern registration task, a statistical approach that detects the maximal alterations between them is applied to their textural descriptors. The hard work of ordering

a set of patterns in a way so that they are comparable is also presented. To find similarities or dissimilarities between pairs of patterns first implies to represent them in a way so they can be compared. There are approaches to solve the registration problem given a set of landmarks. What if this set is not known? What if we only can make assumptions of the region where the landmarks are located? A combined alignment and registration scheme is proposed. Afterwards, change detection in sequences of registered lesion patterns is carried out.

1.2 The Dermatological Images

For the present work, the dermatological images used are *RGB* captures of malignant melanomas and psoriasis lesions. The images of malignant melanomas are single images, this means, one image for each lesion. Some of them are shown in Figure 1.2. Note the presence of diffuse borders and the difficulty to visually determine the number of classes present within the melanoma: their distributions seem to have a high overlap.

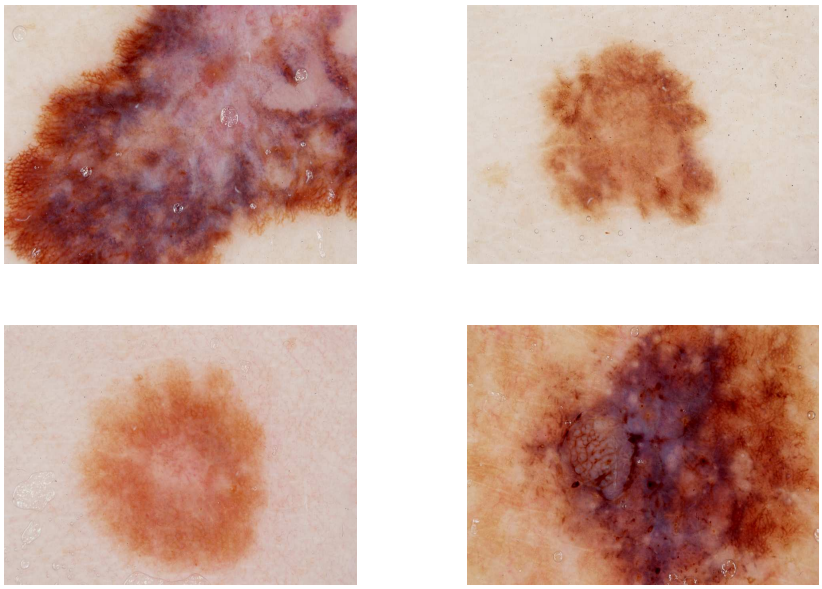


Figure 1.2: Original images of malignant melanomas.

With regard to the psoriasis dataset, we use a set of 175 images of lesions of

this type taken at the Gentofte Hospital, Denmark, during pilot sessions with three invited patients. Three lesions per patient were followed in a sequence of five RGB images of 556 pixels by 748 pixels taken each week for at least three weeks. For each one of those lesions, one image is shown in Figure 1.2. The whole dataset is included in Appendix A. It has to be mentioned that the dataset used plays a special role in the present work, because it is very particular and it has not been published in beforehand. Therefore, some characteristics are described and analyzed in detail.

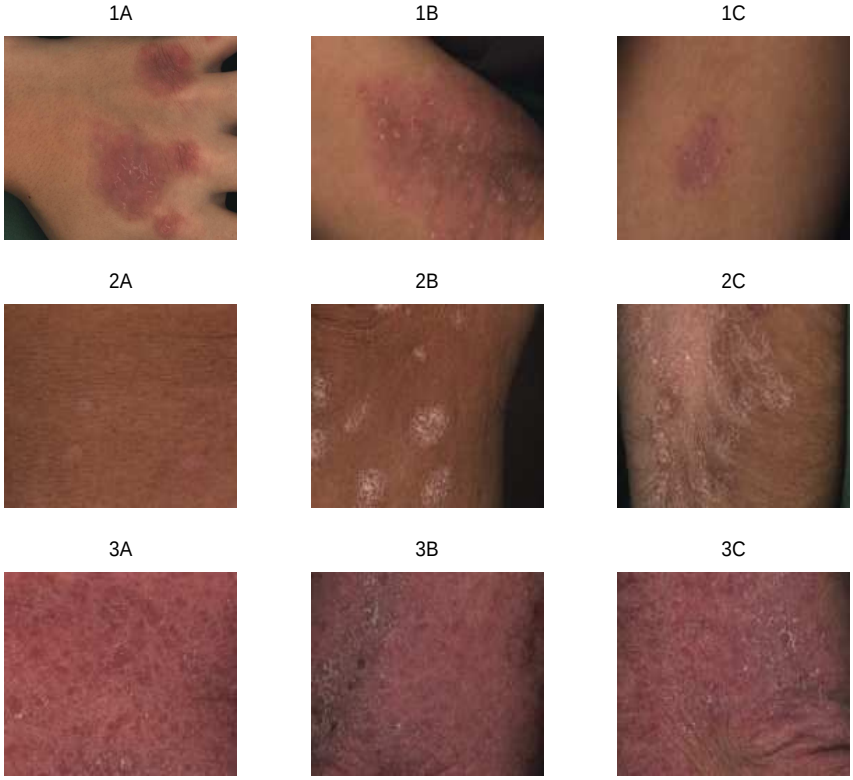


Figure 1.3: Set of psoriasis lesions images. Patient number 1, 2, 3. Lesion A, B, C.

Some comments about this dataset follow. Different problems to be solved in beforehand appear by looking at the images obtained during the pilot sessions: there are incomplete lesions in some images, a whole session for a given lesion was not taken; in some cases, the variation if the background clearly introduces errors in the

estimation of the statistics of the populations; a lesion is partially covered by hair; in some cases, a green curtain used as the background appears, in others, it does not appear; etc.. In the following, one problem at time will be treated. Aspects that could be solved during a future phase at the image analysis level without loosing generality, and situations that should be avoided by re-considering image acquisition standards will be evaluated. Avoiding the user-setting of parameters during system-operation is also an aspect to be considered.

By visual inspection of the pilot dataset, it can be deduced that the size between lesions has a significant variance: in some images, one clear lesion composed by only one connected region can be observed, while in other cases, a lesion could be considered a set of smaller sized but connected regions; finally, there are also examples, where the number of connected regions is large while the size of the regions is small, so that the whole could be taken as only one lesion. Dermatologists have noted that the psoriasis lesions do not significantly change their shape and size along the time: the change happens within the lesions instead. The variance of the shape between different lesions is not considered here.

1.3 Thesis Layout

The present work is composed of two parts. The first part is a summary of papers and technical reports included in the second part. The second part is divided in published papers and technical reports.

CHAPTER 2

Novelty Detection in the Context

2.1 Introduction

In general, a learning process can be treated as a statistical dynamics problem. It passes through different phases[56]: instability, fluctuations, critical slowing down and reorganization. First, "it is impossible to detect structure in the data until a critical number of examples have been presented" [58]. Then, underlying structure in the data can be recognized [3]; afterwards, "additional information provided is redundant" [3]. Finally, the reorganization phase starts when a state of new order emerges [23].

Definition 2.1 Novelty detection in a learning process implies the detection of a change of its phases.

In the present chapter, novelty detection in the learning process is treated as a contextual classification problem. In image analysis, contextual classification schemes place windows on each pixel to be classified, compute descriptors with the values of the pixels belonging to each window and use them to decide whether or not to assign a pixel to a given class. In the learning process using windows of increasing size, we can identify three kinds of optimal windows: the first is related to the minimal amount of information needed to recognize underlying structure in the data; the second is related to the sample size from which no information is added if more data is added to the sample; and the third optimal window size is related to the maximal set for which it is true that all the elements belong to the same class as a given one:

increasing the sample size implies including elements of other classes. Thus, the third window implies the detection of the emergence of a new class.

To detect the third optimal window size, when circular windows of increasing size are used, is the main topic of this chapter. Thus, the optimal window will here be the maximal circular window for which it is true that all the pixels belong to the same class as the center one. Buhot et al. [4] suggested that the emergence of a state of higher order can be detected by the minimization of an optimal learning curve that depends on the fraction of examples. The determination of that optimal learning curve is here based on the uncertainty principle, more specifically, on the role of the covariance as a threshold.

Finally, some classification schemes are proposed. A model that relates separability between classes with the minimal amount of information required to detect structure in the data is specified. The schemes have been applied to dermatological images.

2.2 Context

Contextual classification schemes have shown to have a better performance than punctual classification schemes. A lot of work has been published on the matter. However, in spite of the term "contextual" being widely used in the image analysis area, Sowa [50] noted that there is no widely accepted definition of context. He said, that the term "context" has a basic and derived meaning, given in Definitions 2.2 and 2.3 respectively [50].

Definition 2.2 Context is a section of linguistic text or discourse that surrounds some word or phrase of interest.

Definition 2.3 Context is a non-linguistic situation, environment, domain, setting, background, or milieu that includes some entity, subject, or topic of interest.

The context has the following formal functions:

- Syntax: to group, delimit, or package a section of text.
- Semantics: to describe or refer to some real or hypothetical situation.
- Pragmatics: the reason or purpose for distinguishing the section of linguistic text or non-linguistic situation.

Based on Sowa's work, the following definition of the term window is given:

Definition 2.4 A window defines a subset of an entity with spatially connected instances.

It is a subset because it is a group of instances included in a larger set (syntactic function). An instance is an element of this group, that can be a pixel of an image or a data of a population. The instances are neighbors in a given space (semantic function). The dimensions of this space can be location, time, etc. Something -a theory- needs to be described or concluded regarding this subset: there is a purpose behind it (pragmatic function). A window has properties like size, shape, etc. From here, we are focusing on the size of windows of circular shape. As it was pointed by Hodgson [19], the optimal window size for classification is still an open question. Let us go further with this topic.

2.2.1 Window Size

In order to find an ideal window size for image classification, Hodgson [19] suggested the use of a conceptual model of the human perceptual and cognitive interpretation processes, and -among others-, to take a look at the Feature-Integration Theory of Attention [54] and the Guided Search Model of Human Visual Search Behavior [59], in which the relation between location and identification was discussed. A short review about location and identification in visual attention follows.

2.2.2 Location and Identification in Visual Attention

The aim of including this part, is to take knowledge about the state of the art in psychology about the theories of visual attention. Location and identification are variables that systematically are mentioned in papers related to this topic. Perhaps with an erroneous basis, perhaps with a different point of view, the definition of the relation between these variables is tracked starting with Feature-Integration Theory of Attention proposed in the '80 by Treisman and Gelade [54] and continuing with posterior works, some of them, formulated in mathematical terms.

In the Feature-Integration Theory of Attention, it is assumed that features come first in perception. The features are registered early, automatically, and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention. They claim that attention is necessary for the correct perception of conjunctions. In the absence of focused attention and of effective constraints on top-down processing, conjunctions of features could be formed on a random basis. These unattended couplings will give rise to "illusory conjunctions" [55],[10]. In their work, they explored the relation between identification and spatial localization (for details, see Experiments VIII and IX), and they concluded that:

- In the early and parallel process of registration separable features can be detected and identified across a display. Locating any individual feature requires an additional operation.

- Conjunctions can not be identified without also being spatially localized.

This could be summarized in the following Table 2.1:

Table 2.1: Features and its Conjunctions

Feature-Integration Theory of Attention	Identification and Location	Search
Pre-attentive Stage	Separable Features can be identified without being located	Parallel
Attentive Stage	Identification and Location highly correlated	Serial

Logan [29] mentioned that the initial assumption of independence between the pre-attentive stage and the attentive stage in the Feature Integration Theory of Attention was revised in Treisman and Sato (1990), and an inhibitory interaction between attentive and preattentive processes -which is more effective when targets differed more from distractors- was proposed.

Wolfe [59] argued that the output of the Pre-Attentive Stage, actually guides the attention to the conjunctions. In his "Guided Search Model of Human Visual Search behavior" he considers two types of parallel feature search components: bottom-up and top-down. In the first, the subject does not know the identity of the target, which changes from trial to trial but is always a unique element among homogeneous distractors. In the second, the identity of the target is known (i.e. searching for a red item among distractors of a variety of different colors). Bottom-up activation of an item is based on the differences between the item and its nearest neighbors in each of several different radial directions. Top-down activation is based on the match between an item and the known properties of the target. The calculation of the maps (bottom-up and top-down) produced for each basic feature are relatively independent of each other. The combination of information from these maps can be used to guide the attention to conjunctions of features.

Cave [5] pointed out, that the Feature-Integration Theory of Attention and the Guided Search Model both assume that all selection is done by a single attentional window, so they both have trouble explaining that in some situations visual attention takes the form of inhibition of distractor locations, and that two locations can be selected simultaneously. In fact, he said that, in these models, the selection is applied to objects more than to locations.

Horowitz et. al. [20], in their work about memory-driven search, clearly separated the theories of visual search in two types: serial models and parallel theories. The first ones propose that attention can process the identity of only one item at a time. The second ones assume that identity is computed in parallel for each item, and that an items identity becomes gradually more certain over the course of a trial. There

is also a research line that combines both, suggesting that the conjunction of some features can be conducted in parallel and the same conjunction consisting of other features requires some serial search [53] [16].

Theories in visual attention were sub-divided in space-based attention and object-based attention [29]. In the first ones, it is assumed that the attention selects regions of space independent of the objects they contain, the case of the same Feature-Integration Theory of Attention. The second ones argue that attention selects objects rather than regions of space.

Logan proposed C.T.V.A. (CODE Theory of Visual Attention) [29], which was formulated in mathematical terms. He associated theories to the components of the early and later visual process. The early system -in which location and identity information are combined- is represented by the CO.DE. theory (COntour DEtector theory of perceptual grouping by proximity). The late system that processes identity information is represented by Bundesen's (1990) Theory of Visual Attention and, the late system that processes location information is represented by it's Spatial Relation Theory. In the CO.DE. theory, the location of each item in space is represented by its own -roughly symmetrical and peaked in the center- distribution. Bottom-up processes sum the distributions for the different items producing a CODE surface and make it available to top-down processes. Perceptual groups are produced by applying a threshold to the CODE surface. Items residing in the same above-threshold region of the CODE surface belong to the same perceptual group. Hierarchical grouping is produced by varying the threshold. As an extension of CO.DE. to attention Logan assumes that the location is distributed in the sense that information about the features of the items is distributed over space.

Cave proposed in 1999 the FeatureGate model of Visual Attention, which is built around a hierarchy of spatial maps within a neural network framework: a selection mechanism controls what information -selected according to location- travels from each level to the next; bottom-up and top-down effects interact to identify selection targets. Cave's FeatureGate Model has structural similarities with C.T.V.A..

2.3 Previous Work in Supervised Contextual Classification

Basically, this Section refers to previous work done in supervised contextual classification, using region growing algorithms. In [27], [31], the homogeneity criteria used to generate a statistically valid sample of each class in multi-spectral images was an estimation of its joint normalized histogram. This scheme expected the user to provide one single prototype pixel for each class present in the image in order to generate initial training sets. Afterwards, a region growing algorithm increased

the classes sample sizes, and finally, the whole image was classified. Then, the main problem, was that some parameters were empirically set. An attempt to improve this scheme was the work published in [28], which basically presented a spectral model of separability of Rayleigh distributed classes and included border detection to improve the classification of border pixels. The mentioned model was used to estimate an optimal window size for classification. For the image shown in Figure 4(a) in Section 4.1.5 ([28]), the region growing algorithm of the schemes [27] and [28] both produced the same samples. However, by visual assessment, the thematic maps produced by the classification stage in Figure 2.1 improves when the window size for classification is estimated as a function of class separability (thematic map on the right).

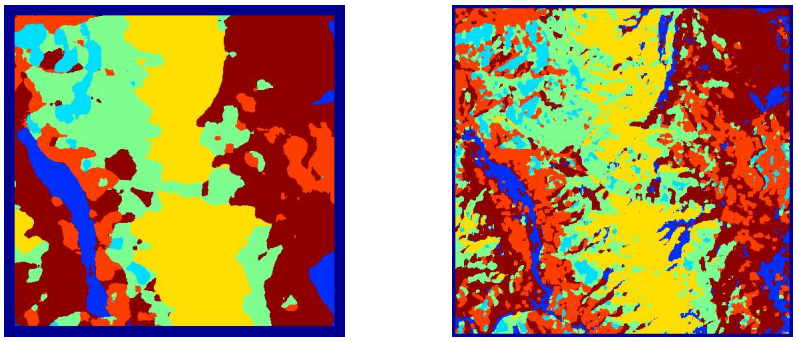


Figure 2.1: Output of the classification stage in [27] (left) and in [28] (right).

However, some aspects were still left for further considerations.

- After seeding the prototype pixel for a given class, the normalized histogram of increasing window sizes centered on the prototype pixel was computed. If the difference of areas of the normalized histograms of two consecutive windows was smaller than a user-provided threshold, then the initial training set was assumed to be found (see Equation 5 in Section 4.1.5). The first aspect was that the user-provided threshold was empirically set. Secondly, the use of this kind of differential was obviously highly dependant on the location and size of the neighborhood containing pixels belonging to the same class on which the prototype was placed.

- To estimate, the thresholds to be used by the region growing algorithm for a given class (see Equation 8 in Section 4.1.5), again, a parameter without any physical explanation was set.

On the other hand, it has been observed that the use of a model of spectral separability based to estimate an optimal window size for classification improved the classification results. In that case, synthetical and real examples were composed of Rayleigh-like distributed classes and square windows were used. Basically, the following was concluded:

- Whenever the classes are completely separated, the maximum precision is obtained using a per-pixel classification.
- Classes with the same separability measure have more or less the same window size for classification.
- An inverse relationship exists between the separability of the classes and the best window size.

From now, the following problems will be considered:

- For a given prototype pixel assumed to be surrounded by pixels belonging to the same class as itself, we want to estimate the maximal neighborhood centered on the prototype pixel for which it is true that all the elements belong to the same class as the center pixel.
- Given the probability density functions of pairs of classes, we want to estimate the optimal window size for classification as a function of the separability between those classes.

2.4 An Initial Training Set Generation Scheme

This Section considers the work [41]. It starts with some definitions, and then an initial training set generation scheme is proposed.

Let a specific image with non-correlated bands $X = \{x[r,c]\}$ be defined over the given lattice:

$$L = \{[r,c] | 1 \leq r \leq r_{max}, 1 \leq c \leq c_{max}\} \tag{2.1}$$

We assume that X is the realization of a two-dimensional random field X. Notice that while X (sans serif) is a set of variables, X (italic) is a set of values of those variables. X is hierarchically defined in terms of the realization $Z = \{z[r,c]\}$ of an underlying random field Z such that: Z represents the partition of the domain L in K regions of different types. Each $z[r,c]$ is a value of the set of labels $B = \{b_1, b_2, ..., b_K\}$ where $z[r,c] = b_k$ indicates that the coordinate $[r,c]$ of a given pixel belongs to the region k.

Let W_o be given by the shortest Euclidean distance from each pixel to the border:

$$W_o = \{w_o[r, c]\} \tag{2.2}$$

The distance function implicitly defines the radii of discs of maximal size $\mathsf{D}_{w_o}[r, c]$ included in X for which is true that all the elements of each given realization belong to the same region k. Then, the optimal window for a given pixel x defines the realization D_{w_o} of the maximal circular set centered on the position $[r, c]$ of x. This is an heuristic approach. An example is shown in Figure 2.2.

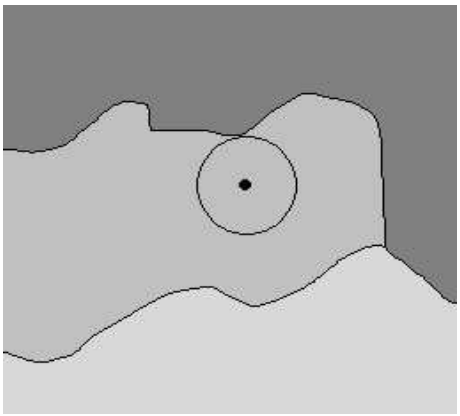


Figure 2.2: The set indicated as a circle is the maximal disc that can be centered on the prototype pixel (the marked point) for which it is true that all the pixels belong to the same class as the center pixel.

We associate a set of weights Ω_w with values between zero and one to a disc D_w, such that $\omega_w[i]$ is the weight corresponding to the $i-th$ element $\mathsf{d}_w[i]$ of the disc D_w:

$$\omega_w[i] \in (0, 1) \subset \Re \tag{2.3}$$

for all $1 \leq i \leq n_w$ such that the sum of the weights is one:

$$\sum_{i=1}^{n_w} \omega_m[i] = 1 \tag{2.4}$$

From now, each time a set of weights named Ω is associated to a disc it is assumed that Equations 2.3 and 2.4 are satisfied.

Each realization D_w of a disc D_w is a group of n_w elements. Therefore, a group is a set of numbers.

We define the estimate y_w of the mean μ_w of a realization D_w of a disc D_w centered on a given position $[r, c]$ as the scalar product of the realization D_w with a set of weights Ω_w:

$$y_w[r,c] =< D_w[r,c], \Omega_w >= \sum_i d_w[i]\omega_w[i] \tag{2.5}$$

where $d_w[i]$ is the $i-th$ element of the realization D_w of the disc corresponding to the $i-th$ position, $\omega_w[i]$ is the weight of the $i-th$ element of D_w corresponding to the same $i-th$ position. The estimated mean $y_w[r,c]$ within each realization $D_w[r,c]$ is an element of a new synthetical image $Y_w = X * \Omega_w$, where $*$ is the convolution operator.

The estimated variance σ_w^2 of a realization D_w of a disc D_w centered on a given position $[r, c]$ is defined as:

$$\sigma_w^2[r,c] =< D_w^2[r,c], \Omega_w > -y_w^2[r,c] \tag{2.6}$$

where each $d_w^2[i] \in D_w^2$ is the square of the element $d_w[i]$ of D_w that is at the same position i.

Let the image X be convolved by the set of weights Ω_v, where the index v indicates a disc of radius v. The output $Y_v = X * \Omega_v$ contains the estimated means within groups computed with the set of weights Ω_v. Let the synthetical image Y_v be convolved by the set of weights Ω_m, where the index m indicates a disc of radius m, such that:

$$Y_{v,m} = Y_v * \Omega_m = X * \Omega_v * \Omega_m \tag{2.7}$$

where the subindexes in $Y_{v,m}$ indicate the original image X is first convolved with a circular set of weights of radius v and then convolved with a circular set of weights of radius m. Each single element in $Y_{v,m}$ is then denoted $y_{v,m}$. For all realizations of discs of radius m in $Y_{v,m}$, $\exists D_t \subset X \wedge \Omega_t$ such that:

$$y_{v,m} \simeq y_{t(v,m)} =< D_t, \Omega_t > \tag{2.8}$$

where

$$t = t(v,m) = v + m \tag{2.9}$$

and Ω_t is approximately equal to the convolution of Ω_v with Ω_m:

$$\Omega_{t(v,m)} \simeq \Omega_v * \Omega_m. \tag{2.10}$$

Equality in Equation 2.10 holds for squares. This is due to the fact that the domain of the convolution of two functions defined over square domains is also square. For the case of circles, some small differences can occur because the image is defined

over a (discrete) lattice (See Equation 2.1). In Figure 2.3, these differences can be seen for the case of a disc radius $t = 8$. According to Equation 2.9, the pairs of discs radius (v, m) that sum to the value 8 are $(1, 7)$, $(2, 6)$, $(3, 5)$, $(4, 4)$, $(5, 3)$, $(6, 2)$ and $(7, 1)$. Let the domain of the convolution of a disc of radius v with a disc of radius m be given by the values different from zero. As it can be seen in the bottom part of Figure 2.3, for a given pair of values (v, m), this domain has small differences compared to the disc of radius t.

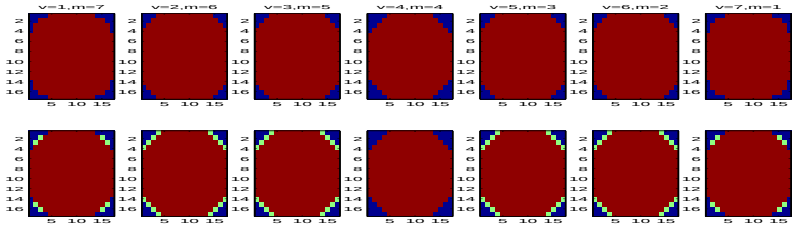


Figure 2.3: On top, the domain of the convolution of discs of radius v with discs of radius m. On bottom, the difference between the a disc of radius t with the domain shown on top.

2.4.1 Redundancy

We want to obtain an optimal initial training set starting from a seeded prototype pixel (See Figure 2.2). In this section, we assume that it is located on the position $[k, l]$ and that its neighbors belong to the same class as it. The neighborhood is delineated by the realization of a disc D_t of (constant) radius t, centered on the position $[k, l]$. For constant disc radius t, we define circular groups of data of the same size. For a given t, the number of groups n_m and the number of elements n_v within a group are variables. We would like to find, for a constant t, the optimal pair of number of groups and number of elements within a group. This is considered here.

The terms location and identification considered in Section 2.2.2 refer us to the terms position and moment used in Heisenberg's Uncertainty Principle, from which we know that the uncertainties σ in the position p and moment q of a particle have a product greater than Planck's constant. This constant can be interpreted, from a statistical point of view, as of the order of the absolute value of the covariance between the position p and the moment q [21]:

$$\sqrt{\sigma q^2 \sigma p^2} \geq |cov(q, p)| \tag{2.11}$$

The threshold established by the absolute value of this covariance is due to the fact that the variance-covariance matrix is positive definite [9]. Obviously, if we find a way to minimize the right side of Equation 2.11, we can think of minimizing the right side too. For instance, we know the minimizer for the classical uncertainty principle inequality is the Gaussian [2]. However, in this case, for a given pixel our knowledge about its neighbors is minimal, in other words, our uncertainty is maximal. This means that, initially, we should give its neighbors the same weight: we should use a uniform model. The connection between a uniform model and a Gauss model can be found in the Central Limit Theorem [46], which can be expressed as a property of convolutions: the convolution of a large number of positive functions is approximately a Gauss function.

The specification of the p and the q for our problem follows after a motivating example.

Motivation

In the present example, we follow with the same notation as before, but the case considered is for one single dimension. Let the vector of data $D_t = \{d_t[1], \ldots, d_t[n_t]\}$ be defined over a lattice $L = \{c | 1 \leq c \leq c_{max} = n_t\}$. The number of elements is the constant $n_t = 7$. We define the following combinations (n_m, n_v) of number of groups n_m and number of elements n_v within a group: $(6, 2)$, $(5, 3)$, $(4, 4)$, $(3, 5)$ and $(2, 6)$ (See Figure 2.4). For each single group, we assign the set of weights $\Omega_v = \{\omega_v[i] | \omega_v[i] = \frac{1}{n_v}, \forall 1 \leq i \leq n_v\}$. We assign to the set of groups the weights

$\Omega_m = \{\omega_m[j] | \omega_m[j] = \frac{1}{n_m}, \forall 1 \leq j \leq n_m\}$.

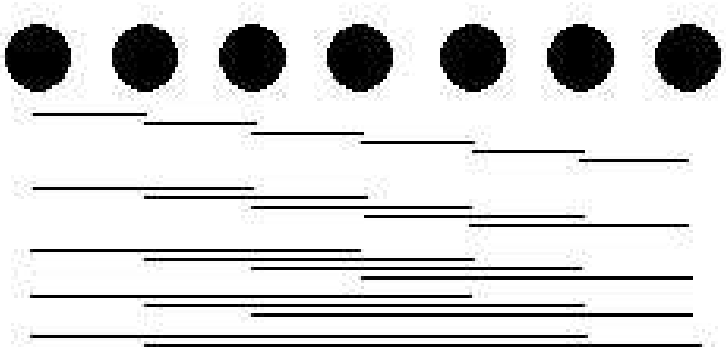


Figure 2.4: Set D_t of 7 data for which the combinations (*number of groups*, *group size*): $(6,2)$, $(5,3)$, $(4,4)$, $(3,5)$ and $(2,6)$ are considered.

The estimated means $y_{v,m}$ between groups (Equation 2.8) can also be obtained by summing the elements of a vector $G_{t,m}$: $y_{v,m} = \sum_{i=1}^{n_m} g_m[i]$. The vector $G_{t,m} = D_t U_{n_t,n_m}$ is the product of the vector of data D_t as a row with a kind of convolution matrix $U_{n_t \times n_m}$. Each column in U represents a weighted group average. For our example, the matrices U are expressed here for each one of the combinations of group sizes and number of groups:

$$U_{7,6} = \begin{bmatrix} \frac{1}{6\times2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6\times2} & \frac{1}{6\times2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6\times2} & \frac{1}{6\times2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6\times2} & \frac{1}{6\times2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6\times2} & \frac{1}{6\times2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{6\times2} & \frac{1}{6\times2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{6\times2} \end{bmatrix}_{7\times6} \tag{2.12}$$

$$U_{7,5} = \begin{bmatrix} \frac{1}{5\times3} & 0 & 0 & 0 & 0 \\ \frac{1}{5\times3} & \frac{1}{5\times3} & 0 & 0 & 0 \\ \frac{1}{5\times3} & \frac{1}{5\times3} & \frac{1}{5\times3} & 0 & 0 \\ 0 & \frac{1}{5\times3} & \frac{1}{5\times3} & \frac{1}{5\times3} & 0 \\ 0 & 0 & \frac{1}{5\times3} & \frac{1}{5\times3} & \frac{1}{5\times3} \\ 0 & 0 & 0 & \frac{1}{5\times3} & \frac{1}{5\times3} \\ 0 & 0 & 0 & 0 & \frac{1}{5\times3} \end{bmatrix}_{7\times5} \tag{2.13}$$

$$U_{7,4} = \begin{bmatrix} \frac{1}{4\times4} & 0 & 0 & 0 \\ \frac{1}{4\times4} & \frac{1}{4\times4} & 0 & 0 \\ \frac{1}{4\times4} & \frac{1}{4\times4} & \frac{1}{4\times4} & 0 \\ \frac{1}{4\times4} & \frac{1}{4\times4} & \frac{1}{4\times4} & \frac{1}{4\times4} \\ 0 & \frac{1}{4\times4} & \frac{1}{4\times4} & \frac{1}{4\times4} \\ 0 & 0 & \frac{1}{4\times4} & \frac{1}{4\times4} \\ 0 & 0 & 0 & \frac{1}{4\times4} \end{bmatrix}_{7\times4} \tag{2.14}$$

$$U_{7,3} = \begin{bmatrix} \frac{1}{3\times5} & 0 & 0 \\ \frac{1}{3\times5} & \frac{1}{3\times5} & 0 \\ \frac{1}{3\times5} & \frac{1}{3\times5} & \frac{1}{3\times5} \\ \frac{1}{3\times5} & \frac{1}{3\times5} & \frac{1}{3\times5} \\ \frac{1}{3\times5} & \frac{1}{3\times5} & \frac{1}{3\times5} \\ 0 & \frac{1}{3\times5} & \frac{1}{3\times5} \\ 0 & 0 & \frac{1}{3\times5} \end{bmatrix}_{7\times3} \tag{2.15}$$

$$U_{7,2} = \begin{bmatrix} \frac{1}{2\times6} & 0 \\ \frac{1}{2\times6} & \frac{1}{2\times6} \\ \frac{1}{2\times6} & \frac{1}{2\times6} \\ \frac{1}{2\times6} & \frac{1}{2\times6} \\ \frac{1}{2\times6} & \frac{1}{2\times6} \\ \frac{1}{2\times6} & \frac{1}{2\times6} \\ 0 & \frac{1}{2\times6} \end{bmatrix}_{7\times2} \tag{2.16}$$

From Equations 2.12 to 2.16, we can observe that we are dealing with matrices U_{n_t,n_m} of different sizes, because the number of groups n_m is variable. Furthermore, a comparison between the input D_t and the output $G_{t,m}$ is difficult, because they do not have the same size. See top part of Figure 2.5.

However, the value $y_{v,m}$ can also be obtained by summing the elements of another vector $G'_{t,m} = D_t U'_{n_t,n_t}$; this means, $y_{v,m} = \sum_{i=1}^{n_t} g'_m[i]$. The matrix U'_{n_t,n_m} is a diagonal matrix such that $u'[i,i] = \sum_{j=1}^{n_t} u[j,i]$, $\forall 1 \leq i \leq n_t$. One column in U' represents how large a weight each data point receives in the grand mean.

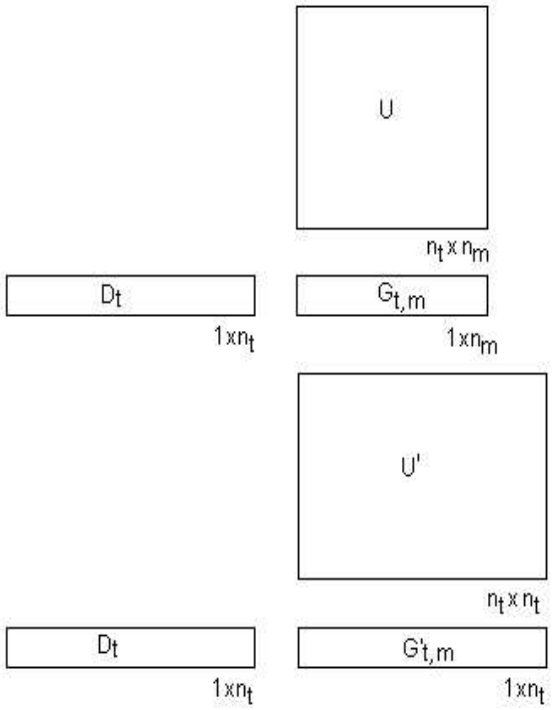


Figure 2.5: The equivalent weighting functions U and U'.

For our example we have:

$$U'_{6,2} = \begin{bmatrix} \frac{1}{6\times 2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{6\times 2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{6\times 2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{6\times 2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{6\times 2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{6\times 2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6\times 2} \end{bmatrix}_{7\times 7} \tag{2.17}$$

$$U'_{5,3} = \begin{bmatrix} \frac{1}{5\times3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{5\times3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{5\times3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{5\times3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{5\times3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{5\times3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{5\times3} \end{bmatrix}_{7\times7} \tag{2.18}$$

$$U'_{4,4} = \begin{bmatrix} \frac{1}{4\times4} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{4\times4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4\times4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{4\times4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4\times4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{4\times4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4\times4} \end{bmatrix}_{7\times7} \tag{2.19}$$

$$U'_{3,5} = \begin{bmatrix} \frac{1}{3\times5} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3\times5} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{3\times5} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{3\times5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{3\times5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{3\times5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3\times5} \end{bmatrix}_{7\times7} \tag{2.20}$$

$$U'_{2,6} = \begin{bmatrix} \frac{1}{2\times6} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{2\times6} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{2\times6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{2\times6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{2\times6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{2\times6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2\times6} \end{bmatrix}_{7\times7} \tag{2.21}$$

The values of the main diagonal of the matrices from Equations 2.17 to 2.21 can be seen in Figure 2.6.

From Equations 2.17 to 2.21 we now have matrices of the same size, which are comparable. As a consequence, the output vector $G'_{t,m}$ also has the same size as the input vector D_t. See bottom part of Figure 2.5.

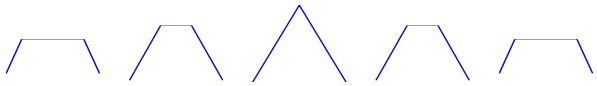


Figure 2.6: From left to right, the plot of the values of the main diagonal of the matrices $U'_{6,2}$, $U'_{5,3}$, $U'_{4,4}$, $U'_{3,5}$ and $U'_{2,6}$.

With this small example we have seen that, for a given data set, the estimated mean between groups depends on the weights assigned to the group elements and to the groups, and also depends on the groups sizes and the number of groups defined. These variables are implicit in a kind of convolution matrix U, which contains the contribution of each single group element to the estimated mean between groups. In an ideal situation we would like to compare the different matrices U, and choose the optimal for a given data set. We have seen that the size of these convolution matrices changes when the number of groups n_m changes; in consequence, they can not be compared. Finally, we have seen that the estimated mean between groups can also be expressed as a function of the contribution of each single element of the original data set. In this way, we could generate same sized matrices U', which are comparable. How the optimal matrix will be chosen for a given data set will be shown later on.

We leave this example for now. A generalization of the approach for discs follows. Basically we now have circular sets: the first with the original data and second, with the weighted data (See Figure 2.7).

Our p

The p in Equation 2.11 is given, for our problem, by a disc D_t of radius t centered on the seeded prototype pixel. A disc D_v is placed at all pixel locations within D_t such that D_v is completely within D_t. Here, we will show how the groups and the group sizes are defined for a given realization D_t. The estimated mean between groups is also expressed as a function of D_t.

Given a realization $D_t[k,l] \subset X$ of a disc of radius t centered on the location $[k,l]$ with n_t elements. We assume that the i-th element $d_t[i]$ is a value of the corresponding random variable $\mathsf{d}_t[i] \in \mathsf{D}_t[k,l]$ and that all the random variables $\mathsf{d}_t[i]$ have a distribution with mean μ.

Given a sliding disc of radius v and a set of weights Ω_v we want to specify each estimated mean y_v (see Equation 2.5) within groups $D_v \subset D_t[k,l]$ as a function of $D_t[k,l]$. Let the realization $D_v[r,c] \subset D_t[k,l]$ be the $j-th$ group (See Figure 2.8).

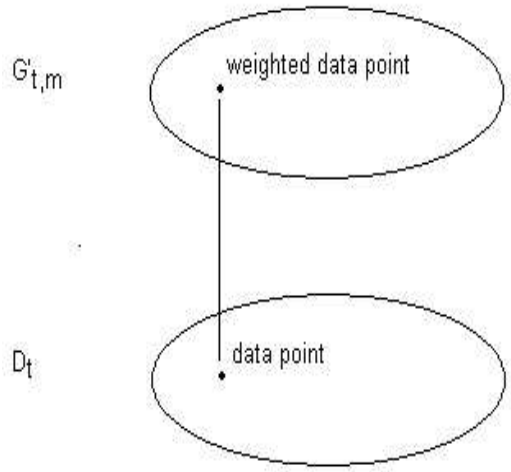


Figure 2.7: The circular sets D_t and $G'_{t,m}$.

For the $j-th$ group ($1 \leq j \leq n_m$) we generate a weighting function $A_j[r,c] = [a_j[1], \ldots, a_j[n_t]]^T$, which contains zeros for the elements of $D_t[k,l]$ not belonging to $D_v[r,c]$ and the corresponding weights Ω_v for the remaining n_v elements. Then

$$y_v[r,c] =< D_t[k,l], A_j[r,c] > \tag{2.22}$$

For simplicity we remove the location indexes $[k,l]$ and $[r,c]$ and rewrite Equation 2.22 as follows:

$$y_j =< D_t, A_j > \tag{2.23}$$

where the index j indicates the $j-th$ group in D_t.

From Equations 2.3 and 2.4 it can be deduced that, for all $1 \leq i \leq n_t$ and $1 \leq j \leq n_m$, the new weights A_j

$$a_j[i] \in [0,1) \subset \Re \tag{2.24}$$

are such that $\forall j$

$$\sum_{i=1}^{n_t} a_j[i] = 1. \tag{2.25}$$

We add the following constraints. Each single element $d_t[i]$ has to belong to at least

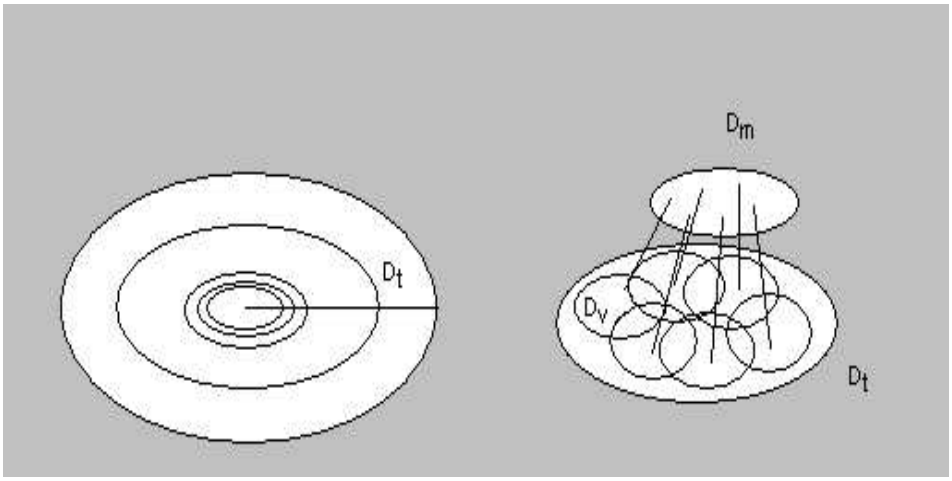


Figure 2.8: The mapping model. To the left, a set of circular discs (in perspective) of increasing radius t centered on a given pixel is shown. To the right, one of these discs is shown in detail: overlapping discs of a fixed radius v are placed into the disc of radius t. A new disc of radius m is constructed with the centers of these overlapping discs of radio v. The mapping model is completely defined when weights are assigned to the elements of each disc.

one group:

$$\sum_{j=1}^{n_m} a_j[i] > 0 \ \forall i. \tag{2.26}$$

Repetitions of groups are not allowed: for all $i \neq j$

$$\mathsf{A}_j \neq \mathsf{A}_i. \tag{2.27}$$

Note that, by Equations 2.24 and 2.3, the groups can not be sets of one single element and $n_m > 1$. By the constraint in Equation 2.26 and by Equation 2.3 we have that pairs of groups could have some overlap. However, by constraint 2.27 we are not allowed to have pairs of groups such that their union is equal to their intersection.

Our q

In this section we will define our q in Equation 2.11. It was, in fact, introduced in the motivating example. The q is given by the random vector G'_t. The connection between the vectors G_t and G'_m is first formulated.

Let the disc G_m be given by n_m functions of D_t:

$$\mathsf{G}_{t,m} = [\mathsf{g}_{t,m}[1](\mathsf{D}_t), \ldots, \mathsf{g}_{t,m}[n_m](\mathsf{D}_t)] \tag{2.28}$$

with m variable. Given a realization D_t, each single value $g_{t,m}[j]$ is obtained as follows:

$$g_{t,m}[j] =< D_t, \omega_m[j]A_j >= \omega_m[j] \sum_{i=1}^{n_t} a_j[i]d_t[i]. \tag{2.29}$$

Note that each single $g_{t,m}[j]$ is a weighted estimation of the mean μ of the random variables of D_t. Rewriting Equation 2.29 we have

$$\mathsf{G}_{t,m} = \mathsf{D}_t U_{t,m} \tag{2.30}$$

with the tensor $U_{t,m}$

$$U_{t,m} = [u_{t,m}[i,j]]_{i,j=1}^{n_t,n_m} = \left[\begin{array}{ccc} \omega_m[1]A_1 & \ldots & \omega_m[n_m]A_{n_m} \end{array} \right] \tag{2.31}$$

of size $n_t \times n_m$ [1] and $\mathbf{\Omega}_m = [\omega_m[1], \ldots, \omega_m[n_m]]^T$. Each single $u_{t,m}[i,j]$ is the weight assigned to the $i-th$ element of D_t is the $j-th$ group.

Doing things in this way, we have that the tensor $U_{t,m}$ (Equation 2.31) depends on an vector $\boldsymbol{s} = [s_1, \ldots, s_{n_m}]$ of group sizes [2] with n_m specifying the number of groups. Thus, $U_{t,m} = u(\boldsymbol{s}, t, m)$. The size of the $j-th$ group is s_j:

$$\sum_{i=1}^{n_t} \delta(a_j[i] > 0) = s_j. \tag{2.32}$$

where δ is a function that takes the value one if the expression $a_j[i] > 0$ is satisfied, and zero, in the contrary case.

Applying Equation 2.10, and like it has been shown in the previous example, the estimated mean $y_{v,m} \in \mathsf{Y}_{v,m}$ between n_m groups of n_v elements is then:

$$y_{v,m} = \sum_{j=1}^{n_m} g_{t,m}[j] \tag{2.33}$$

which, can be rewritten as a function of the elements $d_t[i]$ with some algebraic steps:

$$y_{v,m} = \sum_{i=1}^{n_t} \omega_{t(v,m)}[i]d_t[i] =< D_t, \Omega_t > \tag{2.34}$$

[1] Note that the index n_t is not number of rows in a matrix. It indicates the number of elements in each single $\omega_m[j]A_j$.

[2] For all $i \in [1, n_m] : s_i = n_v$.

with

$$\omega_{t(v,m)}[i] = \sum_{j=1}^{n_m} \omega_m[j] a_j[i] = \sum_{j=1}^{n_m} u_{t,m}[i,j] \tag{2.35}$$

Clearly, $\sum_{i=1}^{n_t} \omega_{t(v,m)}[i] = \sum_{j=1}^{n_m} \sum_{i=1}^{n_t} u_{t,m}[i,j] = 1$.

We create a new "diagonal" tensor $U'_{n_t \times n_t}$ to generate a disc $\mathsf{G}'_{t,m}$ of n_t elements:

$$\mathsf{G}'_{t,m} = \mathsf{D}_t U'_{t,m}. \tag{2.36}$$

The $i-th$ element $g'_{t,m}[i]$ in the realization $G'_{t,m}$ is computed as follows:

$$g'_{t,m}[i] =< D_t, U'_{t,m}[i] > \tag{2.37}$$

where $U'_{t,m}[i]$ is a circular set of weights of radius $t = v + m$. The circular set of weights $U'_{t,m}[i]$ has the value zero for all its elements, except in the $i-th$ location, which has the value:

$$u'_{t,m}[i][i] = \sum_{j=1}^{n_m} u_{t,m}[i,j] \tag{2.38}$$

$u_{t,m}[i,j]$ is the weight that the $i-th$ element of the disc D_t has in the $j-th$ group. Note that $u'_{t,m}[i][i]$ is an element of the "main diagonal" of $U'_{t,m}[i]$. Note also that outside of the main diagonal, the elements of $U'_{t,m}[i]$ are zeros.

Then, by Equation 2.35, the estimated mean between groups can also be obtained as follows:

$$y_{v,m} = \sum_{i=1}^{n_t} g'_{t,m}[i]. \tag{2.39}$$

Therefore, we have shown that, for a disc D_t of constant radius t centered on the prototype pixel, we can define a set of equal sized matrices U'. These matrices depend on the number of circular groups included in D_t and its sizes, and the weighting function set to the elements of each group and to the groups. Now, we are able to compare the input D_t with the output $\mathsf{G}'_{t,m}$.

Our $|cov(p,q)|$

The comparison between the input D_t with the output $\mathsf{G}'_{t,m}$ will be done through the following measure $|cov(\mathsf{D}_t, \mathsf{G}'_{t,m})|$, where the symbol cov indicates the covariance function. According to Equation 2.11, it is the threshold in the product of the uncertainties between D_t and $\mathsf{G}'_{t,m}$. In this section, we first find an expression for this

covariance, and then analyze how it can be minimized.

The $i-th$ element $g'_{t,m}[i]$ is a value of the corresponding random variable $\mathsf{g}'_{t,m}[i]$. It is desired that all the random variables $\mathsf{g}'_{t,m}[i]$ have a distribution with mean $\frac{\mu}{n_m}$. However, this will never be the case, because we do not allow groups of one single element. Furthermore, the weights Ω_t result from a convolution of two weighting functions. Taking into account the limitations of our gross approximation, by the Central Limit Theorem we know that the shape of the convolution of density functions tends to be Gaussian, and is certainly not uniform.

The mathematical expectation of a random variable $\mathsf{g}'_{t,m}[i]$ is:

$$E[\mathsf{g}'_{t,m}[i]] = \omega_t[i]E[\mathsf{d}_t[i]] = \omega_t[i]\mu \tag{2.40}$$

No matter which combination of number of groups and group sizes we have, we will always have a small difference between the mean of D_t and the estimated mean between groups [3]:

$$\mathsf{e}_{t,m} = |E[\sum_{i=1}^{n_t} \mathsf{g}'_{t,m}[i]] - \mu| > 0 \tag{2.41}$$

The variance-covariance matrix of the discs D_t and $\mathsf{G}'_{t,m}$ is:

$$\Sigma[\mathsf{D}_t\mathsf{G}'_{t,m}] = \begin{bmatrix} cov(\mathsf{D}_t, \mathsf{D}_t) & cov(\mathsf{D}_t, \mathsf{G}'_{t,m}) \\ cov(\mathsf{G}'_{t,m}, \mathsf{D}_t) & cov(\mathsf{G}'_{t,m}, \mathsf{G}'_{t,m}) \end{bmatrix} \tag{2.42}$$

where $cov(\mathsf{D}_t, \mathsf{D}_t)$ is the variance-covariance matrix of the disc D_t:

$$cov(\mathsf{D}_t, \mathsf{D}_t) = \begin{bmatrix} cov(\mathsf{d}_t[1], \mathsf{d}_t[1]) & \cdots & cov(\mathsf{d}_t[1], \mathsf{d}_t[i]) & \cdots & cov(\mathsf{d}_t[1], \mathsf{d}_t[n_t]) \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ cov(\mathsf{d}_t[n_t], \mathsf{d}_t[1]) & \cdots & cov(\mathsf{d}_t[n_t], \mathsf{d}_t[i]) & \cdots & cov(\mathsf{d}_t[n_t], \mathsf{d}_t[n_t]) \end{bmatrix} \tag{2.43}$$

The $cov(\mathsf{D}_t, \mathsf{G}'_{t,m})$ is the covariance between the disc D_t and the disc $\mathsf{G}'_{t,m}$ given by:

$$cov(\mathsf{D}_t, \mathsf{G}'_{t,m}) = cov(\mathsf{D}_t, \mathsf{D}_t)U'^T_{t,m} = var(\mathsf{D}_t)U'^T_{t,m} \tag{2.44}$$

This expression is obtained by substituting Equation 2.36 and applying properties of covariance [9].

We would like to minimize $|cov(\mathsf{D}_t, \mathsf{G}'_{t,m})|$. For a given realization D_t, the $|var(\mathsf{D}_t)|$ is constant. Changes in $|cov(\mathsf{D}_t, \mathsf{G}'_{t,m})|$ are introduced by the weights $U'_{t,m}$, which,

[3] except in very special situations, e.g., unless all the elements of D_t are equal

in this case, do not depend on the data D_t. The solution is deduced quite naturally from now. We have to minimize the generalized variance (the determinant) of $U'_{t,m}$, what is equivalent to minimizing the product of its eigenvalues. The eigenvalues are given in Equation 2.38. By Equation 2.35 they are equal to the elements of Ω_t. A useful metric that has the same uncertainty connotation as in certain formulations of statistical mechanics is Shannon's entropy or average amount of information per source output. There are certainly other relevant measures, eg. variance, however, it is interesting that entropy is inversely related to redundancy [43]. In general, the entropy of a set of weights Ω_t can be defined as:

$$E(\Omega_t) = -\sum_{i=1}^{n_t} \omega_t[i] \log(\omega_t[i]) \tag{2.45}$$

where n_t is the number of weights. Following Møller [43], the redundancy of a set of weights Ω_t can be defined as

$$R(\Omega_t) = 1 - \frac{E(\Omega_t)}{\log n_t} \tag{2.46}$$

where n_t has the same meaning as before. It is noted that equal disc weights imply maximal entropy and thus minimal redundancy.

According to Equation 2.46, the point of minimal entropy is the point of maximal redundancy and it corresponds to the combination of number of groups n_m and groups sizes n_v for which the overlap between the n_m groups is maximized. Therefore, for a given sample size n_t, we select the sets of weights Ω_v and Ω_m, number of groups n_m and group sizes n_v for which the redundancy $R(\Omega_{t(v,m)})$ is maximized. It is noted that in this direction the estimation $y_{v,m}$ of the mean μ has the largest error.

Results of the motivating example can be seen in Table 2.2. The values presented can be reproduced by looking at the diagonals from Equations 2.17 to 2.21. From Table 2.2 can be deduced, that the optimal combination of group sizes and number of groups is given by the pair $(4, 4)$.

Table 2.2: Product of the eigenvalues, entropy and redundancy of the matrices U' of the example shown in Section 2.4.1

(v,m)	$\prod$ Eigenvalues	Entropy	Redundancy
(6,2)	9.9306×10^{-7}	1.9073	0.0198
(5,3)	6.3210×10^{-7}	1.8640	0.0421
(4,4)	5.3644×10^{-7}	1.8407	0.0540
(3,5)	6.3210×10^{-7}	1.8640	0.0421
(2,6)	9.9306×10^{-7}	1.9073	0.0198

Last comments

We have seen, through the Uncertainty Principle that the threshold in the product of the uncertainties of two random variables is given by the absolute value of the product of their uncertainties. In the way things were done, by choosing the pair (v, m) for which $|cov(\mathsf{D}_t, \mathsf{G}'_{t,m})|$ is minimized, for a given t, we are able to decrease this threshold. We also have seen that, for t constant the $cov(\mathsf{D}_t, \mathsf{D}_t)$ is constant. Finally, we have to check that the $cov(\mathsf{G}'_{t,m}, \mathsf{G}'_{t,m})$ also decreases. By applying properties of covariance [9] we have that it equals:

$$U'_{t,m} cov(\mathsf{D}_t, \mathsf{D}_t) U'_{t,m} \tag{2.47}$$

We can compute the generalized variance of $cov(\mathsf{G}'_{t,m}, \mathsf{G}'_{t,m})$ as a product of determinants. We have also seen that the eigenvalues in $U'_{t,m}$ are values between zero and one. This means that their product, the determinant of $U'_{t,m}$ will always be smaller than one. Furthermore, the square of this determinant will be smaller than the determinant.

Until now, we have sustained that for a given realization D_t, the $|cov(\mathsf{D}_t, \mathsf{D}_t)|$ is a constant, but we did not compute its determinant. Equation 2.43 becomes:

$$\begin{bmatrix} (d_t[1]-\mu)(d_t[1]-\mu) & \cdots & (d_t[1]-\mu)(d_t[i]-\mu) & \cdots & (d_t[1]-\mu)(d_t[n_t]-\mu) \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ (d_t[n_t]-\mu)(d_t[1]-\mu) & \cdots & (d_t[n_t]-\mu)(d_t[i]-\mu) & \cdots & (d_t[n_t]-\mu)(d_t[n_t]-\mu) \end{bmatrix} \tag{2.48}$$

Applying properties of determinant, we can compute the determinant as a linear combination of the elements of the top row with the determinant of their co-factors. Each co-factor is the array left when the row and column of the given top row element is removed. This rule can be applied iteratively until the co-factor has a size 2×2. Then, for a given realization D_t, determinants we express the determinant of $cov(\mathsf{D}_t, \mathsf{D}_t)$ as a sum of terms containing determinants of size 2×2:

$$\begin{bmatrix} (d_t[i]-\mu)(d_t[j]-\mu) & (d_t[i]-\mu)(d_t[k]-\mu) \\ (d_t[r]-\mu)(d_t[j]-\mu) & (d_t[r]-\mu)(d_t[k]-\mu) \end{bmatrix} \tag{2.49}$$

which have the value zero. Thus, for a given realization D_t, the $|cov(\mathsf{D}_t, \mathsf{D}_t)|$ is singular.

In consequence, by maximizing the redundancy, we could say that we have the "most" singular system. We will see later on, that this is very convenient for detecting the emergence of a new class when windows of increasing size are used.

Examples of weighting functions

The following weighting functions Ω_t were tested in [41]:

- uniform(v)-uniform(m)
- uniform(v)-gauss(m)
- gauss(v)-uniform(m)
- gauss(v)-gauss(m)

where the first element of each pair refers to Ω_v and the second element to Ω_m in Equation 2.10, and

- uniform(h): $\Omega_f[i] = \frac{1}{n_f} \forall i$
- gauss(h): $\Omega_f[d] = exp(-0.5(\frac{d^2}{\sigma^2}))$ $\forall d^2 = (i-r)^2 + (j-c)^2$ and $\sigma = b(f)$. Let us fix $\sigma = g/3$.

The final weighting functions Ω_t using Equation 2.10 can be seen in Figure 7 in Section 4.1.1 for one example. The corresponding redundancy values are shown in Figure 2.9.

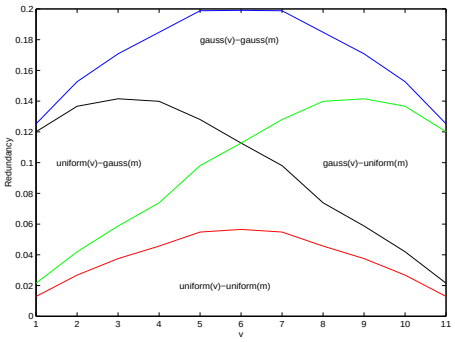


Figure 2.9: Redundancy of a disc $\mathsf{D}_{t=12}$ of fixed radius $t(v,m) = v + m = 12$ using different weighting functions Ω_t in the estimation of the mean of the realization of the disc.

As can be observed comparing top with bottom in Figure 2.10, for each constant radius t, the error in the estimation of the real mean is maximal when the redundancy is maximal. From empirical experiments, we conjecture that this happens no matter which combination of weighting functions considered is used.

$\forall t$ constant

$$\lim_{t(v,m)\to t(v_o,m_o)} |y_{t(v,m)} - \mu| = max(e_t) \tag{2.50}$$

For each fixed disc D_t, there is an optimal pair (v_o, m_o) such that the error in the estimation of the mean of the realization D_t is maximal.

Then for each fixed radius t we choose the pair (v_o, m_o) such that the redundancy is maximal:

$$(v_o, m_o) = max(R(\Omega_{t(v,m)})) \ \forall t(v,m) \ constant \tag{2.51}$$

The sets of pairs (v_o, m_o) for increasing radius t will define the line of maximal redundancy. It has to be mentioned that Møller [43] has noted that learning algorithms will work more efficiently in the minimization if the redundancy is increased.

For our weighting functions example, the position of the line of maximum redundancy of the uniform(v)-gauss(m) weighting function is symmetric to the position of the line of the gauss(v)-uniform(m) weighting function with respect to the plane $v = m$, and the values of maximum redundancy are the same for each t constant.

Again, for our weighting functions example, the position of the line of maximum redundancy of the uniform(v)-uniform(m) weighting function is equal to the position of the line of the gauss(v)-gauss(m) weighting function, but the values of redundancy for each t constant of the second are greater than the values of redundancy of the first.

2.4.2 Theory Revision

In this section we follow with the problem of obtaining an optimal initial training set starting from a seeded prototype pixel (See Figure 2.2) on location $[k, l]$. We have seen in previous section that, for each constant disc radius t, we can find and optimal number of circular groups n_m and an optimal number of elements n_v within the groups. Basically, from now, the disc radius t, which was considered constant in Section 2.4.1, now becomes a variable: we consider windows of increasing size.

Following Donoho [11], theory revision integrates inductive learning and background knowledge by combining training examples with a coarse, perhaps incomplete or incorrect, theory of the problem domain to produce a refined, more accurate theory.

In statistical dynamical problems [17], the theory revision can be carried out through a sub-sampling technique called "windowing". Following Fürnkranz [13], its goal is to reduce the complexity of a learning problem by identifying an appropriate subset of the original data, from which a theory of sufficient quality can be learned. From a statistical point of view, a window defines a sample or the training set for

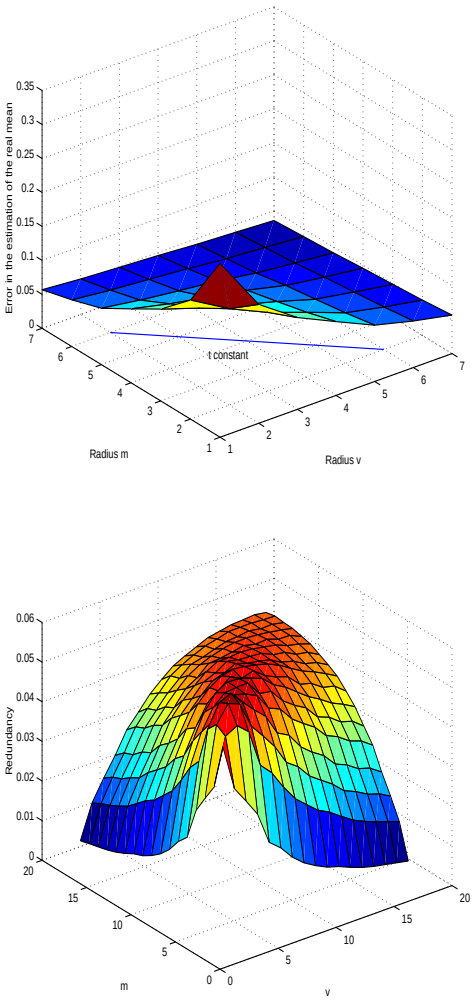


Figure 2.10: On top: Error in the estimation of the real mean using the uniform(v)-uniform(m) model in a Gauss distributed $N(128, 2^3)$ image of 96×96 pixels. On bottom: Redundancy using the uniform(v)-uniform(m) model.

the learning algorithm, which induces a theory from this sample. This theory is then tested on the remaining examples. If the quality of the learned theory is not sufficient, the window is adjusted, usually by adding more examples from the training data, and a new theory is learned. This process is repeated until a theory of sufficient

quality has been found.

The detection of phase transitions in optimal unsupervised learning using windows of increasing size implies the generation of a set of optimal window sizes. The first window is such that it contains the critical number of samples that allow to detect structure in the data. The second window is such that additional samples are redundant. The third window is the maximal set for which it is true that all the elements have the same feature as a given one.

We could think of Fürnkranz's "theory of sufficient quality" as that produced by our first optimal window. Since the main topic of this chapter is to obtain the third window in the set, which is associated with the detection of the emergence of a new class, let us extend Fürnkranz's windowing approach to data-sets composed by homogeneous regions corresponding to more than one class.

The definition of windows of increasing size is here formalized. Let $\mathsf{X}_0 = \{x_0[r,c] \in \mathsf{X}\}$ be the initial subregion pointing out a class of mean μ_A. Let us assume that all the values of the pixels located in the neighborhood of the location $[r,c]$ are generated with the same probability density function as x_0. In general, $\forall s > 0$, let $\mathsf{X}'_{s-\delta s}$ be the set of pixels that do not belong to $\mathsf{X}_{s-\delta s}$ but having at least a neighbor with $\mathsf{X}_{s-\delta s}$ under certain connectivity. Let the set X_s be the region jointly formed by $\mathsf{X}_{s-\delta s}$ and the pixels of $\mathsf{X}'_{s-\delta s}$, with n_s being the number of elements of X_s. For some s, $\mathsf{X}'_{s-\delta s}$ will include pixels with a mean different to μ_A, for instance, μ_B. Let us assume that, $\forall s$, the information provided by $\mathsf{X}'_{s-\delta s}$ to detect the emergence of the new class B is constant and sufficient.

Window size optimization in circular sets

Let us assume that we have a realization $\mathsf{D}_{t_{max}}$ centered on the location (r,c) that changes from an internal class I to an external class E at radius t_o. Let the mean of the internal class μ_I be different from the mean $\mu_{(I \bigcup E)}$ of the union of the internal and the external class. The former is the statistic to be learned. Let the variance of the internal class σ_I^2 be less than the variance of the union $\sigma^2_{(I \bigcup E)}$.

As mentioned before, the index t in Section 2.4.1 becomes a variable. Then, for each vector D_t generated with the elements of the set $\mathsf{X}_t = \mathsf{D}_t$ with $0 \leq t(v_o, m_o) \leq t_{max}$, a corresponding realization G'_{t,m_o} can be generated according to Equation 2.37. Note that the values of $t = t(v_o, m_o)$ are defined over the line of maximum redundancy of the mapping model (See Equation 2.51).

If we compute the estimated means between groups $y_t = y_{v_o,m_o}$ for increasing size of $t(v_o, m_o)$, in each iteration we make an estimation of the real mean of the union of the internal and the external class $\mu_{(I \bigcup E)}$. However, if we avoid adding

instances of the external class to the sample, y_t is really the estimation of the mean μ_I of the internal class. For increasing sample size, we expect the estimated variance of the estimated means within groups to minimize. From a physical point of view, a system that tends to be singular, tends to be more sensitive to perturbations. The emergence of external pixels are a clear perturbation to the system. Some estimated means within groups will be outliers in the set of means. Therefore, it is quite reasonable to expect the estimated variance of the estimated means within groups to increase, if the amount of external pixels is sufficient.

The emergence of a new class can then be detected at the minimum t_o of the estimated variance of the estimated means within groups $\mathtt{Z}$ evaluated along the line of maximum redundancy of the mapping model:

$$\mathtt{Z} = \sigma^2_{v_o,m_o} =< D^2_{v_o,m_o}, \Omega_{m_o} > -y^2_{v_o,m_o} \tag{2.52}$$

where each $d^2_{v_o,m_o}[i] \in D^2_{v_o,m_o}$ is the square of the element $d_{v_o,m_o}[i]$ of D_{v_o,m_o} that is at the same position i. This is equivalent to:

$$\mathtt{Z} = \sum_{i=1}^{n_{m_o}} \frac{g_{t,m_o}[i]^2}{\omega_{m_o}} - (\sum_{i=1}^{n_{m_o}} g_{t,m_o}[i])^2 \tag{2.53}$$

with $m = m_o$, and where $g_{t,m_o}[i]$ is defined in Equation 2.29. Then,

$$t_o = min[\mathtt{Z}] \in \hat{W}_o \tag{2.54}$$

where $\hat{W}_o$ is the set of estimates of the elements of W_o (see Equation 2.2), and the internal entropy to external entropy ratio:

$$IEER = \frac{E(\Omega_{t-s})}{E(\Omega_t - \Omega_{t-s})} \simeq IEER_o \tag{2.55}$$

were s is the variable increment of the radius-, holds $\forall t$. The reason of putting a constant threshold for this ratio is that we would like to hold the effect of the new pixels added to the sample constant. See Section 2.4.2 for more details.

The behavior described before can be observed for the four weighting functions Ω_t tested on an example that was generated by seeding a pixel in the center of a Rayleigh-distributed class[4]. The change of class happens at radius 32. In general, all the models with overlapping discs have a function that minimizes in the neighborhood of the border between the external and the internal class. (See Figure 2.11).

Note in figure 2.12 that the real mean μ_I is estimated faster by the $y_{t(v,m)}$ corresponding to the uniform(v)-uniform(m) model.

[4]This distribution was used in order to evaluate the performance of this model in non-Gauss distributed classes too.

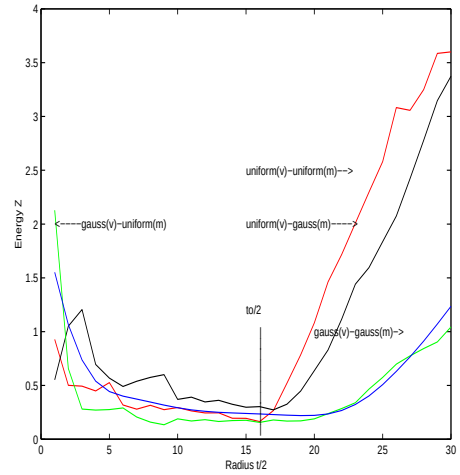


Figure 2.11: Estimated variance of the estimated means within groups using different weighting functions Ω_t in the estimation of the mean

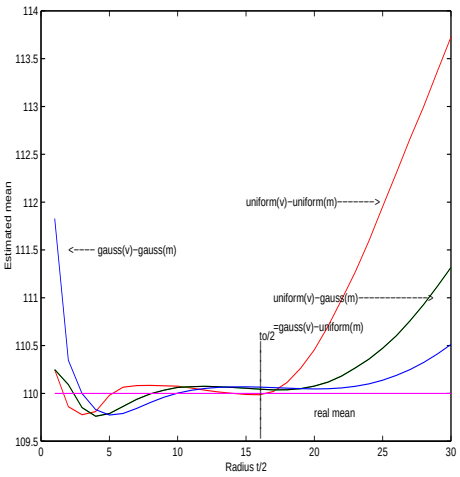


Figure 2.12: The estimated mean y_t using different weighting functions Ω_t

The internal to external entropy ratio

An experiment reported in Section 4.1.1([41]) has shown that, for large expected optimal window sizes in synthetical images, the gauss(v)-gauss(m) generates window

sizes greater than the real ones, which were known. This is due to the fact that the internal entropy to external entropy ratio grows faster in that model than in the others. (See Figure 2.13 and note the difference in the scale of the y-axis.) For this reason, excluding the case in which the internal entropy to external entropy ratio is less than 1, we can conclude that the uniform(v)-uniform(m) weighting function has the best performance at detecting the emergence of a new class. It should be taken into account, that if the ratio is much greater than 1 the emergence of a new class will be detected with a delay, and wrong examples will be included in the initial sample generated. This means, that in the first iterations the amount of information that the new pixels provide is significantly larger than in the next iterations. Therefore, to improve the performance of the scheme proposed, variable and increasing increments of the disc radius have to be used. The increment can depend on a constant internal to external entropy ratio set by the user.

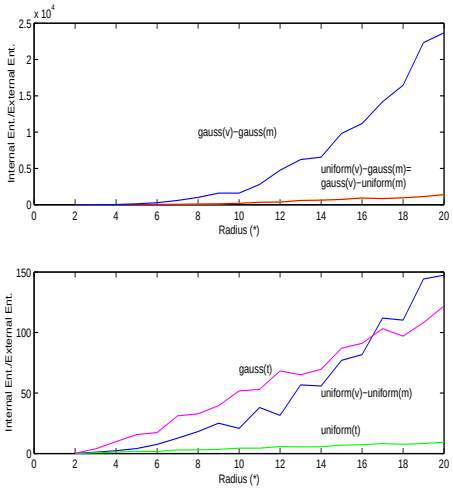


Figure 2.13: Internal entropy to external entropy ratio versus disc radius. (*)The x-axis corresponds to the radius $(t-1)/2$. The gauss(t) and uniform(t) are only included as a reference.

2.4.3 Application: An Extension to Mean Shift Detection

A scheme for detecting heterogeneous regions in dermatological images with malignant melanoma is proposed in [40]. The mean shift detection scheme is divided into two stages: window size optimization and detection. The first stage is carried out according to Section 2.4. Details of the second are included later in this section.

In general, the initial training set generation scheme proposed in [41] (see Section 2.4) works well for prototype pixels placed within homogeneous regions. However, when the optimal window size is estimated for prototype pixels seeded near borders, very large optimal window sizes can be estimated. This is due to the following facts. First, since we are working with a rectangular lattice (See Equation 2.1), the smallest circular 2-D set of weights that can be obtained from a convolution of other two 2-D sets of weights corresponds to a window of 5 pixels by 5 pixels, or -in our notation- a disc of radius $t = 2$. This means, that we can not expect to detect the emergence of a new class having as center of windows pixels that are, for instance, over borders or near them. Secondly, in order to arrive at the same constant minimal value, samples with large values of variance have to be increased in size much more than samples with small values of variance. Since neighborhoods of pixels belonging to borders have a large value of variance, the probability of estimating large optimal window sizes for these kind of pixels increases. The present application takes advantage of this fact.

Plotting the normalized histogram of the synthetical image containing the values of minimum estimated variance of the estimated mean within groups (See Equation 2.52), then a bi-modal density function could be appreciated. The peak on the left corresponding to the probability density function of the homogeneous regions, and the peak on the right, for larger values of a minimizing energy function, to pixels belonging to heterogeneous regions (border pixels). In the second stage, those regions are detected applying, first, an expectation-maximization algorithm and, afterwards, automatically defining a threshold between homogeneous and heterogeneous regions.

Detection

The histogram of the synthetic image containing the minimum estimated variance of the estimated means within groups is constructed. An Expectation-Maximization Algorithm [57] (setting $\epsilon = 0.0001$ in [57]) is expected to provide the parameters (mean μ, covariance matrix Σ and a-priori probability p) of two Gauss distributed classes: the homogeneous region and the heterogeneous region. The threshold between these classes is obtained using the following discrimination function for Gauss-distributed classes with different dispersion [9]:

$$-(h-\mu_1)'\Sigma_1^{-1}(h-\mu_1) + (h-\mu_2)'\Sigma_2^{-1}(h-\mu_2) \geq 2log(\frac{L(2,1)p_2}{L(1,2)p_1}) \tag{2.56}$$

where μ_i is the mean, Σ_i is the covariance matrix and p_i is the a priori probability of the $i-th$ class $\forall i \in \{1,2\}$. $L(1,2)$ is the cost of belonging to class 1 and being classified as belonging to class 2 and viceversa for $L(2,1)$; these can be assumed to be equal. Pixels whose minimum estimated variance of the means within groups is greater than the threshold are considered belonging to the heterogeneous regions,

and viceversa, for the homogeneous regions.

The present scheme was tested on a set of synthetical images. Results can be appreciated for a real image in Figure 2.14. As an extension of the algorithm to textural cases, a synthetical example generated with a set of Brodatz's textures [8] was used (See Figures 5 and 6 in [40]). The function minimized, in this case, was the estimated variance of the fractal dimension within groups [5]. This was computed for discs of radius $t = 2$, 4, 6, 8, 10, 14 and 18.

Figure 2.14: The dermatological image decomposed into homogeneous (left) and heterogeneous (right) regions using estimated means within groups.

2.5 Window Size for Classification

After estimating, for each single class[6], a statistically valid sample, from which the probability density function of that class can be estimated, it could be useful to estimate the minimal sample size needed to classify elements of different classes.

Definition 2.5 Given a pair of classes, the window size for classification is the minimal size of a statistically valid sample to be assigned to the correct class with a certain error rate.

[5]The fractal dimension f_d of a window D_w was defined as:

$$f_d(w) = \frac{\max(D_w) - \min(D_w)}{2w + 1} \tag{2.57}$$

[6]for instance, using the approach presented in Section 2.4, or using region growing algorithms

In the present work, the critical number of samples for classification is estimated as a function of the separability between neighbor classes:

$$w_c = -6.8341 + \frac{10.1541}{\sqrt{2dA}} \tag{2.58}$$

where w_c is the radius of the estimated optimal disc for classification and dA is the weighted sum of the difference of areas of the class density functions:

$$dA_{(1),(2)} = dA_{(2),(1)} = \frac{1}{2}\sum_{n=1}^{Q} |h_{(1)}[n] - h_{(2)}[n]| \tag{2.59}$$

Q is the number of quantization levels, $h_{(1)}$ and $h_{(2)}$ are the estimated density functions of classes 1 and 2 respectively. It is optimal in the sense that it corresponds to the best classification rate for a set of experimental data with similar statistical descriptions. Figure 2.15 shows the graphical behavior of this relation.

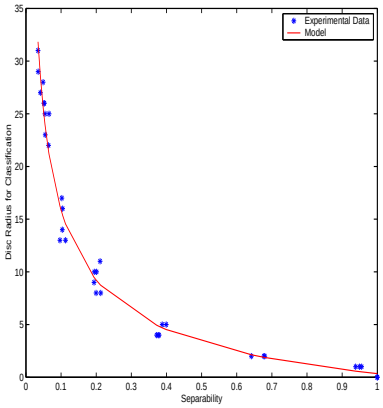


Figure 2.15: Separability versus Critical Sample Size

This empirical model was constructed by performing linear regression of experimental data generated as described bellow. A set of 36 pairs (A, B) of synthetical images of size 128^2 pixels was generated. The values of the pixels were generated using different Gauss weighting functions $N(\mu_A, \sigma_A^2)$ and $N(\mu_B, \sigma_B^2)$. The values of the parameters were $\sigma_A = \sigma_B = 2^{j-1}, j = 1..6$ and $\mu_A = 128 + \frac{2^{(i-1)}}{2}$ and $\mu_B = \mu_A + \frac{2^{(i-1)}}{2}, i = 1..6$. For each pair of images (A, B), neighborhoods (discs) of increasing size were centered on the pixels from rows 33 to 96 and columns 33 to 96

of each one of the images. The radius of the discs varied from 0 to 31.

For each window size:

- the normalized histogram of the centered neighborhood of each pixel was computed;
- the distance dA from this histogram to the normalized histogram of both classes was calculated using Equation 2.59.
- and the pixel was assigned to the closest class.
- The quality of the final classification was measured with Cohen's Kappa coefficient κ (See Equation 7 in Section 4.2.2).

In this case, the optimal radius for each pair of images corresponded to the maximum Kappa value one.

The following tables show the distances between each considered pairs of classes and the corresponding optimal disc radii for classification.

Table 2.3: The distance dA between pairs of classes given by the separation between their means $s\mu$ and their standard deviation σ.

σ / $s\mu$	1	2	4	8	16	32
0.5	0.2130	0.1035	0.0511	0.0349	0.0345	0.0549
1	0.3994	0.1998	0.0968	0.0488	0.0411	0.0543
2	0.6416	0.3720	0.1946	0.1022	0.0659	0.0523
4	0.9383	0.6797	0.3879	0.1938	0.1046	0.0644
8	0.9999	0.9493	0.6772	0.3767	0.2111	0.1127
16	1.0000	1.0000	0.9555	0.6797	0.3790	0.2010

Table 2.4: The disc radius w_c corresponding to the best kappa value (one) for each pair of classes given by the separation between their means $s\mu$ and their standard deviation σ.

σ / $s\mu$	1	2	4	8	16	32
0.5	8	14	26	29	31	23
1	5	8	13	28	27	25
2	2	4	10	17	25	26
4	1	2	5	9	16	22
8	0	1	2	4	11	13
16	0	0	1	2	4	10

2.6 A Contextual Classifier that only requires one Prototype Pixel for each Class

A three stage scheme for classification of multi-spectral images was proposed in [38]. In each stage, statistics of each class present in the image are estimated. The user is required to provide only one prototype pixel for each class to be seeded into a homogeneous region. The algorithm starts by generating optimal initial training sets, one for each class according to Section 2.4. Afterwards a region growing algorithm increases the sample size providing more statistically valid samples of the classes. Final classification of each pixel is done by comparison of the statistical behavior of the neighborhood of each pixel with the statistical behavior of the classes. A critical sample size obtained from the model described in Section 2.5 is used in this stage.

The algorithm was tested with the Kappa coefficient κ (See Equation 6 in Section 4.2.2) on synthetical images and compared with K-means ($\overline{\kappa} = 0.41$) and a similar scheme that uses spectral means ($\overline{\kappa} = 0.75$) instead of histograms ($\overline{\kappa} = 0.90$). Results can be appreciated on a dermatological image with a malignant melanoma in Figures 2.16 and 2.17.

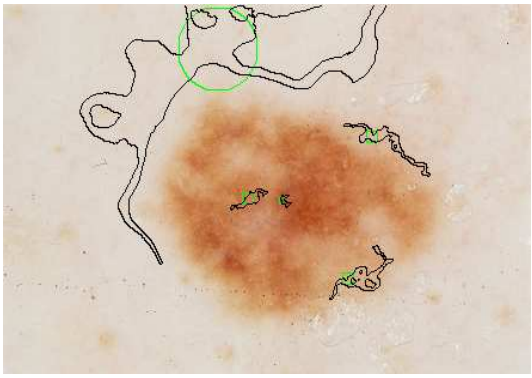


Figure 2.16: The initial training sets (delineated in green) centered on the seeded prototype pixels overlayed with the grown regions (delineated in black) of the image of a malignant melanoma.

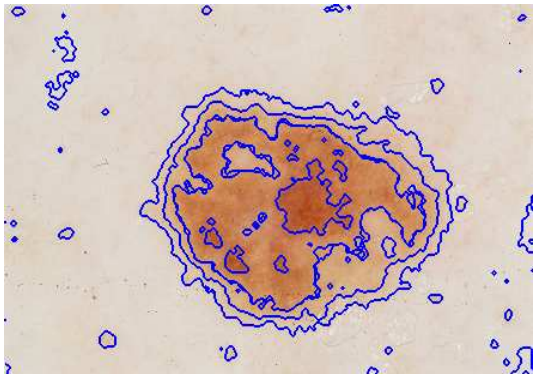


Figure 2.17: The original image overlayed with the thematic map after three iterations of the median filter.

2.7 Iterative Extended Mean Shift Algorithm

For a given class, it is expected that the probability density function of the estimated means (See Figure [7] 2.18) contracts to a single point: "for each pixel of a multi-spectral image, the fixed point is the center of the cluster to which the pixel in question belongs" [39].

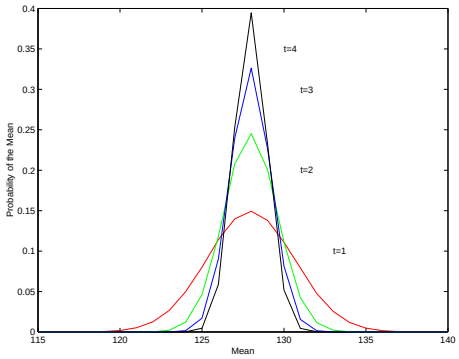


Figure 2.18: Probability density function of the mean of realizations of discs of radius $t = 1, 2, 3, 4$.

We have seen in Section 2.4 that for an image with several classes, the scheme proposed produces a border effect. This behavior can be seen by carrying out the following procedure. Given an image, we generate a set of synthetical images by convolving the original image with windows of increasing size. These windows are weighting functions such that the output of the convolution is, for each single pixel, the mean of the neighborhood centered on it. We can observe that, for windows of increasing size, the histogram of the corresponding synthetical image tends to contract to the class means present in the image. However, for large window sizes, undesired peaks emerge in the histogram. They correspond to the mean values of windows centered on border pixels. Basically, this is a consequence of the fact that the weights do not depend on the data.

For an ideal situation, we would like to have a similar procedure as before, but replacing each single pixel by the mean of the class it belongs to (See Figure 2.19). We would like to clearly detect the mean shift between pairs of border pixels. In other

[7] This Figure was generated with a Gauss distributed synthetical image of 512×512 pixels with a mean of 128 and standard deviation of 8. Circular windows of increasing size were centered on each pixel. The disc radius was varied from 1 to 4. For each single disc radius the normalized histogram of the mean of windows with this disc radius was computed. We can appreciate, that "the variance of the sampling distribution is inversely related to sample size" [22].

words, we would like to enhance the contrast between image regions. The scheme proposed in [39] is an attempt to reach this goal and avoid the problem of the border effect. The main assumption is that the mean of the class a given pixel belongs to, can be estimated from its neighborhood. In comparison with the scheme proposed in Section 2.4, each neighborhood pixel is here weighted with a measure of distance to the neighborhood center pixel.

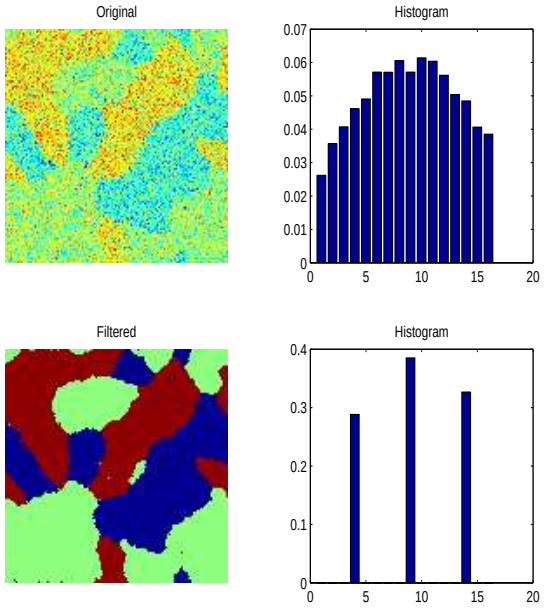


Figure 2.19: On left: a synthetical image (top) and the corresponding synthetical image (bottom) with each pixel replaced by the mean of the class it belongs to. The bottom image is the truth thematic map where each pixel is labelled with its class mean. On right: the corresponding image histograms.

Similarly to the scheme of Section 2.4, but now for each pixel in the image, an optimal circular window size is estimated. The estimated local mean of each single pixel is saved into a new synthetical image which is the input to the next iteration of the process just described. The algorithm stops when the difference between two consecutive synthetical images is smaller than a given threshold.

Also, similarly to the scheme of Section 2.4, for each iteration and for each pixel, the optimal circular window size from which the pixel mean class is estimated is obtained through the minimization of the estimated variance of the estimated means

within groups. However now, the weights used in these functions depend on the data. In the first iteration it is assumed that each group contains one single element (Equation 5 in Section 4.1.3): the original image pixel values. The mean between groups corresponding to an optimal window size is estimated (Equation 9 in Section 4.1.3), and therefore, an optimal set of groups. The optimal set of groups in one iteration is a group in the next iteration.

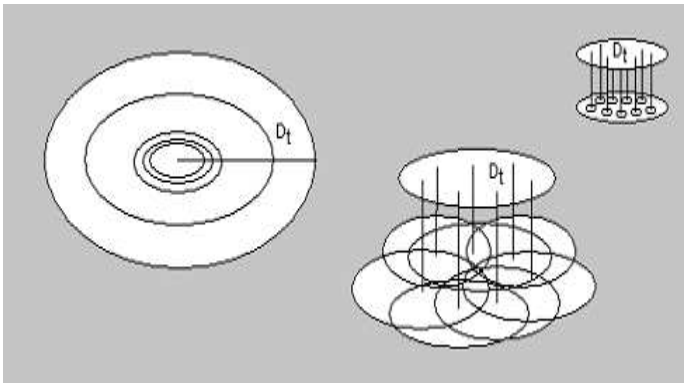


Figure 2.20: The mapping model in the Iterative Extended Mean Shift Algorithm.

Let X be the image to be processed. For each iteration $1 \leq i \leq ite_{max}$, the scheme generates the following synthetical images of the same size as X:

- $Z_{min}^{(i)}$ containing for each pixel the value of minimal estimated variance of the estimated mean within groups in the $i-th$ iteration.
- $Y^{(i)}$, with $1 \leq i \leq ite_{max}$: the estimated means between groups in the $i-th$ iteration.

The algorithm can be summarized as follows:

- 1) Set $Y^{(1)} = X$.
- For each iteration i: window size optimization for each pixel in $Y^{(i)}$.
  - 2) Initialize the the minimal variance image $Z_{min}^{(i)}$ in maximal values.
  - 3) For each pixel, and for each window size: If the estimated variance (Equation 8 in Section 4.1.3) of the estimated means within groups is smaller than or equal to the minimum $Z_{min}^{(i)}$, then update $Z_{min}^{(i)}$ with the corresponding estimated mean between groups (Equation 9 in Section 4.1.3).
  - 4) If $Y^{(i)} = Y^{(i-1)}$ then stop else set $Y^{(i+1)} = Y^{(i)}$ (the optimal set of groups is a groups in the next iteration) and goto 2).

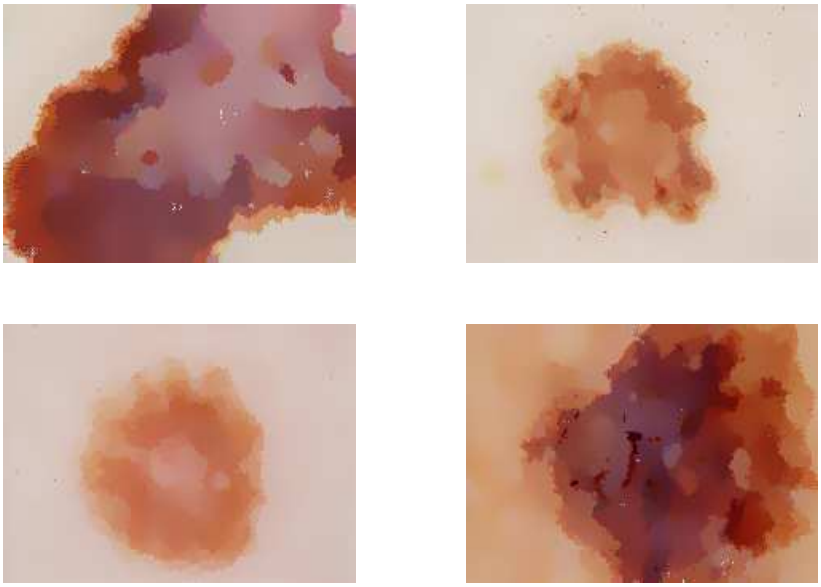


Figure 2.21: Output of the Iterative Extended Mean Shift Algorithm for the images of malignant melanomas shown in Figure 1.2.
From top to bottom and left to right the number of iterations was 7, 7, 7 and 6 respectively.

For the images shown in Figure 1.2, the outputs the Iterative Extended Mean Shift Algorithms can be seen in Figure 2.21.

The algorithm was also tested with the Kappa coefficient κ (See Equation 6 in Section 4.2.2) on the same set of synthetical images shown in Figure 3 in Section 4.1.2. For these images, the monotonically increasing behavior of the coefficient can be seen in Figure 2.22. Compared with the iterative median filter ($\overline{\kappa} = 0.45$), the average Kappa coefficient for the Iterative Extended Mean Shift Algorithm had a value of $\overline{\kappa} = 0.68$.

Conclusion

Let a $K-$class image of different means be composed of homogeneous regions. Let a prototype pixel be seeded in a homogeneous region. We want to find the maximal prototype centered neighborhood for which it is true that all the pixels belong to the

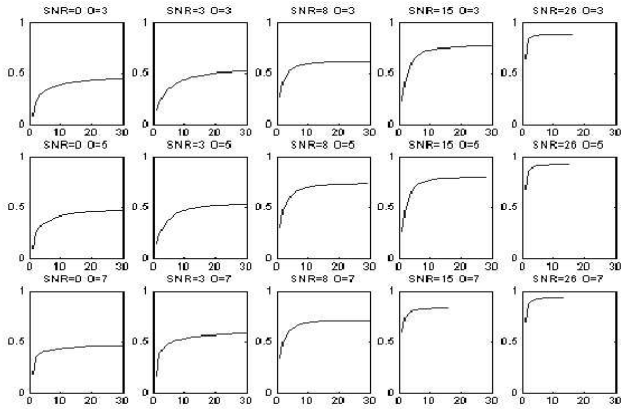


Figure 2.22: Kappa values along the iterations of the Iterative Extended Mean Shift Algorithm for the synthetical images shown in Figure 3 in Section 4.1.3.

same class as the prototype. Let a set of windows of increasing size be centered on the prototype pixel. If we track the estimated variance of the estimated means within groups over the line of maximal redundancy we can find that optimal maximal set in the minimum of this line. The redundancy is computed as a function of the weights finally assigned to the window elements when the estimated mean between groups is computed.

CHAPTER 3

Novelty Detection in Unregistered and Registered Patterns

3.1 Introduction

Novelty detection in textural patterns is considered here. First, the detection of textural changes by avoiding pattern registration is explored. Secondly, changes in registered patterns are analyzed. Problems related to these aspects, both, are also considered. These are the lesions segmentation, the illumination correction and the patterns registration.

The set of psoriasis lesion images mentioned in Section 1.2 is used here. In spite of being redundant, these images correspond to five repetitions of three psoriasis lesions of three patients during at least three weeks. The images taken were labelled with four characters, indicating patient (1, 2, 3), lesion (A, B, C), session (a, b, c, d) and capture (1, 2, 3, 4, 5) respectively. Following the notation used in previous work ([36], [34], [35], [37], [33], [32]), the group of images corresponding to a given patient and lesion is here called "case (patient, lesion)". As an illustrative example, the images of the case $(1, A)$ are shown in Figure 3.1. Each row contains the captures taken during a given session. For this particular case, we have images corresponding to four weekly sessions. From now, results obtained with the different approaches introduced in the present work will be shown for this case. For the remaining cases, details can be found in the technical reports included in Part II.

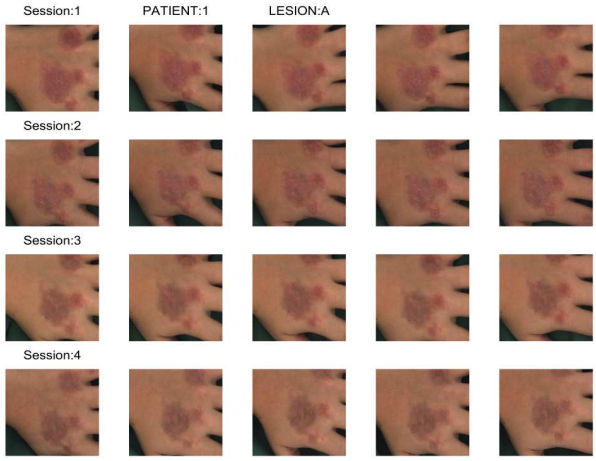


Figure 3.1: Set of Original Images of (Patient 1, Lesion A).

To evaluate lesion changes between sessions, physicians make scores of a four-variable set: redness, scaling, thickness and body area covered by the lesions. Manual scoring depends strongly on the dermatologist, who may show variations of criteria along sessions due to the huge amount of patients and lesions observed during each working day. In order to use a set of lesions like these as input to an automatic system that analyzes their change in time, it could be helpful to express the data in a way so they are comparable, without loosing connection with the original data. Registration of segmented lesion patterns is a possible solution.

In the present work, the feasibility of quantifying textural change between pairs of segmented patterns without registering them is first explored (see Section 3.4). To transform pairs of unregistered original data to a space were they have a perfect registration is possible through a textural descriptor that is invariant to translations and rotations of patterns, like Haralick's co-occurrence matrix [15]. Afterwards, a within and between sessions of segmented patterns registration scheme is proposed (see Section 3.5.1). Then, textural changes after pattern registration are found.

As it is pointed out in [33], not only the way the data are presented to a given distance function is relevant for novelty detection. In order to detect alterations, the measure of distance between pairs of data sets itself is important. A statistical approach that allows detection of alterations is the M.A.D. Transform (Multi-variate Alteration Detection Transform) [44], [45]. This scheme transforms two sets of multi-variate observations into a difference between two linear combinations of the original variables explaining maximal change in all variables simultaneously. In the present

work, the M.A.D. Transform was applied to pairs of registered and unregistered patterns.

This part of the summary starts with a description of the segmentation scheme used to delineate the lesion patterns [34]. Afterwards, a short section describing the M.A.D. Transform is included. This is followed by a section about the textural model used to represent unregistered patterns and the results obtained by applying the M.A.D. Transform to these data [37]. Finally, a combined alignment and registration scheme is proposed [33]. Change detection in registered patterns is carried out [32].

3.2 Lesion Patterns Segmentation

In order to study the variation within and between sessions of lesions, they first have to be found and matched. In order to find a lesion, the set of image bands, or combination of bands where the lesion is expected to be found has to be specified. A comparison of different approaches for unsupervised color image segmentation applied to skin tumors was given in [14]. In this work, it was determined that the median cut and the adaptive thresholding algorithms applied to the first principal component provide the lowest average error. The first is a compression scheme. The second determines valid classes from modes of histograms of scale-space filtered versions of the original image. For all the cases, the tumor object candidate has to be selected. Afterwards, a contour encoding algorithm is used to first vectorize and then smooth the vector data. Border contour are afterwards filled.

Looking at the original images shown in Figure 1.2, we note that, for some cases, the lesion has a light color, while the skin is dark. For other cases, the opposite situation is seen. This means, that in the whole set of original images there is no common lesion pattern homogeneity criteria that can be defined. It could, for instance, be nice, for all the cases, to locate the lesion in the same part of the histogram. For this reason, the linear combination of the original bands given by the principal components decomposition, in which lesion patterns satisfy a common homogeneity criteria, was analyzed in [36].

General results of Principal Components Analysis are presented in Section 3.2.1. The description of a two-stage hierarchical classification scheme follows [34]. Evaluation of the results motivated the design of an illumination correction scheme [35], also included here.

Band	Comp. 1	Comp. 2	Comp. 3
Red	0.9907	0.1170	0.0128
Green	0.9741	0.1664	0.0921
Blue	0.9674	0.2077	0.1078

Table 3.1: Average absolute correlation values between bands of single images and their principal components

3.2.1 Principal Components Analysis

Principal Component Decomposition has been applied to each single psoriasis lesion image. For all the cases, the variance-covariance matrix was diagonalized. The principal components of each single image of our example case, the lesion A of patient 1, are shown in Figure 3.2. From top to bottom, three blocks with the same number of rows are considered. The first block (first third of the rows) is composed by the first principal components of the images of the mentioned case $(1, A)$. The second block (second third of the rows), by the second principal components, and so on, for the third block (remaining rows). Within a block, the row number indicates the number of the session, and the column number indicates the capture number within a session. In general, in order to be able to compare the elements of the sequence given by the components of the images of a given patient and lesion, for each sequence composed by the images taken within and between sessions of a given lesion, all the eigenvectors were pointed to the same sense than the eigenvectors of the first image of that sequence.

Before studying the variation within and between sessions of lesions patterns, we need to find and match the lesion patterns first. The component(s) to use in the search also have to be specified. It has been observed that the lesions look roughly enhanced from the rest in the last principal component. Taking a look at the values of the third eigenvectors, it can be deduced that a rough estimation of the third component can be obtained by subtracting the green band from the blue band, while the red band almost does not have any influence. (See Figure 3.3). The estimation of the third principal components of the images of the cases $(1, A)$ can be seen in Figure 3.4.

Furthermore, for each single image, most of the contribution in its first component is given by the red band, which has an average absolute correlation with the first component of 0.9907. In fact, as it can be seen in Table 3.1, all the bands are highly correlated with the first component. On the other hand, the average over all the cases of the percentages of the total variance explained by the first, second and third principal components is 96.88, 2.61 and 0.51 respectively (These are column mean values of Table 4 in [36].)

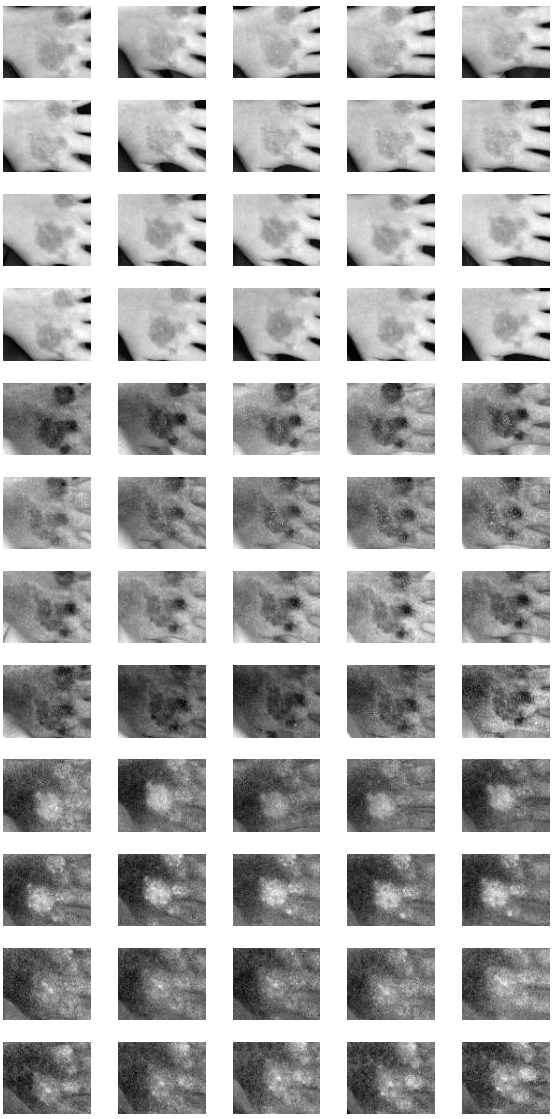


Figure 3.2: Principal Components of the images of the case (Patient 1, Lesion A).

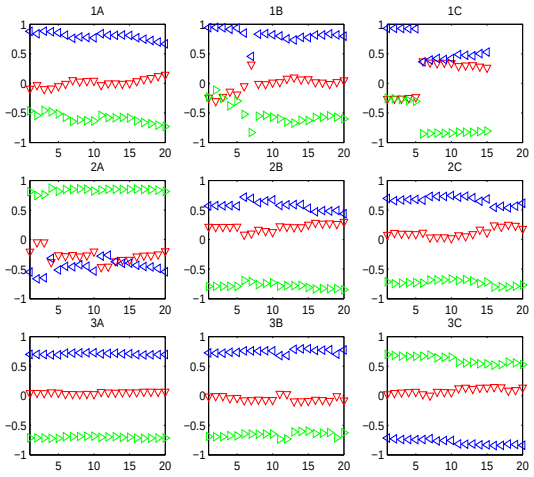


Figure 3.3: Values of the third eigenvectors of the images of the data-set. Triangle down indicates the weight for the red band; triangle right, for the green band; and triangle left, for the blue band.

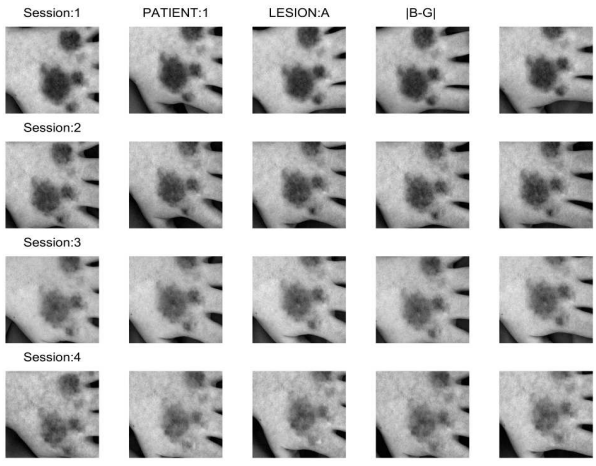


Figure 3.4: Set of |B-G| Images of (Patient 1, Lesion A).

3.2.2 A Hierarchical Classification Scheme

The images of lesions with psoriasis are assumed to contain three classes: background, normal skin and lesion. As was mentioned before, lesions have high contrast with the rest in the third principal component. However, in the estimation of the third principal component, this means, the absolute value of the difference between the blue and the green band $|B - G|$, lesions and background are both enhanced from normal skin. The background has a good contrast with the rest in the red band. Advantage can be taken of the green color of the curtain used as background during the image acquisition process. Its density function is seen to be positioned in one extreme of the red band histogram.

A two-stage hierarchical classification scheme is proposed here. In both stages, the pixels of the classes are assumed to be Gaussian distributed. Although this is almost certainly not the case, Clarke et al. have shown that the quadratic method is fairly robust to deviations from normality, unless we have very skewed marginals [7]. The red band R is used first to discriminate between the background and the skin, and then the output of the previous step is combined with the $|B - G|$ band for segmenting the lesions from the normal skin. In both stages, the use of Wang's Expectation-Maximization Algorithm [57] (setting $\epsilon = 0.0001$ in [57]) is proposed. As was mentioned in Section 2.4.3, this algorithm is expected to provide the parameters (mean μ, covariance matrix Σ and a-priori probability p) of two Gaussian distributed classes: skin and background in the first step, and lesions and normal skin in the second step. The threshold between these classes is obtained using the discrimination function for Gaussian-distributed classes with different dispersion [9] (See Equation 2.56). This function is applied to the output of the convolution of the image with a circular window of optimal size. The critical number of elements needed for classification is computed as a function of the minimal separability between the two classes [38] (See Equation 2.58). The values of this optimal window sum to one; each element of the window has the value $1/n_c$ corresponding to the inverse of the size of the window n_c given by its number of elements. It is clear that after convolving the image, each pixel is replaced with the estimated mean of its neighborhood.

In the first step of this classification scheme, pixels whose estimated local mean is greater than the threshold are considered belonging to the class skin, and vice-versa, for the background. In the second step, pixels whose estimated mean of the neighborhood is greater than the threshold are considered belonging to the class normal skin, and vice-versa, for the lesion. As an illustrative example, the classification outputs of the images of the case $(1, A)$ are shown in Figures 3.5 and 3.6.

After classification evaluation [48] [52] it was obvious that certain important aspects related to the images had not been addressed. This is, for instance, the case of the shadows, which clearly affect the quality of the classification results. Furthermore, in order to automate the lesion registration process, it is required that

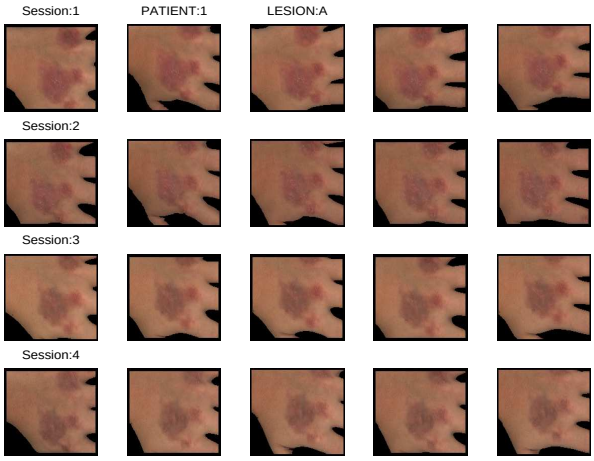


Figure 3.5: Overlay of the outputs of the first classification stage with the original images of the case $(1, A)$.

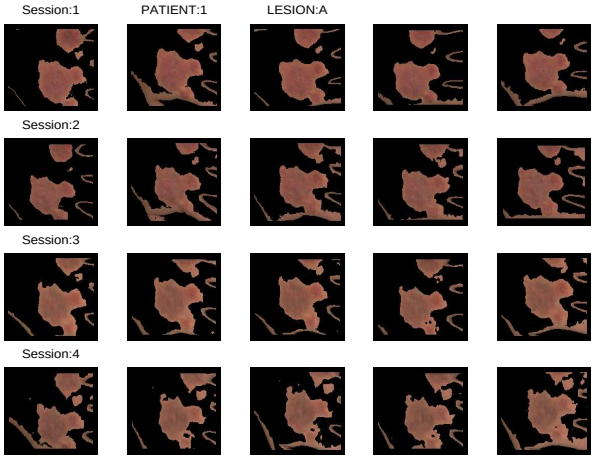


Figure 3.6: Overlay of the outputs of the second classification stage with the original images of the case $(1, A)$.

thematic maps from which the lesion patterns to be registered are selected, contain a repeated pattern within and between sessions. This is in general not the case when

the original data set is used. Illumination correction is required in order to improve the classification output. Hair and lesions occlusions are also aspects mentioned in [34] not considered further here [1]. On the other hand, results shown in [34] indicate that the parameters of the first image of each session could be used to segment the remaining images of the same session, when the situations mentioned before do not appear.

3.2.3 Illumination Correction

Thanks to the use of a sphere [1] (see Figure 3.7) with optimal illumination conditions, -which assumes that the captured objects are plane-, some skin images affected by shadows due to the non-plane shape of the objects can be corrected. Since an arm or a leg are more similar to a cylinder than to a plane, a quadratic model is assumed for the captured objects. Details of the illumination correction scheme proposed follow.

Figure 3.7: Sphere with optimal illumination conditions used to capture the images.

Following [25], let $i(x, y)$ be the illumination function affecting the original scene $f(x, y)$:

$$g(x, y) = i(x, y)f(x, y) \tag{3.1}$$

such that $g(x, y)$ is the illumination affected scene. This model assumes that the output is the product between the amount of arriving light and the original scene reflectivity. Basically, the region representing a used selected class k is used to model the illumination function. Within the k-th class we assume as a first approximation that the original image should be constant:

$$i(x, y) = \frac{g_k(x, y)}{\beta_k} \tag{3.2}$$

[1] See in [42] a scheme for registration in presence of occlusions. A hair removal scheme is proposed in [6].

where g_k is the observed illumination affected scene in the region representing the class k, and β_k is a unknown constant.

Pixel values belonging to the region representing the k-th class are replaced by local means, which are used to estimate an interpolation function. After normalization, the interpolation function is assumed to be the illumination function affecting the scene. This is more precisely defined as follows. Let G, F and I be specific images of $g(x,y)$, $i(x,y)$ and $f(r,c)$ respectively. Let $\mathsf{F} = \{f[r,c]\}$ and $\mathsf{G} = \{g[r,c]\}$ be hierarchically defined in terms of $\mathsf{Z} = \{z[r,c]\}$ that represents the partition of the scene in different classes. Each $z[r,c]$ is a value in $\Omega = \{1,\ldots,K\}$, where $z[r,c] = k$ means that the coordinate $[r,c]$ of a pixel belongs to the $k-th$ class. Let $\mu = \{\mu_1,\ldots,\mu_K\}$ be the class means, and $\mathsf{M} = \{m[r,c] = \mu_k / z[r,c] = k \forall k \in \Omega\}$ be such that $m[r,c]$ is the mean of the class the pixel at location $[r,c]$ belongs to. Let N be defined with the local means of G such that $E[\mathsf{N}] = \mathsf{M}$. Let $\hat{\mathsf{N}}_k$ be the interpolation function of sampled values of $\mathsf{N}_k =< \mathsf{N}, \mathsf{Z} = k >$ greater than zero [2]. Then $\hat{I}[r,c] = \frac{\hat{\mathsf{N}}_k[r,c]}{O_k}$ and O_k is the normalization factor.

Body parts are modelled with a quadratic function. The regression matrix, from which the coefficients of the interpolation function are adjusted, is given by:

$$\boldsymbol{y}_{Px1} = \left[\begin{array}{cccccc} \mathbf{1}_{P\times 1} & \boldsymbol{r} & \boldsymbol{c} & \boldsymbol{rc} & \boldsymbol{rr} & \boldsymbol{cc} \end{array} \right] [\boldsymbol{\theta}_{Px1}] + \boldsymbol{\Sigma} \tag{3.4}$$

where y_i is the local mean value $n_k[r_i,c_i]$ for a given band and $[r_i,c_i]$ is the location of the i-th pixel sampled, which belongs to the k-th class. $\boldsymbol{r} = [r_1,\ldots,r_P]^T$, $\boldsymbol{c} = [c_1,\ldots,c_P]^T$ and $\boldsymbol{rc} = [r_1c_1,\ldots,r_Pc_P]^T$, $\boldsymbol{rr} = [r_1r_1,\ldots,r_Pr_P]^T$ and $\boldsymbol{cc} = [c_1c_1,\ldots,c_Pc_P]^T$. The normalization factor O_k is the maximal value of $(\hat{\mathsf{G}}_k[r,c] + \gamma)$ and γ is the minimal constant added to $\mathsf{G}_k[r,c]$ such that the output is positive.

This procedure can be applied to each single image band. Thus, for multi-spectral images, the output of the present scheme is a multi-spectral estimated illumination function. See for instance, in Figures 3.8 and 3.9 the estimated illumination functions and the illumination corrected images for our example case $(1,A)$.

In Appendix C of [35], it can be seen, that the thematic maps of some illumination corrected images have some repeated patterns within and between sessions. Based on this, it is considered that the output of the illumination correction is a better estimation of the expected data. However, in order to be able to automatically select corresponding objects in the thematic maps of the same lesion, further work needs to

[2]

$$< \mathsf{N}, \mathsf{Z} = k > \tag{3.3}$$

is the scalar product between the tensor N and the thematic map that has, in a given location the value one, if the corresponding pixel belongs to the k-th class and the value zero otherwise.

Figure 3.8: Illumination functions of the original images of (Patient 1, Lesion A).

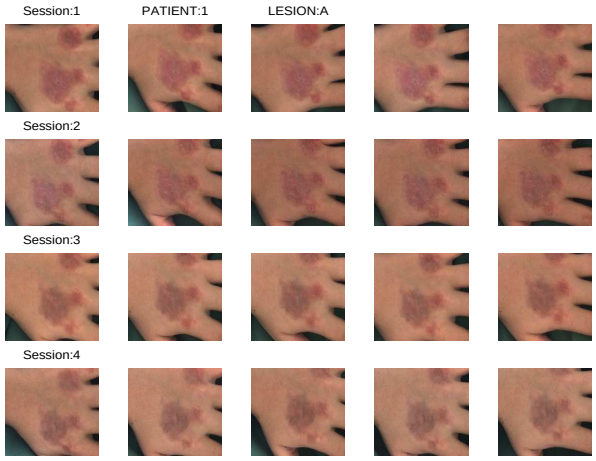


Figure 3.9: Images of (Patient 1, Lesion A) after illumination correction.

be done [3]. Since it is not the main objective of the present work, the cases for which it was not possible to automatically define corresponding objects, manual selection

[3] See, for instance, [49], in which a shell-shaped structure of the skin chromaticities cluster was measured and modelled.

after visual inspection of the images was done.

3.3 The M.A.D. Transform

Texture alteration detection in registered and unregistered patterns is going to be carried out using the Multi-variate Alteration Detection. For completeness, general aspects of this transform are included here.

Following [44], let X_1 and X_2 be two sets of variables of dimensions m_1 and m_2 respectively, with $m_1 \leq m_2$, $E\{X_1\} = E\{X_2\} = 0$ transformed with the coefficients from a standard canonical correlation analysis $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in

$$\begin{array}{l} Y_{1i} = \boldsymbol{a}_{1i}^T X_1, i = 1, ..., m_1 \\ Y_{2i} = \boldsymbol{a}_{2i}^T X_2, i = 1, ..., m_2. \end{array} \tag{3.5}$$

Further Y_1 and Y_2 are positively correlated and with unit variance, such that the variance of their difference is maximized.

$$Var[\boldsymbol{a}_1^T X_1 - \boldsymbol{a}_2^T X_2] = \tag{3.6}$$

$$Var[\boldsymbol{a}_1^T X_1] + Var[\boldsymbol{a}_2^T X_2] - 2Cov[\boldsymbol{a}_1^T X_1, \boldsymbol{a}_2^T X_2] \tag{3.7}$$

$$= 2(1 - Corr[\boldsymbol{a}_1^T X_1, \boldsymbol{a}_2^T X_2]) \tag{3.8}$$

The M.A.D. Transform consists of the variates obtained when the corresponding canonical variates are subtracted in reverse order, what means that the m_1^{th} difference shows maximum variance among such variables, and the $(m_1 - j)^{th}$ difference shows maximum variance subject to the constraint that this difference is not correlated with the previous j ones. The M.A.D. variates are invariant to linear and affine scaling, what is not the case of the Principal Components. The dispersion matrix of the M.A.D. variates is

$$D[\boldsymbol{a}_1^T X_1 - \boldsymbol{a}_2^T X_2] = 2(I - R) \tag{3.9}$$

where I is the $m_1 \times m_1$ unit matrix and R is the $m_1 \times m_1$ matrix containing the sorted canonical correlations on the diagonal and zeros off the diagonal.

3.4 Novelty Detection in Unregistered Patterns

To express pairs of unregistered patterns in a transformed space were there is a perfect descriptors registration is possible through a textural descriptor invariant to translations and rotations of patterns, like Haralick's co-occurrence matrix [15]. The Multivariate Alteration Detection (M.A.D.) Transform is applied here to a co-occurrence texture model. The contribution of each single color band to the textural change is then analyzed.

3.4.1 The Texture Model

Following [26], let δ be a compact and convex set established in the image space as a circular window and let β be a vector position operator relating the relative spatial location of a pair of pixels in the window; for each single band, the co-occurrence matrix C is the estimated probability of having a pair of pixels (λ_p, λ_q) in the relative position given by β. The elements of the $Q \times Q$ co-occurrence matrix $C(\lambda_p, \lambda_q, \beta, \delta)$, with Q being the number of quantization levels, are given by:

$$\frac{O\{r|r, r+\beta \in \sigma(\delta), g(r) = \lambda_p, g(r+\beta) = \lambda_q\}}{O\{r|r+\beta \in \delta\}} \tag{3.10}$$

where $g(r)$ is the subset of the RGB image corresponding to the lesion with psoriasis, $\sigma(\delta)$ is a translation isometry over the window, O is the order of the set, and $r = (i, j)$ is the vector position of a pixel.

In practice, for each single image, a three band synthetical image of 256 by 256 pixels was constructed in the following way: for each single band, the co-occurrence matrix of the region indicating the single lesion included was constructed and normalized. The histogram of the pixels belonging to the area of the thematic map indicating lesion was previously equalized. Logarithms were applied to the normalized co-occurrence matrix values. This was done so, in order to make the histogram look more Gaussian. After applying logarithms the shape of the histogram is slightly modified. After that, the values were again normalized. Cell locations in the co-occurrence matrix were associated to pixel locations of a new synthetical band, and the probability of occurrence of a given pair of pixel values was the pixel value in the new synthetical band. See Equation 3.11:

$$y[r, c, b] = \frac{x[r, c, b]}{\sum_r \sum_c x[r, c, b]} \tag{3.11}$$

and

$$z[r, c, b] = \frac{(log(y[r, c, b] * map[r, c, 1] + (1 - map[r, c, b])) + \beta[b]) * map[r, c]}{\sum_{k,l}(z[k, l, b] * map[k, l, b])} \tag{3.12}$$

where $map[r, c, b]$ is 1 if $y[r, c, b] > 0$ and 0 in the opposite case, and $\beta[b] = max(|z[r, c, b]|)$, $\forall r, c$. In Figure 3.10, the image regions from which the co-occurrence matrixes shown in the upper part of Figure 3.11 were computed are shown.

3.4.2 The M.A.D. Transform to pairs of Co-occurrence Matrixes

The set of 175 captures obtained at Gentofte Hospital allowed the construction of 650 pairs of captures with the same time increment of one week. Two sets of 650

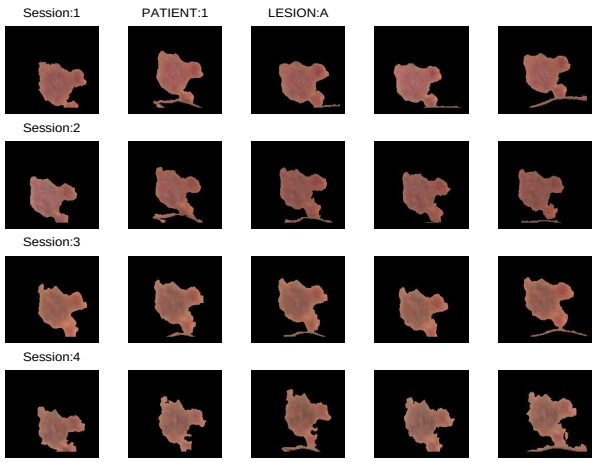


Figure 3.10: Regions indicating the selected lesion A of patient 1 in the illumination-corrected images.

pairs of images were used to apply the present scheme: the first set was given by the original images (See Appendix A.1); the second set, by the illumination-corrected images (See Appendix B in [35]). For each pair of images of the data-set, the union of the maps indicating positive values each co-occurrence matrix was generated. The M.A.D. Transform was only applied to the pixels included in that map. Tables 3.2 and 3.3 show the results obtained for the subsets of the original images indicating lesions. Tables 3.4 and 3.5 show the results obtained for the subsets of illumination corrected images indicating lesion [4].

Each single cell value in Tables 3.2 and 3.4 is the average absolute correlation value between the $b-th$ color bands and the $m-th$ MAD components of all pairs of images of the $l-th$ lesion of the $p-th$ patient. It is computed in the following way:

$$E[\rho_{p,l,b,m}] = \frac{1}{(n_s - \delta t)n_c^2} \sum_{s=1}^{n_s-\delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} a_{p,l,s,i,\delta_t,j,b,m} \tag{3.13}$$

where

$$a_{p,l,s,i,\delta_t,j,b,m} = \frac{1}{2}(|\rho[X_{p,l,s,i,b}, MAD_{p,l,s,i,\delta t,j,b}, m]| + |\rho[X_{p,l,s+\delta t,j,b}, MAD_{p,l,s,i,\delta t,j,b}, m]|)$$

[4]The Tables presented in this section contain values corresponding to what was considered lesion data. Note that it is a combination of results shown in [37].

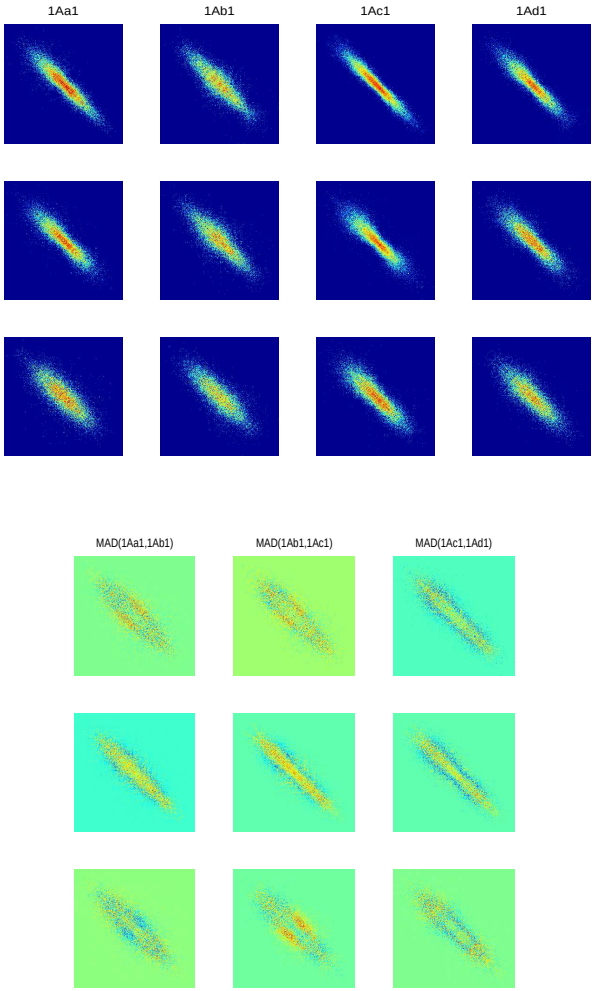


Figure 3.11: Co-occurrence matrixes and M.A.D. Components for the case $(1, A)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 1, Lesion A); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

(3.14)

and ρ is the correlation coefficient, δt is the time increment (in this case, δt

is one week). The indexes i and j indicate the capture number within a session. $MAD_{p,l,s,i,\delta t,j,b,m}$ equals $MAD[X_{p,l,s,i,b}, X_{p,l,s+\delta t,j,b}, m]$.

Each single cell value in Tables 3.3 and 3.5 is the standard deviation of the averaged absolute correlation values between the normalized co-occurrence matrix of the b-th color bands and the m-th MAD components of all pairs of images of the l-th lesion of the p-th patient. It is computed in the following way:

$$S[\rho_{p,l,b,m}] = \sqrt{\frac{1}{(n_s - \delta t)n_c^2} \sum_{s=1}^{n_s-\delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} (a^2_{p,l,s,i,\delta_t,j,b,m}) - E[\rho_{p,l,b,m}]^2} \qquad (3.15)$$

where the symbols have the same meaning as before.

Table 3.2: Average Absolute Correlation Values per Lesion of *RGB* lesion data with the first M.A.D. Component

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.228414	0.241037	0.249670
(1,B)	0.087517	0.291845	0.257182
(1,C)	0.021994	0.444337	0.468673
(2,A)	0.117731	0.151841	0.187481
(2,B)	0.021226	0.524634	0.387158
(2,C)	0.111509	0.133244	0.182223
(3,A)	0.038838	0.195560	0.163730
(3,B)	0.032641	0.206640	0.203355
(3,C)	0.043867	0.228722	0.198393

Table 3.3: Standard Deviation of the Absolute Correlation Values per Lesion of the *RGB* lesion data with the first M.A.D. Component

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.075189	0.051723	0.120361
(1,B)	0.050564	0.057280	0.092927
(1,C)	0.010095	0.050346	0.045895
(2,A)	0.059834	0.030295	0.014737
(2,B)	0.012165	0.078557	0.100383
(2,C)	0.070107	0.042084	0.076832
(3,A)	0.012120	0.010609	0.013386
(3,B)	0.017749	0.009709	0.013645
(3,C)	0.029494	0.014647	0.023083

Table 3.4: Average Absolute Correlation Values per Lesion of the illumination-corrected *RGB* lesion data with the first M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.156864	0.251143	0.306729
(1,B)	0.134858	0.280693	0.287114
(1,C)	0.020100	0.438892	0.477563
(2,A)	0.099910	0.214296	0.151165
(2,B)	0.024732	0.531542	0.375503
(2,C)	0.070576	0.187925	0.195604
(3,A)	0.113855	0.222418	0.142154
(3,B)	0.024691	0.208501	0.202745
(3,C)	0.090282	0.221817	0.190266

Table 3.5: Standard Deviation of the Absolute Correlation Values per Lesion of the illumination-corrected *RGB* lesion data with the first M.A.D.

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.073514	0.043622	0.104251
(1,B)	0.084878	0.075100	0.108610
(1,C)	0.012400	0.049313	0.043281
(2,A)	0.032620	0.019728	0.024057
(2,B)	0.011909	0.085662	0.100957
(2,C)	0.045544	0.024256	0.043097
(3,A)	0.027636	0.008405	0.017570
(3,B)	0.021223	0.014558	0.014621
(3,C)	0.033461	0.023677	0.024106

For the pilot dataset, the most significant contribution in the texture change of lesions with psoriasis is given by the green and blue bands.

3.4.3 Back-projection to the original data

A back-projection scheme from the absolute first M.A.D. component to the original data was proposed in [37]. The output of this scheme for the pair formed by the first images of the third and fourth session of the case $(1, A)$ can be seen in Figure 3.12. The illumination corrected images were used. Note that for each image of the pair, the segmented regions showing most and least textural change are approximately correspond to each other. This gives evidence that it is possible to back-project the M.A.D. Transform output to the original data, such that corresponding regions showing most and least textural change can be delineated. However, in order to be able to generalize the approach, the optimization of each single stage of the whole procedure should be considered.

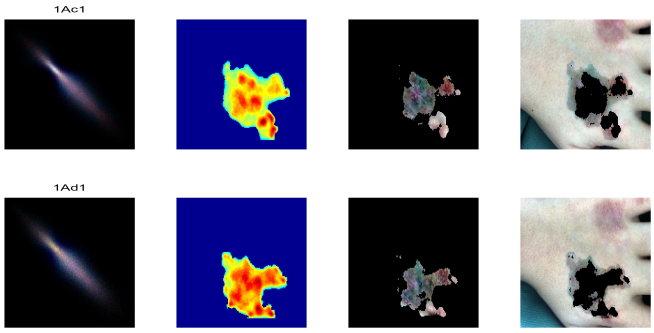


Figure 3.12: From the M.A.D. Transform to the Original Data.
From left to right: co-occurrence matrix, back-projection synthetical map, regions with more textural change and regions with less textural change. From top to bottom: the first images of the third and fourth session of the case $(1, A)$.

3.5 Novelty Detection in Registered Patterns

Procrustes Analysis is a useful approach for aligning two landmark registered shapes [12], [18]. However, when these shapes are not landmark registered, the approach can be extended setting some constraints. For instance, regarding the topology given by the connectivity between the landmarks. Previous work done in registration of lesions in skin images can be found in [47]. In this case, an algorithm considers each lesion as a point. The input to the scheme are multiple lesions skin images. Two initial matches, called the "baseline", are required in order to find the remaining correspondent points.

It also might be interesting to include textural information during the registration process. Previous work that combines shape with textural information was published in [51].

On the other hand, dermatologists experienced that lesions with psoriasis do not significantly change their shape and size along the time: instead the change happens within the lesions. In the present work, it is assumed that the shape of the object is invariant to scale and that for each pattern, the pixels locations are landmarks that have to be matched. It is assumed that the objects of largest size in the thematic map indicating lesions are corresponding, but not aligned. However, they need not have the same size. Therefore, the most convenient set of pixels that will define the set of landmarks to be registered has to be found through an alignment process.

Afterwards, the landmarks can be registered.

This section is composed by the following subsections. A combined alignment and registration scheme is first proposed. Afterwards, a subsection about change detection in pairs of registered psoriasis lesion patterns within and between weekly sessions is included. Color band contribution to the first M.A.D. component is analyzed.

3.5.1 A Combined Alignment and Registration Scheme

An algorithm that finds the n largest size objects belonging to each one of the thematic maps obtained in the segmentation stage was developed. Misclassified pixels forming holes within regions indicating lesions are assigned to the class lesion. Pairs of equally scaled corresponding objects are the input of the present scheme.

The combined alignment and registration scheme is composed of two stages: the first alignment and the registration stage. In the first stage, called first alignment, global translation and rotation effects of two assumed equally scaled patterns are removed. The output of the first alignment stage is used by an internal local registration scheme, which assumes that the local rotation effects in patterns internal region are not significant. Thus, only local translation effects in the interior of the patterns are removed.

Following [30], the scheme could be classified as of spatio-temporal dimensionality: lesions are registered within and between sessions. The nature of registration basis is intrinsic, because "it is based on image information as generated by the patient": "no foreign objects were introduced into the imaged space". It is segmentation-based with a rigid model, because "anatomically the same structures are extracted from both images to be registered and used as sole input for the alignment procedure". In the first alignment, the nature of the transformation is rigid: "only translations and rotations are allowed". In the second stage of the scheme, the nature of the transformation is affine: "parallel lines are mapped onto parallel lines". The domain of the transformation is global: "it applies to the entire image" (lesion). The level of interaction is semi-automatic: "the user needs to initialize the algorithm". The scheme is monomodal: "the images to be registered belong to the same modality" (RGB). It is an intra-subject registration scheme: "all the images involved in the registration task are acquired of a single patient". The object to register is a psoriasis lesion.

The registration scheme is evaluated afterwards. Results for our example case are also shown.

The Initial Alignment

The algorithm that carries out the initial pattern alignment is detailed here. For each given pattern S and a given reference pattern R, each one of the rotations $Rot(R, \hat{\theta})$ is centered on the position of each pixel of S, and the angle $\hat{\theta}_{opt}$ and location $[i, j] \in L$ for which the difference of volume is minimal is found. L is the lattice on which the images including the patterns R and S are defined. In practice, pattern objects are rotated in a range varying from -45 to 45 degrees. For each iteration, this range is divided into five intervals. Neighboring intervals containing the minimum of an evaluating function define the range for the next iteration. This minimum is computed using the Extreme Value Detection Algorithm (E.V.D.A.) detailed in the next paragraph. The algorithm stops when the minimum between two iterations differ by less than a user-provided threshold β.

The Extreme Value Detection Algorithm (E.V.D.A.) is an iterative process that assumes that the expected minimum (extreme value) of the distance function between two objects is near the center of masses of the domains of those objects. In the first iteration, a mapping template centered on the center of mass of the domain of one of the objects is used to define the subset of the domain of the distance function for which their values are going to be estimated. This mapping template includes many points of the neighborhood of the center of mass, and only few that are far away of the center. The extreme value of the calculated values of the distance function is obtained, and the same mapping template is centered on that first guess of the extreme value. This step is repeated until the guessed extreme value does not change anymore.

The output of the first alignment stage is back-projected to the original data and a new set of images containing the aligned shapes is obtained (See Figure 3.13). These images can afterwards be registered and realigned. The most important result produced by this stage is the specification of the number of landmarks to be registered in a next stage, this means, that the area of the objects is finally forced to be constant. Since some lesion patterns are occluded in the original images, it is left up to the user to decide if the number of landmarks is given by the area of the intersection of the whole set or of a subset of the aligned domains (objects in thematic maps indicating lesion).

For a given lesion, the expected correlation value of a randomly selected aligned lesion pattern with any other remaining aligned pattern of the same set (case) is given in Table 3.6. Details of the computation of each single cell value can be found in [33].

Registration

After the first alignment it is assumed that the aligned patterns have the same size and shape. Until now, for a given lesion, elements of pairs of aligned patterns located at the same position were considered corresponding points. This hypothesis is, from

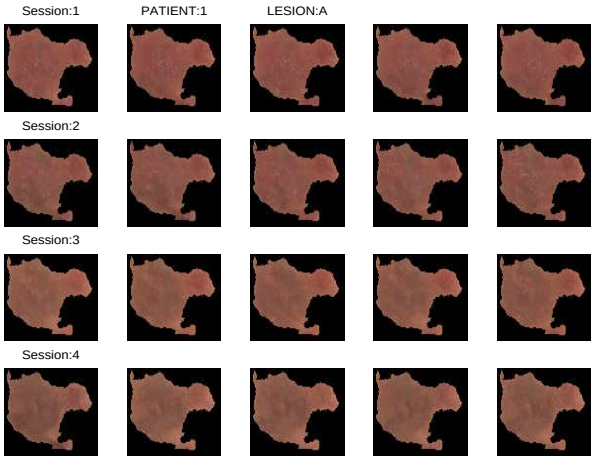


Figure 3.13: First alignment of lesion patterns for the case (Patient 1, Lesion A).

now, more relaxed: we assume that there is some error produced by small internal displacements. The displacements are assumed to be small translations.

The algorithm that carries out the registration is detailed here. In general, in order to define the correspondence, two m-dimensional sets of landmarks T and Y, called the source and the target respectively have to be specified [12]. In the present scheme, $m = 1$ and, for each given image taken during a given session, the source T is given by the locations of pixels belonging to the aligned pattern corresponding to the first image of the first session, and the target Y should be given by their correspondent locations of pixels, but it is unknown. In order to reduce the computational

Table 3.6: Average correlation and standard deviation of a randomly selected aligned lesion pattern with the remaining aligned lesion patterns of the same patient and lesion, per color band, patient and lesion.

(Patient, Lesion	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B	$\mu_{\|B-G\|}$	$\sigma_{\|B-G\|}$
(1,A)	0.6918	0.0468	0.5770	0.0538	0.3679	0.0595	0.6941	0.0323
(1,B)	0.6696	0.0932	0.5241	0.0944	0.4413	0.0843	0.4517	0.1305
(1,C)	0.2534	0.0806	0.1651	0.0642	0.1138	0.0561	0.4211	0.0585
(2,A)	0.5162	0.0606	0.3561	0.0589	0.5720	0.0588	0.6486	0.0367
(2,B)	0.6681	0.0227	0.5736	0.0331	0.7260	0.0219	0.4854	0.0432
(3,C)	0.2550	0.0288	0.0757	0.0350	0.1289	0.0202	0.3524	0.0672

Table 3.7: Average correlation and standard deviation per color band, patient and lesion of a randomly selected registered lesion pattern with the remaining patterns of the same patient and lesion.

(Patient, Lesion	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B	$\mu_{\lvert B-G\rvert}$	$\sigma_{\lvert B-G\rvert}$
(1,A)	0.6853	0.0447	0.5689	0.0505	0.3666	0.0586	0.6728	0.0345
(1,B)	0.6819	0.0790	0.5380	0.0814	0.4608	0.0743	0.4486	0.1209
(1,C)	0.2915	0.0841	0.1993	0.0720	0.1361	0.0586	0.4325	0.0586
(2,A)	0.5781	0.0573	0.4222	0.0553	0.6204	0.0533	0.6499	0.0406
(2,B)	0.6643	0.0258	0.5720	0.0394	0.7147	0.0252	0.4642	0.0369
(3,C)	0.2865	0.0382	0.0999	0.0331	0.1476	0.0203	0.3675	0.0604

time, for each pattern, only a subset of T and Y is considered at first. Pixels placed on a grid composed of square cells with a user-defined number of points form the subset of T for which their corresponding points belonging to the given lesion image will be found. The correspondence between the source subset and the target subset is established as follows. For each pixel belonging to the grid, a user-defined circular neighborhood is considered. At the same location in the second image, another user-defined circular neighborhood[5] - larger than the previous one[6]- is taken into account. The difference of volume values dV of all[7] the translations of the first window over the second window are considered. The position of the pixels belonging to the second image whose neighborhood has the minimal difference of volume compared to the window of the reference pattern is saved into a new grid. This is done for each image of a given session.

In each iteration, an average grid is computed. The set of distance vectors from the grid of the reference pattern image to the average grid is used to translate all the grids to a new position. This process is repeated for a user-provided number of iterations. Then, to generate the whole target set Y for each single image, thin plate splines are applied to the rest of the values of the source T and the rest of the new lesion image can be generated. At the end of this stage, it is expected that the pattern correlation within and between sessions increases. Also, when comparing two registered patterns belonging to the same session, their difference should be randomly distributed Gaussian noise.

Table 3.7 was generated in correspondence with Table 3.6, but using the correlation values of registration outputs.

In Table 3.8, each single cell value is the average per color band of the correlation

[5]These window sizes are established by visual inspection of the image containing the error between pairs of images taken during the same session.

[6]Computed as a function of the spacing between the points on the grid

[7]In practice, only the points defined by the mapping template used in E.C.V.A. were evaluated

Table 3.8: Average correlation values within sessions per color band of the outputs of the first and second stage of the combined alignment and registration scheme.

Stage	RED	GREEN	BLUE	|B-G|
Initial Alignment	0.7288	0.6377	0.6432	0.7040
Registration	0.7412	0.6507	0.6537	0.7412

Table 3.9: Average correlation values within and between sessions per color band of the outputs of the first and second stage of the combined alignment and registration scheme.

Stage	RED	GREEN	BLUE	|B-G|
Initial alignment	0.5938	0.4935	0.5001	0.5780
Registration	0.6114	0.5021	0.5043	0.5741

values within sessions (i.e. rows (a,a), (b,b), (c,c) and (d,d) in Tables 10 and 14 of [33]). In Table 3.9, each single cell value is the average per color band of all the correlation values within and between sessions (this means, from rows (a,a), (a,b) ...to (c,d), (d,d) in Tables 10 and 14 of [33]). As it can be seen, the correlation values increased. However, compared to the output of the initial alignment stage, the improvement done by the registration scheme proposed is not significant.

Note that in Figure 3.14, after the initial alignment for the case $(1,A)$, the kind of structure that the M.A.D. Transform detects in the lesion center tends to disappear after registration. However, a small structure appears after registration on the top left side of the lesion.

3.5.2 The M.A.D. Transform applied to pairs of Registered Patterns

The contribution of each single color band in showing the lesion change in a period of one week is analyzed here. Results are compared to the original image data. Lesion changes between sessions are also described using the absolute value of the first M.A.D. component. For our example case $(1,A)$, it can be seen in Figure 3.15 that most of the change from the first to the second week occurs in the central part of the lesion. Afterwards, from the second to the third week, it spreads out. From the third to the fourth week, most of the change occurs near the borders.

The set of 108 registered images produced by the registration scheme (see Appendix C in [33]) allowed the construction of 377 pairs[8] of registered lesions with the same time increment of one week between them. For each image of a given patient

[8]For each lesion the number of pairs is $(1,A) = 25 \times 3$; $(1,B) = 9 \times 3$; $(1,C) = 25 \times 2$; $(2,A) = 25 \times 3$; $(2,B) = 25 \times 3$ and $(2,C) = 25 \times 3$ respectively.

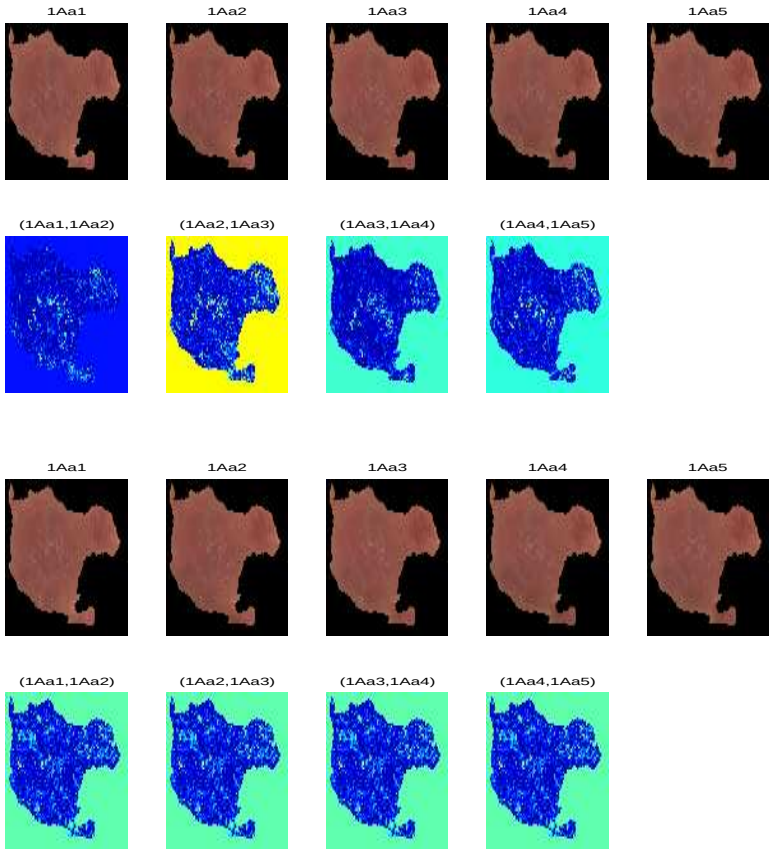


Figure 3.14: Images of the first session of the case $(1, A)$ before (top) and after (bottom) registration, and respective first M.A.D. components

and lesion, each one of the images of the next session were associated. For each pair of synthetic images constructed in the way mentioned, the M.A.D. transform was applied [24]. The M.A.D. Transform was only applied to the pixels belonging to the lesion. Tables 3.10 and 3.11 show the results obtained for the sets of registered images. Each single cell value in Table 3.10 was computed using Equation 3.13. Each single cell value in Table 3.11 was computed using Equation 3.15

Excluding from Table 3.10 the case for which the registered lesion patterns are covered by hair (Patient 2, Lesion B) (see [33] for details), we can see for the pilot data set, most of the contribution to the first M.A.D. component is given by either the green or the blue band. These results can be verified, by visual inspection of the

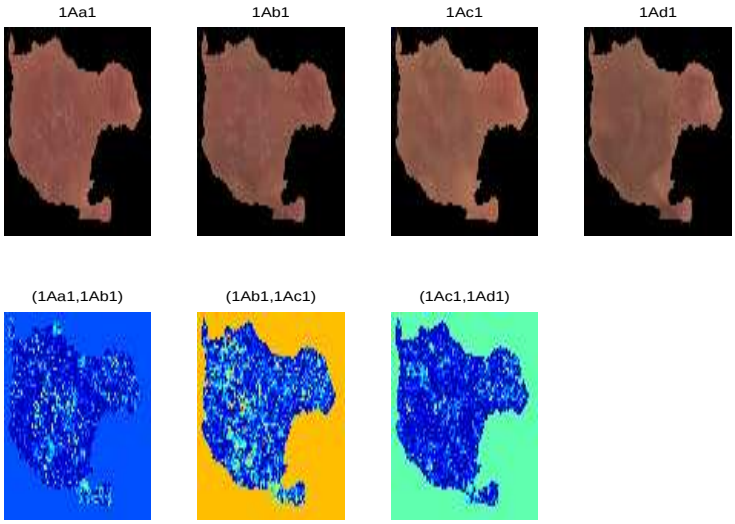


Figure 3.15: Registered Images for the case $(1, A)$ and the First M.A.D. components per week.

registered lesions image patterns decomposed in its RGB bands (See Figure 3.16).

Table 3.10: Average Absolute Correlation Values per Lesion of RGB registered lesion data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.239941	0.312891	0.479555
(1,B)	0.194602	0.342770	0.404496
(1,C)	0.502525	0.582041	0.602087
(2,A)	0.273979	0.444085	0.298466
(2,B)	0.232567	0.325108	0.154075
(3,C)	0.399863	0.536893	0.529572

Table 3.11: Standard Deviation of the Absolute Correlation Values per Lesion of the RGB registered lesion data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.082760	0.088492	0.076644
(1,B)	0.062977	0.088879	0.101755
(1,C)	0.071854	0.051745	0.044919
(2,A)	0.145003	0.130561	0.081781
(2,B)	0.043526	0.078260	0.060818
(3,C)	0.120938	0.129143	0.099749

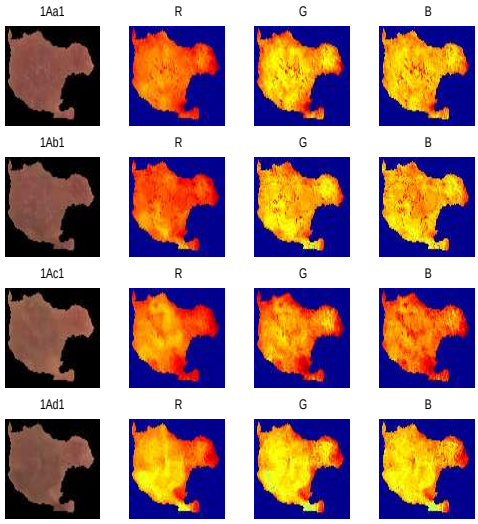


Figure 3.16: RGB bands of registered images of the case $(1, A)$

Chapter 4

Conclusion

In this thesis the novelty detection problem was focused from three different points of view:

- In the first chapter, it was treated as a contextual classification problem. Given a prototype pixel, the maximal set centered on it for which it is true that all the pixels belong to the same class as the center one was estimated.
- Afterwards, an attempt to avoid the lesion pattern registration was made. Textural changes of lesion patterns were studied through the application of the M.A.D. transform to the co-occurrence matrix of lesion patterns.
- Finally, an alignment and registration scheme was developed. Registered lesion pattern changes were studied by applying again the M.A.D. transform.

The set of dermatological images was composed by images of malignant melanomas and images of psoriasis lesions.

Regard the first item, it was assumed that pixels belonging to the neighborhood the prototype pixel belong to the same class as this one. An evaluating function was minimized over the line of maximal redundancy of the weights. This function was the estimated variance of the estimated means within groups. The minimum indicated the maximal set for which it is true that all the pixels belong to the same class as the center one. The scheme was tested with synthetical and real examples. It was shown to be a suitable approach for the generation of an initial training set from a prototype pixel.

A simple model that relates separability between classes with an optimal circular window size for classification was also proposed. It was applied to the output of a developed region growing algorithm that increases the sample size after the initial training set is generated. It was tested with synthetical and real examples. It was also used with the output of an Expectation-Maximization Algorithm. For the set of psoriasis images the results were satisfactory.

With regard to the psoriasis image data set, it has been observed that there is more contrast between the lesion and the rest in the third component. A rough estimation of the third component of the original images can be done by subtracting the blue from the green band. The classification of the captures of a given session, given the parameters of the first image of the first session has shown to have an average Kappa κ error superior to 0.8 in almost all the cases. It has also been shown that the images have an illumination problem that is not irrelevant. An illumination correction scheme that uses the image data has been proposed. By visual appreciation, it can be seen that the classification output improved significantly after illumination correction.

The application of the Multivariate Alteration Detection Transform to the normalized co-occurrence matrix of psoriasis lesions patterns is a novel approach for detecting texture changes that avoids the lesion patterns registration. Results indicate that the most significant contribution in the texture change of pairs of psoriasis lesions patterns is given -for the pilot dataset- by the green and blue bands. Evidence was given that it is possible to back-project the M.A.D. Transform output to the original data, such that corresponding regions showing most and least textural change can be delineated. However, in order to be able to generalize the approach, the optimization of each single stage of the whole procedure has to be considered.

An algorithm that iteratively approximates the extreme value of a smooth function using a known retinal mapping model was proposed. It was applied to the first alignment of lesion patterns and has shown to produce satisfactory results, unless the size of the data sets to be compared present a large variability.

Two lesion pattern registration approaches have been proposed. In both cases, the normalized pixel values of each lesion pattern were considered as the textural descriptor of the pattern. In spite of increasing both the pattern correlation within and between sessions, the improvement of the second stage of the proposed scheme was not significant: the average correlation values within and between sessions per color band of the first and second stage outputs of the combined alignment and registration scheme increased less than 2 percent. The average correlation values within sessions per color band of the first and second stage outputs of the combined alignment and registration scheme with better performance increased for the red band, from 0.7288 (after the first alignment) to 0.7412 (after the second stage). Basically, it can be mentioned that the registration of aligned patterns within and between sessions and

the selection of the measure for registration evaluation are clearly not easy problems.

Visual assessment of the results indicate, at least for the small set of lesions analyzed, that in the beginning, most of the change occurs into the lesion, while it spreads out after a week it spreads out and is distributed on the whole lesion.

Regarding the contribution of each single color band to the first M.A.D. component, using unregistered and registered psoriasis lesions patterns, it has been shown that, in general, it is mostly the green or the blue and to a lesser extent the red band which contributes most in showing the change.

4.1 Published Papers

Papers published during my Ph.D. Studies are included here. This part is organized as follows. Each single paper has a cover page specifying the list of authors and the conference or journal in which the work was published.

In spite of being redundant, but in order to have a general overview of the content of this part, the list of the publications included follows.

- An Initial Training Set Generation Scheme (S.C.I.A. 2001)
- Mean Shift Detection using Active Learning in Dermatological Images (M.I. 2002)
- Iterative Extended Mean Shift Algorithm (C.V.P.R.I.P. 2002)
- A contextual classifier that only requires one prototype pixel for each class. (I.E.E.E. Transactions on Nuclear Science 2002)
- A Supervised Contextual Classifier Based on a Region-Growth Algorithm (Computers & Geosciences 2002).

4.1.1 An Initial Training Set Generation Scheme

Authors: Maletti, G., Ersbøll, B., Conradsen K. and Lira, J.
Conference: 11^{th} Scandinavian Conferences on Image Analysis (S.C.I.A. 2001)
Place: Bergen, Norway
Date: 11-14 June 2001
Pages: 706 – 714
Reference: [41]

AN INITIAL TRAINING SET GENERATION SCHEME

Gabriela Maletti (1), *Bjarne Ersbøll* (1), *Knut Conradsen* (1) *and Jorge Lira* (2)

(1) Section for Image Analysis, Department of Mathematical Modelling
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
gmm@imm.dtu.dk, be@imm.dtu.dk, kc@imm.dtu.dk
(2) Institute for Geophysics, National Autonomous University of Mexico
04510 Ciudad Universitaria, Mexico
lira@tonatiuh.igeofcu.unam.mx

ABSTRACT

Supervised classifiers employ a-priori information of each determined class. This is usually obtained by means of training sets interactively generated or with semi-automatic schemes that still depend on some user input. The purpose of this work is to generate an initial statistically valid sample of a class starting from only one single prototype pixel. This is done by means of a new mapping model that uses redundant contextual information. The initial sample generated is the maximum size neighborhood centered on the prototype pixel, which contains pixels that belong to the same class as the center one. This is defined by a window of optimum size obtained by minimizing the variance of a learning variable over the line of maximal redundancy of the mapping model. The minimum of the function studied corresponds to the estimation of the point in which a new class emerges. Additionally, a supervised classifier was designed and implemented as an application of the model proposed.

1. INTRODUCTION

Supervised classifiers employ a-priori information of each determined class. This is usually obtained by means of training sets defined by an interactive procedure on the image. The quality of the classification can be improved with the implementation of semi-automatic training and validation set generation schemes [6][12]. However, these schemes still require some user-input for the estimation of the parameters of the classes eg. thresholds, window sizes, or require information about an initially assumed correlation structure of the class. It could be an advantage to develop schemes that minimize the amount of information the user is required to provide.

The aim of the present work is to generate an initial statistically valid sample of a class starting from a single prototype pixel. This initial sample could be used for the estimation of parameters of a region growing algorithm, for example. The context of the pixel will be taken into account and, therefore, an initial window of optimum size will be defined. From a semantic point of view [16], a window [7] defines a subset of an entity in which the instances are spatially connected. In order to find an ideal window size, Hodgson [9] suggested the use of a conceptual model of the human perceptual and cognitive interpretation processes.

In the case of visual perception, "human eyes move and successively fixate at the most informative parts of the image which therefore are processed with the highest resolution" [15]. This resolution can be interpreted as the maximum size of a window that contains pixels with the same feature.

With respect to the cognitive process, it is well known that "the learning process can be treated as a statistical dynamical problem" [8]. It has typical characteristics [17] like:

- instability: "It is impossible to detect structure in data until a critical number of examples have been presented" [18].
- fluctuations: underlying structure in the data can be recognized [1].
- critical slowing down: additional information provided is redundant [1].
- reorganization: a state of new order emerges and "the mean-field model exhibits an ergodic to non-ergodic transition where full ensemble averages are no longer equal to time averages"[11].

Obtaining the highest resolution implies detecting the emergence of a state of higher order. This will be done through the minimization of an optimal learning curve that depends of the fraction of examples [4]. The definition of the optimal learning curve will be based on the fact that there is a positive correlation between the redundancy and the efficiency in the minimization of a learning algorithm

[13]. Thus, a mapping model with overlapping discs in order to generate redundant sets will be defined. The optimum size of the neighborhood of a given prototype pixel will be obtained at the minimum of an energy function evaluated over the line of maximum redundancy of the mapping model. This happens when the presence of a new class is detected.

2. WINDOW SIZE OPTIMIZATION

Let us introduce some definitions and notations to be used in the present work.

Let a specific image with non-correlated bands

$$X = \{x[r, c]\} \tag{1}$$

be defined over the given domain L:

$$L = \{[r, c]|1 \leq r \leq max_r, 1 \leq c \leq max_c\} \tag{2}$$

We suppose that X is the realization of a two-dimensional random field X -notice that while X (sans serif) is a set of variables, X is a set of values of those variables-; X (italic) is hierarchically defined in terms of the realization

$$Z = \{z[r, c]\} \tag{3}$$

of an underlying random field Z such that: Z represents the partition of the domain in K regions of different types; each $z[r, c]$ is a value of the set of labels

$$B = \{b_1, b_2, ..., b_K\} \tag{4}$$

where $z[r, c] = b_k$ indicates that the coordinate $[r, c]$ of a given pixel belongs to the region k. In each region of type k the presence of a realization of a discrete and homogeneous random field H_k with a probability density function $P(H_k)$is assumed. Then

$$\forall b_k \in B : x[r, c] = h_k \in H_k \; \forall [r, c]|z[r, c] = b_k \tag{5}$$

Let the probability density functions $P(H_k)$ be unique features – this doesn't imply that the domains of the random variables H_k are disjoint –.

Definition 1: Maximal Disc and Optimum Window

Let Z, in turn, be hierarchically defined in terms of the realization

$$W_o = \{w_o[i, j]\} \tag{6}$$

of an underlying random field W_o such that W_o represents the set of radii of H_k homogeneous maximal discs $\mathsf{D}_{w_o}[r, c]$. From a syntactic point of view [16], a window defines a realization of a disc. The set

$$D_w[r, c] = \{d_w[r - i, c - j]|0 \leq i^2 + j^2 \leq w^2\} \tag{7}$$

is a realization of $\mathsf{D}_w[r, c]$.

The optimum window w_o for a given pixel x defines the maximal set D_{w_o} centered on the position $[r, c]$ of x for which is true that all the elements were generated with the same probability density function. This is an heuristic approach. See an example in Figure 1.

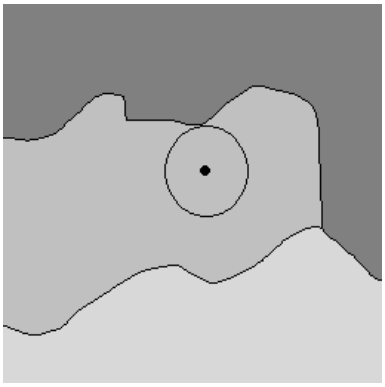


Fig. 1. The set indicated as a circle is the maximal disc that can be centered on the marked point for which is true that all the pixels belong to the same class as the center pixel.

Definition 2: Entropy and Redundancy of a Disc

The disc is a collection of random variables. Based on the Theory of Systems, that considers each system as a subsystem of another system, we now consider the disc itself as a random variable: the random variables that belong to the disc, are now values of this new random variable. Then, we can associate to that new random variable a probability density function or set of weights ω_w. The set of weighting functions used in the present work will be mentioned in section 2.1.1.

The entropy of a disc can be defined as:

$$E(\mathsf{D}_w) = -\sum_{i=1}^{n_w} \omega_w[i] log(\omega_w[i]) \tag{8}$$

where n_w is the number of elements or instances of the disc, $\omega_w[i]$ stands for the i-th instance of the disc, and

$$0 < \omega_w[i] < 1 \tag{9}$$

is the weight assigned to the i-th instance of the disc [1].

[1]From that is deduced that a disc can not be an unitary set, and if an element belongs to the disc, it has a probability greater than zero.

Following Møller [13], the redundancy of a disc of size n is defined as[2]:

$$R(\mathsf{D}_w) = 1 - \frac{E(\mathsf{D}_w)}{logn_w} \tag{11}$$

where – the same as before – n_w denotes the size of a disc given by the number of random variables that it is composed of.

The internal entropy to external entropy ratio $IEER$ for given discs of radii w and $(w-s)$ centered on the same position, where $s < w$, is defined as the quotient between the entropy of their intersection and the entropy of their difference using the set of weights ω_w:

$$IEER[w,s] = \frac{E(\mathsf{D}_w \cap \mathsf{D}_{w-s})}{E(\mathsf{D}_w - \mathsf{D}_{w-s})} \tag{12}$$

2.1. The mapping model

An active learning process with increasing window size along the iterations will be defined. It is expected that at some point, the beginning of the learning phase of reorganization is detected (eg. the class boundary is detected). This point will be associated to the optimum window size. The statistic to be learned is the mean of the maximum window, included in the image, that can be centered on a given prototype pixel. The way this statistic will be estimated is given by the mapping model that is built in the following two steps: the estimation of the mean into groups, and the estimation of the mean between groups.

Several mapping models have been suggested in the last years [19][3]. Models with overlapping contexts [16] are attractive because they contain some degree of redundancy [13]. In the present work a new mapping model based on the Uncertainty Principle and the Central Limit Theorem [14] is proposed. From the Heisenberg Uncertainty Principle we know that, if we want to minimize the uncertainty of the neighborhood of a pixel, the envelope of the neighborhood or the neighborhood itself has to be a Gauss function. However, in that case, for a given pixel our knowledge about its neighbors is minimal, in other words, our uncertainty is maximal. This means that, initially, we should give its neighbors the same weight: we should use an uniform model. The connection between an Uniform model and a Gauss model can be found in the Central Limit Theorem [14], which can be expressed as a property of convolutions. The convolution of a large number of positive functions is approximately a Gauss function. However, strictly speaking, the Central Limit Theorem only applies under certain general conditions and for an infinite number of random variables.

The details of the mapping model follow bellow.

2.1.1. Estimating the mean within the groups

In general, we define the estimation y of the mean of a realization D_w of a disc D_w as the projection [3] of the realization D_w over the weighting function ω_w:

$$y_w = \sum_i d_w[i]\omega_w[i] \in Y_w = \{< D_w, \omega_w >\} \tag{13}$$

where $d_w[i]$ is the i-th element of the realization D_w of the disc corresponding to the i-th position, $\omega_w[i]$ is the weight assigned to the i-th element of D_w corresponding to the same i-th position. Y_w is the realization of a random field Y_w.

Let the realization of a disc $\mathsf{D}_m[r,c]$ be composed by the estimations $y_v[r-i, c-j]$ of the means of n_m groups of size n_v using the same set of weights ω_v.

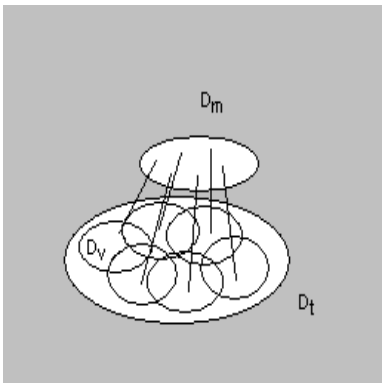


Fig. 2. The mapping model

2.1.2. Estimating the mean between the groups

The dynamics of the learning process will be followed looking at the behavior of the estimation of the mean $y_{v,m}$ of the realizations of discs $\mathsf{D}_{v,m}$. This will be the estimation of the mean between the n_m groups.

$$y_{v,m} =< D_{v,m}, \omega_m > \in Y_{v,m} \tag{14}$$

[2] Our aim is only specifying a trajectory. For this reason, in practical cases we used the following formula as a relative measure:

$$R(\mathsf{D}_w) = 1 - \frac{E(\mathsf{D}_w)}{N} \tag{10}$$

where N is a normalization factor, such that for each w constant N is the maximum value of $E(\mathsf{D}_w)$ computed.

[3] Scalar product is denoted by $< .,. >$

It can be proved that $\forall D_v \subset X \wedge D_{v,m} \subset Y_v \; \exists D_t \subset X \wedge \omega_t$ such that

$$y_{v,m} = \hat{\mathbb{m}}_{v,m} \simeq y_t \tag{15}$$

where

$$t = t(v,m) \simeq v + m \tag{16}$$

and ω_t results from a convolution[4]

$$\omega_t \simeq \omega_m * \omega_v \tag{17}$$

The number of elements in D_t is n_t.

Each $\hat{\mathbb{m}}$ is an estimation of the mean $\mathbb{m}$ of the realization of the disc of maximum size centered on a given position $[r, c]$ that is included in the image X.

$$\lim_{n_t \to \infty} \hat{\mathbb{m}}_{v,m} \simeq \mathbb{m} \tag{18}$$

2.2. The optimization process

The mean field $\mathsf{Y}_{v,m}$ by itself does not describe the fluctuations of each realization of $\mathsf{D}_{v,m}$. A measure like the learner variance [5] is a statistical approach for doing that job. However, let us introduce some previous comments about the redundancy in the mapping model.

2.2.1. The line of maximal redundancy

In general, we know that "the variance of the sampling distribution is inversely related to sample size" [10]. This can be observed in figure 3 for the case of the mean of windows of increasing size. This figure was generated using a Gauss distributed image of 512 x 512 pixels with a mean of 128 and standard deviation of 8. The means of discs with radius varying from 1 to 4 centered on each pixel were computed. The normalized histogram for each radius is the probability density function of the mean of discs with this radius. A similar behavior can be observed if the test is extended to other statistics like the variance, etc..

In our problem, this leads to maximize the size of the discs D_t.

$$\lim_{t \to \infty} e_t = |y_t - \mathbb{m}| = 0 \tag{19}$$

However, for each set of discs $\mathsf{D}_{v,m} \subset \mathsf{Y}_v$ associated to a disc $\mathsf{D}_t \subset \mathsf{X}$ of fixed radius t, if we maximize the size of the discs $\mathsf{D}_v \subset \mathsf{X}$, we minimize the size of the disc $\mathsf{D}_{v,m} \subset \mathsf{Y}_v$ and vice versa (See Equation 16). It has to be taken into account, that the estimation of the mean of D_t depends also on the weighting function ω_t. For each t an optimum 3-tuple (v, m, ω_t) should be found. An uniform weighting function ω_t is not a possible solution here.

[4] Equality holds for square windows.

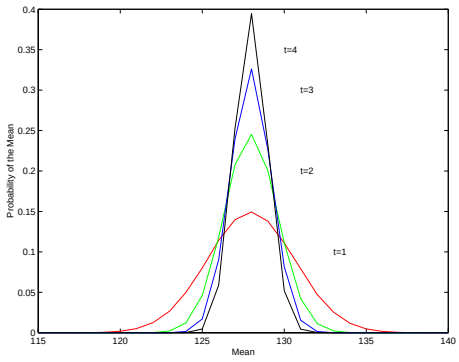


Fig. 3. Probability Density Function of the Mean of Realizations of Discs of radius $t = 1, 2, 3, 4$

The following weighting functions ω_t will be evaluated:

- uniform(v)-uniform(m)
- uniform(v)-gauss(m)
- gauss(v)-uniform(m)
- gauss(v)-gauss(m)

where the first element of each pair refers to ω_v and the second element to ω_m in Equation 17, and

- uniform(g): $\omega_g[i] = \frac{1}{n_g} \forall i$
- gauss(g): $\omega_g[d] = exp(-0.5(\frac{d^2}{\sigma 2})) \; \forall d^2 = i^2 + j^2$ and $\sigma = f(g)$. Let us fix $\sigma = g/3$.

As can be observed comparing Figures 4 and 5, for each constant radius t, the error in the estimation of the real mean is maximal when the redundancy is maximal. This happens no matter which combination of weighting function is used.

$\forall t$ constant

$$\lim_{t(v,m) \to t(v_o,m_o)} |y_{v,m} - \mathbb{m}| = max(e_t) \tag{20}$$

For each fixed disc D_t, there is an optimum pair (v_o, m_o) such that the error in the estimation of the mean of D_t is maximal.

We would like to minimize the error in the worst estimation of the mean and at the same time take advantage of the fact that the learning algorithm will work more efficiently in the minimization if the redundancy is increased [13]. Then for each fixed radius t we choose the pair (v_o, m_o) such that the redundancy is maximal.

$$(v_o, m_o) = max(R(\mathsf{D}_{v,m})) \forall t(v,m) constant \tag{21}$$

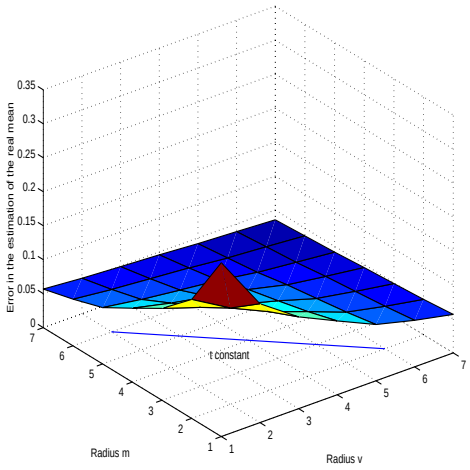


Fig. 4. Error in the estimation of the real mean using the uniform(v)-uniform(m) model in a Gauss distributed $N(\mu = 128, \sigma = 2^3)$ image of 96 x 96 pixels.

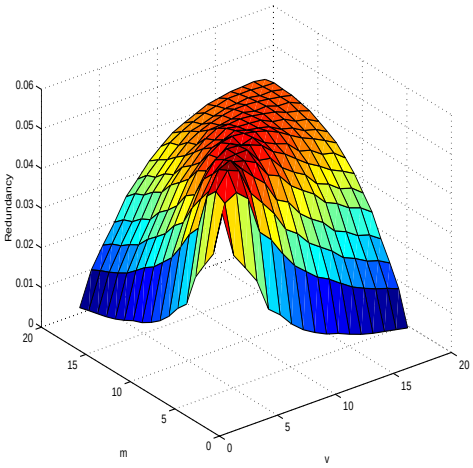


Fig. 5. Redundancy using the uniform(v)-uniform(m) model

The position of the line of maximum redundancy of the uniform(v)-gauss(m) weighting function is symmetric to the position of the line of the gauss(v)-uniform(m) weighting function with respect to the plane $v = m$, and the values of maximum redundancy are the same for each t constant.

The position of the line of maximum redundancy of the uniform(v)-uniform(m) weighting function is equal to the position of the line of the gauss(v)-gauss(m) weighting function, but the values of redundancy for each t constant of the second are greater than the values of redundancy of the first.

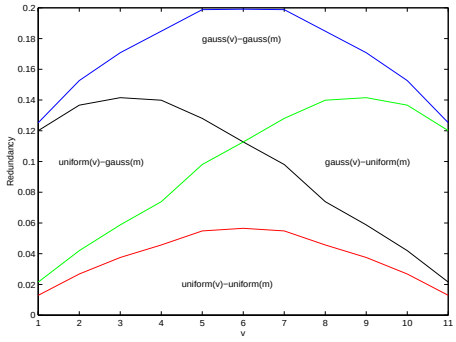


Fig. 6. Redundancy of a disc of fixed radius $t(v, m) = v + m = 12$ $\mathsf{D}_{t=12}$ using different weighting functions ω_t in the estimation of the mean of the realization of the disc.

2.2.2. The energy function

Let us assume that we have a realization D_{max} that changes from an internal class I to an external class E at radius to. Let the mean of the internal class $\mathtt{m}_I$ be different from the mean $\mathtt{m}_{(I \bigcup E)}$ of the union of the internal and the external class. The last one is the statistic to be learned. Let the variance of the internal class σ_I^2 be less than the variance of the union $\sigma^2_{(I \bigcup E)}$.

The emergence of a higher order state will be detected at the minimum to of the following energy function evaluated along the line of maximum redundancy:

$$\mathsf{Z}[t] =< D^2_{v_o,m_o}, \omega_{v_o,m_o} > -y^2_{v_o,m_o} \tag{22}$$

where each $d^2_{v_o,m_o}[i] \in D^2_{v,m}$ is the square of the element $d_{v_o,m_o}[i]$ of $D_{v,m}$ that is at the same position i.

Then

$$\hat{to} = min(\mathsf{Z}[t]) \in \hat{W}_o \tag{23}$$

where $\hat{W}_o$ is the set of estimations of the elements of W_o (see Equation 6), and the ratio

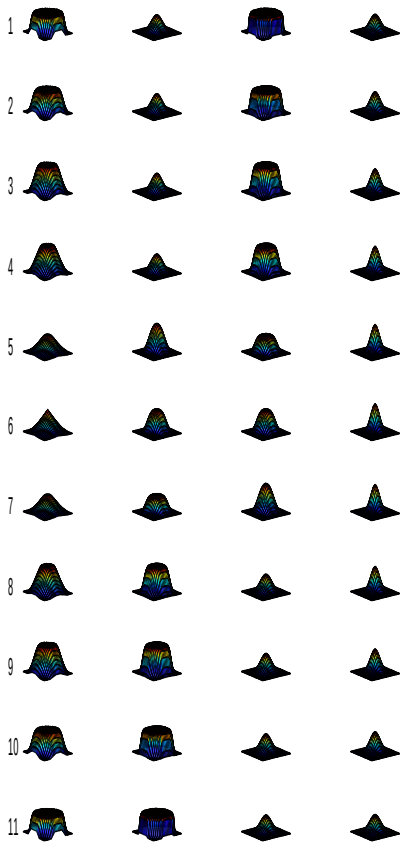
$$IEER[t, s] > 1 \text{ but } not >> 1 \tag{24}$$

holds $\forall t$.

For X p-dimensional we will minimize the trace of a matrix of Z.

The behavior of this energy function will be the following:

- In each iteration the y_t will be the estimation of the real mean of the union of the internal with the external class $\mathtt{m}_{(I \bigcup E)}$.



U(v)-U(m)	U(v)-G(m)	G(v)-U(m)	G(v)-G(m)

Fig. 7. Different weighting functions ω_t for a disc of radius 12 $\mathsf{D}_{t=12}$. From the left to the right, the weighting functions correspond to uniform(v)-uniform(m), uniform(v)-gauss(m), gauss(v)-uniform(m) and gauss(v)-gauss(m) respectively (cf. Figure 6). From the top to the bottom, the number of the row corresponds to the radius v.

- However, if we avoid adding instances of the external class to the sample, we are going to obtain a good estimation of the mean $\mathtt{m}_I$ of the internal class. This means that the variance of the learned mean of the internal class will be minimized.

- From a physics point of view we can think that the system tends to be singular and therefore more sensitive to perturbations[5]. If we add a small quantity of pixels that belongs to the external class this will introduce a perturbation to the system.

- Since "the chances of encountering a more extreme sample value are greater, the smaller the sample size" [10], we can expect that if the number of internal pixels is big enough, the probability of finding an extreme value will be low and all the variation of the learned variable will be introduced when we begin adding external pixels.

- Furthermore, the size of that variation will be increased because the variance of the union of the external and the internal class is greater than the variance of the internal class.

This can be observed for the four ω_t tried on an example that was generated by seeding a pixel in the center of a Rayleigh-distributed class[6]. The change of class happens at radius 32. In general, all the models with overlapping discs have an energy function that minimizes in the neighborhood of the border between the external and the internal class. (See Figure 8).

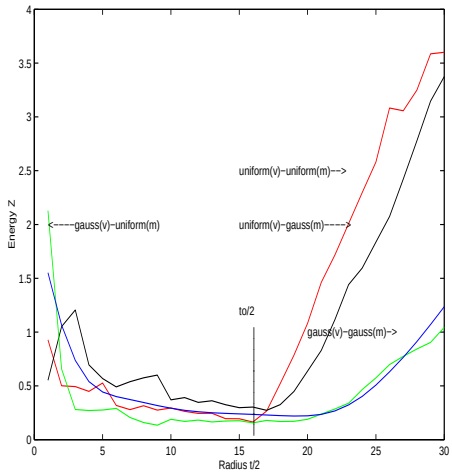


Fig. 8. Energy function using different weighting functions ω_t in the estimation of the mean

Note in figure 9 that the real mean $\mathtt{m}_I$ is estimated faster by the $y_{v,m}$ corresponding to the uniform(v)-uniform(m) model.

3. RESULTS AND DISCUSSION

The energy function proposed was evaluated for each of the four weighting functions ω_t using a set of synthetic images. From these results – and as an application – a supervised

[5]Further work will be done in that sense.

[6]We used this distribution in order to evaluate the performance of this model in non-Gauss distributed classes too.

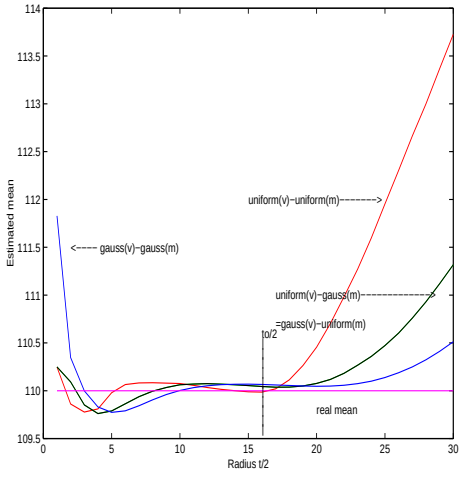


Fig. 9. The estimated mean $y_{v,m}$ using different weighting functions ω_t.

Table 1. Minimum, Average and Maximum Estimated Optimum Window Size using combinations of Uniform(U) and Gauss(G) weighting functions in the Mapping Model. *A set of 144 images with internal class with varying radius (to) from 2 to 18 was used. The difference between the mean of the internal class and the external class was 4 gray levels. The standard deviations of the classes varied from 1 to 4.*

to	U(v)-U(m)	U(v)-G(m)	G(v)-U(m)	G(v)-G(m)
2	2, 7,22	2, 2, 2	2,11,18	2, 2, 2
4	2, 2, 2	2, 2, 2	4, 4, 4	4, 4, 4
6	4, 4, 4	2, 2, 2	6, 6, 6	6, 6, 6
8	6, 8, 8	2, 4, 8	6, 7, 8	8, 8, 8
10	6, 8, 8	2, 8,10	6, 7,10	10,10,10
12	6, 8, 8	2,10,12	6, 7,12	12,12,12
14	6,11,12	2,11,12	6,11,14	14,14,16
16	12,14,14	2,11,12	6,14,16	14,16,18
18	12,14,16	2,11,12	6,17,18	14,18,20

classifier was designed and implemented. The output of this classifier is shown in the second part of this section for a real case.

- The synthetic images:
  Experiment: A set of pairs of Gauss-distributed images of 96 x 96 pixels with separations between means of four gray levels, and varying standard deviations from 1 to 4 was generated. In each pair, the first image was assigned to the internal class, and the second pair was assigned to the external class. With each one of these pairs, a new set of images of the same size was built centering a disc of varying radius from 2 to 18 in steps of 2 containing data of the internal class in the center of the image. The complement of that disc was filled with the data of the external class. In this way a total of 144 different images was generated. Since the optimum window size is known it could be compared with each one of the estimated optimum windows generated by the four ω_t used in this work.

  Looking at the previous table it seems that the gauss(v)-gauss(m) model has the best performance. However, if we observe the last rows, it starts generating window sizes greater than the expected ones. This is in spite of the fact that the internal entropy to external entropy ratio grows faster in that model than in the others. (See Figure 10 and note the difference in the scale of the y-axis.) For this reason, excluding the case in which the internal entropy to external entropy ratio is less than 1, we can conclude that the uniform(v)-uniform(m) weighting function has the best performance at detecting the emergence of a state of higher order. It should be taken into account, that if the ratio is much greater than 1 the emergence of a state of greater order will be detected with a delay, and wrong examples will be included in the initial sample generated.

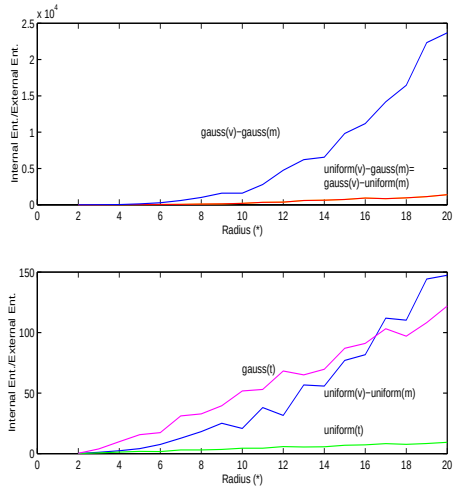


Fig. 10. Internal Entropy to External Entropy Ratio versus Radius of the Disc. (*)The x-axis corresponds to the radius $(t-1)/2$. The gauss(t) and uniform(t) are only included as a reference.

- A real example:
  As an application of the uniform(v)-uniform(m) model, a supervised classifier that requires only one pixel seed for each image class was designed and implemented. The seeds are manually selected and should be located within an homogeneous region. Let us

have K seeds. The algorithm consists of the following steps:

- Initial sample size: For each class k, a disc of maximum radius is centered on the corresponding seed $q_k[r, c]$.For each pixel seed a maximum possible disc still included in the image can be found. In order to treat all classes equally, all the discs are now restricted to have a maximum radius corresponding to the smallest of these. The optimum window size $t_k(v_k, m_k)$ is found for each class k using Equation 22. For each class k, the mean $\mathtt{m}_k^{is}$ and the standard deviation S_k^{is} are computed from its optimum window.
- Region growing algorithm: The thresholds for the region growing algorithm are defined for each class using the previous statistics. A pixel $p(r, s)$ will be assigned to the grown region corresponding to the k-th class if the difference between the mean of a window of size v_k centered on the pixel p and the mean $\mathtt{m}_k^{is}$ of the k-th class is less than a threshold ϵ. This is denoted in the following way:

$$p(r,s) \rightarrow k : |\mathtt{m}_{v_k}(r,s) - \mathtt{m}_k^{is}| < \epsilon \qquad (25)$$

  where

$$\epsilon = \gamma S_k^{is} \qquad (26)$$

  such that $\gamma \in \mathbf{R}^+$. For each class k the mean $\mathtt{m}_k^{gr}$ and the standard deviation S_k^{gr} are computed from its grown region.
- Final classification: A set of windows of sizes $v_1 ... v_K$ are centered on each pixel to be classified. The pixel p is assigned to the closest class.

$$p(r,s) \rightarrow k : min(|\mathtt{m}_{v_k}(r,s) - \mathtt{m}_k^{gr}|) \forall k \quad (27)$$

  An example of an image with a malignant melanoma is presented in Figure 11 and 12. The principal components of that image were used by the classifier.

4. CONCLUSIONS

Preliminary investigations have shown that mapping models with overlapping discs are useful for detecting the emergence of a state of higher order. An optimum window centered on a given pixel and associated to the disc of maximum radius that contains pixels of the same class as the center pixel, can be obtained at the minimum of an energy function that is only evaluated at the line of maximum redundancy of the mapping model. The weighting functions used in the mapping models are only initial suggestions that have a good performance until they are affected by the increasing size of the internal entropy to external entropy ratio. This could be improved by using variable increments

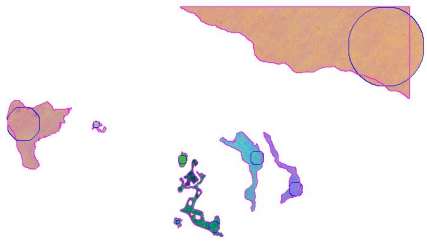


Fig. 11. The initial samples delineated with circles and the grown regions

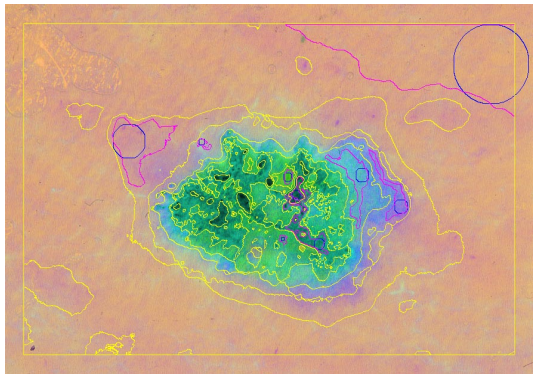


Fig. 12. The different steps in the classification scheme including the final classes generated

of the radius of the discs. It has been shown by introducing these results in the design and implementation of a supervised classifier that, -when the expected optimum window is not too big- the model proposed is a good tool for the generation of an initial statistically valid sample of the class to which the seeded pixel belongs. Further work will be done using the same idea of minimizing the learner variance over the line of maximum redundancy of the mapping model.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number STVF 56-00-0123) for supporting the present work. To The National Hospital of Denmark for providing the dermatological images used. To Associate Professor S. E. Quiñones Cisneros

for his disposition to make the concept of maximization of redundancy understandable.

5. REFERENCES

[1] M. Biehl and A. Mietzner, Statistical mechanics of unsupervised structure recognition, *J. Phys. A.: Math. Gen.*, Vol. 27, 1994, pp. 1885-1897.

[2] M. Bolduc and M. Levine, A Real-time Foveated Sensor with Overlapping Receptive Fields, *Real-Time Imaging*, Vol. 3, 1997, pp. 195-212.

[3] M. Bolduc and M. Levine, A Review of Biologically Motivated Space-Variant Data Reduction Models for Robotic Vision, *Computer Vision and Image Understanding*, Vol. 69, No. 2, 1998, pp. 170-184.

[4] A. Buhot and M. Gordon, Phase transitions in optimal unsupervised learning, *Physical Review E*, Vol. 57, No. 3, March 1998, pp. 3326-3333.

[5] D. Cohn, Z. Ghahramani and M. Jordan, Active Learning with Statistical Models, *Journal of Artificial Intelligence Research*, Vol. 4, 1996 , pp. 129-145.

[6] H. Flesche, A. A. Nielsen, R. Larsen, Supervised Mineral Classification with Semi-automatic Training and Validation Set Generation in Scanning Electron Microscope energy dispersive spectroscopy Images of Thin Sections, *Mathematical Geology*, Vol. 32 , No. 3, 2000, pp. 333-366.

[7] J. Fürnkranz, Integrative Windowing, *Journal of Artificial Intelligence Research*, Vol. 8, 1998, pp. 129-164.

[8] J. Hertz, A. Krogh and G. Thorbergsson, Phase transitions in simple learning, *Journal of Physics A (Mathematical and General)*, Vol. 22, No. 12, 1989, pp. 2133-2150.

[9] M. Hodgson, What Size Window for Image Classification? A Cognitive Perspective , *Photogrammetric Engineering and Remote Sensing*, Vol. 64, No. 8, August 1998, pp. 797-807.

[10] Y. Kareev, Through a narrow window: working memory capacity and the detection of covariation, *Cognition*, Vol. 56, 1995, pp. 263-269.

[11] T. Kirkpatrick and D. Thirumalai, Comparison between dynamical theories and metastable states in regular and glassy mean-field spin models with underlying first-order-like phase transitions, *Physical Review A.*, Vol. 37, No. 11, 1988, pp. 4439-4448.

[12] J. Lira and G. Maletti, A Supervised Classifier for Multispectral and Textured Images Based on an automated Region Growing Algorithm, *European Space Agency Publications*, Vol. SP-434, 1998, pp. 153-158.

[13] M. Møller, Supervised Learning on Large Redundant Training Sets, *International Journal of Neural Systems*, Vol. 4, No. 1, 1993, pp. 15-25.

[14] A. Papoulis, Probability, Random Variables and Stochastic Processes *Mc. Graw Hill Series in Systems Science*, 1965.

[15] I. Rybak, V. Gusakova, A. Golovan, L. Podladchikova and N. Shevtosova , A model of attention-guided visual perception and recognition, *Vision Research*, Vol. 38, 1998, pp. 2387-2400.

[16] J. Sowa, Peircean Fundations for a Theory of Context, *Lecture Notes in Computer Science*, Vol. 1257, No., 1997, pp. 41-64.

[17] G. Vetter, M. Stadler and J. Haynes, Phase Transitions in Learning *The Journal of Mind and Behavior*, Vol. 8, No. 2 and 3, 1997, pp. 335 [233]-350 [248].

[18] T. Watkin and J. Nadal, Optimal unsupervised learning, *J. Phys. A: Math. Gen.*, Vol. 27m, 1994, pp. 1899-1915.

[19] S. Wilson, On the retino-cortical mapping, *International Journal of Man-Machine Studies*, Vol. 18, No. 4, 1983, pp. 361-389.

4.1.2 Mean Shift Detection using Active Learning in Dermatological Images

Authors: Maletti, G., Ersbøll, B. and Conradsen, K.
Conference: The International Society of Optical Engineering International Symposium Medical Imaging 2002 (M.I.), Conference on Image Processing
Date: 23 – 28 February 2002
Place: San Diego, California, USA
Pages 1790 – 1797
Reference [40]

Mean Shift Detection using Active Learning in Dermatological Images

Gabriela M. Maletti, Bjarne K. Ersbøll and Knut Conradsen

Section for Image Analysis and Computer Graphics
Informatics and Mathematical Modelling, Technical University of Denmark

ABSTRACT

A scheme for detecting heterogeneous regions in dermatological images with malignant melanoma is proposed. This is done without setting any parameter. The mean shift detection problem is divided into two stages: window size optimization and detection. In the first stage, the maximum circular neighborhood centered on each pixel for which it is true that all the elements belong to the same class as the central one is estimated using redundant data sets generated with overlapping groups. Statistics are computed from all these neighborhoods and associated to the respective central pixels. As expected, larger values of a minimizing energy function are assigned to pixels belonging to heterogeneous regions. In the second stage, those regions are detected applying first an expectation-maximization algorithm and, afterwards, automatically defining a threshold between homogeneous and heterogeneous regions. The present scheme was tested on a set of synthetical images. Results are shown on synthetical and real images. Extensions of the scheme to textural cases are also shown.

Keywords: Window Size Optimization, Redundancy, Heterogeneous Regions, Dermatological Images

1. INTRODUCTION

Images with malignant melanoma contain non well-defined borders between classes. Generally, the normalized histograms of the classes present in the image are highly overlapped. This causes difficulty in describing the characteristics of the classes used for diagnostics with an uniform criteria. Furthermore, the way the classes present are established is not always obvious. Thus, it is difficult to provide a priori information for classification purposes by generating a statistically valid sample of the classes present in the image. In order to produce more homogeneous regions with more contrast between them in the image, we would like to enhance the borders between the classes by reducing the diffusion that exists in the transition from one class to another. However, a previous problem has to be solved and it will be the main topic of the present work: the detection of the heterogeneous regions. This problem can be divided into two subproblems: the estimation for each pixel of the local mean, and the detection of the mean shift. It will be solved by means of a contextual scheme: centered on each pixel in the image, we first estimate the maximum window that contains pixels belonging to the same class as the center one; data computed from the pixels belonging to each window are then used in order to detect the mean shift.

In the present work, the estimated optimum size of the neighborhood of each pixel is obtained at the minimum of an energy function computed for the most redundant set that can be generated for each window size. This stage will be called "Window Size Optimization". Those minimum values will be used in a "Detection" stage, for the separation of homogeneous regions from heterogeneous regions.

Further author information: (Send correspondence to G. M. M.)
G.M.M.: E-mail: gmm@imm.dtu.dk, Telephone: +45 45 25 34 32
B.K.E.: E-mail: be@imm.dtu.dk, Telephone: +45 45 25 34 13
K.C.: E-mail: kc@imm.dtu.dk, Telephone: + 45 45 25 34 16, Address: Richard Petersens Plads, Building 321, DK-2800 Kongens Lyngby, Denmark

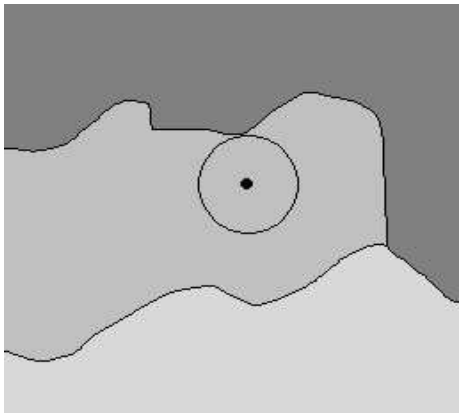


Figure 1. The set indicated as a circle is the maximal disc that can be centered on the marked point for which it is true that all the pixels belong to the same class as the center pixel.

2. RATIONALE OF THE SCHEME

The scheme for mean shift detection consists of two stages: Window Size Optimization and Detection. In the first stage a window of optimum size will be assigned to each pixel of the image. Statistics will be computed from the pixels belonging to these windows. In the second stage, the output of the first stage will be used to define homogeneous and heterogeneous regions. We will first introduce some definitions and notations to be used in the present work.

Let a specific image with non-correlated bands $X = \{x(r,c)\}$ be defined over the given domain $L = \{(r,c)|1 \leq r \leq max_r, 1 \leq c \leq max_c\}$. We assume that X is the realization of a two-dimensional random field X *; X is hierarchically defined in terms of the realization $Z = \{z(r,c)\}$ of an underlying random field Z such that: Z represents the partition of the domain L in K regions of different types. Each $z(r,c)$ is a value of the set of labels $B = \{b_1, b_2, ..., b_K\}$ where $z(r,c) = b_k$ indicates that the coordinate (r,c) of a given pixel belongs to the region k.

Let Z, in turn, be hierarchically defined in terms of the realization

$$W_o = \{w_o(r,c)\} \tag{1}$$

of an underlying random field W_o such that W_o represents the set of radii of discs of maximal size $\mathsf{D}_{w_o(r,c)}(r,c)$ included in X for which is true that all the elements of each given realization belong to the same region k. Then, the optimum window for a given pixel x defines the realization of the maximal set D_{w_o} centered on the position (r,c) of x. This is an heuristic approach. An example is shown in Figure 1.

If we associate a set of weights ω_w with values between zero and one to a disc D_w, such that $\omega_w(i)$ is the weight corresponding to the i-th element $\mathsf{d}_w(i)$ of the disc D_w, and such that the sum of the weights is one, then the entropy of this disc can be defined as:

$$E(\mathsf{D}_w) = -\sum_{i=1}^{n_w} \omega_w(i) log(\omega_w(i)) \tag{2}$$

where n_w denotes the size of a disc given by the number of random variables d_w that it is composed of. Following Møller,[1] the redundancy of a disc D_w can be defined as

$$R(\mathsf{D}_w) = 1 - \frac{E(\mathsf{D}_w)}{log n_w} \tag{3}$$

where n_w has the same meaning as before.

*Notice that while X (sans serif) is a set of variables, X (italic) is a set of values of those variables.

2.1. Window Size Optimization

An active learning process with increasing window size for each iteration will be defined for each pixel in the image (see left side of Figure 2). It is expected that at some point, the beginning of the learning phase of reorganization is detected (eg. the class boundary is detected). This point will be associated to the optimum window size for the given pixel. For each window size, the mean of this window will be estimated in a two-step procedure. First, the estimation of the means of the realizations of all the discs of the same size that can be included in this window is computed; secondly, the estimation of the mean between these groups is computed. The behavior of the estimated variance of the means within groups[2] is then tracked over the line of maximum redundancy. The details follow.

In general, we define the estimate y_w of the mean of a realization D_w of a disc D_w centered on a given position (r, c) as the scalar product of the realization D_w with a set of weights ω_w:

$$y_w(r,c) =< D_w(r,c), \omega_w >= \sum_i d_w(i)\omega_w(i) \tag{4}$$

where $d_w(i)$ is the $i-th$ element of the realization D_w of the disc corresponding to the $i-th$ position, $\omega_w(i)$ is the weight of the $i-th$ element of D_w corresponding to the same $i-th$ position. $y_w(r,c)$ is an element of the realization Y_w of a random field Y_w. We define the estimate σ_w^2 of the variance of a realization D_w of a disc D_w centered on a given position (r, c) as:

$$\sigma_w^2(r,c) =< D_w^2(r,c), \omega_w > -y_w^2(r,c) \tag{5}$$

where each $d_w^2(i) \in D_w^2$ is the square of the element $d_w(i)$ of D_w that is at the same position i.

Now, let $D_{m,v}(r,c)$ be the realization of a disc $\mathsf{D}_m(r,c)$ composed of the estimates of the means $y_v(r-i, c-j)$ of n_m discs of size n_v, all computed with the same set of weights ω_v and such that $0 \leq i^2 + j^2 \leq m^2$ (see right side of Figure 2). The mean of these means $y_{m,v}(r,c)$ is also estimated according to equation 4 using the set of weights ω_m and is called the estimation of the mean between groups. Let us call the discs included in X as low level discs D_{w_L}, and the discs included in Y as high level discs D_{w_H}. For all $D_{v,m}$ $\exists D_t \subset X \wedge \omega_t$ such that $y_{v,m} = y_t$ where

$$t = t(v,m) \simeq v + m \tag{6}$$

and ω_t is approximately equal to the convolution of ω_v with ω_m.

2.1.1. The energy function

Let us assume that we have a realization D_{max} that changes from an internal class I to an external class E at radius *to*. Let the mean of the internal class M_I be different from the mean $M_{(I \bigcup E)}$ of the union of the internal and the external class. The latter is the statistic to be learned. Let the variance of the internal class σ_I^2 be less than the variance of the union $\sigma_{(I \bigcup E)}^2$. If we compute the variance of the realizations of the high level discs for increasing size of t where the mapping model is more redundant, in each iteration the y_t will be the estimation of the real mean of the union of the internal and the external class $M_{(I \bigcup E)}$. However, if we avoid adding instances of the external class to the sample, we are really estimating the mean M_I of the internal class. This means that the variance of the learning mean of the internal class will tend to be minimized. Since "the chances of encountering a more extreme sample value are greater, the smaller the sample size",[3] we can expect that if the number of internal pixels is big enough, the probability of finding an extreme value will be low and all the variation of the learned variable will be introduced when we begin adding external pixels: this will produce a perturbation in the system. Therefore, the emergence of a higher order state will be detected at the minimum *to* of the following energy function Z evaluated along the line of maximum redundancy of the mapping model:

$$\mathsf{Z} = \sigma_{v_o,m_o}^2 =< D_{v_o,m_o}^2, \omega_{m_o} > -y_{v_o,m_o}^2 \tag{7}$$

where each $d_{v_o,m_o}^2(i) \in D_{v,m}^2$ is the square of the element $d_{v_o,m_o}(i)$ of $D_{v,m}$ that is at the same position i. For each fixed radius t we will choose the pair (v_o, m_o) such that the redundancy in D_t is maximal.

$$(v_o, m_o) = max(R(\mathsf{D}_{t(v,m)}))\forall t(v,m) constant \tag{8}$$

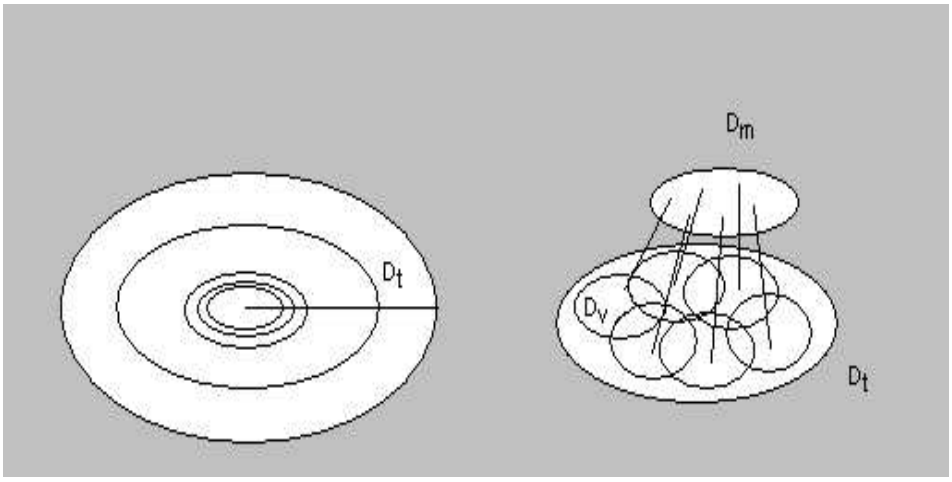


Figure 2. The mapping model. To the left, a set of discs of increasing radius t centered on a given pixel is shown. To the right, one of these discs is shown in detail: overlapping discs of a fixed radius v are placed into the disc of radius t. A new disc of radius m is constructed with the centers of these overlapping discs of radio v. The mapping model is completely defined when weights are assigned to the elements of each disc.

The set of $t = t(v_o, m_o)$ defines the line of maximum redundancy of the mapping model. Then

$$to = min[\mathsf{Z}] \in \hat{W}_o \tag{9}$$

where $\hat{W}_o$ is the set of estimates of the elements of W_o (see Equation 1), and the internal entropy to external entropy ratio

$$R = \frac{E(\mathsf{D}_{t-s})}{E(\mathsf{D}_t - \mathsf{D}_{t-s})} \simeq R_o \tag{10}$$

-were s is the variable increment of the radius-, holds $\forall t$. The reason of putting a constant threshold for this ratio is that we would like to hold the effect of the new pixels added to the sample constant.

The Window Size Optimization stage generates three outputs of the same size as the original image. The first output is a synthetic image with the radii of the estimated optimum window sizes for all the pixels of the original images. The second output is a synthetic image with the estimates of the means between groups corresponding to the estimated optimum windows. The third output is a synthetic image with the values of minimum estimated variance of the estimated mean within groups corresponding to the estimated optimum window sizes. The last one is used in the following stage.

2.2. Detection

The histogram of the synthetic image containing the minimum estimated variance of the estimated means within groups is constructed. An Expectation-Maximization Algorithm[4] (setting $\epsilon = 0.0001$ in[4]) is expected to provide the parameters (mean μ, covariance matrix Σ and a-priori probability p) of two Gauss distributed classes: the homogeneous region and the heterogeneous region. The threshold between these classes is obtained using the following discrimination function for Gauss-distributed classes with different dispersion[5]:

$$-(h-\mu_1)'\Sigma_1^{-1}(h-\mu_1) + (h-\mu_2)'\Sigma_2^{-1}(h-\mu_2) \geq 2log(\frac{L(2,1)p_2}{L(1,2)p_1}) \tag{11}$$

where μ_i is the mean, Σ_i is the covariance matrix and p_i is the a priori probability of the $i-th$ class $\forall i \in \{1,2\}$. $L(1,2)$ is the cost of belonging to class 1 and being classified as belonging to class 2 and viceversa for $L(2,1)$; these can be assumed to be equal.

Pixels whose minimum estimated variance of the means within groups is greater than the threshold are considered belonging to the heterogeneous regions, and viceversa, for the homogeneous regions.

2.3. Extensions of the Algorithm: A small Note

The present scheme can easily be extended to other spectral cases, or inclusive to textural examples. Instead of estimating the mean within the groups, other statistical descriptors[6] can be used. Requirements for the extension are that the homogeneous regions present in the image have to be separable by this new statistical descriptor, and that the size of the regions belonging to the image have to be directly proportional to the dimensionality of that statistical descriptor. See next section for some examples.

3. RESULTS

In order to first evaluate the window optimization stage in images with well defined borders, a set of 15 synthetic images with signal to noise ratios[†] 0, 3, 8, 15 and 26 and number O (from order) of classes 3, 5 and 7 was generated. The separation between consecutive means of the classes was 5 gray levels for all images. The classes are Gauss-distributed. The size of the images is 128 by 128 pixels (See Figure 3). The optimum window size was estimated for each pixel of each synthetic image setting the Internal to External Entropy Ratio R_o to an empirically found value of 10. This means, the energy function was computed for discs of radius t = 2, 4, 6, 8, 10, 14, 18, 24, 32, 42, 56 (the last number of this sequence corresponds to the radius of the disc of maximum size that can be included in the image still preserving the Internal to External Entropy Ratio). The weight assigned to each pixel in each disc was the inverse of the size of the respective disc for all the examples treated. The averaged signal to noise ratio increased from 10.35 in the original images to 30.64 in the estimated means between groups belonging the homogeneous regions.

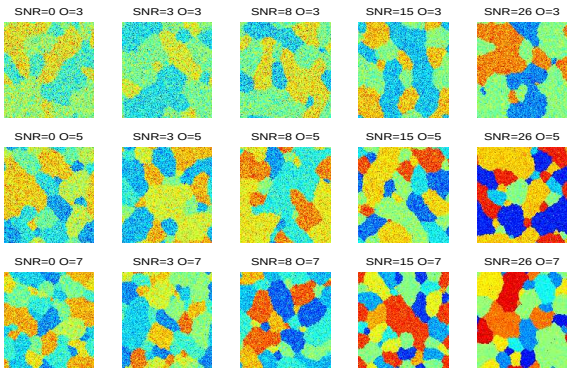


Figure 3: The set of original synthetic images

An Expectation-Maximization Algorithm[4] was applied to the original images and to the images containing the estimated means between groups.[‡] Thresholds between pairs of neighbor classes were estimated using Equation 11, where costs were assumed to be equal. Punctual classification was then applied. The average Cohen's (1960) Kappa coefficient increased from 0.4878 to 0.7737 using data of the whole images, and more, from 0.4625 to 0.9612 only taking into account pixels belonging to homogeneous regions (See Figure 4).

As an extension of the algorithm to textural cases, a synthetical example generated with a set of Brodatz's textures[7] was used (See Figures 5 and 6). The energy function - defined as the variance of the fractal dimension

[†]The signal to noise ratio is defined as $SNR = 20 \log \frac{\sum_{i=1}^{K-1} |\mu_i - \mu_{i+1}|}{(K-1)\sigma}$ where the μ_i is the mean of the i-th class, the means are ordered in increasing size, and σ is the standard deviation of the K classes.

[‡]The number of classes of each image was provided; initialization was done with the K-Means Algorithm; the threshold ϵ was set in 0.0001.

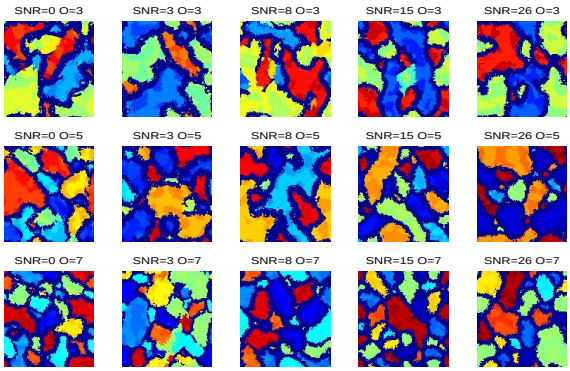


Figure 4: The set of synthetic images containing the estimated means between groups of the homogeneous regions.

within groups [§]-, was computed for discs of radius $t = 2$, 4, 6, 8, 10, 14 and 18. Weighting functions used were defined in the same way as before.

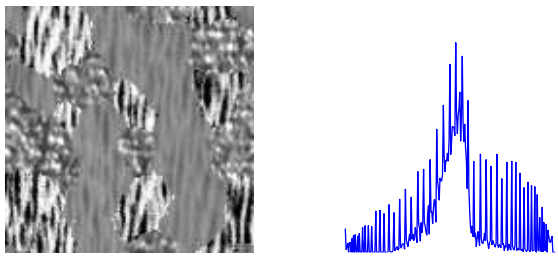


Figure 5: The synthetical image 128t14 with three textural classes (left) and the histogram of this image (right).

The real case presented in this work is a dermatological image of a malignant melanoma. This image is composed of three bands (R, G, B) of 590 rows by 885 columns. Principal component analysis was performed before starting with the proposed scheme. In decreasing order the eigenvalues obtained were: 0.432, 0.132 and 0.001. Only the first principal component was used. The same parameters as before were used. The ratio between costs $\frac{L(2,1)}{L(1,2)}$ in Equation 11 was set in an empirical found value of 40. In order to reduce the computational cost, the scheme was applied to the principal component reduced in scale to 25 percent of its size. The principal component was also divided into normal skin and ill skin using the scheme proposed in Section 2.2 and the output was reduced to 25 percent of its size too. Thus, on the right side of Figure 7 the heterogeneous regions of the ill skin can be appreciated.

4. CONCLUSIONS

A new scheme for mean shift detection in images with diffuse borders between homogeneous regions has been proposed. It has been shown that the maximum size of a window containing pixels of the same class as

[§]The fractal dimension f_d of a window D_w was defined as:

$$f_d(w) = \frac{\max(D_w) - \min(D_w)}{2w + 1} \tag{12}$$

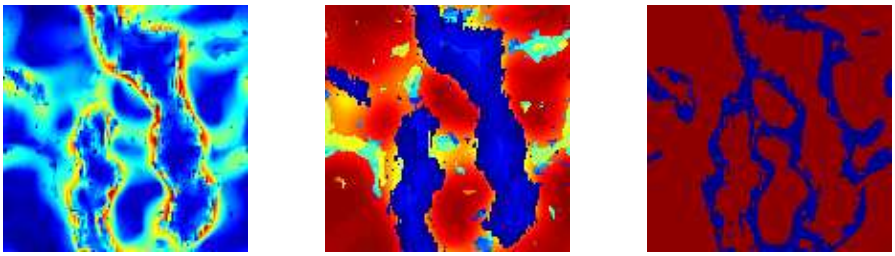


Figure 6. From the left to the right: the estimated minimum variance of the fractal dimension within the groups, the estimated means of the fractal dimension within the groups and the map of homogeneous (in red) and heterogeneous regions (in blue) corresponding to the synthetical image 128t14.

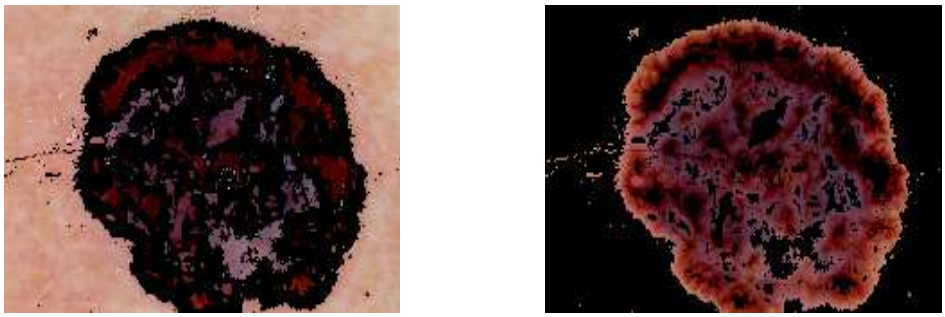


Figure 7. The dermatological image decomposed into homogeneous (left) and heterogeneous (right) regions using estimated means within groups.

the center one can be estimated without setting any parameters at minimizing the estimated variance of the estimated means within groups over the line of maximal redundancy of the mapping model. Large values of minimum estimated variance of the estimated means within groups have been assigned to pixels belonging to heterogeneous regions. This information has been taken into account first by an expectation-maximization algorithm that provides statistics of both regions, and then a threshold has been automatically computed in order to define the regions. The scheme was tested on a set of synthetical images: an improvement in the average signal to noise ratio from 10.35 to 30.64 was obtained and the average Kappa coefficient increased from 0.4878 to 0.7737 using data of the whole images, and from 0.4625 to 0.9612 only taking into account pixels detected as belonging to homogeneous regions. Results were satisfactory both in terms of visual appreciation of the contrast between the classes and in terms of separations of the histograms of the classes in the estimated homogeneous regions. It has been shown that it is possible to apply this scheme to textural images and it is encouraging to extend it to multi-spectral images.

ACKNOWLEDGMENTS

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number $STVF56-00-0123$) for supporting the present work. To The National Hospital of Denmark for providing the dermatological images used.

REFERENCES

1. M. Møller, "Supervised learning on large redundant training sets.," *International Journal of Neural Systems* **4**(1), pp. 15–25, 1993.
2. D. Cohn, Z. Ghahramani, and M. Jordan, "Active learning with statistical models.," *Journal of Artificial Intelligence Research* **4**, pp. 129–145, 1996.
3. Y. Kareev, "Through a narrow window: working memory capacity and the detection of covariation.," *Cognition* **56**, pp. 263–269, 1995.
4. Y. Wang and T. Adali, *Signal Processing for Magnetic Resonance Imaging and SpectroscopyStocastic Model Based Image Analysis*, ch. Stocastic Model Based Image Analysis, 14, pp. 1–34. Marcel Dekker, 2000.
5. K. Conradsen, *En Introduktion til Statistik.*, IMSOR DTH., Lyngby, 1984.
6. J. Lira, *Introducción al Tratamiento Digital de Imágenes*, Fondo de Cultura Económica. México D.F., 2001. In Press.
7. U. B. Computer Vision Group, Computer Science, "Collection of microtextures." http://www-dbv.cs.uni-bonn.de/image/browse.html .

4.1.3 Iterative Extended Mean Shift Algorithm

Authors: Maletti, G., Ersbøll, B. and Conradsen K.
Conference: 4^{th} International Conference on Computer Vision, Pattern Recognition & Image Processing, 2002 (C.V.P.R.I.P. 2002), part of the 6^{th} Joint Conference on Information Sciences (J.C.I.S.)
Date: 8 – 14 March, 2002
Place: Durham, North Carolina, USA
Pages: 793 – 796
Reference: [39]

Iterative Extended Mean Shift Algorithm

Gabriela Maletti, Bjarne Ersbøll and Knut Conradsen
Image Analysis and Computer Graphics, Informatics and Mathematical Modelling
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
gmm@imm.dtu.dk, be@imm.dtu.dk, kc@imm.dtu.dk

Abstract

An iterative filter for locally estimating the center of the clusters present in multi-spectral images is developed. In each iteration and for each pixel, an energy function computed for windows of increasing size is evaluated. The initial window size is defined as a function of the separability between neighboring pixels. The increments of the window size are a function of the internal to external entropy ratio of discs of consecutive radii. For a given pixel, the minimum value of the energy function is preserved and used as the initial guess for the next iteration. This minimum corresponds to the estimated point in which a state of higher order emerges. The scheme proposed was tested on a set of synthetical images, and compared to the output of the iterative median filter and to the k-Means algorithm, showing better performance than these ones. Results are shown on dermatological and fundus digital images.

1 Introduction

Many schemes have been developed in order to define a set of data points that correspond to local means of another set of data points. However, in spite that convergence to a fixed point is guaranteed because the contraction principle [4] was shown to be satisfied ([6], [9], etc.), the user is required to provide some parameter values (i.e. thresholds, initial window sizes, learning rate factor or weighting functions, etc.). This can affect the quality of the results in such a way that for a given class the estimated cluster center could not be the one expected due to border effects introduced by the window size, etc.. Thus, we would like to develop an unsupervised learning approach which first guarantees that the contraction principle is satisfied for each class present in the image: for each pixel of a multi-spectral image, the fixed point is the center of the cluster to which the pixel in question belongs to. Secondly, we would like to reduce the amount of information the user is required to provide.

The local estimation of the center of the cluster can be computed from the maximal set D_{w_o} centered on the position $[r, c]$ of a given pixel x for which it is true that all the elements belong to the same class as the center one [7]. In analogy with statistical mechanics, this corresponds to the point in the learning process for which a state of higher order emerges (for instance, when a class border is detected). The detection of this ergodic to non-ergodic transition – "where full ensemble averages are no longer equal to time averages" – [5] - was suggested to be done through the minimization of an optimal learning curve that depends on the fraction of examples [2]. Using windows of increasing size along iterations seems to be the natural implementation practice. It has to be taken into account that a minimal fraction of examples is required in order to start recognizing structure [1]. The separability between neighbors could provide an initial guess of the variance of the noise present in the image, and this could be used for defining the initial window size, as an extension to [9]. However, during practical implementations in lattices of models with constant increments of window size [9] problems appear when circular neighborhoods are used. This is added to the fact that, because the amount of information provided by the pixels introduced to the sample decreases when the sample size is increased in a constant way, a delay in the detection of higher order states will be introduced. For extension, this also depends on the weighting function used [7]. Automatically controlling the increments of the radii of circular neighborhoods using a previously set constant value of external to internal entropy ratio of discs with consecutive radii, allows for a compromise solution to previously mentioned practical considerations.

The energy function is required to be such that it is sensitive to the emergence of a higher order state. If it depends on sets containing a high degree of redundancy [8], it tends to be more sensitive to perturbations and therefore, more efficient in the minimization process. The variance of the mean within

groups has been shown to be a suitable statistic for this purpose [3].

2 The Iterative Extended Mean Shift Algorithm

For completeness, let us first write some definitions. Given a specific B-dimensional image $X = \{x[r,c] = x_i | 1 \leq i \leq n\}$ defined over the domain $L = \{[r,c] | 1 \leq r \leq max_r, 1 \leq c \leq max_c \wedge n = max_r max_c\}$, X hierarchically defined in terms of the set $Z = \{z[r,c]\} \subseteq X$ that represents the partition of the domain L in regions of different types, $z[r,c]$ such that it is the local estimation of the type mean with a kernel at position $[r,c]$, we would like to find Z.
From a syntactic point of view, a *window* defines a realization $D_w[r,c] = \{d_w[r-i, c-j] \in X | 0 \leq i^2 + j^2 \leq w^2\}$ of a disc D_w of radius w centered on the position $[r,c]$. Given a set of weights $0 < \omega_w[i] < 1$ assigned to a set of pixels $D_w[r,c]$, such that $\sum_{i=1}^{n_w} \omega_w[i] = 1$, where n_w is the number of elements of each one of those sets, and $\omega_w[i]$ is the weight assigned to the $i-th$ element of the set of pixels. The entropy of this set of pixels D_w is defined as:

$$S(D_w) = -\sum_{i=1}^{n_w} \omega_w[i] log(\omega_w[i]) \tag{1}$$

The internal entropy to external entropy ratio $IEER$ for two sets of pixels D_w and D_{w-s} centered on the same position, where $s < w$ is defined as the quotient between the entropy of their intersection and the entropy of their difference using the set of weights ω_w:

$$IEER[w,s] = \frac{S(D_w \cap D_{w-s})}{S(D_w - D_{w-s})} \tag{2}$$

2.1 Rationale of the Algorithm

The user is required to provide the maximal admissible internal to external entropy ratio R_0 (see Equation 2) using an uniform weighting function. The set of disc radii to be used can then be established as follows:

$$V = \{r_v \; | IEER[r_v, r_{v+1} - r_v] < R_0 \forall \; v \in [1; max_v]\} \tag{3}$$

and n_{r_v} is the number of elements of the disc D_{r_v}.[1]
Set $Y = X$ and then iterate varying $t = 1, ...$:

- Initialization:
  - The minimal window size min_v for the actual iteration is the minimal disc size $min_v = min(n_{r_v})$ for which it is true that:

$$\frac{1}{n} \sum_{r,c} min([\|\frac{\delta y}{\delta r}\|^2], [\|\frac{\delta y}{\delta c}\|^2]) \leq n_{r_v} \forall v \tag{4}$$

  - For each pixel $i = 1, 2, ..., n$:

$$m_{i,(min_v - 1)} = y_i \tag{5}$$

  - Initialize the matrix containing the minimal values of energy: $min\psi_{i,t} = max_{value} \forall i$

- For each window size $v = min_v, ..., max_v$ and for each pixel $i = 1, 2, ..., n$:

$$m_{i,v} = \sum_j \omega(y_j - m_{i,(v-1)}; D_v) y_j \tag{6}$$

  where the kernel similar as defined in [9] is:

$$\omega(s; D_v) = exp(-s^2/n_v)/N \tag{7}$$

  N is a normalization factor. Then compute the energy for the given pixel:

$$\psi_{i,v} = \frac{1}{2} \sum_j \omega(y_j - m_{i,(v-1)}; D_v)(y_j - m_{i,(v-1)})^2 \tag{8}$$

  If $\sum_{b=1}^{B} \psi_{i,v} \leq min\psi_{i,t}$ then update:

$$min\psi_{i,t} = \sum_b \psi_{i,v}; z_i = m_{i,v} \tag{9}$$

- Set $Y = Z$.

until between two consecutive iterations Z does not change anymore. The output of the algorithm is Z.

3 Results and Discussion

In order to test the quality of the algorithm outputs along the iterations, Cohen's Kappa Coefficient was used. This global indicator of classification quality varies from minus one to one, where a perfect classification would give a value of one, and a completely erroneous classification would give a value of minus one. The values z of the pixels belonging to the output were considered labels of classes. Pixels whose values were equal to the mean of one of the classes were considered signal pixels and the rest were considered belonging to a noise class. The confusion matrix was then constructed with $K+1$ classes: K signal classes labelled with the mean of the original classes μ_i, $i = 1..K$ and a noise class labelled with values different to the previous ones.

[1] $R_0 = 10$ is a suitable value empirically found. In this case: V={1,2,...,23,25,27,29,...,41,44,...,62,...}.

Table 1: Kappa values of the Original and Filtered Synthetical Images

SNR	0	3	8	15	26
Order $O = 3$					
Original	0.0319	0.0378	0.0596	0.1116	0.2065
K-Means	0.0352	0.0565	0.0911	0.7206	0.9295
Median	0.2567	0.3183	0.3926	0.4896	0.7819
I.E.M.S.	0.4529	0.5194	0.6130	0.7722	0.8786
Order $O = 5$					
Original	0.0468	0.0563	0.0771	0.1313	0.2329
K-Means	0.1186	0.1188	0.3172	0.6964	0.9395
Median	0.2751	0.3049	0.3957	0.5564	0.7642
I.E.M.S.	0.4732	0.5288	0.7326	0.7962	0.9188
Order $O = 7$					
Original	0.0547	0.0650	0.0923	0.1412	0.2536
K-Means	0.1716	0.2537	0.2592	0.4766	0.9347
Median	0.2593	0.3126	0.4012	0.5086	0.7776
I.E.M.S.	0.4645	0.5825	0.7062	0.8274	0.9307

A set of 15 synthetic images with signal to noise ratios [2] SNR of 0, 3, 8, 15 and 26 and number O (from order) of classes 3, 5 and 7 was generated. The separation between consecutive means of the classes was 5 gray levels for all images. The classes are Gaussian-distributed. The size of the images is 128^2 pixels. The averaged Kappa value of the set of synthetical images generated was 0.1066.

The output of the algorithm was compared to the output of K-Means and the Iterative Median Filter. The first one generated in each iteration a map containing the estimated means of the clusters the pixels were assigned to; the real means of the classes were provided as the initial guess. The second one for each iteration used window sizes of 3 pixels by 3 pixels and stopped when the Kappa value did not change anymore. The averaged Kappa value obtained were 0.4080 for K-Means, 0.4530 for the Iterative Median Filter and 0.6798 for Iterative Extended Mean Shift Algorithm (see Table 1 and Figure 1 for details).

As expected, the energy function chosen tends to be asymptotic to the abscissa for neighborhoods containing pixels that belong to the same class as the center one along the iterations, and it drastically separates from it when a higher order state emerges. The Kappa coefficient monotonically increases along the iterations, until no more changes occur in the image. This guarantees the noise to be reduced at the same time that the contraction principle is satisfied for each class present in the image.

A set of real examples can be appreciated in Fig. 2.

[2] The signal to noise ratio is defined as $SNR = 20 \log \frac{\sum_{i=1}^{K-1} |\mu_i - \mu_{i+1}|}{(K-1)\sigma}$ where the μ_i is the mean of the i-th class, the means are ordered in increasing size, and σ is the standard deviation of the K classes.

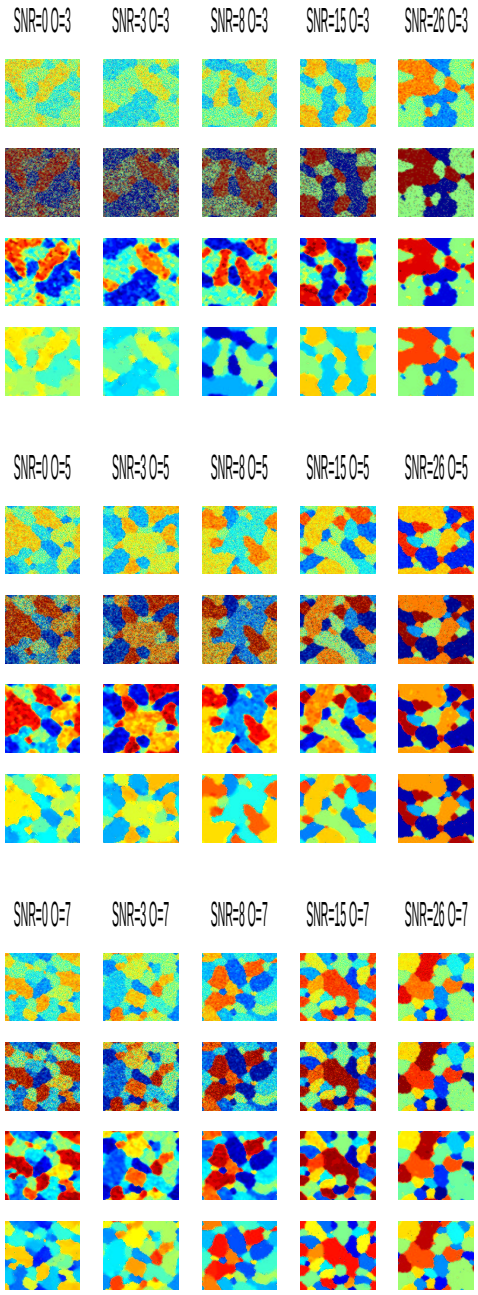


Figure 1: The original images (1^{st}, 5^{th} and 9^{th} row), and the respective K-Means (2^{nd}, 6^{th} and 10^{th} row), Iterative Median (3^{rd}, 7^{th} and 11^{th} row) and Iterative Extended Mean Shift (4^{th}, 8^{th} and 12^{th} row) Filtered Images.

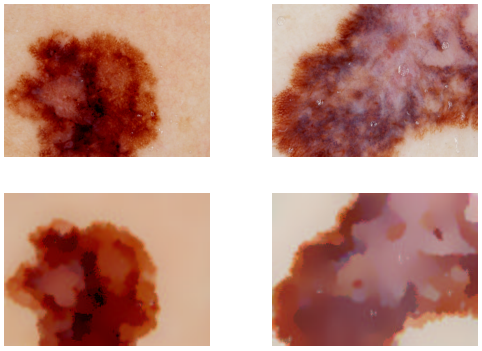


Figure 2: The original dermatological images 31056706 and 05041807 (top) and the respective outputs of the Iterative Extended Mean Shift Algorithm after 7 iterations (bottom).

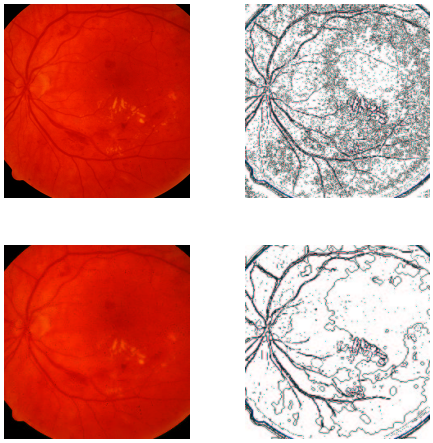


Figure 3: On the left, the original fundus image $f2577d7479f2$(top) and the IEMS filtered image 36 iterations (bottom). On the right, respective images after Contour Detection and Histogram Equalization using Paint Shop Pro 6.0.

4 Conclusions

A new scheme for local estimation of center of clusters has been developed. It has been tested on a set of synthetical images with varying signal to noise ratios from 0 to 26 and number of classes from 3 to 7 showing better performance than other filters with the same amount of information the user is required to provide. The average Kappa value increased from 0.1066 for the original images to 0.6798 for the output of the scheme proposed. The initialization and control of the window parameter is shown to be appropriate. For each pixel, the proposed energy function tends to be asymptotic to the abscissa in the homogenous region along the iterations, and it separates from it when a new class appears. The quality of the output monotonically increases along the iterations.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number STVF 56-00-0123) for supporting the present work. To The National Hospital of Denmark, to Prof. David Owens and to Torsana Diabetes Diagnosis for providing the digital images used.

References

[1] M. Biehl and A. Mietzner, Statistical mechanics of unsupervised structure recognition, *J. Phys. A.: Math. Gen.*, Vol. 27, 1994, pp. 1885–1897.

[2] A. Buhot and M. Gordon, Phase transitions in optimal unsupervised learning, *Physical Review E*, Vol. 57, No. 3, March 1998, pp. 3326–3333.

[3] D. Cohn, Z. Ghahramani and M. Jordan, Active Learning with Statistical Models, *Journal of Artificial Intelligence Research*, Vol. 4, 1996, pp. 129–145.

[4] Cambridge: MA: MIT Press, *Encyclopedic Dictionary of Mathematics*, Vol. 1–2, 1977.

[5] T. Kirkpatrick and D. Thirumalai, Comparison between dynamical theories and metastable states in regular and glassy mean-field spin models with underlying first-order-like phase transitions, *Physical Review A.*, Vol. 37, No. 11, 1988, pp. 4439–4448.

[6] Kohonen, T., The self-organizing map, *Neurocomputing*, Vol. 21, 1998, pp. 1–6.

[7] Maletti, G., Ersbøll, B., Conradsen, K. and Lira, J.: An Initial Training Set Generation Scheme. Scandinavian Conference on Image Analysis. (2001) 706–714

[8] M. Møller, Supervised Learning on Large Redundant Training Sets, *International Journal of Neural Systems*, Vol. 4, No. 1, 1993, pp. 15–25.

[9] T. Wakahara and K. Ogura, Extended Mean Shift in Handwriting Clustering, *Proceedings of the Foutheenth International Conference on Pattern Recognition*, Vol. 1, 1998, pp. 384–388.

4.1.4 A Contextual Classifier that only requires one Prototype Pixel for each Class.

Authors: Maletti, G., Ersbøll, B. and Conradsen K .
Conference: 2001 IEEE Nuclear Science Symposium (N.S.S.) and Medical Imaging Conference (M.I.C.)
Date: November 4 – 10, 2001
Place: San Diego, CA, USA
Pages: 1385 – 1389

Journal: IEEE Transactions on Nuclear Science
Volume: 49
Number: 3
Date: June 2002
Pages: 700 – 706
Reference: [38]

A Contextual Classifier That Only Requires One Prototype Pixel for Each Class

Gabriela Maletti, Bjarne Ersbøll, and Knut Conradsen

***Abstract*—A three-stage scheme for the classification of multispectral images is proposed. In each stage, statistics of each class present in the image are estimated. The user is required to provide only one prototype pixel for each class to be seeded into a homogeneous region. The algorithm starts by generating optimum initial training sets, one for each class, maximizing the redundancy in the data sets. These sets are the realizations of the maximal discs centered on the prototype pixels for which it is true that all the elements belong to the same class as the center one. Afterwards, a region-growing algorithm increases the sample size, providing more statistically valid samples of the classes. Final classification of each pixel is done by comparison of the statistical behavior of the neighborhood of each pixel with the statistical behavior of the classes. A critical sample size obtained from a model constructed with experimental data is used in this stage. The algorithm was tested with the Kappa coefficient κ on synthetical images and compared with K-means ($\bar{\kappa} = 0.41$) and a similar scheme that uses spectral means ($\bar{\kappa} = 0.75$) instead of histograms ($\bar{\kappa} = 0.90$). The results are shown on a dermatological image with a malignant melanoma.**

***Index Terms*—Prototypes, redundancy, region growing, supervised classification, window size optimization.**

I. Introduction

A PREMISE to a supervised classifier is that the training sets provided are statistically valid samples of the classes. Many semi-automatic training and validation set generation schemes by means of region-growing algorithms [1], [2] have been proposed. However, these schemes still require some user-input for the estimation of the parameters of the classes. It could be an advantage to develop schemes that minimize the amount of information the user is required to provide. On the other hand, contextual classifiers produce higher classification accuracy. However, the definition of an optimum window size for each step of the image analysis is still a problem [4].

From a semantic and pragmatic point of view [5], a window defines a subset of an entity in which the instances are spatially connected. From a heuristic point of view, the initial optimum window for a given prototype pixel defines the realization of the maximal disc centered on it for which it is true that all the elements belong to the same class as the center one. This can be obtained by detecting the emergence of a state of higher order through the minimization of an optimal learning curve that depends on the fraction of examples [6]. Based on the Heisenberg Uncertainty Principle, it can be deduced that this optimal learning curve corresponds to cases for which the envelope of the signal or the signal itself is a Gaussian function. From the Central Limit Theorem, we know that the convolution of a large number of positive functions is approximately a Gaussian function. Since consecutive values of the function resulting from convolution of two other functions contain a high degree of redundancy [7], we choose as trajectory for the learning curve the points for which the redundancy in the envelope of the fraction of examples is maximized. In the present work, the estimated optimum size of the neighborhood of each prototype pixel is obtained at the minimum of an energy function computed for the most redundant set that can be generated for each window size. This stage will be called "the initial training set generation scheme" [8].

A statistically valid sample of the class can then be obtained by increasing the sample size by means of a "region growing" algorithm [9]. Once the samples are defined, the minimum amount of information needed for classifying each pixel in the image has to be established. This is done based on the relationship between the window size and separability between classes obtained from experimental data, which provides a way of computing the critical number of examples needed for inferring an underlying structure in the data. Pixels are finally classified by comparison of the statistical behavior of its neighborhood with the statistical behavior of the classes [2].

II. Rationale of the Algorithm

Let a specific image with B noncorrelated bands $X = \{x[r,c]\}$ defined over the given domain $L = \{[r,c] \,|\, 1 \leq r \leq M, 1 \leq c \leq N\}$ with Q quantization levels. Let K be the number of prototype pixels $x_k^0 \in X, 1 \leq k \leq K$, seeded into homogeneous regions.

From a syntactic point of view [5], let a window $D_w[r,c]$ define the realization of a disc $\mathsf{D}_w[r,c]$ of radius w centered on the position $[r,c]$ of a given pixel

$$D_w[r,c] = \{d_w[p] = x[r-i, c-j] \,|\, 0 \leq i^2 + j^2 \leq w^2 \wedge 1 \leq p \leq n_w\} \quad (1)$$

where n_w is the number of pixels belonging to the set $D_w[r,c]$.

A. The Initial Training Set Generation Scheme

For each prototype pixel x_k^0, an active learning process with increasing window size for each iteration is defined (see the left side of Fig. 1). It is expected that, at some point, the class boundary is detected. This point will be associated with the optimum window size n_{to} for the given pixel. For each window

Manuscript received November 23, 2001; revised March 22, 2002. This work was supported by the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number STVF 56-00-0123).

The authors are with the Section for Image Analysis and Computer Graphics, Informatics and Mathematical Modeling (IMM), Technical University of Denmark (DTU), 2800-DK Kgs. Lyngby, Denmark (e-mail: gmm@imm.dtu.dk; be@imm.dtu.dk; kc@imm.dtu.dk).

Publisher Item Identifier S 0018-9499(02)06167-1.

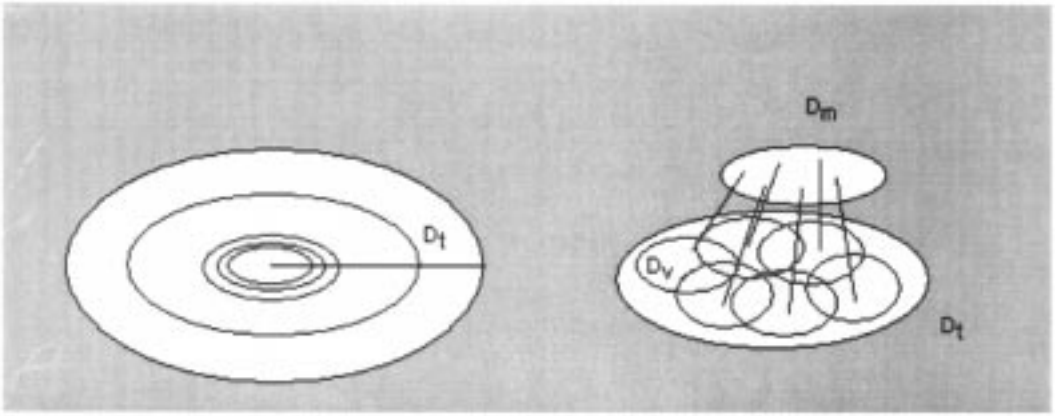


Fig. 1. The mapping model. To the left, a set of discs of increasing radius t centered on a given pixel is shown. To the right, one of these discs is shown in detail: overlapping discs of a fixed radius v are placed into the disc of radius t. A new disc of radius m is constructed with the centers of these overlapping discs of radius v. The mapping model is completely defined when weights are assigned to the elements of each disc.

size, the means of the realizations of all the discs of the same size that can be placed into this window are estimated. The estimated variance of the estimated means within groups is then tracked and minimized over the line of maximum redundancy of the mapping model [8].

In general, let the estimation y of the mean of a realization D_w of a disc D_w be the projection[1] of the realization D_w over the weighting function ω_w

$$y_w = \sum_p d_w[p]\omega_w[p] \tag{2}$$

where $d_w[p]$ is the pth pixel belonging to D_w and $\omega_w[p]$ is the weight assigned to it.

Let the realization of a disc $\mathsf{D}_{m,v}[r,c]$ be composed by the estimations $y_v[r-i,c-j] \,|\, 0 \le i^2+j^2 \le m^2$ of the means within n_m groups $D_v[r-i,c-j]$ of size n_v using the same set of weights ω_v. Let $y_{m,v}[r-i,c-j]$ be the estimated mean between these groups using the set of weights ω_m.

Since, for all realization $D_{v,m} \exists D_t \wedge \omega_t$ such that $t = t(v,m) \simeq v+m$ and $\omega_t \simeq \omega_v * \omega_m$,[2] the redundancy of a given pair $t(v,m)$ is defined as

$$R(\mathsf{D}_t) = 1 - \frac{S(\mathsf{D}_t)}{\log n_t} \tag{3}$$

where n_t is the number of pixels in the disc D_t and $S(\mathsf{D}_t)$ is the entropy of the disc defined as

$$S(\mathsf{D}_t) = -\sum_{p=1}^{n_t} \omega_t[p]\log(\omega_t[p]) \tag{4}$$

where n_t is the number of pixels of the disc, and $0 < \omega_t[p] < 1$ is the weight assigned to the pth pixel of the disc.

As can be observed, the measure of redundancy used depends on the group size n_v and the number of groups n_m, and it is inversely related to the entropy of the weights ω_v and ω_m assigned to the elements of the groups and to the groups, respectively.

[1]Scalar product.

[2]The equality holds for squares but small differences can be obtained in the case of discs.

For a given prototype pixel x_k^0 with position $[a,b]$, the size of the initial training set n_{to} corresponds with the minimum estimated variance of the estimated means within groups. This function is computed for each window size for the most redundant set

$$\mathsf{Z}^2[t(v_o,m_o)] = \sum_{p=1}^{n_{m_o}} \omega_{m_o} \| y_{v_o}[p] - y_{v_o,m_o}[a,b] \|^2 \tag{5}$$

where (v_o, m_o) is the pair for which the redundancy is maximal for each fixed radius $t(v,m)$, $\|\,\|$ is the Euclidian norm, $y_{v_o}[p] \in D_{m_o,v_o}[a,b]$ are the estimated means within groups, and $y_{v_o,m_o}[a,b]$ is the estimated mean between groups. The radius t is increased in such a way that the internal-to-external entropy ratio [8] is smaller than or equal to a constant R_0 $\forall t$ used. Then

$$t_o = \min(\mathsf{Z}^2[t(v_o,m_o)]). \tag{6}$$

Note that, in order to simplify the notation, a small change will be made. Hereafter the notations v and m are substituted for the notations v_o and m_o of the optimum radius t_o.

B. The Region-Growing Algorithm

Once the initial training sets are defined, the sample size can be increased by means of an optimized region-growing algorithm. Following the approach originally proposed in [2], [9], and [10], for each class, pixels satisfying a homogeneity criterion are included in the grown region. This criterion is evaluated in a window of optimum size n_v obtained in the Initial Training Set Generation Stage. A pixel is aggregated into the region if the difference between the homogeneity value of the class and the homogeneity value of the pixel centered window does not exceed a certain threshold. The growing of a region stops, once the homogeneity criterion is no longer satisfied. Following the notation previously introduced, let $G_0 = \{x^0[a,b]\}$ be the initial subregion pointing out a class. In general, $\forall g > 0$ let G'_{g-1} be the set of pixels that do not belong to G_{g-1} but having at least a neighbor with G_{g-1} under a certain connectivity. The set G_g is the region jointly formed by G_{g-1} and the pixels of G'_{g-1} such that the distance from the estimated mean of the class and the

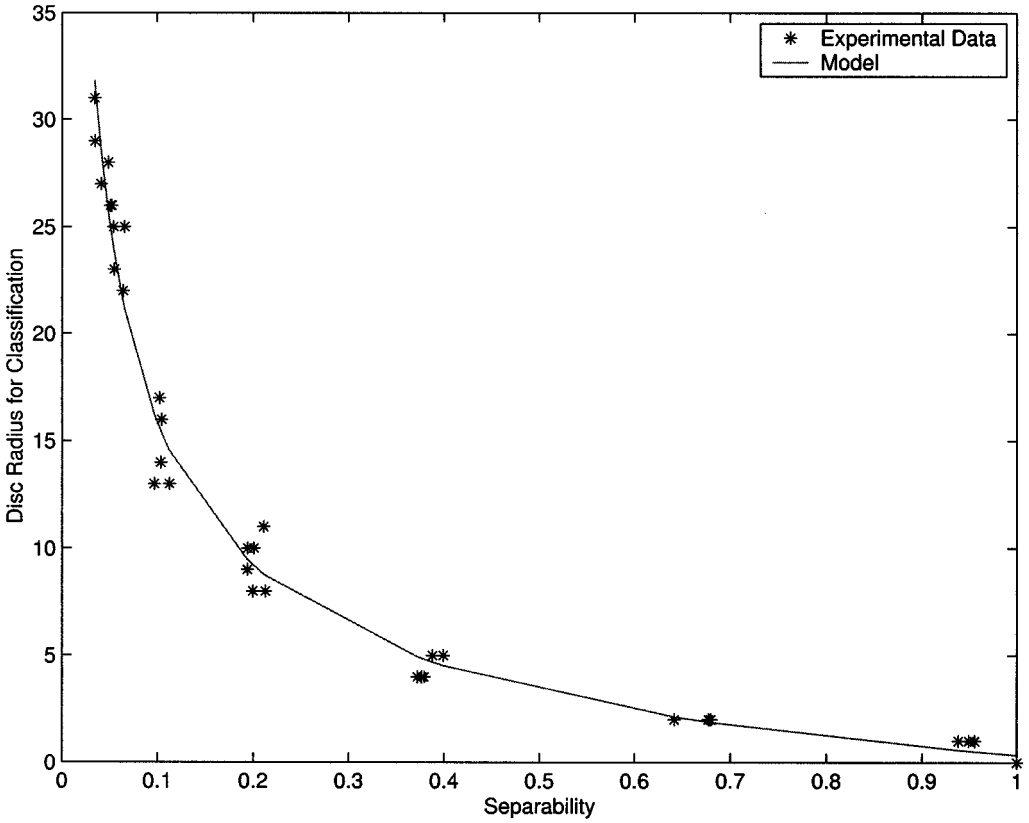


Fig. 2. Separability dA versus optimum disc radius w_c for classification.

estimated mean of the neighborhood centered on those pixels does not differ more than a certain threshold

$$G_g \equiv \{x[q,r]^g : ||y_v[q,r] - y_{m,v}[a,b]|| \leq \beta Z[t_o]\} \quad (7)$$

where $||\,||$ is the Euclidian norm and $\beta \in R^+$. The growing of a region continues until $G_{l+1} = G_l$.

C. The Final Classification

In this last stage, the normalized histograms of the grown regions of each class are used. An optimum window size for image classification is defined using these histograms. Afterwards, the following measure of distance from each pixel to each class is computed; it is the pondered sum of the difference of areas of pairs of class density functions:

$$dA_{(i),(j)} = \frac{1}{2B} \sum_{b=1}^{B} \sum_{n=1}^{Q} |h_{(i)}[b,n] - h_{(j)}[b,n]| \quad (8)$$

where B is the number of bands of the image, Q is the number of quantization levels, and $h_{(i)}$ and $h_{(j)}$ are the per band normalized histograms of the ith and jth class, respectively.

The critical number of samples for classification is computed as a function of the minimal separability between neighbor classes. This empirical model was constructed doing linear regression of experimental data (see Appendix A)

$$w_c = -6.8341 + \frac{7.1800}{\sqrt{dA}} \quad (9)$$

where dA has the same meaning as before and w_c is the radius of the estimated optimum disc for classification. It is optimum in the sense that it corresponds to the best classification rate for a set of experimental data with similar statistical descriptions. Fig. 2 shows the graphical behavior of this relation.

A pixel $x[s,t]$ belonging to the multispectral image X is classified according to the following criterion:

$$x[s,t] \rightarrow k : dA_{(x),(k)}, \text{ minimum}, \forall k \quad (10)$$

where $_{(k)}$ represents the normalized histogram $h_{(k)}$ of the grown region corresponding to the kth class and $_{(x)}$ represents the normalized histogram $h_{(x)}$ of the neighborhood of radius w_c centered on the position $[s,t]$ of the pixel x.

III. Results and Discussion

In order to evaluate the algorithm, a set of 15 synthetic images with signal-to-noise ratios (SNR)[3] SNR of 0, 3, 8, 15 and 26 and number O (from order) of classes of 3, 5, and 7 was generated. The separation between consecutive means of the classes was five gray levels for all images. The classes are Gauss-distributed. The size of the images is 128^2 pixels (see Fig. 3). The prototype pixels were seeded in suitable places. The initial optimum window size was estimated for each prototype pixel of each synthetic image setting the internal-to-external-entropy ratio R_o [8] to an empirically found value of 10. The weight assigned to each pixel in each disc was the inverse of the size of the respective

[3]The SNR is defined as SNR $= -20[(K-1)\sigma]^{-1} \log \sum_{i=1}^{K-1} |\mu_i - \mu_{i+1}|$ where the μ_i is the mean of the ith class, the means are ordered in increasing size, and σ is the standard deviation of each one of the K classes.

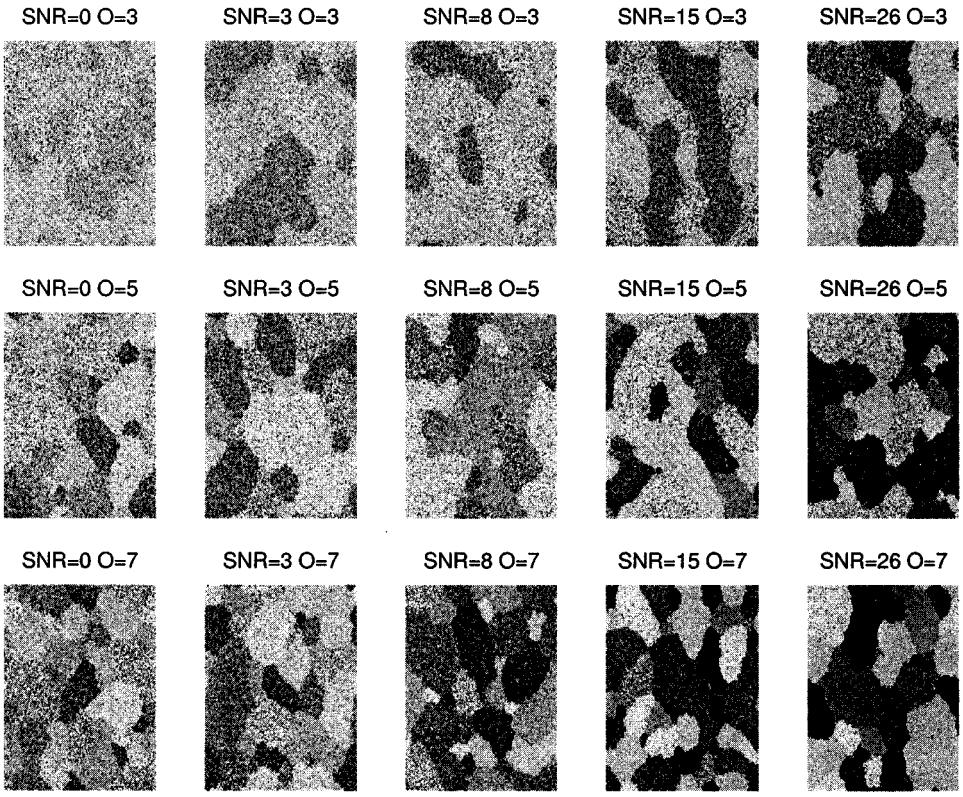


Fig. 3. The original set of synthetical images.

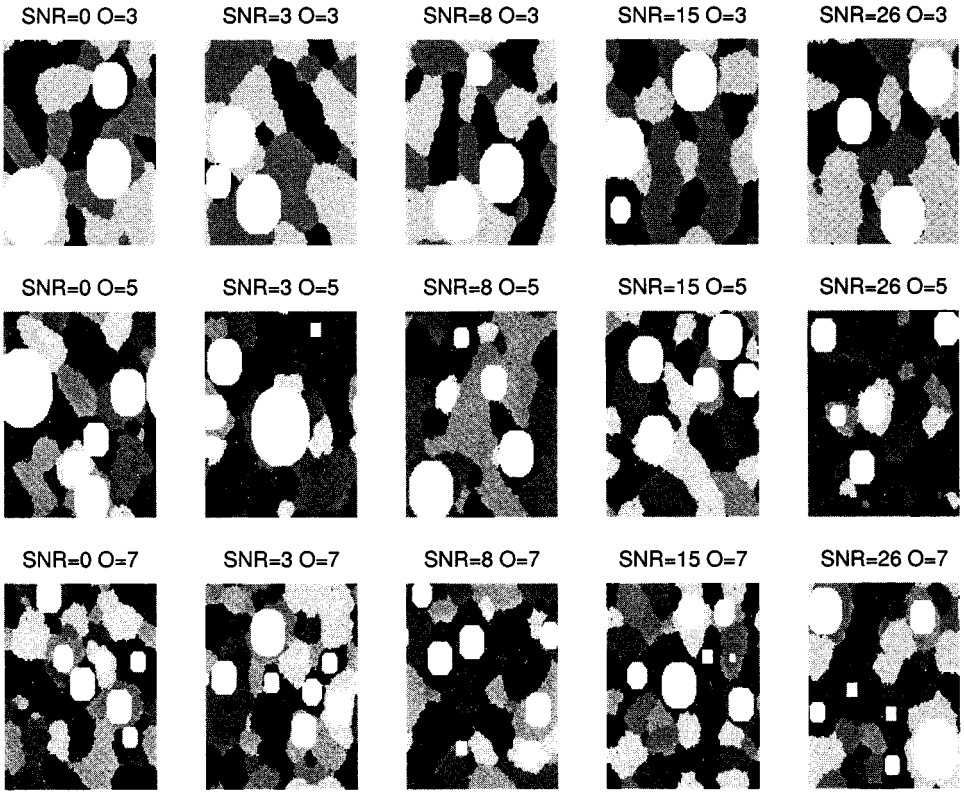


Fig. 4. The initial training sets (in white) centered on the seeded prototype pixels overlayed with the real thematic maps of the synthetical images shown in Fig. 3.

disc for all the examples treated. The initial training sets generated are shown in Fig. 4. The outputs of the second and third stage of the present scheme are presented in Figs. 5 and 6, respectively.

Fig. 5. The grown regions (in white) generated using the initial training sets delineated in Fig. 4 overlayed with the real thematic maps of the synthetical images shown in Fig. 3.

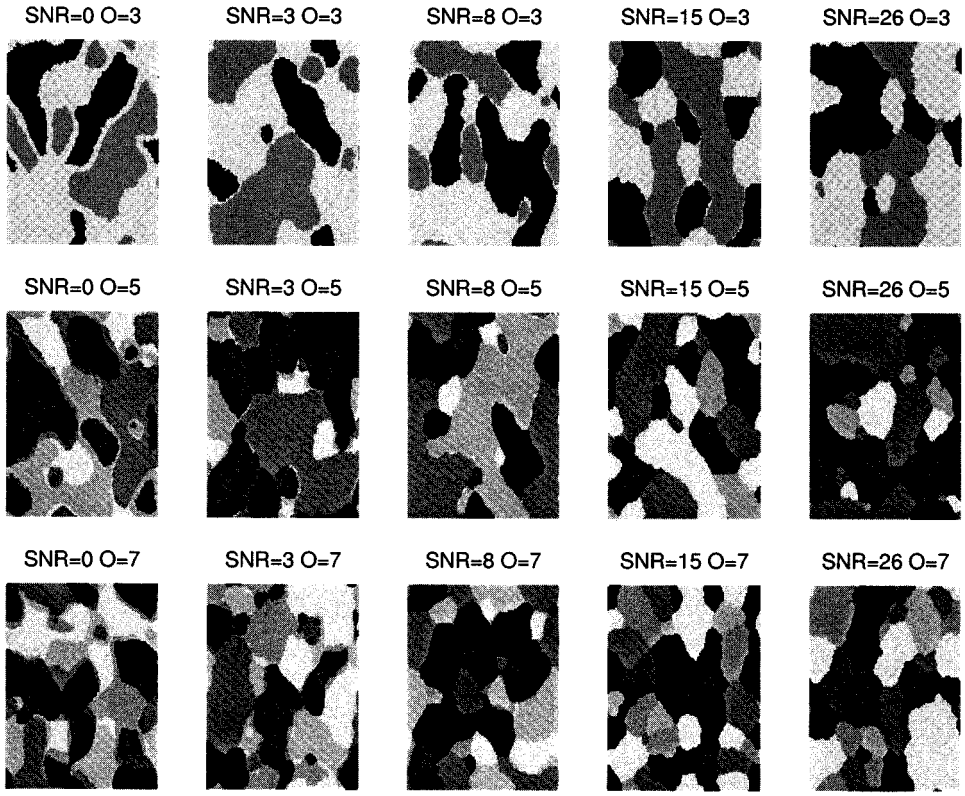


Fig. 6. The estimated thematic maps generated by the present classifier for the set of synthetical images shown in Fig. 3 using the training sets delineated in Fig. 5.

The output of the algorithm for this set of synthetical images was tested with Cohen's Kappa coefficient κ. It was compared to K-means and a scheme similar to the present one, except that in the final classification stage the spectral means of each pixel

TABLE I
KAPPA VALUES OF THE DIFFERENT CLASSIFICATION OUTPUTS

SNR	0	3	8	15	26
Order $O = 3$					
K-Means	0.0352	0.0565	0.0911	0.7206	0.9295
Spectral Means	0.6991	0.8718	0.8692	0.8902	0.9565
Norm. Histograms	0.7967	0.9021	0.9164	0.9498	0.9799
Order $O = 5$					
K-Means	0.1186	0.1188	0.3172	0.6964	0.9395
Spectral Means	0.5484	0.5925	0.7812	0.8281	0.8672
Norm. Histograms	0.7350	0.8884	0.9414	0.9583	0.9738
Order $O = 7$					
K-Means	0.1716	0.2537	0.2592	0.4766	0.9347
Spectral Means	0.3995	0.6312	0.6465	0.7225	0.8741
Norm. Histograms	0.7211	0.8511	0.9204	0.9543	0.9749

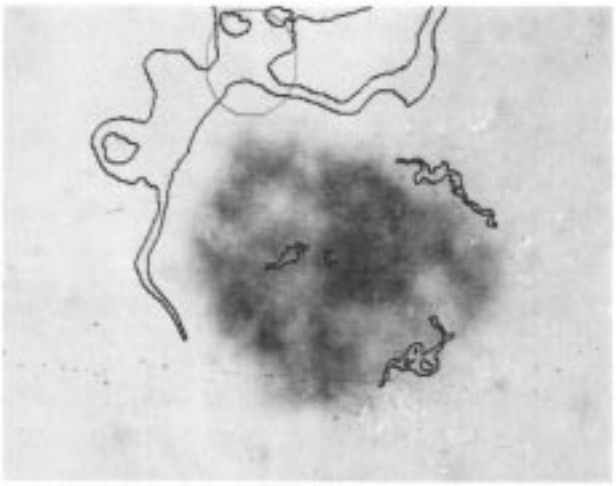


Fig. 7. The initial training sets (delineated with circles) centered on the seeded prototype pixels overlayed with the grown regions (delineated in black) of the image of a malignant melanoma.

centered neighborhood are computed and the distance between these means and the estimated means of the classes are used for assigning the pixels to the classes (see Table I for details).

Testing the algorithm for dermatological examples introduced some practical considerations that follow. First, the main problem in dermatological images is precisely defining all the classes and their homogeneity. The seed of the prototype pixels is crucial in the sense that this will set the parameters for the region-growing algorithm. Too homogeneously grown regions will produce classes with almost total separability and, therefore, the final classification will be done with too small window sizes not defining representative real large variation neighborhoods. On the other hand, if the pixels are seeded into regions with large heterogeneity, a high overlap in the statistics of the grown regions can occur, and the final classification stage will have to handle with large neighborhoods in order to make a good discrimination, but border effects can be introduced.

Secondly, since all the pixels are classified, thematic maps produced for real examples can contain misclassified pixels (healthy skin classified as ill and vice versa). This can be considered as a kind of noise that can be reduced by applying an iterative median filter to the thematic map.

An image of a malignant melanoma is the real case presented in this paper (see Fig. 7). The size of the original image was

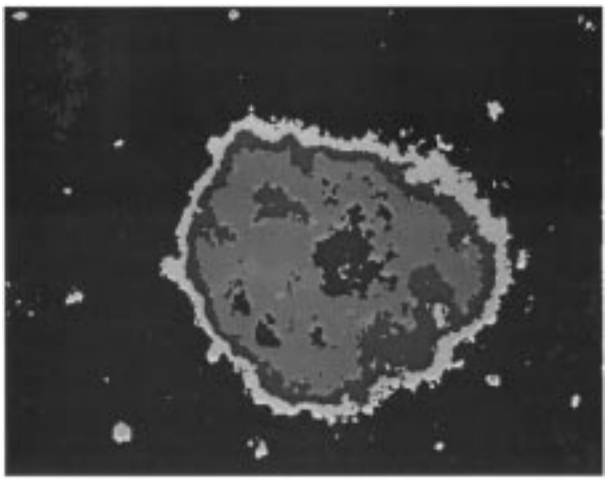


Fig. 8. The thematic map produced by the algorithm for the image.

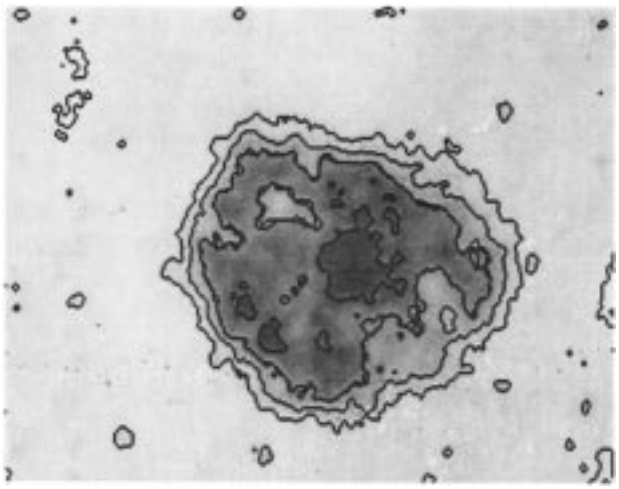


Fig. 9. The original image overlayed with the thematic map after three iterations of the median filter.

885 by 590 pixels. The eigenvalues of the principal components were in descending order: 0.209 849, 0.001 498, and 0.000 137. The two principal components of the original image reduced 50% in size were used as input to the algorithm. Five prototype pixels pointing each one to a class were seeded into homogeneous regions. The scheme produced a thematic map containing misclassified pixels (see Fig. 8), the number of which was reduced by applying an iterative median filter to the thematic map (see the overlay with the original image reduced by 50% in Fig. 9). Pixels with no overlap to one of the classes were assigned to a reject class.

IV. CONCLUSION

A new contextual classifier for multispectral images has been developed. The scheme has been tested for synthetic and real examples. It has been shown that the application of the iterative median filter to the thematic map can improve the quality of the results in real examples. In general, results are satisfying with respect to both numerical evaluation and visual appreciation.

TABLE II
DISTANCE dA BETWEEN PAIRS OF CLASSES GIVEN BY THE SEPARATION BETWEEN THEIR MEANS $s\mu$ AND THEIR STANDARD DEVIATION σ

$s\mu/\sigma$	1	2	4	8	16	32
0.5	0.2130	0.1035	0.0511	0.0349	0.0345	0.0549
1	0.3994	0.1998	0.0968	0.0488	0.0411	0.0543
2	0.6416	0.3720	0.1946	0.1022	0.0659	0.0523
4	0.9383	0.6797	0.3879	0.1938	0.1046	0.0644
8	0.9999	0.9493	0.6772	0.3767	0.2111	0.1127
16	1.0000	1.0000	0.9555	0.6797	0.3790	0.2010

TABLE III
DISC RADIUS w_c CORRESPONDING TO THE BEST KAPPA VALUE (ONE) FOR EACH PAIR OF CLASSES GIVEN BY THE SEPARATION BETWEEN THEIR MEANS $s\mu$ AND THEIR STANDARD DEVIATION σ

$s\mu/\sigma$	1	2	4	8	16	32
0.5	8	14	26	29	31	23
1	5	8	13	28	27	25
2	2	4	10	17	25	26
4	1	2	5	9	16	22
8	0	1	2	4	11	13
16	0	0	1	2	4	10

APPENDIX A

A. The Empirical Relation Window Size for Classification—Separability Between Classes

A set of 36 pairs (A, B) of synthetic images 128^2 pixels in size was generated. The values of the pixels were generated using different Gauss probability density functions $N(\mu_A, \sigma_A^2)$ and $N(\mu_B, \sigma_B^2)$. The values of the parameters were $\sigma_A = \sigma_B = 2^{j-1}, j = 1 \ldots 6$ and $\mu_A = 128 + ((2^{(i-1)})/2)$ and $\mu_B = \mu_A + ((2^{(i-1)})/2), i = 1 \ldots 6$.

For each pair of images (A, B), neighborhoods (discs) of increasing size were centered on the pixels from rows 33 to 96 and columns 33 to 96 of each one of the images. The radius of the discs varied from 0 to 31.

For each window size:

- the normalized histogram of each pixel centered neighborhood was computed;
- the distance dA from this histogram to the normalized histogram of both classes was calculated [see (8)];
- the pixel was assigned to the closest class;
- the quality of the final classification was measured with Cohen's Kappa coefficient κ.

The optimum radius for each pair of images corresponded to the maximum Kappa value of one.

Tables II-III show the distances between each considered pairs of classes and the corresponding optimum disc radii for classification.

ACKNOWLEDGMENT

The authors would like to thank the National Hospital of Denmark for providing the digital images used and J. L. Chávez from the Laboratory of Remote Sensing of the Institute for Geophysics of the National Autonomous University of Mexico for providing source code of the package SANDI (Digital Image Analysis System) used in the development of the present scheme.

REFERENCES

[1] H. Flesche, A. A. Nielsen, and R. Larsen, "Supervised mineral classification with semi-automatic training and validation set generation in scanning electron microscope energy dispersive spectroscopy images of thin sections," *Math. Geol.*, vol. 32, no. 3, pp. 333–366, 2000.

[2] J. Lira and G. Maletti, "A supervised classifier for multispectral and textured images based on an automated region growing algorithm," *Eur. Space Agency Publ.*, vol. SP-434, pp. 153–158, 1998.

[3] J. Lira and L. Frulla, "An automated region growing algorithm for segmentation of texture regions in SAR images," *Int. J. Remote Sensing*, vol. 19, no. 18, pp. 3595–3606, 1998.

[4] M. Hodgson, "What size window for image classification? a cognitive perspective," *Photogramm. Eng. Remote Sensing*, vol. 64, no. 8, pp. 797–807, 1998.

[5] J. Sowa, "Peircean fundations for a theory of context," *Lecture Notes Comput. Sci.*, vol. 1257, pp. 41–64, 1997.

[6] A. Buhot and M. Gordon, "Phase transitions in optimal unsupervised learning," *Phys. Rev. E*, vol. 57, no. 3, pp. 3326–3333, 1998.

[7] M. Møller, "Supervised learning on large redundant training sets," *Int. J. Neural Syst.*, vol. 4, no. 1, pp. 15–25, 1993.

[8] G. Maletti, B. Ersbøll, K. Conradsen, and J. Lira, "An initial training set generation scheme," in *Proc. 11th Scandinavian Conf. Image Analysis*, 2001, pp. 706–714.

[9] J. Lira and G. Maletti, "A supervised contextual classification scheme based on an automated region growing algorithm," in *5th Iberoamerican Symp. Pattern Recognition (SIARP 2000)*, Lisbon, Portugal, Sept. 11–13, 2000.

[10] ——, "A supervised contextual classifier based on a region-growth algorithm," *Comput. Geosci.*, vol. 8, no. 8, pp. 951–959, Oct. 2002.

4.1.5 A Supervised Contextual Classifier Based on a Region-Growth Algorithm

Authors: Lira, J. and Maletti, G.
Conference: 5^{th} Ibero-American Symposium on Pattern Recognition (S.I.A.R.P. 2000)
Place: Lisbon, Portugal
Date: 11 – 13 September 2000

Journal: Computers & Geosciences
Volume: 28
Issue: 8
Date: October 2002
Pages: 951 – 959
Reference: [28]

PERGAMON

Computers & Geosciences 28 (2002) 951–959

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

A supervised contextual classifier based on a region-growth algorithm

Jorge Lira[a,*], Gabriela Maletti[b]

[a] Instituto de Geofisica-UNAM, Circuito Institutos, Cd. Universitaria, 04510 México D.F., Mexico
[b] Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark

Received 1 May 2001; received in revised form 15 February 2002; accepted 18 February 2002

Abstract

A supervised classification scheme to segment optical multi-spectral images has been developed. In this classifier, an automated region-growth algorithm delineates the training sets. This algorithm handles three parameters: an initial pixel seed, a window size and a threshold for each class. A suitable pixel seed is manually implanted through visual inspection of the image classes. The best value for the window and the threshold are obtained from a spectral distance and heuristic criteria. This distance is calculated from a mathematical model of spectral separability. A pixel is incorporated into a region if a spectral homogeneity criterion is satisfied in the pixel-centered window for a given threshold. The homogeneity criterion is obtained from the model of spectral distance. The set of pixels forming a region represents a statistically valid sample of a defined class signaled by the initial pixel seed. The grown regions therefore constitute suitable training sets for each class. Comparing the statistical behavior of the pixel population of a sliding window with that of each class performs the classification. For region-growth, a window size is employed for each class. For classification, the centered pixel of the sliding window is labeled as belonging to a class if its spectral distance is a minimum to the class. The window size used for classification is a function of the best separability between the classes. A series of examples, employing synthetic and satellite images are presented to show the value of this classifier. The goodness of the segmentation is evaluated by means of the κ coefficient and a visual inspection of the results.

Keywords: Region growing; Context classification; Homogeneity criterion; Spectral; Distance; Spectral separability

1. Introduction

Segmentation is a partition of the image into a number of regions (Cohen and Fam, 1992), each region related to class objects of the scene. The regions may be labeled as pertaining to a certain class of objects, hence generating a classification. The final product is therefore a thematic map useful for scene understanding. The first classifiers labeled the pixels of the image into classes using only their spectral properties and ignoring their context. This approach, named per-pixel classification, proved to be limited in nature and applicable only to spectrally well-differentiated cases. In the last years, efforts have been devoted to develop contextual classifiers (Arai, 1993; Khazenie and Crawford, 1990; Gong and Howarth, 1992; Kontoes and Rokos, 1996). Classifiers that incorporate contextual information into the classification have been reported in the literature as well (Chica-Olmo and Abarca-Hernández, 2000; Atkinson and Lewis, 2000). A contextual classifier consistently produces higher classification accuracies than the per-pixel classifier (González Alonso et al., 1991; Kontoes and Rokos, 1996; Stuckens et al., 2000). In the approach of this paper, a pixel is labeled to a class by taking into account its spectral properties and the context of its location.

A supervised classifier employs a priori information of each determined class; this is usually done by means of

*Corresponding author. Tel.: +52-55-5622-4211; fax: +52-55-5550-2486.

E-mail addresses: lira@tonatiuh.igeofcu.unam.mx (J. Lira), gmm@imm.dtu.dk (G. Maletti).

0098-3004/02/$ - see front matter
PII: S0098-3004(02)00017-1

training sets. These training sets are defined through closed polygons outlined on the image by some interactive procedure. In this definition, there is not a clear criterion to assume that the training sets are valid statistical samples of the classes (Cetin and Levandowsky, 1991). However, some work on the purification of training samples has been reported that improves the quality of classification (Buchheim and Lillesand, 1989; Bolstad and Lillesand, 1991; Arai, 1992). Region-growth-based algorithms for image segmentation and region classification have been reported in literature Raafat and Wong (1988) and Gahegan and Flack (1999). However, further work is required to use an optimized region-growth algorithm as a preliminary step in a contextual classification scheme.

As a basic premise to a classifier, a procedure should be established to assure that the training sets are representative samples of the classes. A second premise is that the classification of a pixel should be performed by direct comparison between the statistical behavior of the classes and that of the pixel-centered window. A third premise is to make no assumption on the particular statistical behavior of the density function of the classes. This model-based approach yields the automated determination of the set of parameters handled by the classifier; the only exception is the locations of the initial seeds.

In this work, a new contextual classifier is proposed that determines statistical samples of defined classes as a result of an automated region-growth algorithm. A pixel is then classified by comparing the density function of the pixel-centered window population to those of the classes. The comparison is done by means of a measure of similarity between such density functions. The growth of each region is performed by employing a window size and a threshold value suitable for each class. The measure of similarity is utilized to calculate best values for such parameters. The window used for classification is a function of the best separability between the classes. As explained in the next sections, this scheme of classification is valid for a spectral contextual classifier, even though extension to a textural classifier is easy. In brief the goals of this research are the following: (a) develop a new supervised contextual classifier to segment an image into a number of classes, (b) derive the statistics of the classes from a region-growth algorithm, (c) determine the goodness of such classifier.

2. Contextual classifier

2.1. Region-growth scheme

The contextual classifier uses the training sets determined by an automated region-growth algorithm (Lira and Frulla, 1998). This algorithm begins by seeding pixels in suitable places of the image where the existence of a class is known. This task is done manually by visual inspection of the image with the support of ancillary data. Once the seeds are determined, one per class, the growth of the class regions starts. The growth is performed by pixel aggregation satisfying a homogeneity criterion. The criterion is evaluated in a window with the best size for each class. A pixel is aggregated into the region provided the difference between the homogeneity value of the seed- and the pixel-centered window does not exceed a certain threshold. The growth of a region is terminated when this homogeneity criterion is no longer satisfied. The homogeneity criterion and the threshold are both derived from a measure of separability and heuristic criteria; details of this are given in the next paragraphs.

Let $g(\mathbf{r})$ be a multi-spectral image and $\mathbf{p}_{ij}^0 \in \mathrm{g}$, and let $\{R_0 = \mathbf{p}_{ij}^0\}$ be the initial sub-region signaling a given class. The pixel $\mathbf{p}_{ij}^0$ is known as the seed related to R_0. The vector $\mathbf{r}$ defines the coordinates of the pixel $\mathbf{p}_{ij}^0$ in the image, i.e. $\mathbf{r} = (i,j)$. In general, the pixel $\mathbf{p}_{ij}^0$ is formed by the tuple $\{p_1, p_2, \ldots, p_\gamma\}_{ij}$, where γ is the number of bands of the multi-spectral image. Let R_0' be the set of pixels that do not belong to R_0 but having at least a neighbor with R_0 under certain connectivity. Let $E(R_0)_v$ be the value of the homogeneity criterion applied to the window v of R_0. The set R_1 is the region jointly formed by R_0 and the pixels $\mathbf{p}_{kl}^1 \in R_0'$ for which $E(\mathbf{p}_{kl}^1)_v$ differs from $E(\mathbf{p}_{ij}^0)_v$ in less than a threshold ε. In other words, R_1 is the following set:

$$R_1 \equiv \{\mathbf{p}_{kl}^1 : |E(\mathbf{p}_{kl}^1)_v - E(\mathbf{p}_{ij}^0)_v| \leqslant \varepsilon\}. \tag{1}$$

The real number ε is known as the parameter of uniformity, directly related to the homogeneity criterion. Once R_1 has been determined the previous step is repeated; so, in general the region R_m is given by

$$R_m \equiv \{\mathbf{p}_{kl}^m : |E(\mathbf{p}_{kl}^m)_v - E(\mathbf{p}_{ij}^0)_v| \leqslant \varepsilon\}. \tag{2}$$

The homogeneity criterion E is always tested facing the original sub-region window of R_0. The growth of a region continues until no change occurs from one step to the next: $R_{k+1} = R_k$. The above is easily generalized for a number of initial regions, i.e. for a set of classes. Thus, aggregation of a tested pixel $\mathbf{p}_{kl}$ is carried out as

$$\mathbf{p}_{kl} \rightarrow \text{region } t : |E(\mathbf{p}_{kl})_v^t - E(\mathbf{p}_{ij}^0)_v^t| \leqslant \varepsilon, \tag{3}$$

where $E(\mathbf{p}_{kl})_v^t$ is the homogeneity criterion applied to the tested pixel, and the region t is identified as the training set of class t.

2.2. Estimation of region-growth parameters

The best values for the window v and the threshold ε are obtained as follows. An odd-sized window is assumed for each seeded pixel. Beginning from $v = 3$

pixels, the windows are systematically incremented in size. Only squared odd windows are considered. The density function is estimated. Let $\mathscr{S}_v^a(i)$ be the density function of the pixel population in a window v for spectral class a. Such a function is normalized to unity according to

$$\sum_{\mathbf{i}} \mathscr{S}_v^a(\mathbf{i}) = 1. \tag{4}$$

The above is assuming 256 quantization levels in the image, and $\mathbf{i}$ is a vector with dimension equal to the number of bands of the multi-spectral image. Let a be any spectral class in the image, then

$$\delta_{\Delta v}^a = \sum_{\mathbf{i}} |\mathscr{S}_v^a(\mathbf{i}) - \mathscr{S}_{v+2}^a(\mathbf{i})|. \tag{5}$$

On the grounds of extensive tests with several images, the best window for class a is set whenever $\delta_{\Delta v}^a \leqslant 0.3$. A similar condition is established for the set of classes. The size of the window must satisfy a two-fold condition: large enough for statistical validity, and small enough for maximum spatial resolution. In this sense, the window size is adjusted according to class heterogeneity (Hodgson, 1998). A homogeneous class requires a small window size; a heterogeneous class requires a larger window size. Therefore, a window size is determined for each class. As v increments, the density function of a class experiments a great variation, until the number of pixels is enough to represent a valid statistical sample of the class. While the window size is further incremented, the density function is approximately constant as long as the window is entirely included in the class. If the window population is a mixture of two or more classes, the density function changes again. The condition $\delta_{\Delta v}^a \leqslant 0.3$ is a heuristic rule to measure such change.

The second parameter handled in the region-growth algorithm is a threshold ε, named the uniformity parameter. This parameter is set in the following way. Let d_v^{ab} be the distance between any two classes a and b for a given window size v:

$$d_v^{ab} = \sum_{\mathbf{i}} \left|\mathscr{S}_v^a(\mathbf{i}) - \mathscr{S}_v^b(\mathbf{i})\right|, \quad \forall\, a \neq b. \tag{6}$$

The distance d_v^{ab} is of spectral separability. Let a be a spectral class in the image; thus, given the set $\{v_a, v_b, \ldots, v_u\}$ of windows of the classes, the following minimum is determined:

$$d^a = \min\left[d^{ab} = \sum_{\mathbf{i}} \left|\mathscr{S}_{v_a}^a(\mathbf{i}) - \mathscr{S}_{v_b}^b(\mathbf{i})\right|\right], \quad \forall a \neq b. \tag{7}$$

On the grounds of extensive tests with several images, the threshold corresponding to the class j is given by the following criterion:

$$\varepsilon^j = 0.75 d^j \tag{8}$$

and so on for the set of window classes: $\{v_a, v_b, \ldots, v_u\}$.

The region-growth is performed on the grounds of a homogeneity criterion defined in terms of the distance determined by Eq. (6). Thus, given the best window for class j:

$$\mathbf{p}_{\mathrm{kl}} \rightarrow \text{class } j: \ \sum_{\mathbf{i}} \left|\mathscr{S}_{\mathbf{p}_{kl}}^j(\mathbf{i}) - \mathscr{S}_{\mathbf{p}_0}^j(\mathbf{i})\right| \leqslant \varepsilon, \tag{9}$$

where $\mathbf{p}_0$ is the pixel seed for class j. In brief, Eqs. (5), (8) and (9) determine the best window, the threshold and the homogeneity criterion, respectively. This means that, once the seed pixels are determined, the classification process is fully automated.

2.3. Rationale of spectral classifier

The contextual classifier uses a pixel-centered window to estimate the density function associated to a pixel. This function is compared to the density function of the classes determined by the training classes in the region-growth process. A pixel is classified to that class where the comparison is the best. However, a tested pixel window may share pixels from two or more classes. In this situation the density function is a composite of various classes. The pixels having windows located in more than one class are named border pixels. To avoid misclassification, the border pixels must be treated separately, labeled as such, and classified to the nearest class in a separate process. Once classified, the border pixels are encoded with the rest of the pixels to form the final classification of the image. Border pixels are identified as follows: The pixel-centered window is displaced on every pixel of the image. In each position of the window, the density function is obtained. This function is compared with the density function of the classes and to the density function of pair of classes. The result of this comparison leads to the labeling of a border pixel; details of this are in the classifier algorithm steps (v) and (viii).

The rationale of the spectral classifier consists of three stages: (1) construct a model of spectral separability to calculate the best window to perform the classification, (2) design an overall block diagram of the classifier, and (3) list the details of the algorithm for classification. The construction of such a model now follows.

A set of 36 images with 6 classes is generated. The density function of the classes is Rayleigh-like (Dougherty, 1999). The size of the images is 256×192 pixels. The difference of the mean among these classes ranges in the interval: [1,2,...,32]. The difference of the standard deviation ranges in the same interval. This produces a set of 216 classes with varying spectral separability among them. The normalized density function of each class is obtained according to Eq. (4). In each class, a pixel-centered window is considered. This window is systematically incremented from 1×1 to 23×23 pixels. The distance between the density function of this

Table 1
Separability as a function of window size

Window	Separability	Window	Separability	Window	Separability
1	1.9997–2.0000	7	1.1606–1.6299	17	0.3406–0.3746
3	1.9187–1.9997	11	0.6772–1.1606	21	0.2293–0.3405
5	1.6299–1.9186	15	0.3746–0.6771	23	0.2193–0.2293

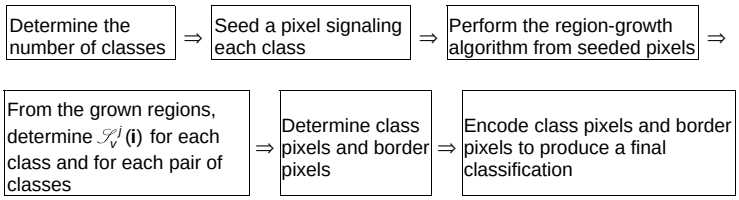


Fig. 1. Block diagram of contextual classifier.

window and the one from each other class is derived (Eq. (6)). A pixel-centered window is classified into a class where this distance is the least. The precision of classification is calculated for each window size. The best window is the one yielding the highest precision. Table 1 condenses this model of spectral separability. In this model, the following is observed:

(i) Whenever the separability is complete, the maximum precision is obtained using a per-pixel classification, i.e. a 1×1 pixel window.
(ii) Classes with similar separability have a similar window size for classification.
(iii) An inverse relationship exists between the separability of the classes and the best window size.

The overall block diagram of the classifier is given in Fig. 1. According to this, the details of the classifier algorithm are the following:

(i) Let $g(\mathbf{r})$ be the image to be classified. The bands selected for classification are loaded into RAM memory. Decorrelated bands resulting from principal component analysis are usually employed in this step.
(ii) Pixels are seeded in selected places of each spectral class defined for segmentation. Let $\xi \equiv \{a, b, ..., u\}$ be the set of spectral classes.
(iii) The best window is derived for each class according to Eq. (5). Let $\eta \equiv \{v_a, v_b, .., v_u\}$ be the set of windows.
(iv) On the grounds of the region-growth algorithm, the normalized density function $\mathscr{S}_{v_j}^j(\mathbf{i})$ for each training set is obtained $\forall j \in \xi$.
(v) A matrix of the average of density functions for a pair of classes is constructed. Let $\mathscr{S}^{ts}(\mathbf{i})$ be the elements of such a matrix for any two spectral classes t and s, $\forall t \neq s$. The function $\mathscr{S}^{ts}(\mathbf{i})$ is named as the density function of a border formed by a pair of distinct classes.
(vi) The distance $d^{ab} = \sum_{\mathbf{i}} \left| \mathscr{S}_{v_a}^a(\mathbf{i}) - \mathscr{S}_{v_b}^b(\mathbf{i}) \right|$, $\forall a \neq b \in \xi$ is obtained as a function of class window v_i, $\forall v_i \in \eta$. For classification, the best window v is obtained from Table 1 using $d = \min[d^{ab}]$, $\forall a \neq b \in \xi$.
(vii) For every pixel $\mathbf{p}_{kl}$ to be classified, the normalized density function $\mathscr{S}_v^{kl}(\mathbf{i})$ is generated, where (k, l) are the coordinates of the pixel in the image: $\mathbf{r} = (k, l)$, and v is the best window obtained in step (vi).
(viii) A pixel $\mathbf{p}_{kl}$ of the multi-spectral image is classified according to the following procedure:

(a) The distance between the density function $\mathscr{S}_v^{kl}(\mathbf{i})$ of the pixel-centered window and the density function of the classes $\mathscr{S}_{v_j}^j(\mathbf{i})$ is calculated

$$d_{kl}^j = \sum_{\mathbf{i}} \left| \mathscr{S}_v^{kl}(\mathbf{i}) - \mathscr{S}_{v_j}^j(\mathbf{i}) \right|, \quad \forall k,\ l \in \mathrm{g},\ l = \forall j \in \xi. \tag{10}$$

(b) The distance between the density function $\mathscr{S}_v^{kl}(\mathbf{i})$ of the pixel-centered window and the density function of the pair of classes $\mathscr{S}^{ts}(\mathbf{i})$ is calculated

$$d_{kl}^{ts} = \sum_{\mathbf{i}} \left| \mathscr{S}_v^{kl}(\mathbf{i}) - \mathscr{S}^{ts}(\mathbf{i}) \right|, \quad \forall k, l \in g,\ \forall (t, \mathrm{s}) \in \xi. \tag{11}$$

(c) The distances d_{kl}^j and d_{kl}^{ts} are sorted out by value. Let $d = \{d_{kl}^j, d_{kl}^{ts}\}v$ be the ordered set of such distances.

(d) If $\min\{d_{kl}^{j}, d_{kl}^{ts}\}v$, is a distance to a border, a new window v' for $\mathbf{p}_{kl}$ is considered. The size of this window is the one providing the best separability (Table 1) between the density function of such border and the density function corresponding to the second smallest distance of δ.

(e) A new set of distances $\delta' = \{d_{kl}^{j}, d_{kl}^{ts}\}_{v'}$ is calculated for v'. If $\min\{d_{kl}^{j}, d_{kl}^{ts}\}_{v'}$ is a distance to the border found in step (d), then $\mathbf{p}_{kl} \rightarrow$ border.

(f) If $\min\{d_{kl}^{j}, d_{kl}^{ts}\}_{v}$ is the distance to a class t, then $\mathbf{p}_{kl} \rightarrow$ class t.

(g) If $\mathbf{p}_{kl} \rightarrow$ border, its average density function is considered. Let this border be the one formed by classes q and r. The following distances are calculated:

$$d_{kl}^{l} = \sum_{\mathbf{i}} \left| \mathscr{S}_{v}^{kl}(\mathbf{i}) - \mathscr{S}_{v}^{l}(\mathbf{i}) \right|, \quad \forall k, 1 \in g, \ 1 = (q, r) \in \xi. \quad (12)$$

Then $\mathbf{p}_{kl}$ is labeled as border-class q if $d_{kl}^{q} = \min\{d_{kl}^{l}\}$, $1 = (q, r) \in \xi$.

This algorithm generates a segmentation of the image in classes and borders. The encoding of class pixels and border pixels produces a final classification of the image as shown in Section 4. This algorithm is not computationally demanding: the density functions and the images are loaded into the RAM memory, the whole procedure for the multi-spectral classification presented in this research takes approximately 2 min on a Pentium IV @ 700 MHz. The following section describes the test images used in the classifier.

3. Test images

Two types of examples are presented in this paper. These examples are worked out on the grounds of:

(a) A set of two synthetic images (Figs. 2a and 3a) with well-known statistical parameters for each class. These synthetic images are single band and contain five (Fig. 2a) and three (Fig. 3a) barely discernible classes. The class separation, in both images, is five gray levels between means. In Fig. 2a, each class has a variance of 11.26. In Fig. 3a, each class has a variance of 24.92. The dimension of these images is 128×128 and 96×96 pixels, respectively. The density function of the classes in both the images is Gaussian-like. Fig. 2a shows an image with low content of noise: snr = 7.9586 dB. Fig. 3a depicts a relatively high level of noise: snr = 0.3719 dB. The formula to calculate the noise is

$$\text{snr} = \frac{1}{N_{\text{C}} - 1} \sum_{i=1}^{N_{\text{C}}} 20\log\left[\frac{\mu_{i+1} - \mu_i}{\sigma}\right], \quad (13)$$

where N_{C} is the number of classes, μ_i and σ are the mean and the standard deviation of the classes,

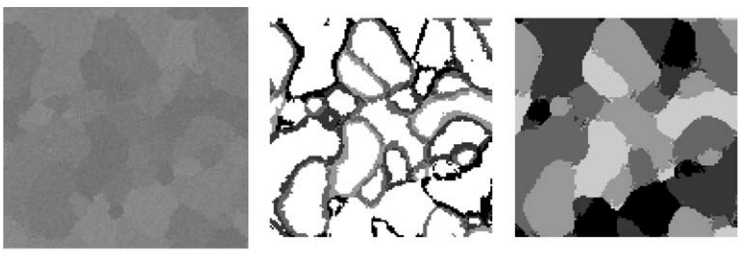


Fig. 2. (a) Synthetic image with five classes and signal-to-noise ratio of 7.9586 d**B**, (b) border pixels of synthetic image depicted in Fig. 2a, (c) encoding of border pixel and class pixels of the synthetic image depicted in Fig. 2a.

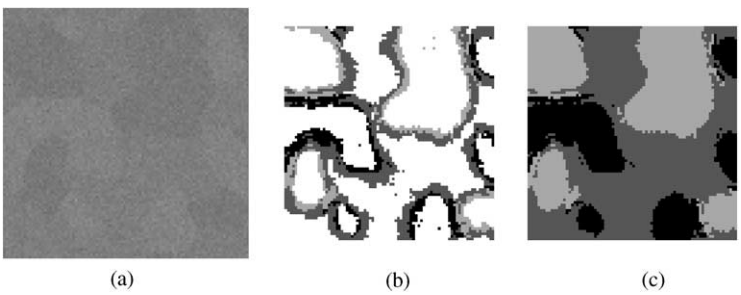


Fig. 3. (a) Synthetic image with three classes and signal-to-noise ratio of 0.3719 d**B**, (b) border pixels of the synthetic image depicted in Fig. 3a, (3) encoding of border pixel and class pixels of synthetic image depicted in Fig. 3a.

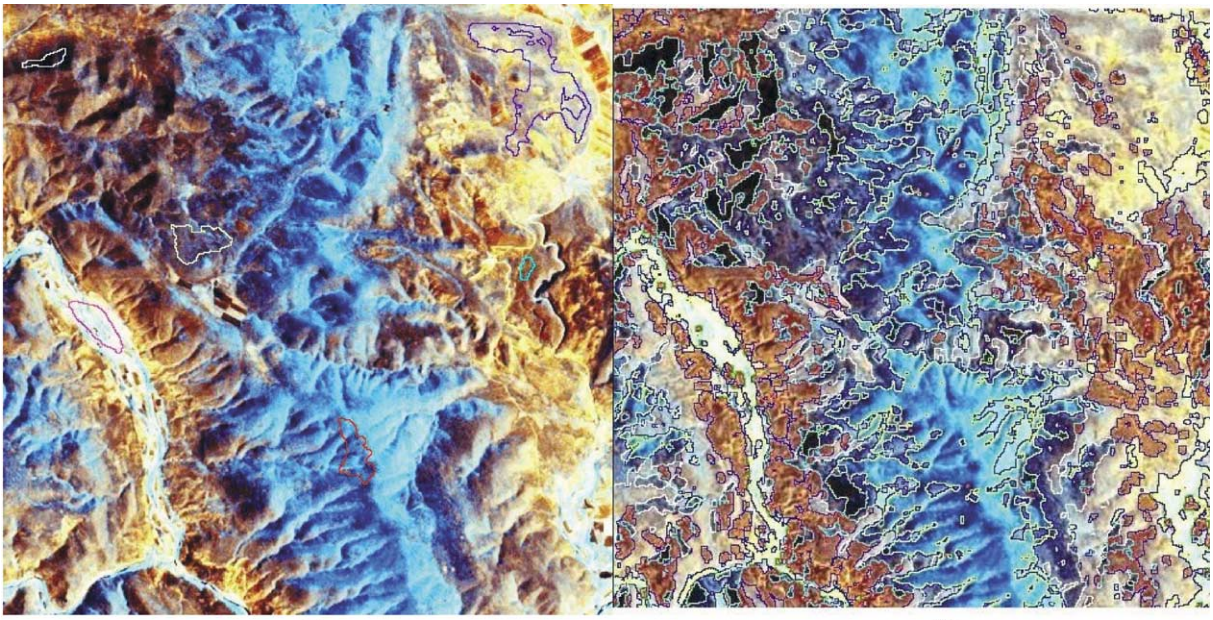


Fig. 4. (a) RGB composite of SPOT image overlain with training fields defined by region-growth, (b) RGB composite of the SPOT image overlain with borders of spectral classes defined in classification process.

respectively. The means are sorted by the value: $\mu_{i+1} \geqslant \mu_i$.

(b) A satellite SPOT multi-spectral (Fig. 4a) image. The dimension of this image is 512×512 pixels, covering a region of Central México. This image is not geocoded.

4. Results and discussion

Results are organized in two categories: for the synthetic images final results are provided. For the SPOT image full details are given.

The synthetic images are considered an extreme case to test the validity of the classification scheme. Fig. 2b shows the border pixels of Fig. 2a, while Fig. 2c shows the encoding of class and border pixels with the final segmentation achieved. The finite size of the window employed in this process reduces the size of the images by $(v—1)/2$ pixels per side. The border pixels and the segmentation achieved for Fig. 3a are presented in Fig. 3b and 3c, respectively. The high content of noise in Fig. 3a produces a less precise segmentation than in Fig. 2a. The synthetic nature of the images used in this work allows a quick quantification of the classification accuracy. The region of the image occupied by each class is well known. The confusion matrices related to the segmentation of Figs. 2a and 3a are given in Tables 2 and 3. From these matrices, the κ coefficient is easily calculated (Landis and Koch, 1977; Smits et al., 1999). For Fig. 2a, κ is 0.8586, and for Fig. 3a, it is 0.7937.

The second example deals with a multi-spectral SPOT image covering a region of Central México. A principal component analysis was applied to this image, only the first two components were retained for the classification process.

By visual inspection on this image, and with the help of ancillary data, six pixels were seeded signaling six spectral classes. Fig. 4a shows the region growth for the six classes overlain with an RGB color composite of the image. The color composite is, [RGB] = [Principal component 1, Principal component 2, Band 3]. These regions define the training fields for the spectral classes. From upper left in Fig. 4a, and going in clockwise direction, the classes defined in the classification process are: shadows, soil/herbage, microphylum thicket, oak/pine woods, river bed/herbage, and submontane thicket/oak. The distance (Eq. (6)) among the density functions of the regions measures the separability matrix of such classes. This separability matrix is shown in Table 4. The least separability in Table 4 is 1.8028; therefore, based on the values of Table 1, this classification was performed using a window of 5×5 pixels. To classify the border pixels, the distance among the pair of classes is calculated. The least separability in Table 5 is 0.9014; therefore, based on the values of Table 1, this classification is performed using a window of 11×11 pixels.

Table 2
Confusion matrix of classification for synthetic image 1

Class	1	2	3	4	5	Region size (pixels)
1	1862	175	73	43	49	2202
2	76	3410	133	104	41	3764
3	62	178	2710	86	20	3056
4	33	48	85	2556	58	2780
5	13	41	80	94	1426	1654
Total	2046	3852	3081	2883	1594	13456

κ Coefficient: 0.8586.

Table 3
Confusion matrix of classification for synthetic image 2

Class	1	2	3	4	5	6	Region size (pixels)
1	6436	158	189	0	0	0	6783
2	253	6367	47	42	0	0	6709
3	100	57	7189	222	81	0	7649
4	0	227	127	7045	161	75	7635
5	0	0	94	23	6543	167	6827
6	0	0	0	43	239	6461	6743

κ Coefficient: 0.7973.

Table 4
Distance among the class density functions in SPOT image

Class	1	2	3	4	5	6
1	0.0000	2.0000	2.0000	2.0000	1.9895	1.8028
2	2.0000	0.0000	1.8359	1.9873	2.0000	2.0000
3	2.0000	1.8359	0.0000	1.8931	1.9499	1.9919
4	2.0000	1.9873	1.8931	0.0000	2.0000	2.0000
5	1.9895	2.0000	1.9499	2.0000	0.0000	1.9925
6	1.8028	2.0000	1.9919	2.0000	1.9925	0.0000

The classification procedure generates a segmentation of the multi-spectral image. Fig. 4b shows the result of the segmentation once the class and border pixels are encoded. The overlay of the RGB composite plus the border of the regions (Fig. 4b) shows a precise segmentation of the spectral classes defined in this process. The evaluation of the goodness of this segmentation is as follows:

(i) A visual inspection of the overlay of the class borders to the RGB composite. This inspection shows a precise segmentation of the image according to the classes defined.
(ii) A verification of each region growth with ancillary data. The ancillary data was used to corroborate that each region contains the class signaled by the seed pixels. The resulting classification map was also corroborated with ancillary data.
(iii) A comparison of the κ coefficient for the contextual vs. a per-pixel classification (Richards and Jia, 1999). The κ coefficient is calculated from the confusion matrix resulting from the contextual classification process (Table 6). The value of the κ coefficient is 0.98823. On the other hand, using the grown regions as training sets as input to a per-pixel classifier, the first two principal components of the SPOT image were classified into six classes. A mode filter employing windows of 5×5 and 7×7 pixels was applied to this classification. The resulting κ coefficient for this set of images is: 0.81198 for no filter, 0.91169 for the 5×5 filter, and 0.93382 for the 7×7 filter. Increasing the size of the window in the mode filter did not bring a significant improvement to the κ coefficient. The contextual classification was performed with a window of 5×5 pixels. The spatial detail achieved in this classification is comparable with that obtained from the per-pixel classifier employing a mode filter of 5×5 pixels.

Table 5
Distance among pair of classes and class density functions in SPOT image

Class, Class pair	1	2	3	4	5	6
1,2	1.0000	1.0000	1.8887	1.9875	1.9948	1.8773
1,3	1.0000	1.8913	1.0000	1.9204	1.9697	1.8701
1,4	1.0000	1.9936	1.9262	1.0000	1.9948	1.8773
1,5	0.9948	2.0000	1.9504	2.0000	0.9948	1.8710
1,6	0.9014	2.0000	1.9960	1.9078	1.9910	0.9014
2,3	2.0000	0.9180	0.9180	0.9936	1.9749	1.9928
2,4	2.0000	0.9936	1.8148	1.9875	2.0000	2.0000
2,5	1.9895	1.0000	1.8391	1.9875	1.0000	1.9925
2,6	1.8507	1.0000	1.8846	1.9875	1.9963	1.0000
3,4	2.0000	1.8849	0.9466	0.9466	1.9749	1.9928
3,5	1.9895	1.8913	0.9749	1.9204	0.9749	1.9882
3,6	1.8507	1.8913	0.9960	1.9204	1.9712	0.9960
4,5	1.9895	1.9936	1.8766	1.0000	1.0000	1.9925
4,6	1.8507	1.9936	1.9222	1.0000	1.9963	1.0000
5,6	1.8407	2.0000	1.9478	2.0000	0.9963	0.9963

Table 6
Confusion matrix for classification of SPOT image

Class	1	2	3	4	5	6	Region size (pixels)
1	730	0	0	0	18	15	763
2	0	199	6	0	0	0	205
3	0	0	918	0	0	0	918
4	0	0	13	772	0	0	785
5	0	0	0	0	122	0	122
6	0	0	0	0	1	3475	3476
Total	730	199	937	772	141	3490	6269
κ Coefficient: 0.98823							

κ Coefficient: 0.98823.

5. Conclusions

A new contextual classifier based upon an automated region-growth algorithm has been developed and tested. This algorithm provides valid statistical samples of defined classes as input into a contextual spectral classifier. Even though the grown regions do not encircle pure spectral objects, the resulting classification is better compared to the one from a per-pixel classifier. The contextual classifier is semi-automatic, embracing only two parameters derived from heuristic criteria. More work is needed regarding the estimation of these parameters though. In order to avoid the heuristic determination of their value, a model-based approach may be adequate. The best location to seed the initial pixels for region growth requires some attention too.

The classification and growth of the regions are performed employing best windows for each class. No a priori assumptions are made concerning the density functions of the classes. This is a basic premise since, based on experimentation, in a multi-spectral image some classes show a Gaussian behavior and some a Rayleigh behavior. A classifier is prone to produce wrong classification results whenever a particular density function is assumed and the image classes show a different statistical behavior. The synthetic images have a set of classes barely discernible due to the noise content. However, the contextual classifier produces good results, generating segmentation where the classes are clearly differentiated. The classification results concerning the SPOT image are good as well. The κ coefficient shows a greater spectral separability of the contextual classifier compared to a per-pixel classifier. Our contextual classifier performs well for multi-spectral images; a direct application might be in the classification of satellite images for remote sensing of the environment. The classification results in our classifier show a κ coefficient competitive with any other contextual classifier. Finally, the rationale of classification presented in this work is of general nature and might be adapted to new models of texture and spectral separability.

Acknowledgements

This work is supported by project CONACYT 27838A.

References

Arai, K., 1992. A supervised thematic mapper classification with a purification of training samples. International Journal of Remote Sensing 13 (11), 2039–2049.

Arai, K., 1993. A classification method with a spatial-spectral variability. International Journal of Remote Sensing 14 (14), 699–709.

Atkinson, P.M., Lewis, P., 2000. Geostatistical classification for remote sensing: an introduction. Computers & Geosciences 26 (4), 361–371.

Bolstad, P.V., Lillesand, T.M., 1991. Semi-automated training approaches for spectral class definition. International Journal of Remote Sensing 13 (16), 3157–3166.

Buchheim, M.P., Lillesand, T.M., 1989. Semi-automated training field extraction and analysis for efficient digital image classification. Photogrammetric Engineering & Remote Sensing 55 (9), 1347–1355.

Cetin, H., Levandowsky, D.W., 1991. Interactive classification and mapping of multi-dimensional remotely sensed data using *n*-dimensional probability density functions (*n*pdf). Photogrammetric Engineering & Remote Sensing 57 (12), 1579–1587.

Chica-Olmo, M., Abarca-Hernández, F., 2000. Computing geostatistical image texture for remotely sensed data classification. Computers & Geosciences 26 (4), 373–383.

Cohen, F.S., Fam, Z., 1992. Maximum likelihood unsupervised textured image segmentation. CVGIP: Graphical Models and Image Processing 54 (3), 239–251.

Dougherty, E.R., 1999. Random Processes for Image and Signal Processing. SPIE Optical Engineering Press, Bellingham, WA, 592pp.

Gahegan, M., Flack, J.C., 1999. Recent developments towards integrating scene understanding within a geographic information system for agricultural applications. Transactions in GIS 3 (1), 31–50.

Gong, P., Howarth, P.J., 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification. Photogrammetric Engineering & Remote Sensing 58 (4), 423–437.

González Alonso, F., Lopez Soria, S., 1991. Using contextual information to improve land use classification of satellite images in Central Spain. International Journal of Remote Sensing 12 (11), 2227–2235.

Hodgson, M.E., 1998. What size window for image classification? A cognitive perspective. Photogrammetric Engineering & Remote Sensing 64 (8), 797–807.

Khazenie, N., Crawford, M.M., 1990. Spatial-temporal autocorrelated model for contextual classification. IEEE Transactions on Geoscience and Remote Sensing 28 (4), 529–539.

Kontoes, C.C., Rokos, D., 1996. The integration of spatial context information in an experimental knowledge-based system and the supervised relaxation algorithm—two successful approaches to improving spot-xs classification. International Journal of Remote Sensing 17 (16), 3093–3106.

Landis, J., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Lira, J., Frulla, L., 1998. An automated region-growing algorithm for segmentation of texture regions in SAR images. International Journal of Remote Sensing 19 (18), 3595–3606.

Raafat H, .M., Wong, A.K., 1988. A texture information-directed region-growth algorithm for image segmentation and region classification. Computer Vision, Graphics and Image Processing 43 (1), 1–21.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis, An Introduction, 3rd Edition. Springer, Berlin, 363pp.

Smits, P.C., Dellepiane, S.G., Schowengerdt, R.A., 1999. Quality assesment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. International Journal of Remote Sensing 20 (8), 1461–1486.

Stuckens, J., Coppin, P.R., Bauer, M.E., 2000. Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing of Environment 71 (3), 282–296.

4.2 Technical Reports

Technical reports written during my Ph.D. Studies are included here. The list of reports follows.

- Principal Component Analysis of Psoriasis Lesions Images
- A Hierarchical Classification Scheme of Psoriasis Images
- Illumination Correction in Psoriasis Lesions Images
- Texture Alteration Detection in Bitemporal Images of Lesions with Psoriasis
- A Combined Alignment and Registration scheme of Psoriasis Lesions Images
- Change Detection in Registered Psoriasis Lesions Image Patterns

4.2.1 Principal Component Analysis of Psoriasis Lesions Images

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 5
Pages: 1 – 28
Reference: [36]

Principal Component Analysis of Psoriasis Lesions Images

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. A set of *RGB* images of psoriasis lesions is used. By visual examination of these images, there seem to be no common pattern that could be used to find and align the lesions within and between sessions. It is expected that the principal components of the original images could be useful during future lesion segmentation and alignment purposes.

1 Introduction

We use a set of 175 *RGB* psoriasis lesions images, of size 556×748 pixels, taken at the Gentofte Hospital, Denmark, during pilot sessions with three invited patients. For each patient, three lesions were were followed once a week during at least three weeks. In each session, five images of each lesion were taken. The images were labelled with four characters, indicating patient (1, 2, 3), lesion (A, B, C), session (a, b, c, d) and capture (1, 2, 3, 4, 5) respectively.

In order to study the variation within and between sessions of a lesions, they first have to be found and matched. In order to find a lesion, the set of image bands, or combination of bands where the lesion is expected to be found has to be specified. Looking at the original images, we can appreciate that for some cases, the lesion has a light color, while the skin is dark, while, for other cases, the opposite situation is seen. This means, that in the whole set of original images there is no common pattern of lesions that could be defined. It could be nice, for instance, to locate the lesion in the same part of the histogram for all the cases. For this reason, in order to be able to define a common pattern of lesion, the linear combination of the original bands given by the principal components is here explored.

Much of the present work was done with the program *maf* [1], which is available at the server of the Institute of Informatics and Mathematical Modelling.

2 Principal Component Analysis

The borders of the original images were cut by 10 pixels on each side on before hand. This was done, because of camera problems. Some artefacts appeared on the image borders.

Principal Component Decomposition has been applied to each single psoriasis lesion image. For all the cases, the variance-covariance matrix was diagonalized. For each single image, most of the contribution in its first component is given by the red band, which has an average absolute correlation with the first component of 0.9907. When the correlation matrix is diagonalized, all the bands have almost the same weight in the first component. In Figure 1 and Table 1, it can be seen that all the bands are highly correlated with the first component.

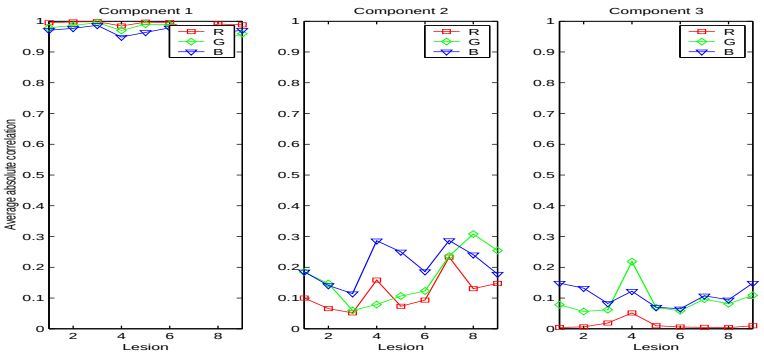


Fig. 1. Average absolute correlation per lesion between bands of single images and their principal components.

Band	Comp. 1	Comp. 2	Comp. 3
Red	0.9907	0.1170	0.0128
Green	0.9741	0.1664	0.0921
Blue	0.9674	0.2077	0.1078

Table 1. Average absolute correlation values between bands of single images and their principal components

For a given patient and lesion, the captures taken within and between sessions are ordered in a sequence. In order to be able to compare the elements of the sequence given by the corresponding principal components, all the eigenvectors

were oriented to the same sense than the eigenvectors of the first image of that sequence[1]. For each image in the sequence, two scalar products were computed and compared. The first one, was the scalar of the eigenvectors provided by the program *maf* by the eigenvectors of the first image of the first session of that patient and lesion. The second scalar product, was the same as before, but the first eigenvector had the sign [2] changed. The eigenvectors sign, for which the scalar product was positive was selected. The images whose components therefore had to be flipped were: $1Cb1$, $1Cb2$, $1Cb3$, $1Cb4$, $1Cb5$, $1Cc2$, $1Cc3$, $2Bd1$, $2Bd3$, $2Bd4$, $2Bd5$, $3Ca3$, $3Ca4$, $3Ca5$, $3Cb1$, $3Cb2$, $3Cb3$, $3Cb4$, $3Cb5$, $3Cc1$, $3Cc2$, $3Cc3$, $3Cc4$, $3Cc5$, $3Cd1$, $3Cd2$, $3Cd3$, $3Cd4$ and $3Cd5$.

2.1 Eigenvector Values

Tables 2 and 3 show the average and standard deviation values of the eigenvectors. This can be found in detail as one table per patient and lesion, in Appendix B. For a given patient and lesion, each single cell value in Table 2 is the corresponding column data average of the table, with the eigenvector values of the images for this patient and lesion, included in the Appendix B. For a given patient and lesion, each single cell value in Table 3 is the corresponding column data standard deviation of the table, with the eigenvector values of the images for this patient and lesion, included in Appendix B.

Table 2. Average Eigenvector Values

(Patient, Lesion)	Red $1^{st}C.$	Green $1^{st}C.$	Blue $1^{st}C.$	Red $2^{nd}C.$	Green $2^{nd}C.$	Blue $2^{nd}C.$	Red $3^{rd}C.$	Green $3^{rd}C.$	Blue $3^{rd}C.$
(1,A)	0.8083	0.4781	0.3429	-0.5847	0.6401	0.4845	-0.0517	-0.5922	0.7975
(1,B)	0.8279	0.4660	0.3111	-0.5422	0.6864	0.4192	-0.0977	-0.5171	0.8207
(1,C)	0.7814	0.4999	0.3732	-0.5463	0.4249	0.5698	-0.2980	-0.6469	0.6070
(2,A)	0.7334	0.4909	0.4650	-0.6098	0.2071	0.7475	-0.2738	0.8339	-0.4552
(2,B)	0.8252	0.4582	0.3271	-0.5191	0.4052	0.7443	0.2107	-0.7843	0.5744
(2,C)	0.7781	0.4859	0.3964	-0.6119	0.4708	0.6260	0.1190	-0.7293	0.6655
(3,A)	0.7303	0.5049	0.4497	-0.6741	0.4937	0.5403	0.0504	-0.7052	0.7068
(3,B)	0.8260	0.3830	0.3932	-0.5452	0.6376	0.5251	-0.0539	-0.6606	0.7464
(3,C)	0.7781	0.4480	0.4370	-0.6181	0.6483	0.4331	0.0868	0.6077	-0.7843

2.2 Eigenvalues

In order to evaluate how much of the total variance is explained by each of the eigenvalues, Table 4 was constructed using the data of Table 16. Each cell in

[1] The eigenvectors give a direction, which has two senses or orientations.
[2] This means, that the other orientation of the eigenvector was tested.

Table 3. Standard Deviation of the Eigenvector Values

(Patient, Lesion)	Red $1^{st}C.$	Green $1^{st}C.$	Blue $1^{st}C.$	Red $2^{nd}C.$	Green $2^{nd}C.$	Blue $2^{nd}C.$	Red $3^{rd}C.$	Green $3^{rd}C.$	Blue $3^{rd}C.$
(1,A)	0.0138	0.0153	0.0126	0.0177	0.0736	0.0423	0.0666	0.0765	0.0589
(1,B)	0.0129	0.0164	0.0149	0.0312	0.1270	0.2125	0.1023	0.1738	0.1072
(1,C)	0.0076	0.0045	0.0103	0.0162	0.2958	0.3502	0.0445	0.2727	0.2385
(2,A)	0.0459	0.0244	0.0498	0.0488	0.1421	0.0731	0.1076	0.0314	0.1048
(2,B)	0.0243	0.0283	0.0277	0.0355	0.0932	0.0548	0.0592	0.0421	0.0758
(2,C)	0.0220	0.0258	0.0128	0.0216	0.0906	0.0627	0.0726	0.0446	0.0666
(3,A)	0.0684	0.0470	0.0563	0.0731	0.0429	0.0573	0.0148	0.0127	0.0133
(3,B)	0.0585	0.0869	0.0791	0.1177	0.0651	0.0582	0.0403	0.0402	0.0344
(3,C)	0.0330	0.0260	0.0337	0.0445	0.0724	0.0580	0.0452	0.0653	0.0460

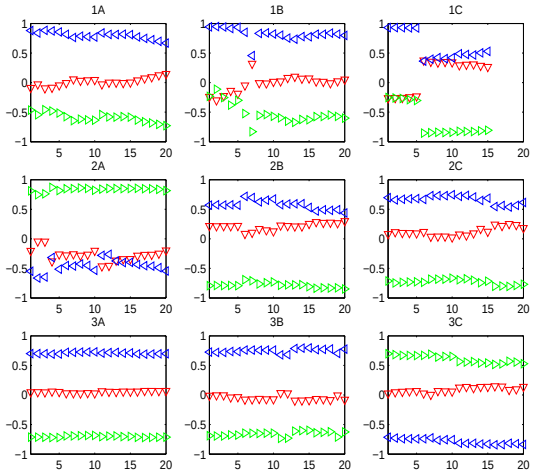


Fig. 2. Values of the third eigenvectors of the images of the data-set. Triangle down indicates the weight for the red band; triangle right, for the green band; and triangle left, for the blue band.

Table 4 is computed as follows:

$$p_{p,l,r} = 100\frac{1}{n_s n_c}\sum_{s=1}^{n_s}\sum_{c=1}^{n_c}\frac{\lambda_{p,l,s,c,r}}{\sum_{t=1}^{3}\lambda_{p,l,s,c,t}} \quad (1)$$

where p is the patient number, l corresponds to the lesion and r to the number of the principal component. The index s varies over the sessions and the index c, over captures from the given session. $\lambda_{p,l,s,c,r}$ is the eigenvalue of the $r-th$ principal component of the $c-th$ capture taken during the $s-th$ session of lesion l of patient p.

Table 4. Average and standard deviation per patient and lesion of the percentage of the explained total variance of the principal components of the original images

(Patient, Lesion)	Mean 1^{st} C.	Std. Dev. 1^{st} C.	Mean 2^{nd} C.	Std. Dev. 2^{nd} C.	Mean 3^{rd} C.	Std. Dev. 3^{rd} C.
(1,A)	97.65	0.74	1.92	0.62	0.43	0.15
(1,B)	98.65	0.72	1.08	0.74	0.27	0.08
(1,C)	99.19	0.19	0.56	0.16	0.25	0.07
(2,A)	94.70	1.96	3.47	1.13	1.83	0.90
(2,B)	98.42	0.76	1.41	0.73	0.17	0.06
(2,C)	98.26	1.01	1.58	0.97	0.16	0.05
(3,A)	93.42	2.20	6.10	2.13	0.48	0.11
(3,B)	95.76	2.52	3.97	2.52	0.27	0.06
(3,C)	95.88	1.38	3.41	1.23	0.71	0.31

3 Discussion

From Table 4 it can be deduced, that, for all the cases, almost all the total variance is explained by the first principal component.

If we think of the images as being composed of three classes: background, normal skin and lesion, looking at the images shown in Appendix A we can deduce that for the cases $(1, A)$, $(1, B)$ and $(1, C)$ the first component histogram peak corresponding to the lesion is the central one, while for the remaining cases, it is in one of the extremes. On the other hand, it is interesting to observe that there is more contrast between the lesions and the rest in the third principal component. In fact, they are roughly enhanced from the rest. This allowed us, for instance, to find out, that the visual registration of some lesions, during the image acquisition process, was of regular quality. This detail that can not easily seen in the first principal component. Compare, for instance, the images of the first and third session of the case $(2, C)$.

Taking a look at the values of the third eigenvectors, it can be deduced that a rough estimation of the third component can be obtained by subtracting the green band from the blue band, while the red band almost does not have any influence (See Figure 2). Note also in Table 5, that the large variance of the third eigenvector values for the case $(1, C)$ is due to the fact that the magnitude of the weights for the B and G band are kind of flipped. For the blue bands of the first session, the magnitude of the weights -in the third component- is more similar to the magnitude of the weights for the green band in the next sessions, and viceversa.

The $|B - G|$ bands do not have, for all the cases, the probability density function of the lesion class located in one extreme of the normalized histogram. However, the contrast between the lesion and the normal skin is still significant. See, for instance, the case $(2, C)$, or note in the case $(2, A)$ how clear the lesions are enhanced from the rest.

4 Conclusions

The R, G, B bands are highly correlated with the first principal component, where the largest contribution is given by the red band. There is high contrast between the lesions and the rest in the third principal component. This component could roughly be estimated by subtracting the blue from the green band. In this estimation, there is a high contrast between the lesion and normal skin. It is tempting to use this estimation of the third principal component as input for a classification scheme. On the other hand, it has been shown that it could be useful to provide the person who carries out the image acquisition with the third principal component in order to do a better visual registration of the lesion to be captured.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A Principal Components

In this section, for the set of images corresponding to a given patient and lesion, the principal components are shown in one single figure. Each figure is arranged

as follows. From top to bottom, three blocks are considered. All the blocks have the same number of rows. The number of rows of the block is the number of sessions for which images of the same patient and lesion has been taken. The first block (first third of the rows) is composed by the first principal components of the images of the given patient and lesion. The second block (second third of the rows), by the second principal components, and so on, for the third block (remaining rows). Within a block, the row number indicates the number of the session, and the column number indicates the number of capture within a session.

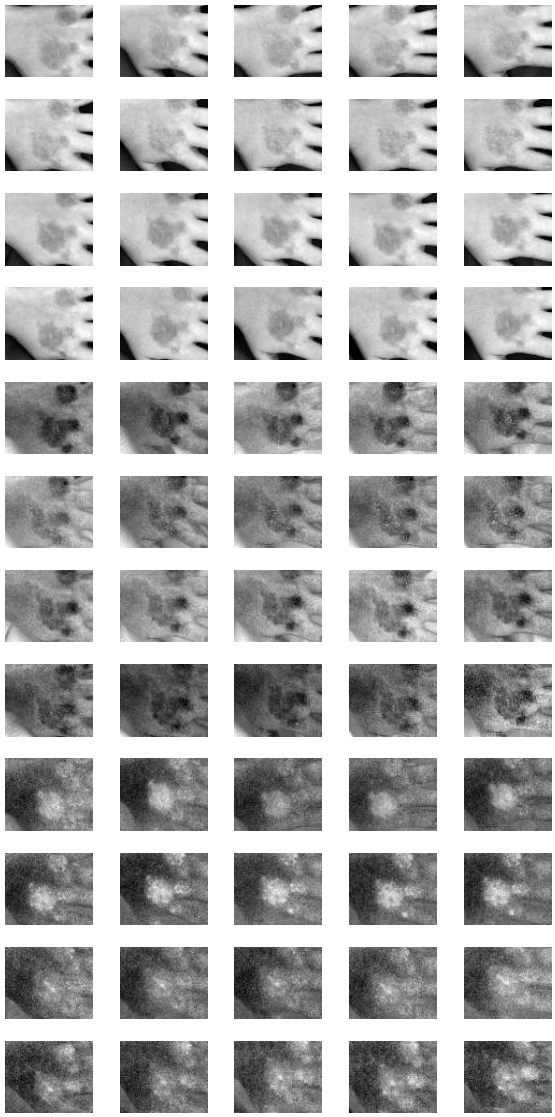


Fig. 3. Principal Components of the images of the case (Patient 1, Lesion A).

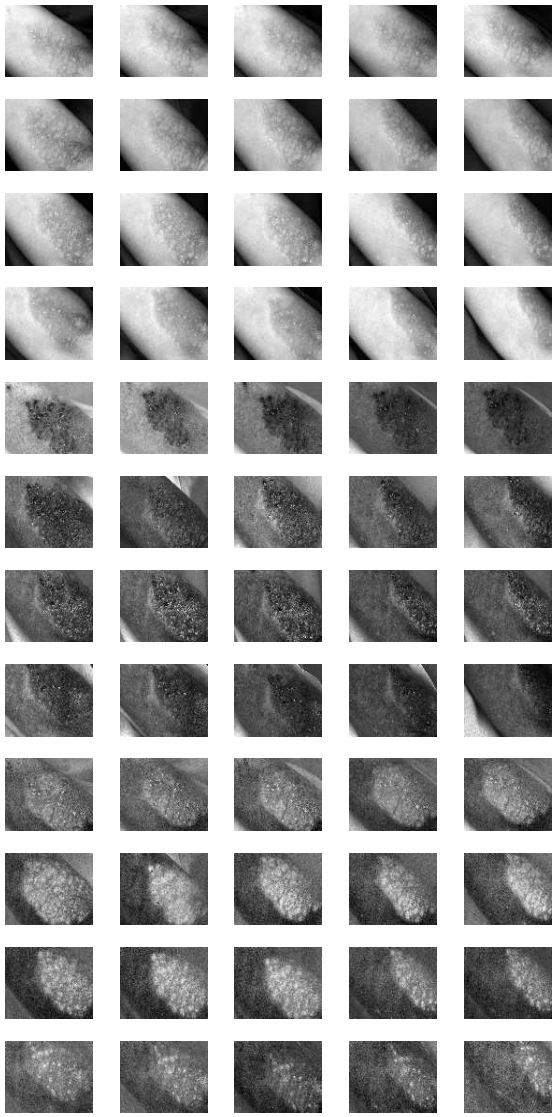


Fig. 4. Principal Components of the images of the case (Patient 1, Lesion B).

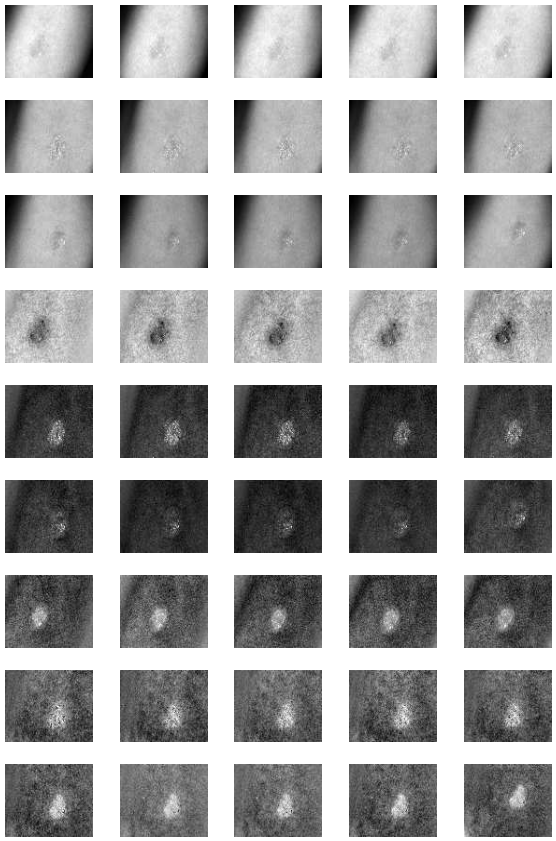


Fig. 5. Principal Components of the images of the case (Patient 1, Lesion C).

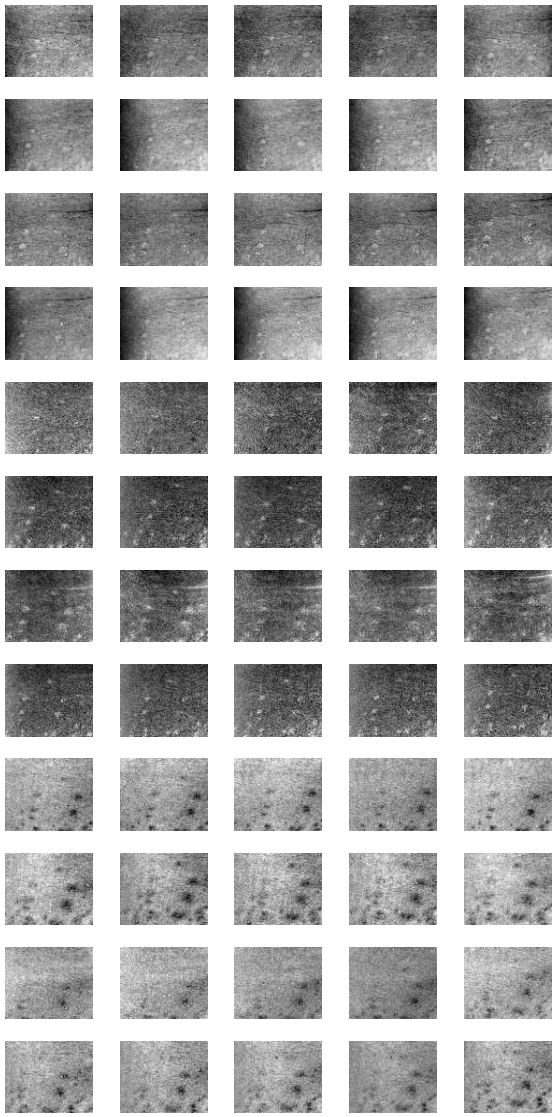


Fig. 6. Principal Components of the images of the case (Patient 2, Lesion A).

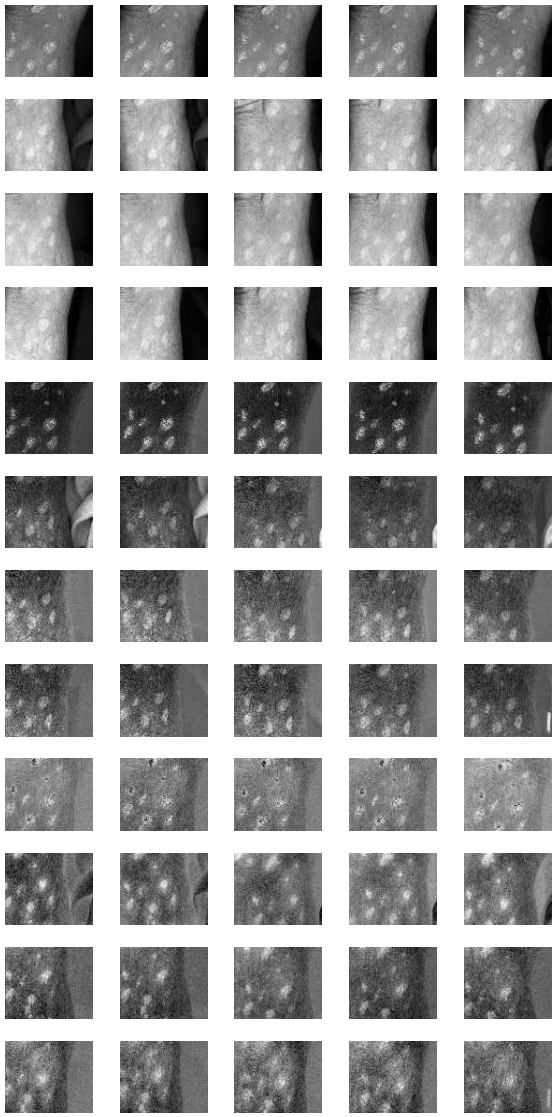


Fig. 7. Principal Components of the images of the case (Patient 2, Lesion B).

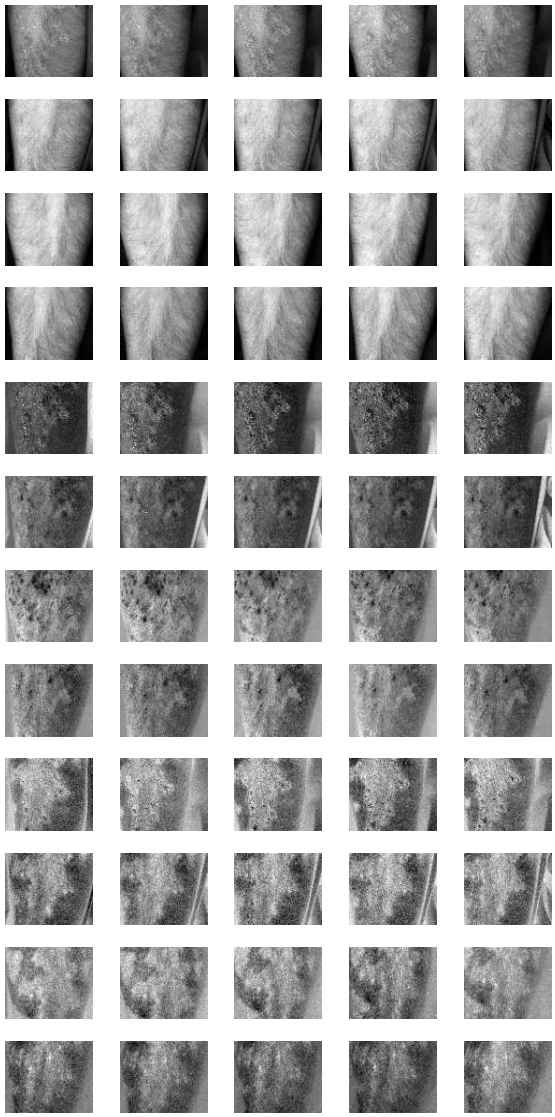


Fig. 8. Principal Components of the images of the case (Patient 2, Lesion C).

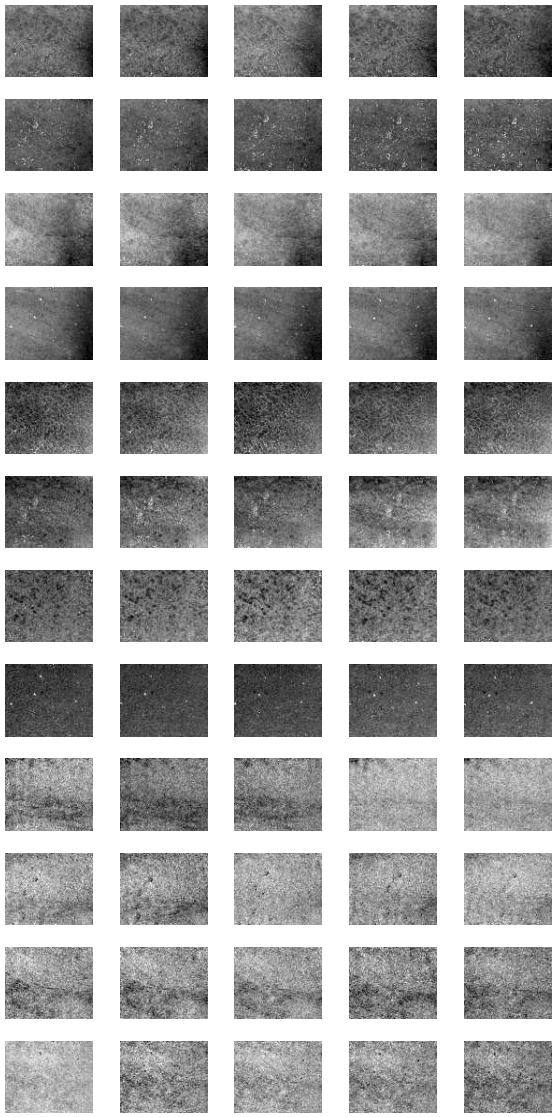


Fig. 9. Principal Components of the images of the case (Patient 3, Lesion A).

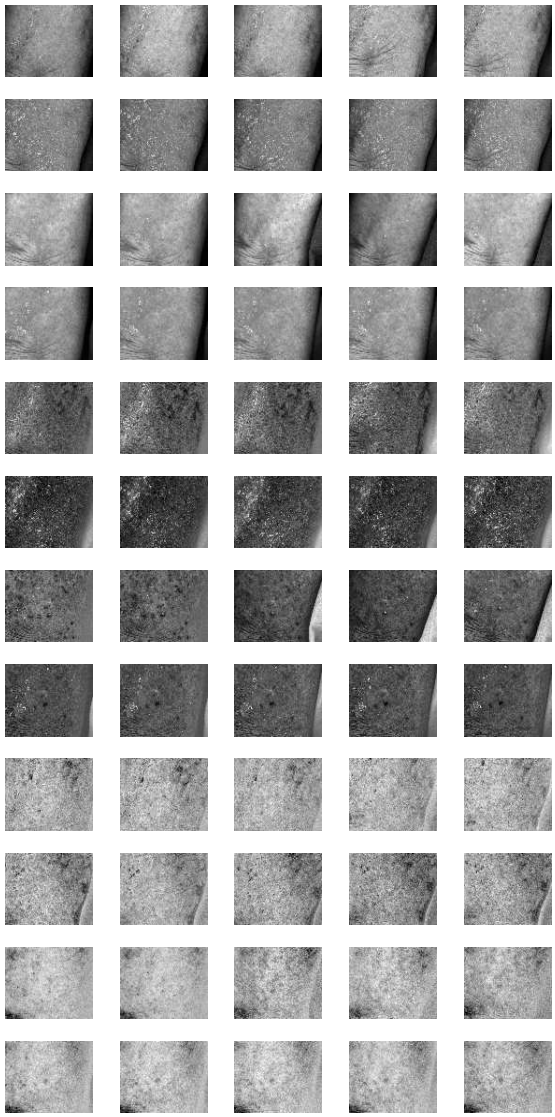


Fig. 10. Principal Components of the images of the case (Patient 3, Lesion B).

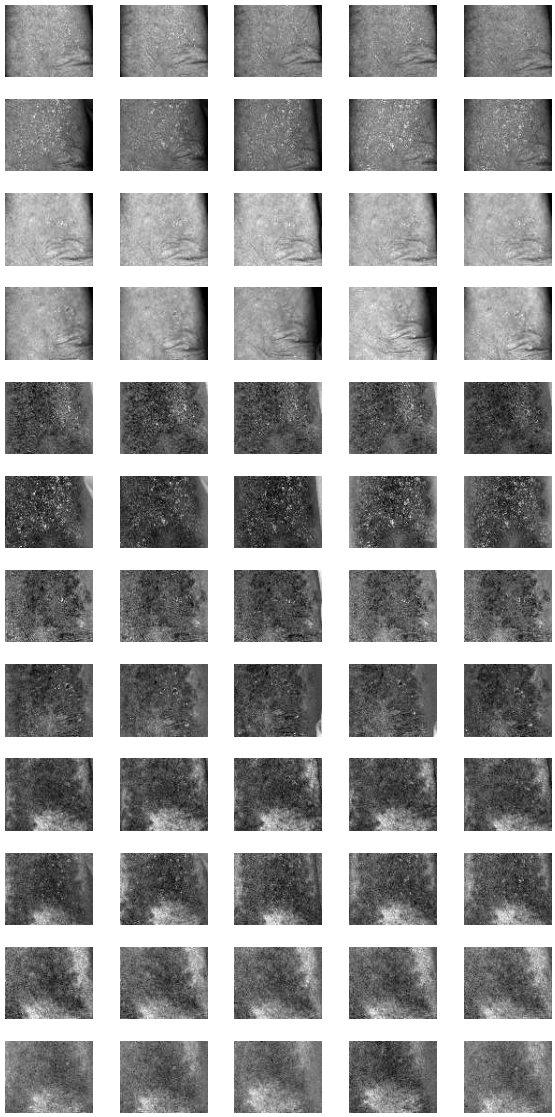


Fig. 11. Principal Components of the images of the case (Patient 3, Lesion C).

B Eigenvector Values of the Original Images

Table 5. Eigenvectors Values of the original images of (Patient 1, Lesion A)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7938	0.5007	0.3452	-0.6015	0.7302	0.3241	-0.0898	-0.4649	0.8808
a2	0.8216	0.4642	0.3308	-0.5692	0.6986	0.4336	-0.0298	-0.5445	0.8382
a3	0.8048	0.4871	0.3392	-0.5858	0.7440	0.3215	-0.0957	-0.4574	0.8841
a4	0.8051	0.4827	0.3446	-0.5869	0.7321	0.3457	-0.0854	-0.4806	0.8728
a5	0.8398	0.4434	0.3133	-0.5408	0.7333	0.4120	-0.0470	-0.5154	0.8556
b1	0.8017	0.4847	0.3498	-0.5977	0.6576	0.4586	-0.0077	-0.5767	0.8169
b2	0.8277	0.4564	0.3265	-0.5584	0.6131	0.5588	0.0549	-0.6449	0.7623
b3	0.8273	0.4570	0.3267	-0.5609	0.6398	0.5254	0.0311	-0.6179	0.7856
b4	0.8133	0.4712	0.3413	-0.5805	0.6177	0.5306	0.0392	-0.6297	0.7759
b5	0.8179	0.4662	0.3371	-0.5735	0.6145	0.5417	0.0455	-0.6364	0.7700
c1	0.8009	0.4904	0.3437	-0.5981	0.6830	0.4193	-0.0291	-0.5414	0.8403
c2	0.8024	0.4863	0.3459	-0.5968	0.6532	0.4661	0.0008	-0.5804	0.8143
c3	0.8084	0.4784	0.3431	-0.5887	0.6566	0.4715	0.0003	-0.5832	0.8124
c4	0.7968	0.4911	0.3521	-0.6043	0.6545	0.4545	-0.0073	-0.5749	0.8182
c5	0.8189	0.4665	0.3344	-0.5739	0.6638	0.4795	0.0017	-0.5846	0.8113
d1	0.8061	0.4760	0.3516	-0.5906	0.6092	0.5292	0.0377	-0.6342	0.7722
d2	0.7904	0.4942	0.3620	-0.6085	0.5650	0.5573	0.0709	-0.6607	0.7473
d3	0.8041	0.4803	0.3504	-0.5875	0.5517	0.5920	0.0910	-0.6819	0.7258
d4	0.7852	0.4993	0.3662	-0.6069	0.5029	0.6155	0.1232	-0.7055	0.6979
d5	0.7988	0.4857	0.3550	-0.5835	0.4820	0.6536	0.1464	-0.7292	0.6684

Table 6. Eigenvectors Values of the original images of (Patient 1, Lesion B)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.8089	0.4881	0.3279	-0.5361	0.8413	0.0700	-0.2416	-0.2324	0.9421
a2	0.8169	0.4822	0.3165	-0.4924	0.8688	-0.0529	-0.3005	-0.1126	0.9471
a3	0.8170	0.4759	0.3257	-0.5299	0.8423	0.0987	-0.2274	-0.2532	0.9403
a4	0.8244	0.4669	0.3199	-0.5485	0.7985	0.2480	-0.1396	-0.3799	0.9144
a5	0.8218	0.4737	0.3166	-0.5379	0.8282	0.1573	-0.1877	-0.2996	0.9354
b1	0.8472	0.4299	0.3122	-0.5292	0.7353	0.4235	-0.0475	-0.5240	0.8504
b2	0.8351	0.4739	0.2793	-0.4483	0.2921	0.8448	0.3188	-0.8307	0.4564
b3	0.8372	0.4512	0.3092	-0.5469	0.6992	0.4605	-0.0085	-0.5546	0.8321
b4	0.8366	0.4544	0.3062	-0.5477	0.7078	0.4461	-0.0141	-0.5408	0.8410
b5	0.8452	0.4453	0.2954	-0.5342	0.6901	0.4882	0.0135	-0.5705	0.8212
c1	0.8271	0.4611	0.3215	-0.5617	0.6577	0.5020	0.0200	-0.5957	0.8029
c2	0.8184	0.4741	0.3248	-0.5692	0.5908	0.5718	0.0792	-0.6528	0.7534
c3	0.8188	0.4745	0.3230	-0.5652	0.5680	0.5983	0.1004	-0.6725	0.7333
c4	0.8268	0.4719	0.3062	-0.5584	0.6227	0.5481	0.0680	-0.6241	0.7784
c5	0.8276	0.4704	0.3062	-0.5568	0.6193	0.5536	0.0708	-0.6287	0.7745
d1	0.8387	0.4548	0.2996	-0.5443	0.6804	0.4907	0.0193	-0.5746	0.8182
d2	0.8297	0.4633	0.3113	-0.5580	0.6751	0.4826	0.0134	-0.5741	0.8187
d3	0.8090	0.4870	0.3291	-0.5877	0.6801	0.4383	-0.0104	-0.5480	0.8364
d4	0.8151	0.4852	0.3165	-0.5789	0.6615	0.4767	0.0220	-0.5718	0.8201
d5	0.8568	0.4366	0.2745	-0.5131	0.6688	0.5380	0.0513	-0.6018	0.7970

Table 7. Eigenvectors Values of the original images of (Patient 1, Lesion C)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7917	0.4946	0.3585	-0.5512	0.8314	0.0703	-0.2633	-0.2533	0.9309
a2	0.7923	0.4928	0.3598	-0.5519	0.8302	0.0783	-0.2601	-0.2606	0.9298
a3	0.7913	0.4922	0.3626	-0.5541	0.8281	0.0850	-0.2584	-0.2682	0.9281
a4	0.7860	0.4991	0.3647	-0.5673	0.8168	0.1050	-0.2455	-0.2895	0.9252
a5	0.7889	0.4967	0.3620	-0.5687	0.8132	0.1236	-0.2329	-0.3034	0.9240
b1	0.7741	0.5027	0.3848	-0.5107	0.1369	0.8488	0.3740	-0.8536	0.3627
b2	0.7763	0.5015	0.3819	-0.5222	0.1723	0.8353	0.3531	-0.8478	0.3956
b3	0.7714	0.5054	0.3867	-0.5387	0.1950	0.8196	0.3388	-0.8406	0.4227
b4	0.7717	0.5049	0.3866	-0.5311	0.1774	0.8285	0.3498	-0.8447	0.4051
b5	0.7702	0.5071	0.3868	-0.5382	0.1915	0.8208	0.3421	-0.8404	0.4204
c1	0.7842	0.4977	0.3706	-0.5473	0.2736	0.7909	0.2922	-0.8231	0.4870
c2	0.7796	0.5016	0.3751	-0.5482	0.2567	0.7960	0.3030	-0.8262	0.4750
c3	0.7802	0.5009	0.3747	-0.5445	0.2491	0.8009	0.3078	-0.8289	0.4671
c4	0.7815	0.5005	0.3724	-0.5557	0.2871	0.7802	0.2836	-0.8167	0.5025
c5	0.7817	0.5012	0.3711	-0.5642	0.3148	0.7633	0.2657	-0.8060	0.5289

Table 8. Eigenvectors Values of the original images of (Patient 2, Lesion A)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7357	0.4960	0.4611	-0.6460	0.3094	0.6978	-0.2035	0.8113	-0.5480
a2	0.7226	0.4835	0.4940	-0.6898	0.4574	0.5613	-0.0454	0.7463	-0.6640
a3	0.7002	0.4862	0.5227	-0.7123	0.4275	0.5566	-0.0472	0.7621	-0.6457
a4	0.6893	0.4940	0.5299	-0.6160	0.0147	0.7876	-0.3812	0.8694	-0.3144
a5	0.7205	0.5208	0.4578	-0.6413	0.2493	0.7257	-0.2638	0.8164	-0.5137
b1	0.7558	0.4836	0.4415	-0.5923	0.2174	0.7758	-0.2792	0.8479	-0.4507
b2	0.7917	0.4609	0.4009	-0.5524	0.2601	0.7920	-0.2607	0.8485	-0.4605
b3	0.7804	0.4649	0.4182	-0.5571	0.2133	0.8026	-0.2839	0.8593	-0.4254
b4	0.7805	0.4655	0.4173	-0.5652	0.2400	0.7893	-0.2672	0.8519	-0.4504
b5	0.7434	0.4822	0.4636	-0.6380	0.3029	0.7080	-0.2010	0.8220	-0.5328
c1	0.6783	0.5409	0.4974	-0.5692	-0.0414	0.8212	-0.4647	0.8401	-0.2798
c2	0.6758	0.5230	0.5194	-0.5789	-0.0596	0.8132	-0.4563	0.8503	-0.2625
c3	0.6790	0.5200	0.5182	-0.6357	0.0633	0.7693	-0.3672	0.8518	-0.3735
c4	0.6831	0.5192	0.5136	-0.6447	0.0981	0.7581	-0.3433	0.8490	-0.4018
c5	0.6524	0.5192	0.5521	-0.6771	0.0721	0.7323	-0.3404	0.8516	-0.3986
d1	0.7510	0.4804	0.4530	-0.5957	0.1968	0.7788	-0.2850	0.8547	-0.4339
d2	0.7844	0.4682	0.4068	-0.5583	0.2473	0.7919	-0.2702	0.8483	-0.4554
d3	0.7884	0.4686	0.3986	-0.5533	0.2570	0.7923	-0.2689	0.8452	-0.4619
d4	0.7838	0.4697	0.4062	-0.5687	0.2802	0.7733	-0.2494	0.8372	-0.4868
d5	0.7716	0.4712	0.4273	-0.6049	0.3358	0.7220	-0.1967	0.8156	-0.5441

Table 9. Eigenvectors Values of the original images of (Patient 2, Lesion B)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.8076	0.4736	0.3514	-0.5486	0.3846	0.7423	0.2164	-0.7923	0.5705
a2	0.8015	0.4789	0.3582	-0.5583	0.3847	0.7351	0.2142	-0.7891	0.5757
a3	0.8050	0.4757	0.3547	-0.5528	0.3842	0.7395	0.2155	-0.7913	0.5722
a4	0.8002	0.4788	0.3612	-0.5607	0.3836	0.7338	0.2128	-0.7897	0.5754
a5	0.7937	0.4831	0.3696	-0.5678	0.3706	0.7350	0.2181	-0.7932	0.5685
b1	0.8928	0.3729	0.2527	-0.4424	0.6201	0.6479	0.0849	-0.6903	0.7185
b2	0.8767	0.3961	0.2729	-0.4698	0.5835	0.6625	0.1032	-0.7090	0.6976
b3	0.8305	0.4508	0.3273	-0.5312	0.4637	0.7091	0.1679	-0.7627	0.6245
b4	0.8316	0.4499	0.3256	-0.5360	0.4964	0.6829	0.1455	-0.7424	0.6540
b5	0.8460	0.4342	0.3095	-0.5178	0.5305	0.6711	0.1272	-0.7280	0.6737
c1	0.8167	0.4751	0.3275	-0.5312	0.3973	0.7483	0.2254	-0.7851	0.5769
c2	0.8158	0.4761	0.3282	-0.5360	0.4094	0.7383	0.2172	-0.7782	0.5892
c3	0.8148	0.4737	0.3342	-0.5406	0.4123	0.7334	0.2096	-0.7782	0.5920
c4	0.8122	0.4770	0.3359	-0.5444	0.4127	0.7303	0.2097	-0.7760	0.5949
c5	0.8179	0.4718	0.3293	-0.5165	0.3499	0.7815	0.2535	-0.8093	0.5299
d1	0.8334	0.4572	0.3107	-0.4745	0.3033	0.8264	0.2836	-0.8361	0.4697
d2	0.8299	0.4589	0.3173	-0.4883	0.3223	0.8110	0.2699	-0.8279	0.4916
d3	0.8282	0.4598	0.3204	-0.4884	0.3119	0.8150	0.2747	-0.8315	0.4829
d4	0.8241	0.4606	0.3299	-0.5002	0.3183	0.8053	0.2659	-0.8286	0.4926
d5	0.8263	0.4599	0.3250	-0.4768	0.2642	0.8384	0.2997	-0.8478	0.4376

Table 10. Eigenvectors values of the original images of (Patient 2, Lesion C)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7914	0.4704	0.3904	-0.6060	0.5195	0.6025	0.0806	-0.7133	0.6962
a2	0.8022	0.4589	0.3820	-0.5863	0.4841	0.6496	0.1131	-0.7451	0.6573
a3	0.8000	0.4597	0.3857	-0.5917	0.4971	0.6347	0.1000	-0.7359	0.6697
a4	0.7922	0.4686	0.3910	-0.6023	0.4972	0.6244	0.0982	-0.7302	0.6762
a5	0.8019	0.4596	0.3819	-0.5903	0.5101	0.6256	0.0927	-0.7270	0.6803
b1	0.7794	0.4841	0.3977	-0.6155	0.4730	0.6305	0.1171	-0.7362	0.6666
b2	0.8021	0.4558	0.3859	-0.5962	0.5723	0.5631	0.0358	-0.6817	0.7308
b3	0.8006	0.4591	0.3850	-0.5979	0.5696	0.5639	0.0396	-0.6817	0.7306
b4	0.8011	0.4582	0.3851	-0.5972	0.5703	0.5639	0.0388	-0.6818	0.7305
b5	0.8144	0.4459	0.3715	-0.5798	0.5985	0.5528	0.0241	-0.6656	0.7459
c1	0.7603	0.5052	0.4083	-0.6453	0.5156	0.5637	0.0742	-0.6921	0.7180
c2	0.7580	0.5083	0.4087	-0.6498	0.5342	0.5408	0.0565	-0.6755	0.7352
c3	0.7564	0.5125	0.4064	-0.6474	0.4981	0.5769	0.0932	-0.6995	0.7086
c4	0.7618	0.5162	0.3914	-0.6254	0.4284	0.6522	0.1690	-0.7416	0.6492
c5	0.7573	0.5138	0.4032	-0.6419	0.4718	0.6045	0.1204	-0.7166	0.6870
d1	0.7614	0.5056	0.4057	-0.6016	0.3179	0.7328	0.2416	-0.8020	0.5463
d2	0.7622	0.4993	0.4121	-0.6095	0.3390	0.7166	0.2181	-0.7974	0.5627
d3	0.7563	0.5098	0.4101	-0.6039	0.3029	0.7372	0.2517	-0.8052	0.5370
d4	0.7506	0.5173	0.4111	-0.6181	0.3299	0.7135	0.2334	-0.7897	0.5674
d5	0.7533	0.5105	0.4146	-0.6319	0.3872	0.6714	0.1822	-0.7678	0.6142

Table 11. Eigenvectors Values of the original images of (Patient 3, Lesion A)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.6940	0.5265	0.4911	-0.7185	0.4628	0.5192	0.0461	-0.7132	0.6995
a2	0.6707	0.5414	0.5070	-0.7402	0.4449	0.5042	0.0474	-0.7135	0.6991
a3	0.7409	0.4924	0.4567	-0.6701	0.4968	0.5514	0.0447	-0.7146	0.6981
a4	0.6766	0.5371	0.5037	-0.7341	0.4383	0.5186	0.0577	-0.7207	0.6909
a5	0.6724	0.5378	0.5085	-0.7383	0.4390	0.5120	0.0522	-0.7197	0.6923
b1	0.6686	0.5498	0.5007	-0.7431	0.4680	0.4783	0.0287	-0.6919	0.7214
b2	0.6618	0.5542	0.5048	-0.7491	0.4627	0.4742	0.0292	-0.6920	0.7213
b3	0.6518	0.5608	0.5106	-0.7580	0.4593	0.4631	0.0252	-0.6889	0.7245
b4	0.6220	0.5830	0.5227	-0.7824	0.4354	0.4453	0.0321	-0.6859	0.7270
b5	0.6196	0.5839	0.5245	-0.7844	0.4375	0.4396	0.0273	-0.6838	0.7291
c1	0.7944	0.4652	0.3907	-0.6040	0.5368	0.5891	0.0643	-0.7039	0.7074
c2	0.7911	0.4687	0.3931	-0.6092	0.5449	0.5762	0.0559	-0.6952	0.7166
c3	0.7896	0.4693	0.3952	-0.6114	0.5475	0.5714	0.0518	-0.6928	0.7192
c4	0.7802	0.4744	0.4078	-0.6228	0.5279	0.5774	0.0587	-0.7045	0.7073
c5	0.7929	0.4644	0.3946	-0.6067	0.5410	0.5824	0.0570	-0.7011	0.7107
d1	0.7985	0.4512	0.3986	-0.5992	0.5316	0.5986	0.0582	-0.7168	0.6949
d2	0.7962	0.4565	0.3970	-0.6011	0.5227	0.6045	0.0684	-0.7200	0.6906
d3	0.7955	0.4583	0.3964	-0.6021	0.5242	0.6022	0.0682	-0.7177	0.6930
d4	0.7957	0.4587	0.3955	-0.6019	0.5267	0.6003	0.0670	-0.7157	0.6952
d5	0.7928	0.4634	0.3959	-0.6056	0.5262	0.5969	0.0683	-0.7130	0.6978

Table 12. Eigenvectors Values of the original images of (Patient 3, Lesion B)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7766	0.4496	0.4412	-0.6298	0.5683	0.5295	-0.0127	-0.6891	0.7245
a2	0.7863	0.4414	0.4323	-0.6178	0.5680	0.5438	-0.0055	-0.6946	0.7193
a3	0.7884	0.4395	0.4304	-0.6151	0.5695	0.5453	-0.0054	-0.6946	0.7193
a4	0.8175	0.4012	0.4132	-0.5744	0.6188	0.5358	-0.0408	-0.6754	0.7364
a5	0.8167	0.4029	0.4132	-0.5760	0.6143	0.5394	-0.0365	-0.6785	0.7337
b1	0.7753	0.4399	0.4532	-0.6268	0.6240	0.4666	-0.0775	-0.6458	0.7596
b2	0.7775	0.4386	0.4506	-0.6244	0.6238	0.4701	-0.0748	-0.6469	0.7589
b3	0.7534	0.4648	0.4652	-0.6543	0.6002	0.4601	-0.0653	-0.6510	0.7563
b4	0.7923	0.4254	0.4374	-0.6059	0.6330	0.4819	-0.0719	-0.6468	0.7593
b5	0.7959	0.4219	0.4342	-0.6012	0.6356	0.4844	-0.0716	-0.6466	0.7594
c1	0.8104	0.4167	0.4119	-0.5851	0.5370	0.6077	0.0320	-0.7335	0.6790
c2	0.8126	0.4139	0.4104	-0.5822	0.5440	0.6042	0.0268	-0.7299	0.6830
c3	0.8914	0.2994	0.3404	-0.4434	0.7321	0.5171	-0.0944	-0.6118	0.7853
c4	0.8933	0.3009	0.3338	-0.4397	0.7392	0.5102	-0.0932	-0.6025	0.7926
c5	0.9297	0.2380	0.2811	-0.3577	0.7651	0.5353	-0.0877	-0.5983	0.7965
d1	0.8128	0.4098	0.4140	-0.5786	0.6507	0.4917	-0.0679	-0.6392	0.7660
d2	0.8134	0.4105	0.4122	-0.5772	0.6573	0.4846	-0.0720	-0.6321	0.7715
d3	0.8438	0.3701	0.3886	-0.5298	0.6899	0.4933	-0.0855	-0.6221	0.7782
d4	0.9875	0.1063	0.1164	-0.1575	0.6951	0.7015	-0.0063	-0.7111	0.7031
d5	0.8461	0.3693	0.3843	-0.5271	0.6870	0.5003	-0.0793	-0.6258	0.7759

Table 13. Eigenvectors Values of the original images of (Patient 3, Lesion C)

Image	Red 1^{st} C.	Green 1^{st} C.	Blue 1^{st} C.	Red 2^{nd} C.	Green 2^{nd} C.	Blue 2^{nd} C.	Red 3^{rd} C.	Green 3^{rd} C.	Blue 3^{rd} C.
a1	0.7756	0.4380	0.4546	-0.6307	0.5665	0.5303	0.0252	0.6980	-0.7156
a2	0.7624	0.4475	0.4675	-0.6455	0.5776	0.4997	0.0464	0.6827	-0.7292
a3	0.7834	0.4302	0.4485	-0.6190	0.6050	0.5009	0.0558	0.6700	-0.7403
a4	0.7655	0.4446	0.4652	-0.6405	0.5957	0.4847	0.0616	0.6690	-0.7407
a5	0.7691	0.4399	0.4637	-0.6357	0.6022	0.4830	0.0668	0.6663	-0.7427
b1	0.7525	0.4748	0.4564	-0.6580	0.5699	0.4922	0.0264	0.6707	-0.7412
b2	0.7385	0.4868	0.4665	-0.6742	0.5375	0.5064	0.0042	0.6885	-0.7252
b3	0.7608	0.4624	0.4554	-0.6455	0.6121	0.4569	0.0675	0.6415	-0.7641
b4	0.7097	0.5041	0.4921	-0.7017	0.5678	0.4303	0.0625	0.6507	-0.7567
b5	0.7117	0.5038	0.4896	-0.7001	0.5664	0.4348	0.0582	0.6522	-0.7558
c1	0.7836	0.4473	0.4311	-0.6083	0.6935	0.3861	0.1263	0.5648	-0.8155
c2	0.7816	0.4488	0.4332	-0.6089	0.6999	0.3734	0.1356	0.5556	-0.8203
c3	0.8154	0.4151	0.4034	-0.5669	0.7134	0.4119	0.1168	0.5646	-0.8170
c4	0.7977	0.4367	0.4160	-0.5875	0.7186	0.3722	0.1364	0.5412	-0.8297
c5	0.7939	0.4410	0.4186	-0.5924	0.7162	0.3690	0.1370	0.5409	-0.8298
d1	0.7993	0.4397	0.4096	-0.5825	0.7343	0.3485	0.1475	0.5172	-0.8431
d2	0.8095	0.4262	0.4038	-0.5696	0.7367	0.3644	0.1422	0.5250	-0.8391
d3	0.8280	0.4149	0.3772	-0.5540	0.7087	0.4367	0.0861	0.5706	-0.8167
d4	0.8212	0.4236	0.3823	-0.5623	0.7146	0.4160	0.0970	0.5566	-0.8251
d5	0.8036	0.4352	0.4060	-0.5791	0.7290	0.3649	0.1372	0.5284	-0.8379

B.1 The estimations of the third component using the |Blue-Green| band

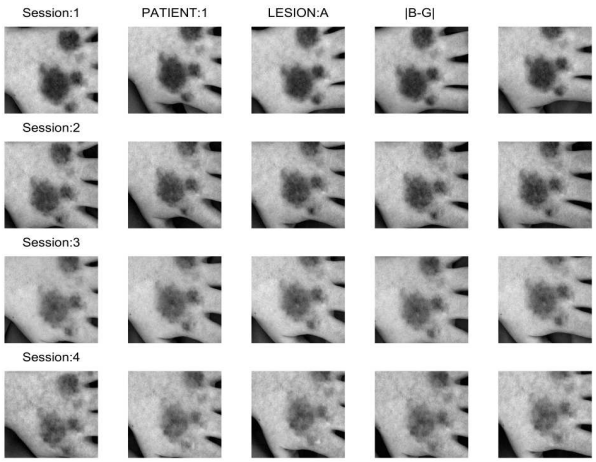


Fig. 12. Set of |B-G| Images of (Patient 1, Lesion A).

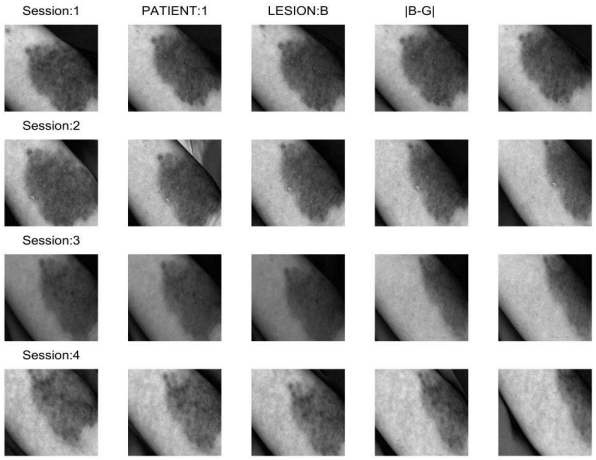


Fig. 13. Set of |B-G| Images of (Patient 1, Lesion B).

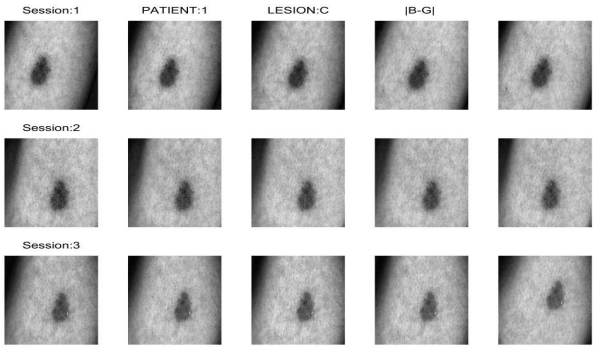


Fig. 14. Set of |B-G| Images of (Patient 1, Lesion C).

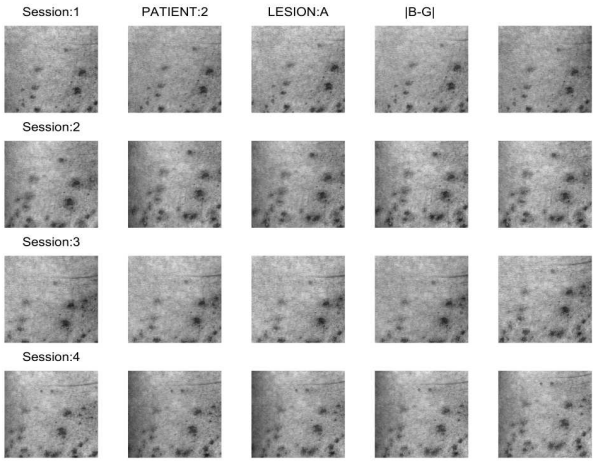


Fig. 15. Set of |B-G| Images of (Patient 2, Lesion A).

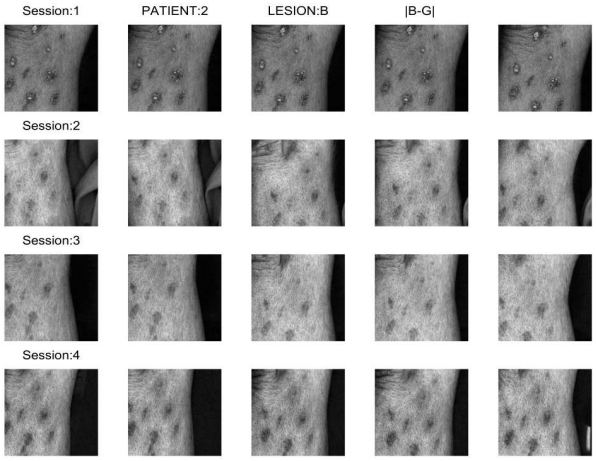


Fig. 16. Set of |B-G| Images of (Patient 2, Lesion B).

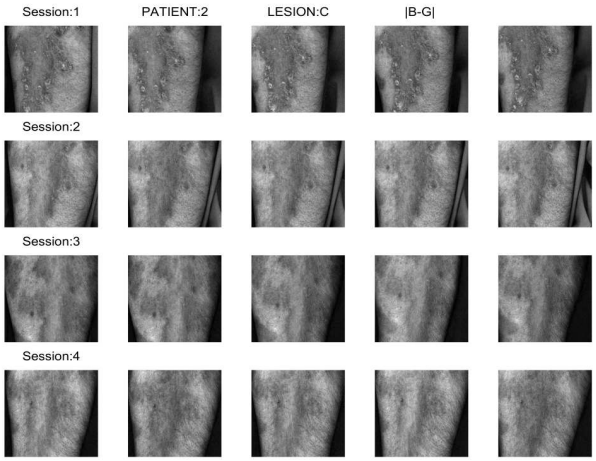


Fig. 17. Set of |B-G| Images of (Patient 2, Lesion C).

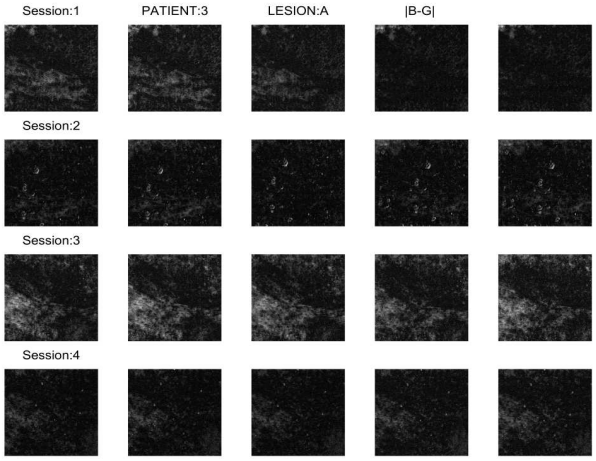


Fig. 18. Set of |B-G| Images of (Patient 3, Lesion A).

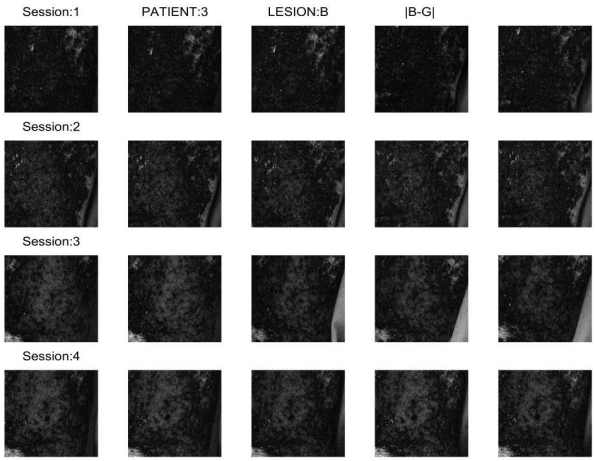


Fig. 19. Set of |B-G| Images of (Patient 3, Lesion B).

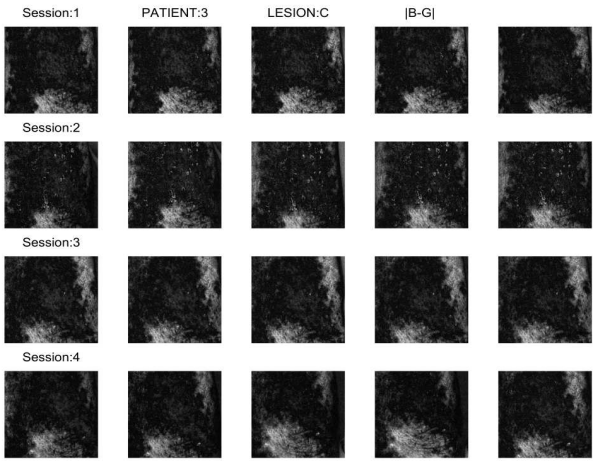


Fig. 20. Set of |B-G| Images of (Patient 3, Lesion C).

C Eigenvalues

Table 14. Eigenvalues of the original images of Patient 1

Image	Lesion A			Lesion B			Lesion C		
	1^{st} C.	1^{st} C.	1^{st} C.	2^{nd} C.	2^{nd} C.	2^{nd} C.	3^{rd} C.	3^{rd} C.	3^{rd} C.
a1	1444.88	54.42	12.75	3596.48	25.48	8.42	2285.39	7.31	4.86
a2	2526.04	70.61	12.69	3106.99	20.41	9.27	2005.87	7.32	4.44
a3	1749.43	57.12	9.73	3388.21	28.69	8.09	1874.55	7.30	5.01
a4	1992.79	47.79	9.95	2880.91	34.18	11.33	1372.19	8.47	5.64
a5	2372.88	52.13	9.93	3138.65	22.62	10.86	1346.64	8.75	5.44
b1	1691.89	36.36	9.69	2401.09	34.23	9.49	1675.95	12.83	3.96
b2	2458.85	39.03	9.38	3148.66	36.86	10.86	1764.73	12.44	3.83
b3	1739.71	33.00	9.82	2776.68	23.46	9.34	1559.24	12.29	3.93
b4	1499.85	30.12	10.46	3188.71	21.02	7.69	1588.07	12.27	3.80
b5	1729.61	25.50	9.33	2682.42	32.03	8.49	1510.40	12.10	3.94
c1	2208.10	44.47	7.72	2703.60	31.40	10.15	1824.84	7.52	3.39
c2	2282.69	39.09	7.80	2835.99	21.76	9.30	1597.96	7.41	3.24
c3	2484.93	42.24	7.41	3227.58	17.70	8.18	1597.69	7.47	3.18
c4	1772.81	34.85	7.41	3146.52	20.15	5.72	1637.90	7.53	3.20
c5	2352.96	41.59	7.68	3075.80	21.01	5.83	1499.40	7.05	3.41
d1	2333.76	35.34	6.41	1888.44	28.28	5.62			
d2	1851.65	28.93	6.75	2363.85	30.98	5.26			
d3	2856.83	40.40	6.70	2834.36	24.25	4.53			
d4	2138.39	26.13	6.67	2726.42	24.39	3.88			
d5	2849.13	29.47	6.37	2766.16	117.89	3.25			

References

1. R. Larsen. Maf and other transformations applied to remote sensing. Master's thesis, Institute for Mathematical stastistics and Operations Research (IMSOR), Technical University of Denmark, DTU, 1991. pp. xii+130+205.

Table 15. Eigenvalues of the original images of Patient 2

Image	Lesion A			Lesion B			Lesion C		
	1^{st} C.	1^{st} C.	1^{st} C.	2^{nd} C.	2^{nd} C.	2^{nd} C.	3^{rd} C.	3^{rd} C.	3^{rd} C.
a1	172.33	7.81	3.93	2231.70	46.91	3.21	2513.00	106.44	8.49
a2	152.58	6.55	4.19	1819.39	43.95	3.32	3383.01	81.88	4.73
a3	148.49	5.64	4.88	2042.38	45.60	3.19	3324.11	76.87	5.82
a4	147.52	5.66	4.86	1783.37	44.10	3.20	3327.13	75.38	6.18
a5	156.78	7.12	4.00	1686.52	44.73	3.35	3316.42	82.44	5.58
b1	321.17	9.76	4.75	2346.98	55.01	3.64	2898.91	37.79	4.93
b2	553.82	10.70	4.36	2040.16	44.42	3.72	2778.70	67.35	5.62
b3	389.53	9.83	4.38	1763.97	26.64	5.30	2857.95	69.98	5.08
b4	395.31	10.89	4.21	1728.86	30.87	4.94	2823.23	68.87	4.99
b5	226.90	9.76	4.84	2079.96	35.50	5.87	3238.57	87.96	5.20
c1	154.41	9.17	4.51	3790.00	20.12	3.21	3262.82	27.35	4.61
c2	157.91	8.50	4.54	3445.74	20.91	3.55	3536.77	24.82	4.70
c3	154.56	8.36	4.58	2434.32	24.07	4.28	4381.46	24.73	4.41
c4	160.29	8.33	4.60	2347.57	24.22	4.01	4176.64	27.17	4.29
c5	175.15	7.76	4.85	3457.68	24.17	3.66	5274.74	26.34	4.09
d1	339.97	11.56	4.19	3054.50	16.41	3.58	3244.15	24.45	4.09
d2	624.10	11.73	3.66	3466.72	16.08	3.37	2823.77	22.74	5.52
d3	655.37	12.13	3.71	2346.85	17.96	3.84	3222.69	25.79	3.86
d4	582.09	12.45	3.72	2407.31	20.74	3.99	3216.34	30.16	3.78
d5	459.18	12.29	3.98	2602.88	25.12	4.13	3388.39	28.01	5.53

Table 16. Eigenvalues of the original images of Patient 3

Image	Lesion A			Lesion B			Lesion C		
	1^{st} C.	1^{st} C.	1^{st} C.	2^{nd} C.	2^{nd} C.	2^{nd} C.	3^{rd} C.	3^{rd} C.	3^{rd} C.
a1	432.79	38.48	2.39	1118.61	32.33	2.23	945.99	39.22	6.37
a2	383.62	36.70	2.38	1283.75	32.68	2.32	810.32	34.82	6.43
a3	598.55	41.30	1.99	1261.05	33.08	2.35	749.23	44.51	6.75
a4	350.66	38.99	1.88	1444.51	56.44	2.35	698.18	39.69	6.81
a5	352.76	38.41	1.84	1451.91	50.91	2.46	815.85	39.76	5.70
b1	545.49	31.44	2.52	1353.42	67.05	3.61	2080.46	57.52	6.34
b2	523.02	31.61	2.59	1347.74	65.91	3.63	1713.35	52.22	7.51
b3	525.11	32.61	2.12	986.75	54.96	3.57	1587.72	83.14	6.21
b4	443.85	31.12	2.32	1267.77	62.97	3.31	1196.66	47.50	7.07
b5	438.80	31.44	2.36	1344.24	64.67	3.39	1234.96	46.53	7.22
c1	388.04	19.31	2.15	1054.96	11.68	3.62	588.61	20.91	7.73
c2	339.61	21.80	2.24	916.67	12.92	3.47	527.21	21.09	7.69
c3	316.48	23.07	2.17	874.05	77.91	2.63	796.43	29.10	7.17
c4	246.08	23.08	1.97	835.44	95.21	2.76	694.37	24.10	7.31
c5	311.00	22.75	1.91	923.03	93.51	3.19	585.96	24.98	7.14
d1	616.06	15.92	1.98	1387.41	19.47	3.66	825.42	16.95	6.34
d2	527.81	16.37	1.90	1384.15	20.19	3.61	1230.42	17.53	5.56
d3	491.69	16.99	1.92	1095.60	29.06	3.70	1940.85	27.61	6.49
d4	472.31	17.18	1.93	1130.25	29.10	3.69	1968.43	24.25	6.59
d5	440.91	17.75	2.01	1038.86	28.94	3.67	770.58	20.26	5.84

4.2.2 A Hierarchical Classification Scheme of Psoriasis Images

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 6
Pages: 1 – 23
Reference: [34]

A Hierarchical Classification Scheme of Psoriasis Images

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. A two-stage hierarchical classification scheme of psoriasis lesion images is proposed. These images are basically composed of three classes: normal skin, lesion and background. The scheme combines conventional tools to separate the skin from the background in the first stage, and the lesion from the normal skin in the second stage. These tools are the Expectation-Maximization Algorithm, the quadratic discrimination function and a classification window of optimal size. Extrapolation of classification parameters of a given image to other images of the set is evaluated by means of Cohen's Kappa coefficient.

1 Introduction

We use a set of 175 *RGB* psoriasis lesions images, of size 556 pixels by 748 pixels, taken at the Gentofte Hospital, Denmark, during pilot sessions with three invited patients. For each patient, three lesions were followed once a week during at least three weeks. In each session, five images of each lesion were taken. Examples can be seen in Figure 1.

The images of lesions with psoriasis can be assumed to contain three classes: background, normal skin and lesion. Lesions have high contrast with the rest of the image in the third principal component. Principal Component Analysis has shown that a rough estimation of the third component can be obtained subtracting the green band from the blue band, while the red band almost does not have any influence [6]. Alternatively, in the absolute difference $|B - G|$, lesions and background are both enhanced from the normal skin. The background has a good contrast with the rest in the red band: advantage can be taken of the green color of the curtain used as background during the image acquisition process, because its density function is positioned in one extreme of the red band histogram.

The present classification scheme consists of two stages. In the first stage, the skin is separated from the background using the red band, and, in the second stage, the lesion is separated from the normal skin using the absolute value of

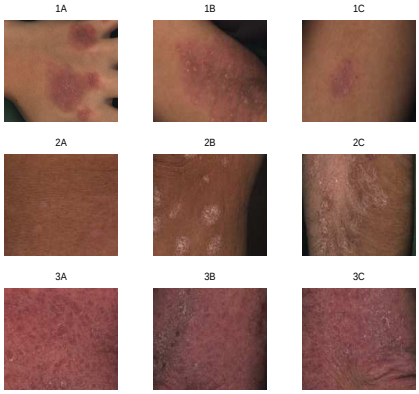


Fig. 1. Set of Psoriasis Lesions. From top to bottom, the row number corresponds with the patient number. From left to the right, columns correspond to the labelled lesions A, B and C, respectively.

the difference between the blue and the green bands.

For future alignment purposes, it is expected that, for a given lesion, the classification scheme produces thematic maps with a repeated pattern indicating that lesion within and between sessions.

2 Lesion Classification

As mentioned before, the hierarchical classification scheme is composed of two stages. In both steps, the pixels of the classes are assumed to be Gaussian distributed. Although this is almost certainly not the case, Clarke et al. have shown that the quadratic method is fairly robust to deviations from normality, unless we have very skewed marginals [2]. The red band R is first used to discriminate between the background and the skin, then the output of the previous step is combined with the $|B - G|$ band for segmenting the lesions from the normal skin. In both stages, the use of Wang's Expectation-Maximization Algorithm [11] (setting $\epsilon = 0.0001$ in [11]) is proposed. This algorithm is expected to provide the parameters (mean μ, covariance matrix Σ and a-priori probability p) of two Gaussian distributed classes: skin and background in the first step, and lesions and normal skin in the second step. The threshold between the classes is obtained using the following discrimination function for Gaussian distributed classes with different dispersion [3]:

$$-(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)+(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\geq 2log(\frac{L_{(2,1)}p_2}{L_{(1,2)}p_1}) \quad (1)$$

where μ_i is the mean, Σ_i is the covariance matrix and p_i is the a priori probability of the $i-th$ class $\forall i \in \{1,2\}$. $L_{(1,2)}$ is the cost of belonging to class 1 and being classified as belonging to class 2 and vice-versa for $L_{(2,1)}$.

The discrimination function is applied to the output of the convolution of the image with a circular window of optimal size. The critical number of elements needed for classification is computed as a function of the minimal separability between the two classes [7]:

$$w_{(1),(2)} = -6.8341 + \frac{7.1800}{\sqrt{dA_{(1),(2)}}} \quad (2)$$

where w is the radius of the estimated optimal disc for classification and dA is the weighted sum of the difference of areas of the class density functions:

$$dA_{(1),(2)} = dA_{(1),(2)} = \frac{1}{2}\sum_{n=1}^{Q}|h_{(1)}[n]-h_{(2)}[n]| \quad (3)$$

Q is the number of quantization levels, $h_{(1)}$ and $h_{(2)}$ are the estimated density functions of classes 1 and 2 respectively.

The values of this optimal window sum to one; each element of the window has the value $1/n_c$ corresponding to the inverse of the size of the window n_c given by its number of elements. It is clear that after convolving the image, each pixel is replaced with the estimated mean of its neighborhood.

In the first step of this classification scheme, pixels whose estimated local mean is greater than the threshold are considered belonging to the class skin, and vice-versa, for the background. In the second step, pixels whose estimated mean of the neighborhood is greater than the threshold are considered belonging to the class normal skin, and vice-versa, for the lesion. Note that, in both cases, the discrimination function is constructed with the original data, but it is applied to the pixels belonging to the output of the convolution. A set of thematic maps indicating the clases, the lesions and the rest, is obtained at the end of this stage.

2.1 Results and Discussion

During the classification stage: for the first step, the costs were $L_{(A,B)} = 500$ and $L_{(B,A)} = 1$; for the second step, $L_{(A,B)} = 1$ and $L_{(B,A)} = 1$ was set. These

values were empirically found. Setting individual parameters depending on the lesions could obviously improve the results obtained. However, to set the same values for all the cases was helpful for comparison purposes. The Expectation-Maximization Algorithm was applied to each single image. Classification outputs are presented for each single image and classification stage in Appendix A.

By visual assessment of the classification output of the first stage, it can be noted that the results are satisfactory, except for the set of images of the third session of the case $(2, C)$, which is a lesion partially covered by hair. With regard to the output of the second classification stage, which is the desired output, comments to each particular case follow. For the first patient, the output of the classification scheme proposed has shown to be enough to define a smaller searching and alignment area of the lesion patterns with satisfactory results, in spite of some border effects produced by the window size. On the other hand, it could be said that for the case $(2, A)$ the expected common pattern roughly appears in the thematic maps, however, it does not correspond to the lesion. The shadows present on the image borders clearly are influencing the classification results. In fact, the presence of shadows did not affect the output of the first classification stage very much, which contained for a given lesion, a repeated skin pattern within and between sessions. However, it clearly affects the output of the second classification stage for the cases $(2, B)$, $(2, C)$, $(3, A)$, $(3, B)$ and $(3, C)$.

An aspect that has to be put under consideration during the alignment process, is the size of the lesion pattern. By visual inspection of the original data-set, it can be deduced that the size between lesions has a significant variance: in some images, one clear lesion composed by only one connected region can be observed, while in other cases, a lesion can consist of a set of smaller size connected regions; finally, there are also examples, where the number of connected regions is so large while the size of the regions is so small, that the whole can be taken as only one lesion. It is clear that a good definition of the lesion does not avoid the illumination problem.

In order to reduce the processing time, one aspect considered was the use of the output of the first image of each session as input for the remaining images of the session. In the following, the uncertainty of using the parameters of a randomly selected image for a given patient, lesion and session, to segment the remaining images of the same session is evaluated. Since the images are not registered, a descriptor that is invariant to translations and rotations is required. This is the case of the normalized histogram. Table 1 is based on data presented in Table 4. Each single μ cell value in Table 1 was computed according to the following Equation:

$$\mu_{p,l,s} = \frac{1}{n_c} \sum_{i=1}^{n_c} E[c[h_{p,l,s,i,b}, h_{p,l,s,:,b}]] \tag{4}$$

where $E[corr(h(nX_{p,l,s,i,b}), h(nX_{p,l,s,:,b}))]$ is calculated using Equation 13 and p is the index for patient, l, for lesion, s, for session, i, for capture, b for color band. The symbol n_c is the number of captures of the $s-th$ session. The session s takes values in {'a','b','c','d'}, where the order in the sequence indicates the session week. The color band b takes values in {'R','G','B'}.

In a similar fashion, for the standard deviation, each single σ cell value is obtained by computing the following:

$$\sigma_{p,l,s} = \sqrt{\frac{1}{n_c}\sum_{i=1}^{n_c}(E[c[h_{p,l,s,i,b}, h_{p,l,s,:,b}]] - \mu_{p,l,s})^2} \tag{5}$$

where the indexes have the same meaning as in Equation 4.

Table 1. Average and dispersion correlation value between the normalized histogram of a normalized image red band and the normalized histogram of the normalized red bands of the images belonging to the same session.

(Patient, Lesion)	μ_a	σ_a	μ_b	σ_b	μ_c	σ_c	μ_d	σ_d
(1,A)	0.8525	0.0260	0.8618	0.0328	0.8521	0.0391	0.8448	0.0258
(1,B)	0.8055	0.0248	0.8107	0.0348	0.7264	0.0210	0.7342	0.0305
(1,C)	0.8413	0.0061	0.8876	0.0192	0.8650	0.0156		
(2,A)	0.7991	0.0113	0.8777	0.0174	0.8486	0.0215	0.8774	0.0673
(2,B)	0.8937	0.0110	0.7312	0.0438	0.8326	0.0100	0.8841	0.0464
(2,C)	0.8345	0.0151	0.7924	0.0231	0.8535	0.0145	0.7906	0.0514
(3,A)	0.8477	0.0094	0.9072	0.0155	0.8675	0.0204	0.8731	0.0186
(3,B)	0.7708	0.0372	0.8776	0.0292	0.7498	0.0310	0.9062	0.0329
(3,C)	0.8749	0.0265	0.9057	0.0098	0.8799	0.0362	0.8074	0.0233

Table 2 was constructed in the same fashion as in Table 1, but using the skin data of the |B-G| band.

Results presented in Tables 1 show that in general, the normalized histograms of red bands of images belonging to the same sessions are have a correlation superior to 0.8. Exceptions are, for instance, the cases for which the lesion presents occlusions. See the last captures of the the third and fourth session of the case $(1, B)$. See also the first and fifth capture of the second session of the case $(2, B)$. Note that in the first session of the case $(3, B)$, the two last captures show a large translation on the focused region. For the images of the third session of the case $(3, B)$, it can be observed that the shadows are not presented in the same way, for all the captures. Shadows, hairs and occlusions seem not to be irrelevant.

With regard to the results presented in Table 2, which was generated using only the values of pixels belonging to the class skin, it has to be mentioned,

Table 2. Average and standard deviation correlation value between the normalized histogram of a normalized $|B-G|$ band skin data and the normalized histogram of the normalized $|B-G|$ bands skin data of the images belonging to the same session.

(Patient, Lesion)	μ_a	σ_a	μ_b	σ_b	μ_c	σ_c	μ_d	σ_d
(1,A)	0.6472	0.0819	0.5790	0.0353	0.5011	0.1946	0.5387	0.0735
(1,B)	0.6170	0.1594	0.7657	0.1718	0.3759	0.0752	0.5461	0.0437
(1,C)	0.6166	0.1483	0.7061	0.0502	0.6569	0.0272		
(2,A)	0.7306	0.0474	0.6279	0.0227	0.6916	0.1287	0.6956	0.0678
(2,B)	0.5881	0.1014	0.5786	0.1156	0.5373	0.0678	0.6598	0.1591
(2,C)	0.5877	0.1490	0.5755	0.1130	0.6440	0.1524	0.6296	0.1287
(3,A)	0.6199	0.0732	0.6602	0.0604	0.5344	0.0787	0.5798	0.0727
(3,B)	0.6797	0.0457	0.5004	0.0695	0.4920	0.1040	0.4405	0.0782
(3,C)	0.8134	0.1557	0.7117	0.0653	0.7988	0.1681	0.5462	0.0851

that the average correlation values decrease for all the cases, which is not very encouraging. However, as mentioned before, there is more contrast between the lesion and the rest in these bands.

The application of the discrimination function generated with the parameters produced by the Expectation-Maximization Algorithm for the first image of each session to the remaining images of the same session is evaluated here. Cohen's Kappa coefficient [9], [10] was used to compare pairs of thematic maps produced for each given single image. For each pair, the first thematic map was produced with the discrimination function depending on the given image data. The second thematic map was produced with the discrimination function of the first image of the session the given image belongs to. For each pair, the first thematic map was used as the ground truth. The Kappa coefficient is a global indicator of classification quality, which varies from minus one to one, where a perfect classification would give a value of one, and a completely erroneous classification would give a value of minus one. For each pair of corresponding thematic maps, the confusion matrix C [5] was constructed assuming two classes ($k = 2$). For the first classification stage, these classes were the curtain and the skin. For the second classification stage, the normal skin and the lesion.
The Kappa coefficient is defined as:

$$\kappa = \frac{(d-q)}{N-q} \tag{6}$$

where

$$d = \sum_{i=1}^{k} c_{i,i}$$
$$q = \frac{\sum_{i=1}^{k}((\sum_{j=1}^{k} c_{i,j})(\sum_{j=1}^{k} c_{i,j}))}{N} \quad (7)$$
$$N = \sum_{i=1}^{k}\sum_{j=1}^{k} c_{i,j}$$

and $c_{i,j}$ is the element of the confusion matrix located at the $i-th$ row and $j-th$ column.

Each single cell value in Tables 6 and 7 is a Kappa value $\kappa_{p,l,s,c,r}$ where p means patient, l, lesion, s, session, c, capture and r is for Table 6 equal to "SKIN" and for Table 7 equal to "LESION".

The average and standard deviation of the Kappa coefficient was computed for each patient and lesion, first using the thematic maps indicating skin (the output of the first classification stage), and then using the thematic maps indicating lesions (the output of the second classification stage). The estimation of the average of Kappa is defined as:

$$E[\kappa_{p,l,r}] = \frac{1}{n_s(n_c-1)} \sum_{s=1}^{n_s}\sum_{c=2}^{n_c} \kappa_{p,l,s,c,r} \quad (8)$$

where p is the patient, l, the lesion, s, the session, c the capture and r takes values in $\{'SKIN','LESION'\}$. Note that the Kappa value of the first image of each session is not included, because it is supposed to be one. The standard deviation of $\kappa_{p,l,r}$ is defined as:

$$S[\kappa_{p,l,r}] = \sqrt{\frac{1}{n_s(n_c-1)} \sum_{s=1}^{n_s}\sum_{c=2}^{n_c} \kappa^2_{p,l,s,c,r} - E^2[\kappa_{p,l,r}]}. \quad (9)$$

Results can be seen in Table 3, which is based on data of Tables 6 and 7.

Excluding the six captures where the lesion $(1,B)$ is occluded (the last two captures from the second to the fourth session), the mean Kappa (see Equation 8) increases from 0.7805 to 0.9081 and from 0.6687 to 0.8769, and its standard deviation reduces from 0.1974 to 0.0601 and from 0.3223 to 0.0969, for the first and second classification stage respectively[1]. The bad classification outputs in the case $(2,C)$ are partly due to the fact that the lesion is partially covered by hair. On the other hand, by visual assessment of the incomplete thematic maps

[1] See in [8] a scheme for registration in presence of occlusions.

Table 3. Classification Stage: Kappa values κ of the classification outputs using the parameters of each single image compared with the classification output by using the parameters of the first image of each session.

N is the number of processed images.

(Patient, Lesion)	N	$E[\kappa_{p,l,SKIN}]$	$S[\kappa_{p,l,SKIN}]$	$E[\kappa_{p,l,LESION}]$	$S[\kappa_{p,l,LESION}]$
(1,A)	20	0.9171	0.0580	0.8966	0.0710
(1,B)	20	0.7805	0.1974	0.6687	0.3223
(1,C)	15	0.9428	0.0470	0.8224	0.0902
(2,A)	20			0.7244	0.1100
(2,B)	20	0.8916	0.0848	0.6124	0.1641
(2,C)	20	0.7630	0.2013	0.4579	0.4043
(3,A)	20			0.8606	0.1005
(3,B)	20	0.9082	0.0810	0.7551	0.1955
(3,C)	20	0.9082	0.0810	0.9297	0.0516

of the third session of the same case $(2, C)$, they could, in fact, be considered not qualified for being ground truth: the revision of the standards of the image acquisition procedure is suggested[2]. Related work on controls during the image acquisition time can be found in [4]. Note, that, in general, in spite of being related to descriptors completely different, the results presented in Tables 1 and 2 predict the classification results presented in Table 3 quite well.

3 Conclusions

The combination of conventional tools like the Expectation-Maximization Algorithm, the quadratic discriminant function and a classification window of optimal size has shown to be a suitable approach to segment images of psoriasis lesions. There are aspects related to the image data that are not irrelevant. This is, for instance, the case of the shadows, which clearly affect the quality of the classification results. For a correct lesions registration, it is required that thematic maps to be used to select the lesion patterns to be registered, -if they are automatically generated- have to contain a repeated pattern within and between sessions. This is, in general, not the case when the original data set is used. Illumination correction is required in order to improve the classification output. Hair and lesions occlusion are aspects to consider during the revision of the image acquisition standards.

On the other hand, results indicate that the parameters of the first image of each session could be used to segment the remaining images of the same session, when the situations mentioned before do not appear.

[2] A hair removal scheme is proposed in [1].

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A Expectation-Maximization Algorithm to each image

A.1 The output of the first classification stage

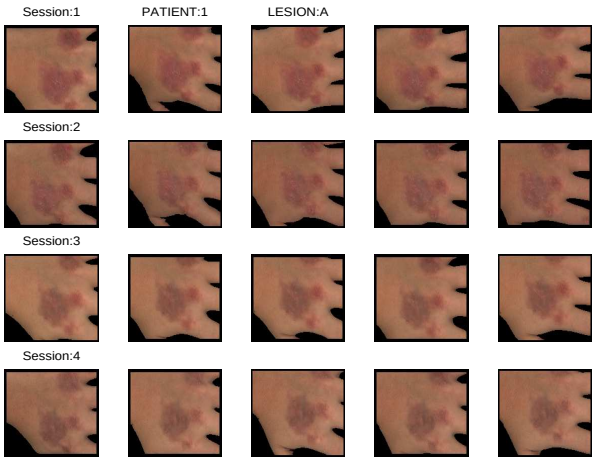


Fig. 2. Overlay of thematic maps with original images for (Patient 1, Lesion A).

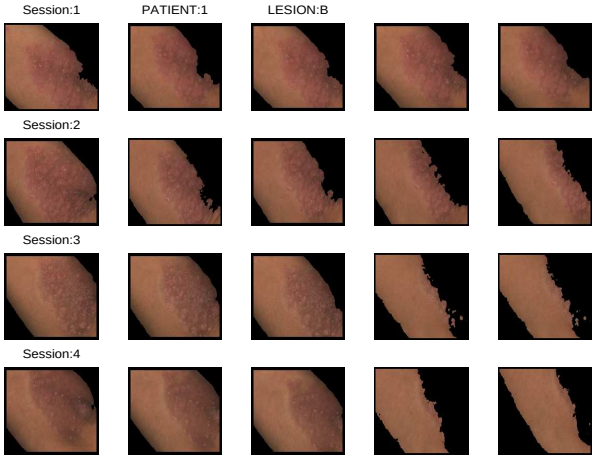


Fig. 3. Overlay of thematic maps with original images for (Patient 1, Lesion B).

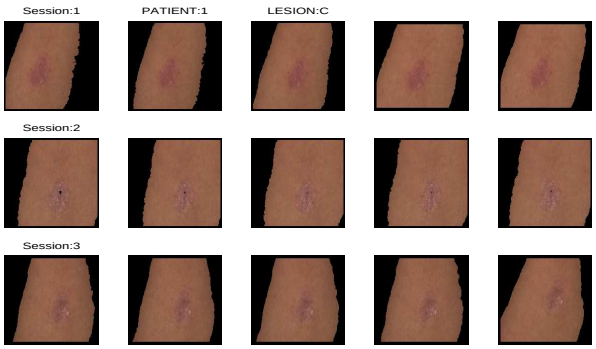


Fig. 4. Overlay of thematic maps with original images for (Patient 1, Lesion C).

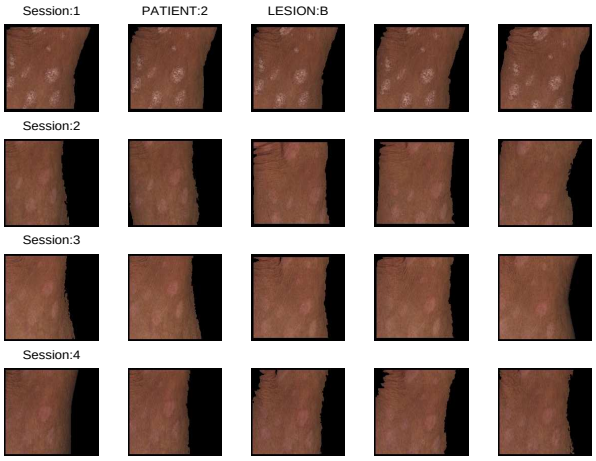


Fig. 5. Overlay of thematic maps with original images for (Patient 2, Lesion B).

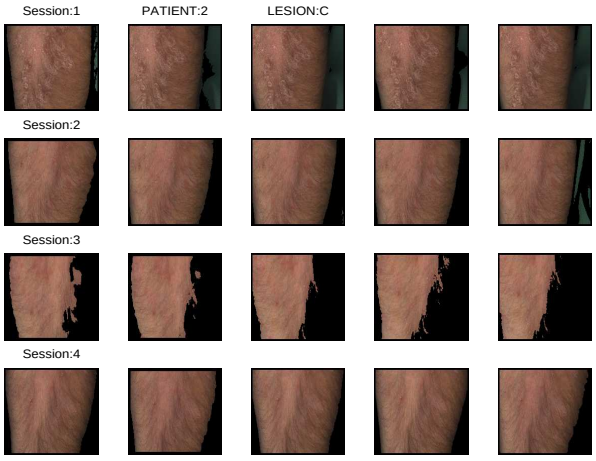


Fig. 6. Overlay of thematic maps with original images for (Patient 2, Lesion C).

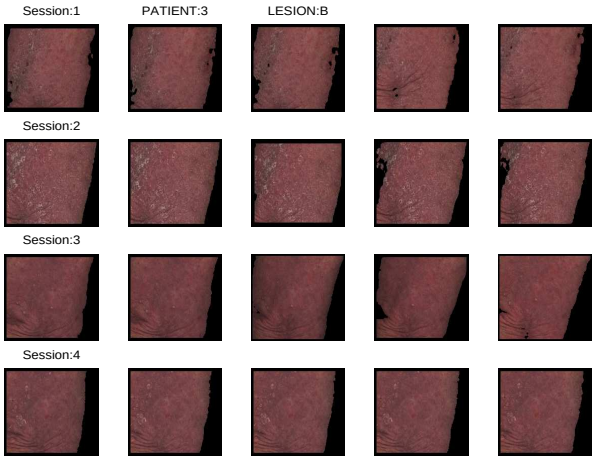


Fig. 7. Overlay of thematic maps with original images for (Patient 3, Lesion B).

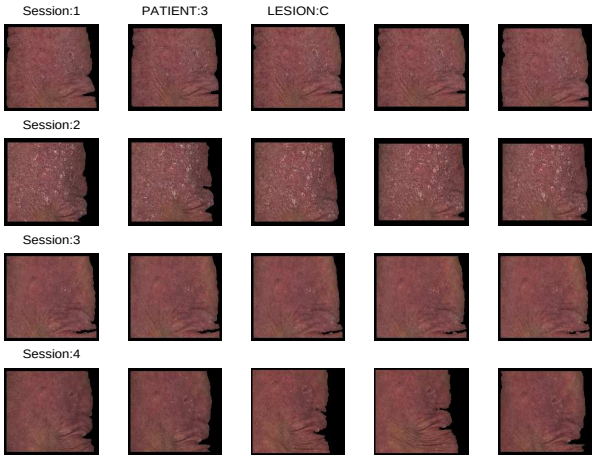


Fig. 8. Overlay of thematic maps with original images for (Patient 3, Lesion C).

A.2 The output of the second classification stage

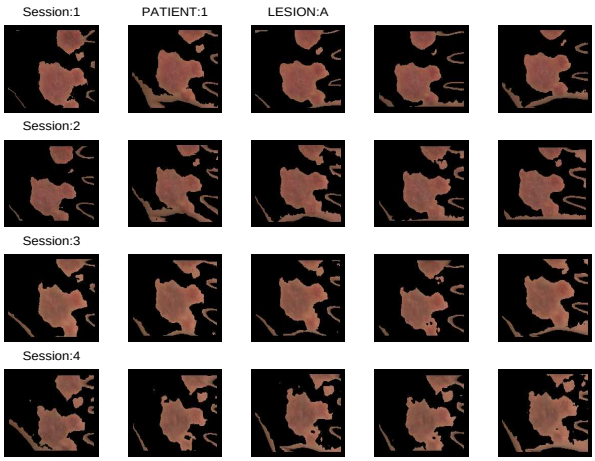


Fig. 9. Overlay of thematic maps with original images for (Patient 1, Lesion A).

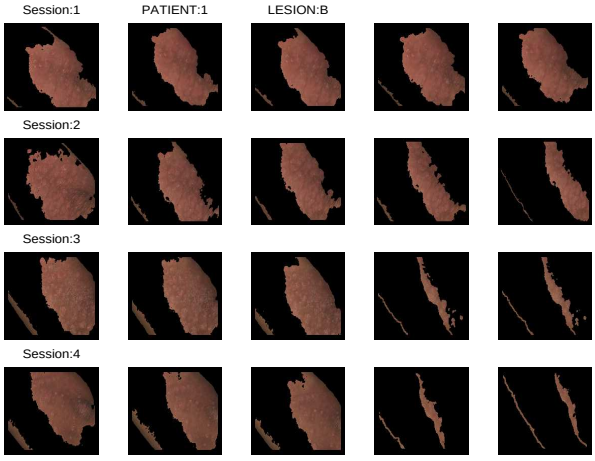


Fig. 10. Overlay of thematic maps with original images for (Patient 1, Lesion B).

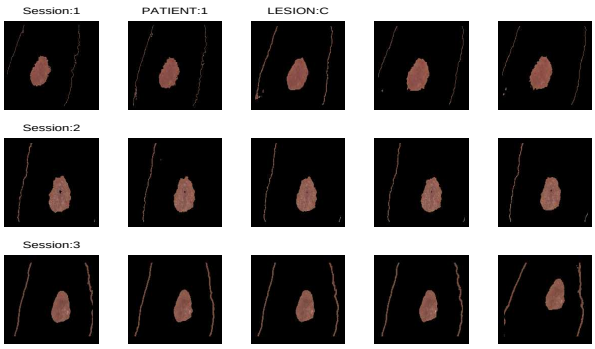


Fig. 11. Overlay of thematic maps with original images for (Patient 1, Lesion C).

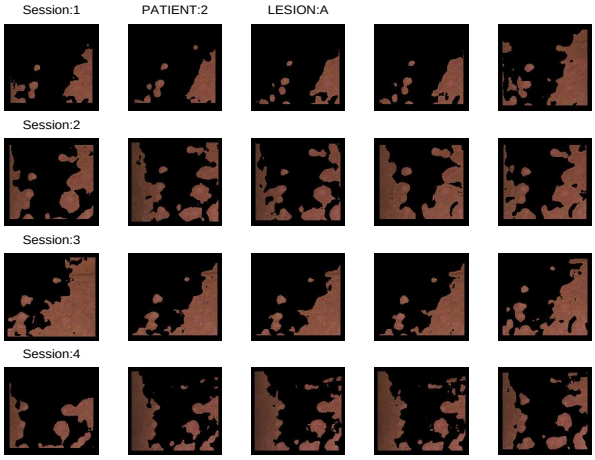


Fig. 12. Overlay of thematic maps with original images for (Patient 2, Lesion A).

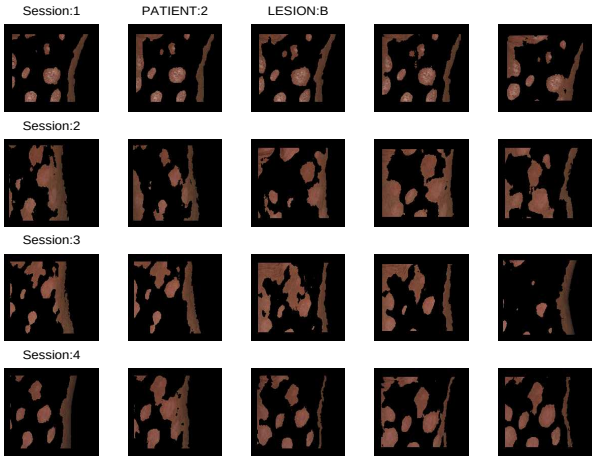


Fig. 13. Overlay of thematic maps with original images for (Patient 2, Lesion B).

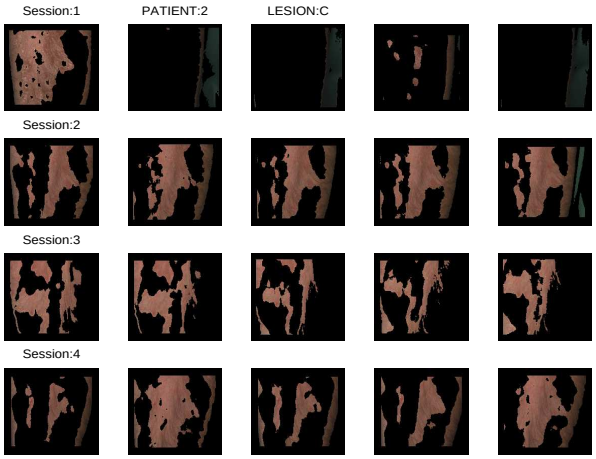


Fig. 14. Overlay of thematic maps with original images for (Patient 2, Lesion C).

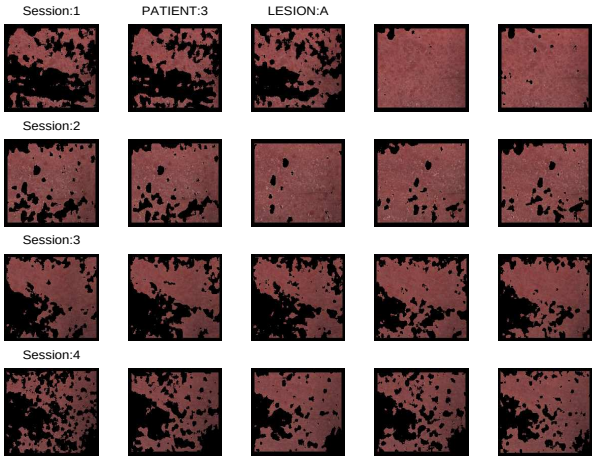


Fig. 15. Overlay of thematic maps with original images for (Patient 3, Lesion A).

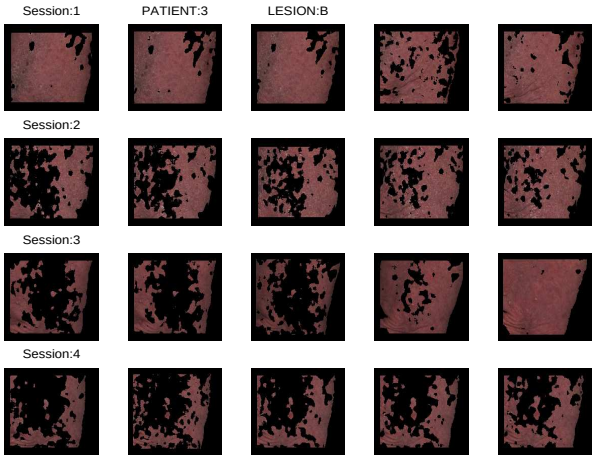


Fig. 16. Overlay of thematic maps with original images for (Patient 3, Lesion B).

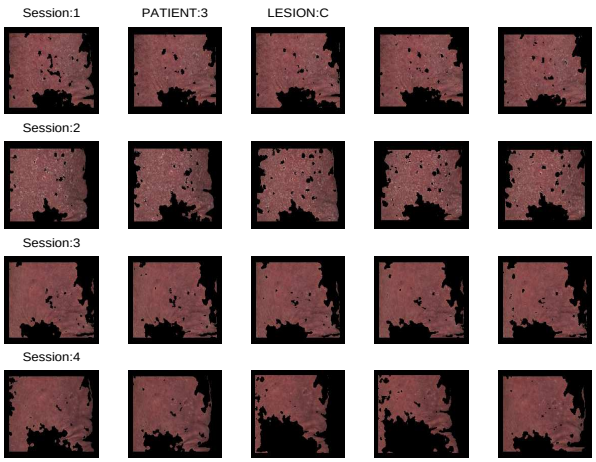


Fig. 17. Overlay of thematic maps with original images for (Patient 3, Lesion C).

B Classification output analysis

In Table 4 the average correlation values of the normalized histogram of the normalized red band of each single image with the normalized histogram of the normalized red bands of the rest of the images of the same session are shown.

Details of the computation of each single cell value follow. The discrete correlations between pairs of normalized histograms of images of the same session was first computed:

$$c[h_{p,l,s,i,b}, h_{p,l,s,j,b}] = \frac{max(h_i \circ h_j)}{\sqrt{max(h_i \circ h_i) max(h_j \circ h_j)}} \tag{10}$$

where

$$(h_i \circ h_j)[n] = \sum_{k=-\infty}^{\infty} h_{p,l,s,i,b}[k], h_{p,l,s,j,b}[n+k] \tag{11}$$

where $h_{p,l,s,i,b} = h(nX_{p,l,s,i,b})$ and $h_{p,l,s,j,b} = h(nX_{p,l,s,j,b})$ are the normalized histograms of the normalized original images according to Equation 12:

$$nX_{p,l,s,c,b} = \frac{round(255. * (X_{p,l,s,c,b} - min(X_{p,l,s,c,b}))}{(max(X_{p,l,s,c,b}) - min(X_{p,l,s,c,b}))) + 1}. \tag{12}$$

The nX images are the input to the Expectation-Maximization Algorithm. Each single cell value in Table 4 is the average correlation between each single image band and the rest of the images of the same session:

$$E[c[h_{p,l,s,i,b}, h_{p,l,s,:,b}]] = \frac{1}{n_c - 1}(\sum_{j=1} n_c c[h_{p,l,s,i,b}, h_{p,l,s,j,b}] - 1). \tag{13}$$

It has to be mentioned that the X images used to compute these values were the "TIFF" format version of the original acquired "HIPS" format images, from which 10 rows and columns were removed from the borders, and then subsampled every four columns and rows.

Afterwards, with the same meaning as before, but now using the data of the region indicating skin Table 5 was generated.

Table 4. Average correlation values of the normalized histogram of the red band of each single image with the normalized histogram of the red band of the images that belong to the same session than that.

Capture	$(1,A)$	$(1,B)$	$(1,C)$	$(2,A)$	$(2,B)$	$(2,C)$	$(3,A)$	$(3,B)$	$(3,C)$
a1	0.8189	0.7957	0.8316	0.8024	0.8816	0.8191	0.8530	0.8086	0.8500
a2	0.8516	0.8369	0.8481	0.7829	0.9027	0.8578	0.8353	0.7687	0.8919
a3	0.8871	0.7765	0.8421	0.8143	0.9027	0.8346	0.8410	0.8082	0.8433
a4	0.8388	0.7934	0.8404	0.7969	0.8819	0.8237	0.8585	0.7322	0.9027
a5	0.8662	0.8251	0.8442	0.7990	0.8999	0.8376	0.8505	0.7361	0.8866
b1	0.8929	0.7868	0.9063	0.8581	0.7149	0.8088	0.9177	0.8734	0.9052
b2	0.8894	0.7624	0.9104	0.8867	0.7719	0.8190	0.8999	0.8406	0.9053
b3	0.8327	0.8357	0.8765	0.8975	0.6648	0.7632	0.8835	0.9065	0.9124
b4	0.8216	0.8242	0.8756	0.8855	0.7673	0.7753	0.9213	0.8603	0.8901
b5	0.8722	0.8442	0.8693	0.8607	0.7373	0.7955	0.9136	0.9074	0.9157
c1	0.8358	0.6901	0.8783	0.8683	0.8240	0.8644	0.8755	0.7955	0.8318
c2	0.8855	0.7340	0.8633	0.8155	0.8275	0.8636	0.8865	0.7167	0.9079
c3	0.8799	0.7445	0.8829	0.8387	0.8396	0.8308	0.8842	0.7560	0.9007
c4	0.8680	0.7322	0.8535	0.8591	0.8252	0.8473	0.8428	0.7553	0.8504
c5	0.7913	0.7310	0.8467	0.8613	0.8466	0.8613	0.8487	0.7257	0.9087
d1	0.8026	0.6896		0.9136	0.9044	0.7447	0.8956	0.9084	0.8189
d2	0.8452	0.7172		0.8950	0.8018	0.7256	0.8541	0.8490	0.8321
d3	0.8689	0.7646		0.9085	0.9078	0.8180	0.8614	0.9275	0.7794
d4	0.8450	0.7534		0.9122	0.9111	0.8344	0.8641	0.9278	0.8207
d5	0.8623	0.7463		0.7577	0.8953	0.8304	0.8905	0.9181	0.7860

Table 5. Average correlation values of the normalized histogram of the |B-G| band skin data of each single image with the normalized histogram of the |B-G| band skin data of the images that belong to the same session than that.

Capture	$(1, A)$	$(1, B)$	$(1, C)$	$(2, A)$	$(2, B)$	$(2, C)$	$(3, A)$	$(3, B)$	$(3, C)$
a1	0.6747	0.7358	0.7299	0.6528	0.6746	0.6895	0.6899	0.7057	0.8840
a2	0.6866	0.4658	0.7075	0.7703	0.5487	0.6706	0.6906	0.6957	0.5348
a3	0.6573	0.7413	0.7321	0.7358	0.6035	0.3545	0.5158	0.5983	0.8833
a4	0.5051	0.7211	0.4225	0.7671	0.4346	0.7017	0.6030	0.7016	0.8812
a5	0.7122	0.4212	0.4908	0.7272	0.6792	0.5224	0.6002	0.6973	0.8837
b1	0.5920	0.4590	0.6418	0.6617	0.5848	0.3785	0.6737	0.3939	0.7666
b2	0.5681	0.8282	0.7505	0.6029	0.4627	0.6018	0.6868	0.4781	0.7572
b3	0.5227	0.8355	0.7631	0.6373	0.4604	0.5970	0.7039	0.5069	0.6391
b4	0.6121	0.8588	0.6823	0.6230	0.6854	0.6525	0.5539	0.5619	0.7541
b5	0.6000	0.8470	0.6929	0.6147	0.6996	0.6477	0.6828	0.5610	0.6415
c1	0.2272	0.3549	0.6809	0.7719	0.4226	0.7562	0.6255	0.3718	0.4981
c2	0.3616	0.3713	0.6213	0.5366	0.5482	0.7568	0.5076	0.4608	0.8747
c3	0.6340	0.2635	0.6814	0.7889	0.5825	0.4534	0.6097	0.6005	0.8747
c4	0.6382	0.4448	0.6356	0.7944	0.5925	0.5029	0.4510	0.6004	0.8747
c5	0.6445	0.4449	0.6653	0.5660	0.5404	0.7504	0.4782	0.4266	0.8717
d1	0.4177	0.4937		0.7435	0.7712	0.7371	0.6559	0.4613	0.5106
d2	0.5616	0.5653		0.6118	0.7674	0.6809	0.6586	0.3015	0.6386
d3	0.6008	0.5882		0.7371	0.4272	0.7338	0.5211	0.4771	0.4917
d4	0.5256	0.5784		0.6323	0.5592	0.5559	0.5552	0.4761	0.6350
d5	0.5877	0.5049		0.7532	0.7739	0.4405	0.5085	0.4865	0.4552

C Expectation-Maximization Algorithm to the first image of each section

Table 6. Kappa values $\kappa_{p,l,s,c,SKIN}$ of the classification outputs pointing skin using the Expectation-Maximization Algorithm for all the images of each sessions and for only the first image of each session.

Capture	(1,A)	(1,B)	(1,C)	(2,B)	(2,C)	(3,B)	(3,C)
a1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
a2	0.8716	0.9087	0.9846	0.9898	0.5030	0.8906	0.8906
a3	0.8717	0.9039	0.9519	0.9904	0.3547	0.8463	0.8463
a4	0.9786	0.8591	0.8795	0.9778	0.7077	0.9011	0.9011
a5	0.8707	0.8332	0.8645	0.9429	0.3407	0.9081	0.9081
b1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
b2	0.8976	0.8192	0.9925	0.9734	0.8306	0.9995	0.9995
b3	0.8775	0.8846	0.9792	0.8318	0.8233	0.8718	0.8718
b4	0.9967	0.7799	0.9609	0.7741	0.8199	0.9251	0.9251
b5	0.9891	0.7217	0.8593	0.9483	0.7019	0.9242	0.9242
c1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
c2	0.8714	0.9695	0.9328	0.9748	0.9033	0.9761	0.9761
c3	0.8795	0.9355	0.9916	0.8448	0.8353	0.6875	0.6875
c4	0.8368	0.4858	0.9433	0.8595	0.8788	0.7995	0.7995
c5	0.9791	0.4762	0.9730	0.9195	0.8372	0.8946	0.8946
d1	1.0000	1.0000		1.0000	1.0000	1.0000	1.0000
d2	0.9889	0.9759		0.8966	0.7557	0.9712	0.9712
d3	0.8823	0.9912		0.7813	0.9807	0.9773	0.9773
d4	0.9935	0.3843		0.7408	0.9960	0.9776	0.9776
d5	0.8878	0.5585		0.8196	0.9400	0.9808	0.9808

Table 7. Kappa values $\kappa_{p,l,s,c,LESION}$ of the classification outputs pointing lesions using the Expectation-Maximization Algorithm for all the images of each sessions and for only the first image of each session.

Capture	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(2,C)	(3,A)	(3,B)	(3,C)
a1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
a2	0.9122	0.9405	0.8717	0.9225	0.6138	-0.2457	0.9812	0.9234	0.9072
a3	0.9638	0.9483	0.7589	0.6850	0.8012	-0.2721	0.7832	0.8886	0.9790
a4	0.9245	0.9044	0.8364	0.9156	0.3756	0.1033	0.9351	0.6703	0.9382
a5	0.9235	0.8978	0.7821	0.5262	0.8730	-0.3099	0.9713	0.8554	0.9356
b1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
b2	0.8202	0.7448	0.9560	0.8165	0.8760	0.5099	0.9406	0.7988	0.9552
b3	0.7717	0.7573	0.9104	0.5856	0.3908	0.5731	0.9093	0.7724	0.8910
b4	0.7699	0.6711	0.8282	0.7383	0.6498	0.6802	0.7837	0.7712	0.8828
b5	0.7684	0.6160	0.7628	0.8145	0.5664	0.6141	0.8890	0.7509	0.8627
c1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
c2	0.9449	0.9388	0.6406	0.6720	0.7954	0.8951	0.9111	0.9803	0.9822
c3	0.9087	0.7187	0.9521	0.6532	0.4063	0.6506	0.9807	0.6579	0.9858
c4	0.9334	0.2057	0.7234	0.7160	0.4181	0.5518	0.7379	0.2495	0.9844
c5	0.9100	0.1663	0.8468	0.7315	0.6587	0.6263	0.8478	0.3693	0.9747
d1	1.0000	1.0000		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
d2	0.9308	0.9565		0.7249	0.5826	0.7522	0.6218	0.7755	0.9456
d3	0.9606	0.9624		0.5918	0.6335	0.6579	0.7978	0.8724	0.8503
d4	0.9520	0.1097		0.7147	0.6157	0.6424	0.7971	0.8483	0.8335
d5	0.9515	0.1614		0.7818	0.5416	0.8976	0.8822	0.8970	0.9665

References

1. D. Chung and G. Sapiro. Segmenting skin lesions with partial differential equation based image processing algorithms. *IEEE International Conference on Image Processing*, 3(24):404–407, 2000.
2. W. Clarke, P. Lachenbruch, and B. Broffitt. How non-normality affects the quadratic discriminant function. *Communications in Statistics*, A8:1285–1301, 1979.
3. K. Conradsen. *En Introduktion til Statistik. Bind 2.* IMSOR DTH., Lyngby, 1984.
4. A. Gutenev, V. Skaldnev, and D. Varvel. Acquisiton-time quality control in digital dermatoscopy of skin lesions. *Computerized Medical Imaging and Graphics*, 25:495–499, 2001.
5. A. Hay. The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9(8):1395–1398, 1988.
6. G. Maletti and B. Ersbøll. Principal component analysis of psoriasis lesions images. Technical Report 5, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
7. G. Maletti, B. Ersbøll, and K. Conradsen. A contextual classifier that only requires one prototype pixel for each class. *IEEE Transactions on Nuclear Science*, 49(3):700–706, June 2002.
8. M. McGuire and H. Stone. Techniques for multiresolution image registration in the presence of occlusions. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1476–1479, May 2000.
9. G. Rosenfield and K. Fitzpatrick-Lins. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2):223–227, February 1986.
10. S. Stehman. Thematic map accuracy assessment from the perspective of finite population sampling. *International Journal of Remote Sensing*, 16(3):589–593, 1995.
11. Y. Wang and T. Adali. *Signal Processing for Magnetic Resonance Imaging and Spectroscopy. Stocastic Model Based Image Analysis*, chapter 14, pages 1–34. 2000. (Topics in Biomedical Imaging Course Notes.).

4.2.3 Illumination Correction in Psoriasis Lesions Images

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 7
Pages: 1 – 25
Reference: [35]

Illumination Correction in Psoriasis Lesion Images

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. An approach to automatically correct illumination problems in dermatological images is presented. The illumination function is estimated after combining the thematic map indicating skin -produced by an automated classification scheme- with the dermatological image data. The user is only required to specify the class for which its thematic map is most suitable to be used in the illumination correction. Results are shown for real examples. It is also shown that the classification output improves after illumination correction.

1 Introduction

We use a set of 175 *RGB* psoriasis lesion images, of size 556 pixels by 748 pixels, taken at the Gentofte Hospital, Denmark, during pilot sessions with three invited patients. For each patient, three lesions were followed once a week during at least three weeks. In each session, five images of each lesion were taken. Examples can be seen in Figure 3. The skin images are affected by shadows due to the non-plane shape of the objects. Thanks to the use of an integrating sphere [1] (see Figures 1 and 2) with optimal illumination conditions some skin images [6] could be corrected. The sphere assumes that the captured objects are plane. A quadratic model is assumed for the objects, because, for instance, an arm or a leg, as a first approximation, is similar to a cylinder than to a plane.

On the other hand, in an ideal situation, for a given lesion, the classification output should contain a repeated pattern within and between sessions. It could be nice to count with an algorithm that selects automatically, from the thematic maps, the objects to be aligned. However, as it was reported in [4], for many examples without any illumination correction, this is not a possibility. Thus, alignment and registration could turn to be a difficult -if not impossible- task. In order to be able to produce valid results regarding the actual data-set, an illumination correction scheme is proposed here.

This report is organized as follows. The first part is composed by the proposed Illumination Correction Scheme and the general results. The second part

Fig. 1. Integrating sphere with optimal illumination conditions used to capture the images.

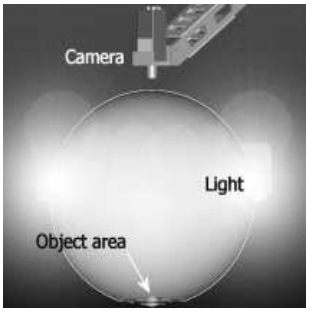


Fig. 2. Sketch of the sphere integrating lighting technology with camera technology. Thanks to Videometer [1] for providing this Figure.

is composed by the Appendix, which contains results for each psoriasis image belonging to the data set. It is included for the reader interested on analyzing the results in detail.

2 Illumination Correction Scheme

A set of data for which it can be assumed that in each neighborhood the data have the same distribution is required. The images of lesions with psoriasis can be assumed to contain three classes: background, normal skin and lesion. In order to segment the lesions, a two-stage hierarchical classification scheme -originally reported in [4]- has been proposed. The user is required to select the class which he considers most suitable to be used as input to the Illumination Function Estimation scheme. Neighborhoods of pixels belonging to that class are assumed to satisfy the same homogeneity criteria. Data are interpolated and the illumination function is estimated from these data. Details follow.

2.1 Classification Scheme

For completeness, general aspects of the scheme proposed in [4] are given here. Basically, in the first stage, the skin is separated from the background, and, in the second stage, the lesion is separated from the normal skin. In both stages, the classes are assumed to be Gaussian distributed. In each stage, means, standard deviations and prior probabilities of the classes were estimated with an Expectation-Maximization Algorithm [7]. These statistics were introduced in a discrimination function [2], which was applied to the output of the convolution of the image with a circular window of optimal size [5].

2.2 Illumination Function Estimation

Following [3], let $i(x, y)$ be the illumination function affecting the original scene $f(x, y)$:

$$g(x, y) = i(x, y)f(x, y) \tag{1}$$

such that $g(x, y)$ is the illumination affected scene. This model assumes that the output is the product between the amount of arriving light and the original scene reflectivity. Basically, the region representing a used selected class k is used to model the illumination function. Within the k-th class we assume as a first approximation that the original image should be constant:

$$i(x, y) = \frac{g_k(x, y)}{\beta_k} \tag{2}$$

where g_k is the observed illumination affected scene in the region representing the class k, and β_k is a unknown constant.

Pixel values belonging to the region representing the k-th class are replaced by local means, which are used to estimate an interpolation function. After normalization, the interpolation function is assumed to be the illumination function affecting the scene. This is more precisely defined as follows. Let G, F and I be specific images of $g(x, y)$, $i(x, y)$ and $f(r, c)$ respectively. Let $\mathsf{F} = \{f[r, c]\}$ and $\mathsf{G} = \{g[r, c]\}$ be hierarchically defined in terms of $\mathsf{Z} = \{z[r, c]\}$ that represents the partition of the scene in different classes. Each $z[r, c]$ is a value in $\Omega = \{1, \ldots, K\}$, where $z[r, c] = k$ means that the coordinate $[r, c]$ of a pixel belongs to the $k - th$ class. Let $\mu = \{\mu_1, \ldots, \mu_K\}$ be the class means, and $\mathsf{M} = \{m[r, c] = \mu_k / z[r, c] = k \forall k \in \Omega\}$ be such that $m[r, c]$ is the mean of the class the pixel at location $[r, c]$ belongs to. Let N be defined with the local means of G such that $E[\mathsf{N}] = \mathsf{M}$. Let $\hat{\mathsf{N}}_k$ be the interpolation function of sampled values of $\mathsf{N}_k =< \mathsf{N}, \mathsf{Z} = k >$ greater than zero [1]. Then $\hat{\mathsf{I}}[r, c] = \frac{\hat{\mathsf{N}}_k[r,c]}{O_k}$ and O_k is the normalization factor.

[1]

$$< \mathsf{N}, \mathsf{Z} = k > \tag{3}$$

Body parts are modelled with a quadratic function. The regression matrix, from which the coefficients of the interpolation function are adjusted, is given by:

$$\boldsymbol{y}_{Px1} = \left[\mathbf{1}_{P\times 1}\ \boldsymbol{r}\ \boldsymbol{c}\ \boldsymbol{rc}\ \boldsymbol{rr}\ \boldsymbol{cc}\right]\left[\boldsymbol{\theta}_{Px1}\right] + \boldsymbol{\Sigma} \tag{4}$$

where y_i is the local mean value $n_k[r_i, c_i]$ for a given band and $[r_i, c_i]$ is the location of the i-th pixel sampled, which belongs to the k-th class. $\boldsymbol{r} = [r_1, \ldots, r_P]^T$, $\boldsymbol{c} = [c_1, \ldots, c_P]^T$ and $\boldsymbol{rc} = [r_1c_1, \ldots, r_Pc_P]^T$, $\boldsymbol{rr} = [r_1r_1, \ldots, r_Pr_P]^T$ and $\boldsymbol{cc} = [c_1c_1, \ldots, c_Pc_P]^T$. The normalization factor O_k is the maximal value of $(\hat{G}_k[r, c] + \gamma)$ and γ is the minimal constant added to $G_k[r, c]$ such that the output is positive.

This procedure can be applied to each single image band. Thus, for multispectral images, the output of the present scheme is a multi-spectrally estimated illumination function.

3 Results and Discussion

The algorithm that estimates the illumination function requires that, for each single image, a map of an expected homogeneous region must be provided. On the other hand, the problem is precisely that the classification can not be done in a uniform way and for this reason, the illumination correction is needed. As was mentioned before, it is assumed that the images of lesions with psoriasis contain three classes: curtain, normal skin and lesion. The best candidate to be interpolated is the normal skin, because it usually appears in different parts of the image. However, there are cases for which we could consider that the whole image is pure lesion. For this reason, each set of images corresponding to the same lesion was treated separately. For the images of the first patient, it was possible to provide a rough estimation of the maps corresponding to the normal skin. These maps were the output of the second stage of the classification scheme used. However, for the remaining cases, the maps of the skin, including lesions[2], were used to estimate the illumination function. The thematic maps produced in the classification scheme combined with the original images used as input for the illumination correction stage are shown in Figure 3. As it can be seen, for the cases $1A$, $1B$ and $1C$, the region mapping normal skin was used. For the cases $2A$ and $3A$, a map indicating the whole image was generated. For the cases $2B$, $2C$, $3B$ and $3C$ the thematic map indicating skin was selected.

The image region to be modelled was convolved with a circular window with a diameter of eleven pixels, empirically found. In order to reduce, during the

is the scalar product between the tensor N and the thematic map that has, in a given location the value one, if the corresponding pixel belongs to the k-th class and the value zero otherwise.

[2] This means the output of the first stage of the classification scheme

convolution, the influence of pixel values belonging to other classes in the region borders, pixel values to be excluded from the model were previously replaced with the mean value of the data to be modelled. This was also done in order to preserve the correct shadow information of the region to be modelled. An artificial grid was placed on the whole image, and grid pixels belonging to the region to be modelled were sampled.

The estimated multi-spectral illumination functions and the corresponding estimated illumination-corrected images can be seen in Figure 4. The estimated illumination function for each single image of the psoriasis image set can be found in Appendix A. Note that the extrapolation errors mostly occur in the background, which really does not contain information of interest anyhow.

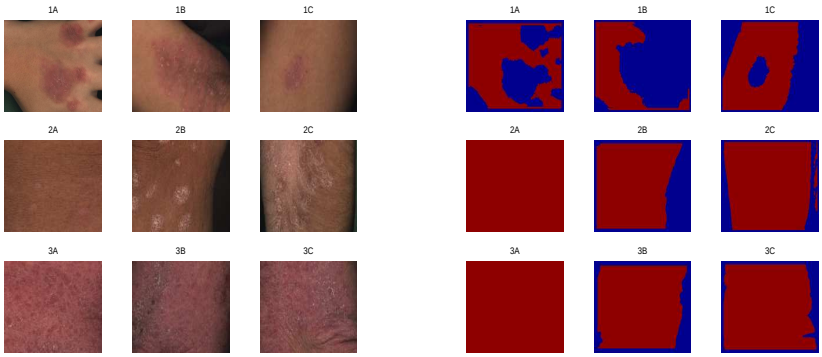


Fig. 3. Examples of original images of lesions with psoriasis (left) and the thematic maps (right) outputs of the classification scheme used in the illumination function estimation.

Comments based on visual assessment of the illumination functions of each group of images of the same lesion follow. It is quite obvious that a quadratic model is not very representative of an object like a hand. This is reflected in the illumination function, which does not include the shadows between the fingers. Continuing with the next lesion, we can observe that part of the lesion B of patient 1 is covered by shadow in the original images. In spite of the illumination function being almost the same for all the captures, it has to be noted, that the lesion presents occlusions in some cases, which means that, along the captures, the shadow is not affecting a given part of the lesion in the same way. With respect to lesion C of patient 1, which is placed in the center of the image, it seems that the illumination correction would not improve future alignment and registration outputs very much. However, note in the illumination corrected images,

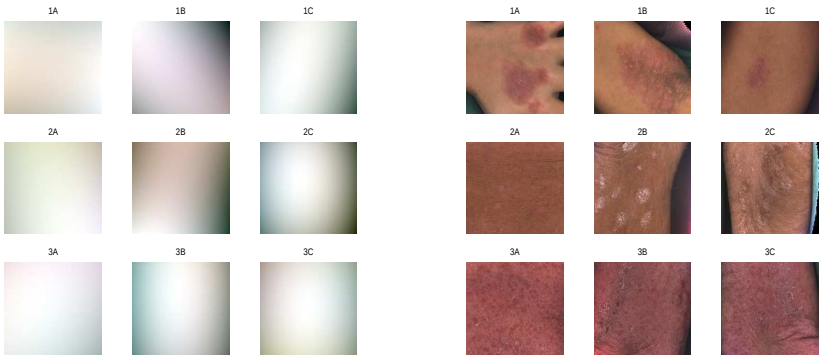


Fig. 4. Estimated illumination functions of the images of Figure 3 (left) and the corresponding illumination corrected images (right).

that the shadows are not removed. Note in [4] that the provided thematic maps indicating skin do not include the regions covered by shadows. Results could be improved finding better the cost values for the discrimination function in the classification scheme. For the lesion A of patient 2 it is noted that the illumination model represents the shadows that appear on the left of the images in the second and fourth session quite well. Comparing the classification output after illumination correction with the classification output using the original data (see [4]), it can be deduced, that, for this case, the illumination correction is clearly necessary. Looking at the illumination corrected images of patient 2, lesion B it can be noted that the illumination model does not correct the shadows in some captures well. This is due to the fact that the thematic map of the skin does not include parts of the shadows, which were assigned to the class curtain in the first classification stage: changes in the cost values could improve the results obtained. For the case $(2, C)$, it is noted that the classification output improves when the illumination is corrected. In spite of being extrapolation errors in the curtain, they obviously do not affect the classification between normal skin and lesion. The illumination functions of the case $(3, A)$ seem to represent the shadows on the original images well. This case is quite difficult, because it is hard to identify correspondent points in the images by visual assessment. Therefore, to evaluate the classification output is difficult. However, for same pairs of sessions, some kind of repeated structure can be observed. The classification output within sessions after illumination correction of the images of the case $(3, B)$ does not look very convincing. This may be due to the spherical shape of the illumination function, which is not representative of the body part modelled. Note in the illumination corrected images the light areas on the image corners. A better thematic map of the region to be modelled could improve the results obtained. For the case $(3, C)$, skin displacements attained after folding the elbow allow to

identify corresponding points visually. By comparing the classification outputs after illumination correction within sessions, in general, they contain a repeated pattern. However, for some pairs of sessions, particularly the last ones, is not quite obvious how the correspondence should be.

The normalized histograms of the region indicating skin was again calculated for each single |B-G| band. Table 1, comparable to Table 2 in [4], was constructed. [3]. Details can be found in Table 2. As it can be observed, in general, the correlation values are smaller than before. The large correlation values before illumination correction could be due to the illumination itself and the actual correlation values are more real for the skin data. It is suggested to extrapolate statistics from one single image to another only after normalization.

Table 1. Average and standard deviation correlation value between the histogram of a normalized illumination-corrected image band and the normalized histogram of the normalized bands of the illumination-corrected images belonging to the same session.

(Patient, Lesion)	μ_a	σ_a	μ_b	σ_b	μ_c	σ_c	μ_d	σ_d
(1,A)	0.6031	0.1133	0.5716	0.0760	0.5757	0.1365	0.5989	0.1250
(1,B)	0.5354	0.1547	0.3669	0.1081	0.3671	0.0933	0.3551	0.0745
(1,C)	0.6550	0.1090	0.6850	0.0869	0.7539	0.1096		
(2,A)	0.6149	0.1154	0.7712	0.0542	0.4616	0.0846	0.4831	0.1299
(2,B)	0.4000	0.0547	0.5591	0.1029	0.4901	0.0693	0.3961	0.0262
(2,C)	0.3421	0.0441	0.4627	0.1523	0.5447	0.0915	0.4285	0.0573
(3,A)	0.5196	0.0717	0.6408	0.1374	0.3968	0.0896	0.4179	0.0456
(3,B)	0.5913	0.0636	0.4585	0.0818	0.3425	0.0239	0.3450	0.0821
(3,C)	0.4813	0.1172	0.4678	0.0337	0.5518	0.0994	0.4702	0.0626

The same thematic maps indicating skin generated in the previous classification using the original images, were assumed to be the output of the first classification stage using the images with illumination correction. The outputs of the second stage of the classification scheme are shown in Subsection C. In order to avoid extrapolation errors, the discrimination function was computed for each single image separately.

As it can be seen, thematic maps of some lesions, have some repeated patterns in common. Based on this, it is considered that the output of the illumination correction is a better estimation of the expected data. However, in order to be able to automatically select corresponding objects in the thematic maps of the same lesion, further work needs to be done. Since it is not the main objective of the present work, the cases for which it was not possible to automatically define corresponding objects, manually selection after visual inspection of the images

[3] For computational details, see Equations 4 and 5 in [4]

is suggested.

It should to be mentioned that in many cases, shadows in the skin were assigned to the class curtain. This did not allow the introduction of the correct slope of the surface that represents the skin in the illumination model. Finding better cost values to use use in the discrimination functions is encouraged in order to improve the obtained results.

4 Conclusions

A new scheme for illumination correction in dermatological data that combines conventional tools has been proposed. The classification scheme used has been shown to produce suitable thematic maps to be used in the estimation of the illumination function affecting the scene. It has also been shown that the illumination corrected images are a better input for the second stage of the classification scheme. After illumination correction, the classification results improved in cases for which it was not possible to produce thematic maps with a repeated pattern previously. By visual assessment of the results, it is encouraging to extend the scheme to other application areas.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A Illumination functions

Fig. 5. Illumination functions of the original images of (Patient 1, Lesion A).

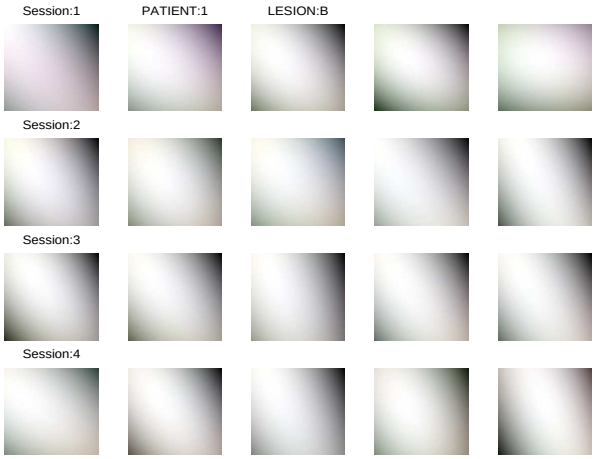


Fig. 6. Illumination functions of the original images of (Patient 1, Lesion B).

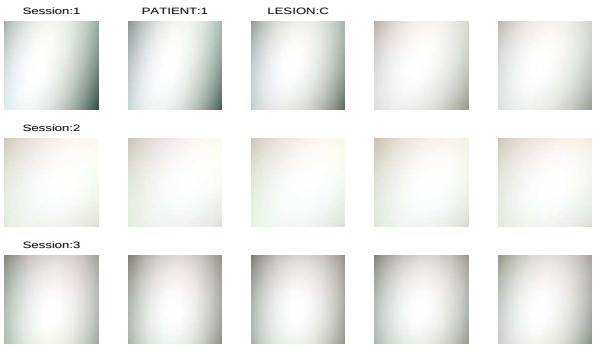


Fig. 7. Illumination functions of the original images of (Patient 1, Lesion C).

Fig. 8. Illumination functions of the original images of (Patient 2, Lesion A).

Fig. 9. Illumination functions of the original images of (Patient 2, Lesion B).

Fig. 10. Illumination functions of the original images of (Patient 2, Lesion C).

Fig. 11. Illumination functions of the original images of (Patient 3, Lesion A).

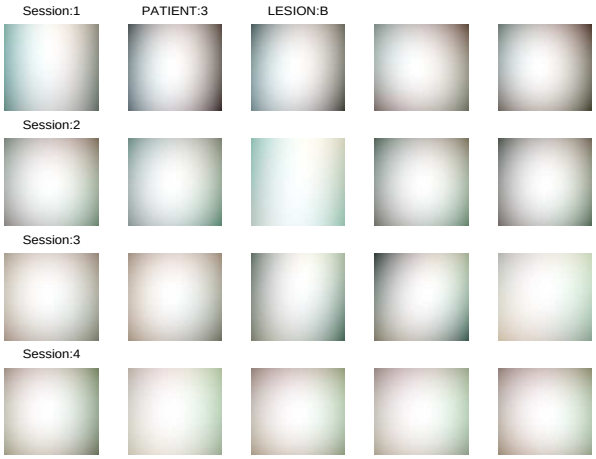


Fig. 12. Illumination functions of the original images of (Patient 3, Lesion B).

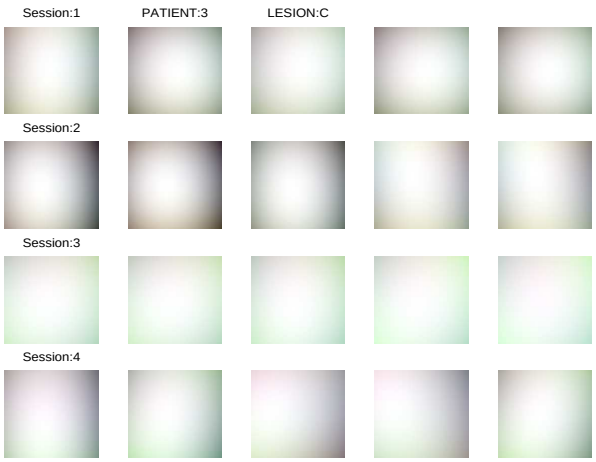


Fig. 13. Illumination functions of the original images of (Patient 3, Lesion C).

B Illumination corrected images

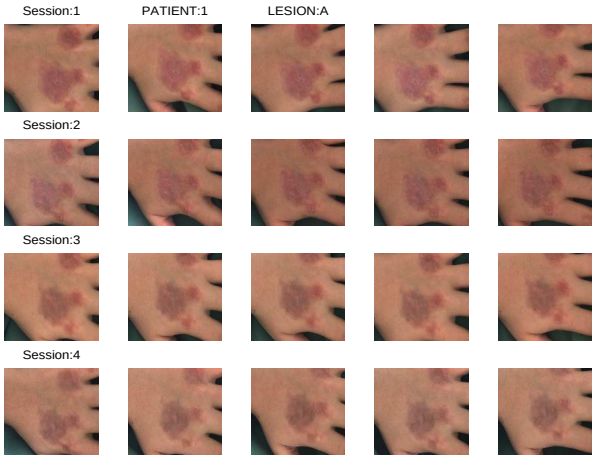


Fig. 14. Images of (Patient 1, Lesion A) after illumination correction.

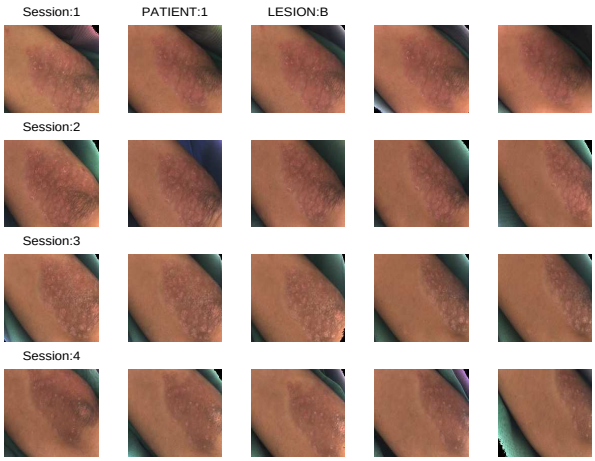


Fig. 15. Images of (Patient 1, Lesion B) after illumination correction.

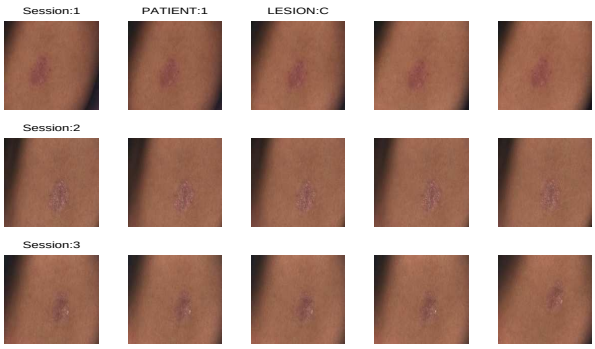


Fig. 16. Images of (Patient 1, Lesion C) after illumination correction.

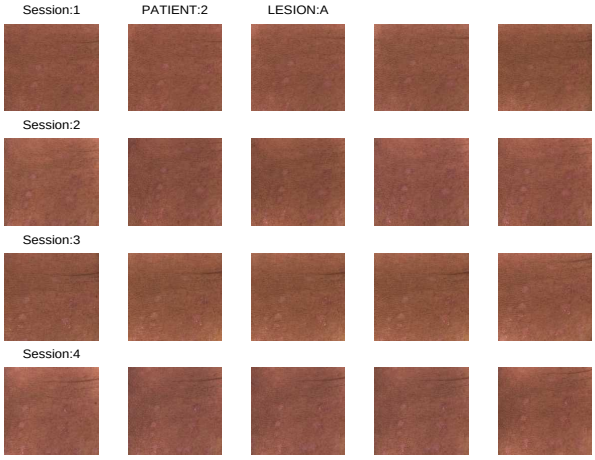


Fig. 17. Images of (Patient 2, Lesion A) after illumination correction.

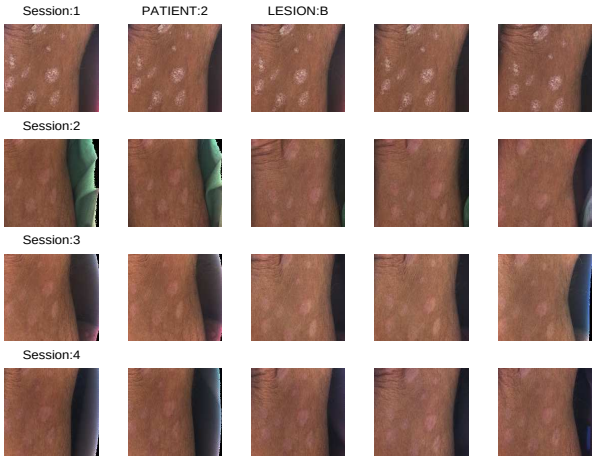


Fig. 18. Images of (Patient 2, Lesion B) after illumination correction.

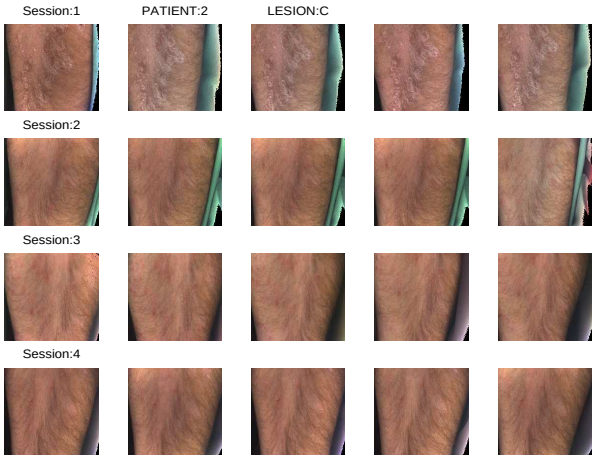


Fig. 19. Images of (Patient 2, Lesion C) after illumination correction.

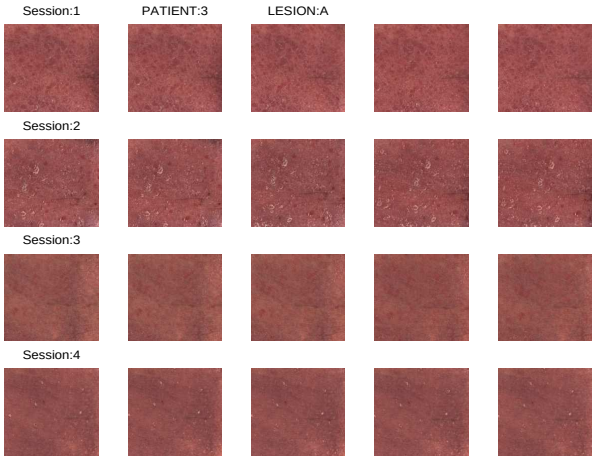


Fig. 20. Images of (Patient 3, Lesion A) after illumination correction.

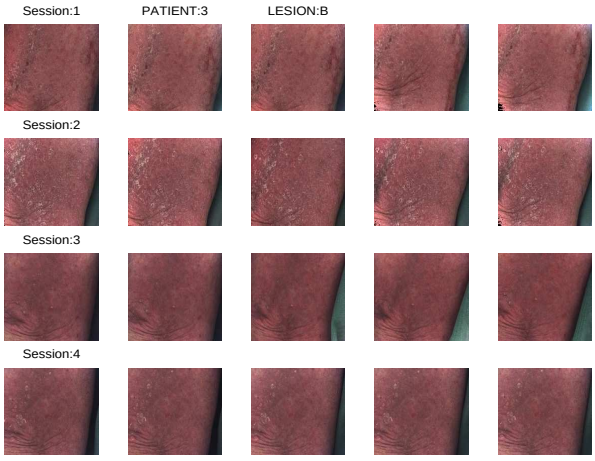


Fig. 21. Images of (Patient 3, Lesion B) after illumination correction.

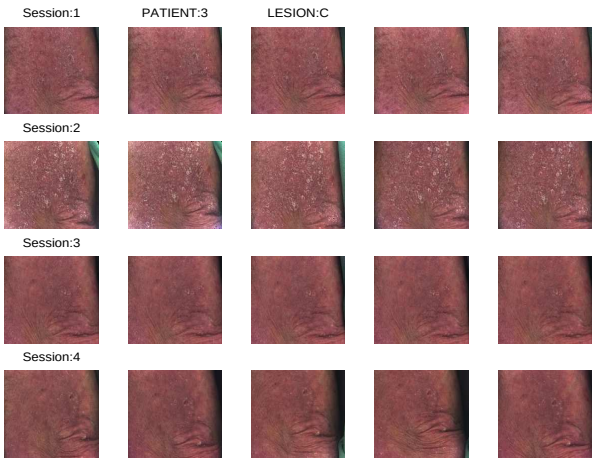


Fig. 22. Images of (Patient 3, Lesion C) after illumination correction.

Table 2. Average correlation values of the normalized histogram of the |B-G| band skin data of each single corrected image with the normalized histogram of the |B-G| band skin data of the corrected images that belong to the same session than that.

Capture	$(1,A)$	$(1,B)$	$(1,C)$	$(2,A)$	$(2,B)$	$(2,C)$	$(3,A)$	$(3,B)$	$(3,C)$
a1	0.6155	0.3172	0.7399	0.7006	0.4071	0.3594	0.5819	0.6593	0.5928
a2	0.6107	0.6505	0.7344	0.6982	0.4515	0.2797	0.4596	0.5422	0.4386
a3	0.6944	0.4272	0.7096	0.6887	0.4162	0.3514	0.4640	0.6583	0.6006
a4	0.4121	0.6614	0.6049	0.5386	0.3069	0.3222	0.6117	0.5698	0.4538
a5	0.6828	0.6205	0.4861	0.4483	0.4184	0.3978	0.4810	0.5267	0.3208
b1	0.5682	0.1986	0.7548	0.7964	0.6628	0.3162	0.7402	0.4282	0.4367
b2	0.4687	0.4596	0.6169	0.8109	0.6593	0.5622	0.7367	0.3680	0.4961
b3	0.6497	0.4624	0.7436	0.7443	0.4637	0.6063	0.4906	0.5456	0.5097
b4	0.5311	0.3413	0.7426	0.6886	0.4475	0.5486	0.4900	0.5441	0.4369
b5	0.6405	0.3727	0.5669	0.8157	0.5621	0.2799	0.7464	0.4065	0.4595
c1	0.3731	0.2717	0.8276	0.3583	0.5686	0.4148	0.3528	0.3160	0.5978
c2	0.6779	0.3465	0.6987	0.4786	0.3974	0.6229	0.4282	0.3344	0.5948
c3	0.4952	0.2880	0.5855	0.5422	0.4491	0.4847	0.4726	0.3351	0.5897
c4	0.6554	0.4677	0.8287	0.5392	0.5425	0.6186	0.2616	0.3461	0.6024
c5	0.6768	0.4615	0.8288	0.3900	0.4931	0.5826	0.4685	0.3807	0.3742
d1	0.6438	0.3840		0.3499	0.3912	0.4589	0.3807	0.4331	0.4318
d2	0.6685	0.3044		0.3363	0.4080	0.4788	0.4702	0.3092	0.5418
d3	0.6429	0.2590		0.5912	0.3533	0.4002	0.4609	0.2485	0.4083
d4	0.6630	0.3779		0.5395	0.4066	0.4633	0.4072	0.3055	0.5336
d5	0.3763	0.4502		0.5988	0.4214	0.3410	0.3706	0.4289	0.4355

C Image classification after illumination correction

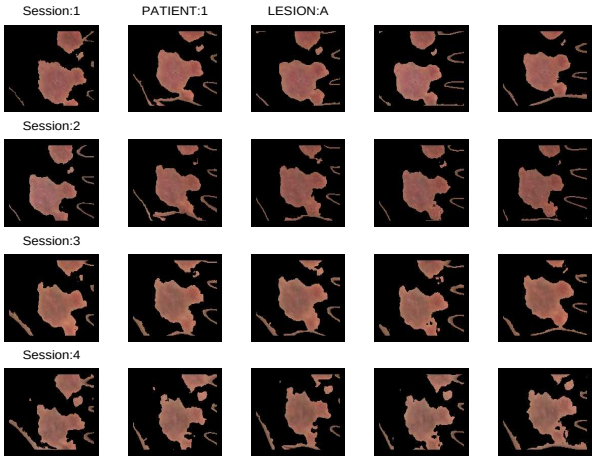


Fig. 23. Overlay of the thematic maps with the images for (Patient 1, Lesion A) after illumination correction.

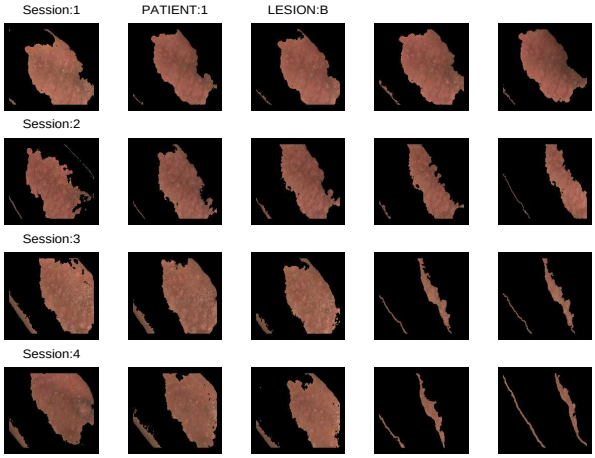


Fig. 24. Overlay of the thematic maps with the images for (Patient 1, Lesion B) after illumination correction.

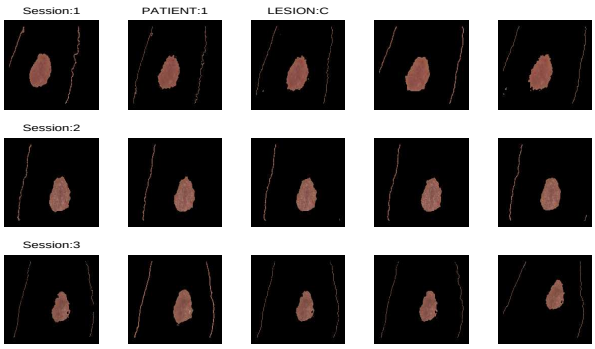


Fig. 25. Overlay of the thematic maps with the images for (Patient 1, Lesion C) after illumination correction.

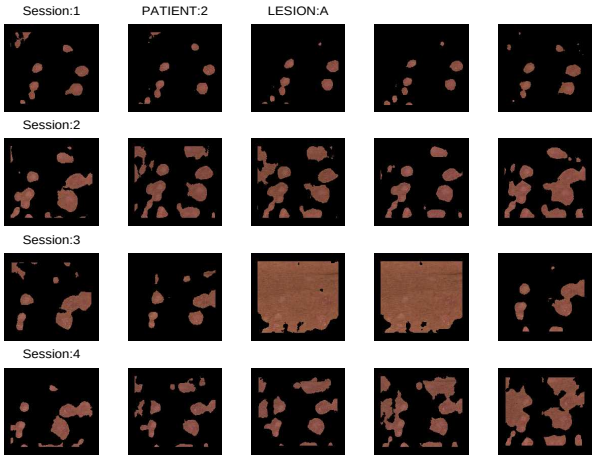


Fig. 26. Overlay of the thematic maps with the images for (Patient 2, Lesion A) after illumination correction.

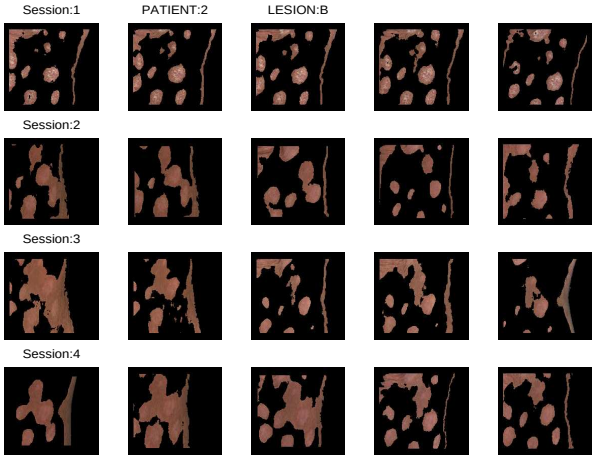


Fig. 27. Overlay of the thematic maps with the images for (Patient 2, Lesion B) after illumination correction.

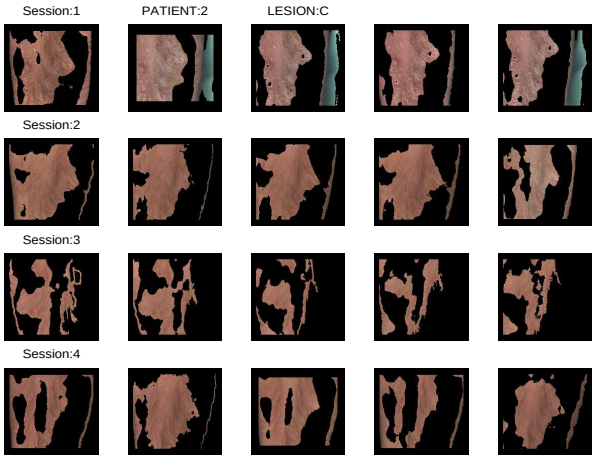


Fig. 28. Overlay of the thematic maps with the images for (Patient 2, Lesion C) after illumination correction.

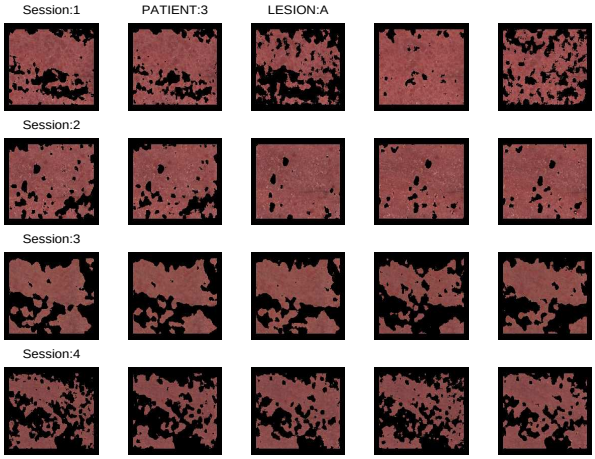


Fig. 29. Overlay of the thematic maps with the images for (Patient 3, Lesion A) after illumination correction.

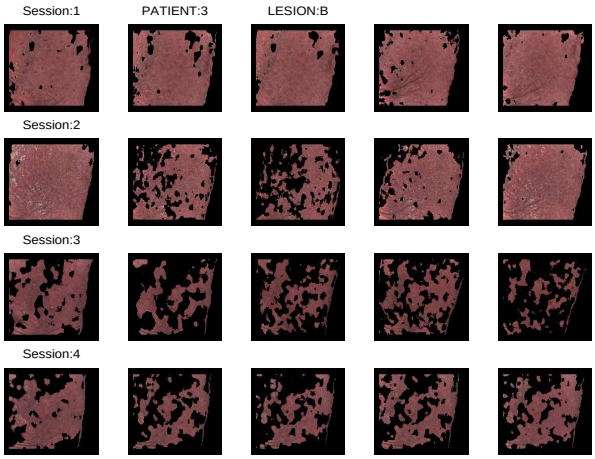


Fig. 30. Overlay of the thematic maps with the images for (Patient 3, Lesion B) after illumination correction.

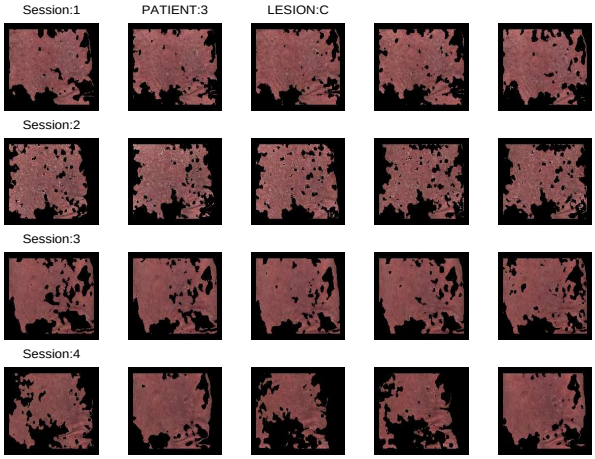


Fig. 31. Overlay of the thematic maps with the images for (Patient 3, Lesion C) after illumination correction.

References

1. Videometer. http://www.videometer.dk.
2. K. Conradsen. *En Introduktion til Statistik. Bind 2.* IMSOR DTH., Lyngby, 1984.
3. J. Lira. *Introducción al Tratamiento Digital de Imágenes.* Instituto Politécnico Nacional, Universidad Nacional Autónoma de México, Fondo de Cultura Económica., México D.F., 1 edition, September 2002.
4. G. Maletti and B. Ersbøll. A hierarchical classification scheme of psoriasis images. Technical Report 6, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
5. G. Maletti, B. Ersbøll, and K. Conradsen. A contextual classifier that only requires one prototype pixel for each class. *IEEE Transactions on Nuclear Science*, 49(3):700–706, June 2002.
6. U. Mattsson, A Jönsson, M. Jontell, and J. Cassuto. Digital image analysis (dia) of color changes in human skin exposed to standardized thermal injury and comparison with laser doppler measurements. *Computer Methods and Programs in Biomedicine*, 50:31–42, 1996.
7. Y. Wang and T. Adali. *Signal Processing for Magnetic Resonance Imaging and Spectroscopy. Stocastic Model Based Image Analysis*, chapter 14, pages 1–34. 2000. (Topics in Biomedical Imaging Course Notes.).

4.2.4 Texture Alteration Detection in Bitemporal Images of Lesions with Psoriasis

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 8
Pages: 1 – 30
Reference: [37]

Texture Alteration Detection in Bitemporal Images of Lesions with Psoriasis

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. The objective of this work is to explore the feasibility of quantifying textural change between pairs of segmented patterns without registering them. The Multi-variate Alteration Detection (M.A.D.) Transform is applied to a texture model constructed with the data of segmented psoriasis lesions images. The texture model is Haralick's co-occurrence matrix, which is computed and normalized for each single band with the equalized data of a given lesion. The contribution of each single color band to the textural change is analyzed.

1 Introduction

We use a database consisting of a set of 175 images of lesions with psoriasis taken at the Gentofte Hospital of Denmark during four pilot sessions with three invited patients. For each patient, three lesions were captured five times during each session, along four weeks. The images were labelled with four characters, indicating patient (1, 2, 3), lesion (A, B, C), session (a, b, c, d) and capture (1, 2, 3, 4, 5) respectively. The original *RGB* images were reduced in size to $(576 - 20)/4$ rows by $(768 - 20)/4$ columns of pixels. The images can be assumed to contain three classes (background, normal skin and lesion), which were segmented with a two-stage hierarchical classification scheme [9]. This scheme separates, in the first stage, the skin from the background, and, in the second stage, the lesion from the normal skin. Since the images are affected by shadows, an illumination correction scheme was proposed [10].

To evaluate lesion changes between sessions, physicians make scores of a four-variables set: the redness, the scaling, the thickness and the body area covered by the lesions. Manual scoring highly depends on the dermatologist, who can show variations in criteria along sessions due to the huge amount of patients and lesions observed during each working day. In order to use a set of lesions like these as input to an automatic system that analyzes their change in time, it could be helpful to express the data in a way where they are comparable, without loosing connection with the original data. Registration of classified lesions patterns appear as options to possible solutions. However, they could have

a high computational cost.

To transform pairs of not registered original data to a space were the transformed data have a perfect registration is possible through a textural descriptor [17], [8], [13], [1], [7], [16], [15] that is invariant to translations and rotations of patterns, like Haralick's co-occurrence matrix [3]. On the other hand, a statistical approach that allows to detect alterations is the M.A.D. Transform (Multivariate Alteration Detection Transform) [11], [12]. This scheme transforms two sets of multi-variate observations into a difference between two linear combinations of the original variables explaining maximal change in all variables simultaneously.

The present work consists of the description of the following aspects: the textural model, the two-set canonical correlation analysis and the M.A.D. Transform. Results show the contribution of each color band to each M.A.D. component. Finally, a back-projection approach from the M.A.D. output to the original data is proposed.

2 The Texture Model

Following [6], let δ be a compact and convex set established in the image space as a circular window and let β be a vector position operator relating the relative spatial location of a pair of pixels in the window; for each single band, the co-occurrence matrix C is the estimated probability of having a pair of pixels (λ_p, λ_q) in the relative position given by β. The elements of the $Q \times Q$ co-occurrence matrix $C(\lambda_p, \lambda_q, \beta, \delta)$, with Q being the number of quantization levels, are given by:

$$\frac{O\{r|r, r+\beta \in \sigma(\delta), g(r) = \lambda_p, g(r+\beta) = \lambda_q\}}{O\{r|r+\beta \in \delta\}} \tag{1}$$

where $g(r)$ is the subset of the RGB image corresponding to the lesion with psoriasis, $\sigma(\delta)$ is a translation isometry over the window, O is the order of the set, and $r = (i, j)$ is the vector position of a pixel.

3 Two-Set Canonical Correlations Analysis

Conceptually, two-set canonical correlations analysis is a technique that finds corresponding sets of linear combinations -called canonical variables- of two groups of variables, such that the first canonical variables are the ones with the largest correlation, and higher order canonical variables are maximally correlated subject to orthogonality or uncorrelatedness with lower order canonical variables [11], [4].

Formally, let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two sets of variables of dimensions m_1 and m_2 respectively, with $m_1 \leq m_2$. The two-set canonical correlation analysis chooses $m_1 + m_2$ linear transforms

$$\begin{array}{c} \boldsymbol{y}_{1i} = \boldsymbol{a}_{1i}^T \boldsymbol{x}_1, i = 1, ..., m_1 \\ \boldsymbol{y}_{2i} = \boldsymbol{a}_{2i}^T \boldsymbol{x}_2, i = 1, ..., m_2, \end{array} \tag{2}$$

such that the correlations within groups are

$$\begin{array}{c} Corr[\boldsymbol{a}_{1i}^T \boldsymbol{x}_1, \boldsymbol{a}_{1j}^T \boldsymbol{x}_1] = \delta_{ij} \\ Corr[\boldsymbol{a}_{2i}^T \boldsymbol{x}_2, \boldsymbol{a}_{2j}^T \boldsymbol{x}_2] = \delta_{ij} \end{array} \tag{3}$$

and the correlations between groups are

$$Corr[\boldsymbol{a}_{1i}^T \boldsymbol{x}_1, \boldsymbol{a}_{2j}^T \boldsymbol{x}_2] = \rho_j \delta_{ij} \tag{4}$$

where δ is the Kronecker delta.

Let the sets of variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be described by the $m_1 + m_2$ dimensional variable

$$X = [\boldsymbol{x}_1^T \boldsymbol{x}_2^T] \tag{5}$$

and assume

$$E[\boldsymbol{X}] = 0 \wedge D[\boldsymbol{X}] = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{6}$$

with $\Sigma_{ij} \in \Re^{m_i x m_j}$ and

$$\Sigma_{ij} = \Sigma_{ji}^T, \forall i, j = 1, 2 \tag{7}$$

and non-singular.

We would like to find the transforms

$$\begin{array}{c} \boldsymbol{y}_1 = \boldsymbol{a}_1^T \boldsymbol{x}_1 \\ \boldsymbol{y}_2 = \boldsymbol{a}_2^T \boldsymbol{x}_2 \end{array} \tag{8}$$

with

$$D\{Y\} = \Sigma_Y = \begin{bmatrix} Var[\boldsymbol{y}_1] & Cov[\boldsymbol{y}_1, \boldsymbol{y}_2] \\ Cov[\boldsymbol{y}_2, \boldsymbol{y}_1] & Var[\boldsymbol{y}_1] \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^T \Sigma_{11} \boldsymbol{a}_1 & \boldsymbol{a}_1^T \Sigma_{12} \boldsymbol{a}_2 \\ \boldsymbol{a}_2^T \Sigma_{21} \boldsymbol{a}_1 & \boldsymbol{a}_2^T \Sigma_{22} \boldsymbol{a}_2 \end{bmatrix} \tag{9}$$

under the constraints

$$\begin{array}{c} Var[\boldsymbol{y}_1] = 1 \\ Var[\boldsymbol{y}_2] = 1 \end{array} \tag{10}$$

such that the correlation

$$\rho = Corr[\boldsymbol{y}_1, \boldsymbol{y}_2] = \frac{Cov[\boldsymbol{y}_1, \boldsymbol{y}_2]}{\sqrt{Var[\boldsymbol{y}_1] Var[\boldsymbol{y}_2]}} = \boldsymbol{a}_1^T \Sigma_{12} \boldsymbol{a}_2 \tag{11}$$

is maximized.

Let $\lambda_1/2$ and $\lambda_2/2$ be Lagrange multipliers introduced in

$$\psi = \boldsymbol{a}_1^T \Sigma_{12} \boldsymbol{a}_2 - \frac{\lambda_1}{2}(\boldsymbol{a}_1^T \Sigma_{11} \boldsymbol{a}_1 - 1) - \frac{\lambda_2}{2}(\boldsymbol{a}_2^T \Sigma_{22} \boldsymbol{a}_2 - 1). \tag{12}$$

We would like to maximize ψ with respect to $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. Due to the symmetry condition in Equation 7, the following matrix differentiation property can be applied

$$\frac{\partial(\boldsymbol{a}_i^T \Sigma_{ii} \boldsymbol{a}_i)}{\partial \boldsymbol{a}_i} = 2\Sigma_{ii}\boldsymbol{a}_i \tag{13}$$

and the partial derivatives of ψ with respect to $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are

$$\frac{\partial \psi}{\partial \boldsymbol{a}_1} = \Sigma_{12}\boldsymbol{a}_2 - \lambda_1 \Sigma_{11}\boldsymbol{a}_1 \tag{14}$$

$$\frac{\partial \psi}{\partial \boldsymbol{a}_2} = \Sigma_{21}\boldsymbol{a}_1 - \lambda_2 \Sigma_{22}\boldsymbol{a}_2 \tag{15}$$

Note that Equations 14 and 15 could also be obtained by rewriting Equation 12 with λ_1 and λ_2 as dependent variables: λ_1 is derived with respect to $\boldsymbol{a}_1$ and λ_2, with respect to $\boldsymbol{a}_2$; the derivative of the quotient and the symmetry condition 7 is applied and the huge expressions obtained can easily be simplified because the main part of each one is only λ_1 and λ_2 respectively.

Setting the partial derivatives of ψ to zero and multiplying respectively by $\boldsymbol{a}_1^T$ and $\boldsymbol{a}_2^T$, we have

$$\boldsymbol{a}_1^T \Sigma_{12}\boldsymbol{a}_2 - \lambda_1 \boldsymbol{a}_1^T \Sigma_{11}\boldsymbol{a}_1 = 0 \tag{16}$$

$$\boldsymbol{a}_2^T \Sigma_{21}\boldsymbol{a}_1 - \lambda_2 \boldsymbol{a}_2^T \Sigma_{22}\boldsymbol{a}_2 = 0. \tag{17}$$

Due to $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are scalars, the following property can be applied $\boldsymbol{a}_1^T \Sigma_{12}\boldsymbol{a}_2 = (\boldsymbol{a}_1^T \Sigma_{12}\boldsymbol{a}_2)^T$. The transpose of a product is the product of the transposes with inverted order, and using Equation 7, the correlation ρ between the transformed variables turns to be

$$\rho = \boldsymbol{a}_1^T \Sigma_{12}\boldsymbol{a}_2 = \boldsymbol{a}_2^T \Sigma_{21}\boldsymbol{a}_1 = \lambda_1 = \lambda_2 \tag{18}$$

Substituting Equation 18 in Equations 14 and 15

$$\Sigma_{12}\boldsymbol{a}_2 - \rho \Sigma_{11}\boldsymbol{a}_1 = 0 \tag{19}$$

$$\Sigma_{21}\boldsymbol{a}_1 - \rho \Sigma_{22}\boldsymbol{a}_2 = 0 \tag{20}$$

and multiplying Equation 20 by Σ_{22}^{-1} gives

$$\rho = \Sigma_{22}^{-1}\Sigma_{21}\frac{\boldsymbol{a}_1}{\boldsymbol{a}_2}. \tag{21}$$

Pre-multiplying Equation 19 with ρ, substituting it with Equation 21 and re-writing everything with respect to $\boldsymbol{a}_1$ we have

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\boldsymbol{a}_1 = \rho^2\Sigma_{11}\boldsymbol{a}_1, \tag{22}$$

which is recognized as an eigenvalue problem. On the other hand, multiplying Equation 19 by Σ_{11}^{-1} gives

$$\rho = \Sigma_{11}^{-1}\Sigma_{12}\frac{\boldsymbol{a}_2}{\boldsymbol{a}_1}. \tag{23}$$

Substituting Equation 23 in - previously multiplied by the scalar ρ- Equation 20 we have

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\boldsymbol{a}_2 = \rho^2\Sigma_{22}\boldsymbol{a}_2. \tag{24}$$

Rewriting Equations 22 and 24 we have the generalized equations of eigenvalues and eigenvectors:

$$[\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1} - \rho^2 I]\boldsymbol{a}_1 = 0 \tag{25}$$

and

$$[\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} - \rho^2 I]\boldsymbol{a}_2 = 0. \tag{26}$$

The sets $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are obtained by projecting the sets $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ onto the subspaces spanned by the eigenvectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ respectively with the corresponding largest eigenvalue equal to the square correlation.

If $m_1 = m_2$ the $\boldsymbol{y}_{1i}, i = 1, ..., m_1$ are obtained by projecting $\boldsymbol{x}_1$ onto the subspaces spanned by the eigenvectors $\boldsymbol{a}_{11}, ..., \boldsymbol{a}_{1m_1}$ corresponding to the eigenvalues $\rho_1^2 \geq ... \geq \rho_{m_1}^2$ of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ with respect to Σ_{11} , and the $\boldsymbol{y}_{2i}, i = 1, ..., m_2$ are obtained by projecting $\boldsymbol{x}_2$ onto the subspaces spanned by the eigenvectors $\boldsymbol{a}_{21}, ..., \boldsymbol{a}_{2m_2}$ corresponding to the same eigenvalues. If $m_1 \leq m_2$ the eigenvalue problem in Equation 22 degenerates since the last eigenvalue will equal zero with $(m2 - m1)$ multiplicity.

4 The M.A.D. Transform

Again, for completeness, let us follow [11] for describing the M.A.D. Transform. Let X_1 and X_2 be two sets of variables of dimensions m_1 and m_2 respectively, with $m_1 \leq m_2$, $E\{\mathsf{X}_1\} = E\{\mathsf{X}_2\} = 0$ transformed with the coefficients from a standard canonical correlation analysis $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in

$$\begin{array}{l} \mathsf{Y}_{1i} = \boldsymbol{a}_{1i}^T\mathsf{X}_1, i = 1, ..., m_1 \\ \mathsf{Y}_{2i} = \boldsymbol{a}_{2i}^T\mathsf{X}_2, i = 1, ..., m_2. \end{array} \tag{27}$$

Further, Y_1 and Y_2 are positively correlated and with unit variance, such that the variance of their difference is maximized.

$$Var[\boldsymbol{a}_1^T\mathsf{X}_1 - \boldsymbol{a}_2^T\mathsf{X}_2] = \tag{28}$$

$$Var[\boldsymbol{a}_1^T \boldsymbol{X}_1] + Var[\boldsymbol{a}_2^T \boldsymbol{X}_2] - 2Cov[\boldsymbol{a}_1^T \boldsymbol{X}_1, \boldsymbol{a}_2^T \boldsymbol{X}_2] \tag{29}$$

$$= 2(1 - Corr[\boldsymbol{a}_1^T \boldsymbol{X}_1, \boldsymbol{a}_2^T \boldsymbol{X}_2]) \tag{30}$$

The M.A.D. Transform consists of the variates obtained when the corresponding canonical variates are subtracted in reverse order, what means that the m_1^{th} difference shows maximum variance among such variables, and the $(m_1 - j)^{th}$ difference shows maximum variance subject to the constraint that this difference is not correlated with the previous j ones. The M.A.D. variates are invariant to linear and affine scaling, what is not the case of the Principal Components. The dispersion matrix of the M.A.D. variates is

$$D[\boldsymbol{a}_1^T \boldsymbol{X}_1 - \boldsymbol{a}_2^T \boldsymbol{X}_2] = 2(\boldsymbol{I} - \boldsymbol{R}) \tag{31}$$

where $\boldsymbol{I}$ is the $m_1 \times m_1$ unit matrix and $\boldsymbol{R}$ is the $m_1 \times m_1$ matrix containing the sorted canonical correlations on the diagonal and zeros off the diagonal.

5 Results and Discussion

In order to evaluate the color band contribution in the textural change of the lesions, the following experiment was designed. The set of 175 captures obtained at Gentofte's Hospital allowed the construction of 650 pairs of captures with the same time increment of one week between them. For each image of a given patient and lesion, each one of the five images of the next session were associated. For each one of the 175 images, the histogram of the pixels belonging to the area of the thematic map indicating a lesion was first equalized. For each single image, a three band synthetical image of 256 by 256 pixels was constructed in the following way: for each single band, the co-occurrence matrix of the region indicating the single lesion included was constructed and normalized.

Two sets of 650 pairs of images were used to apply the present scheme: the first set was given by the original images; the second set, by the illumination-corrected images (see Figures 13 to 21 in [10]). For each single image, the co-occurrence matrix of the equalized region indicating skin was generated. For each pair of original and illumination-corrected images of the same capture, the same thematic map indicating skin was used (See Figures 2 to 8 in [9]). In order to reduce the influence of very high values, logarithms were applied to the normalized co-occurrence matrix values:

$$y[r, c, b] = \frac{x[r, c, b]}{\sum_r \sum_c x[r, c, b]}. \tag{32}$$

After that, the values were again normalized:

$$z[r, c, b] = \frac{(log(y[r, c, b] * map[r, c, 1] + (1 - map[r, c, b])) + \beta[b]) * map[r, c]}{\sum_{k,l}(z[k, l, b] * map[k, l, b])} \tag{33}$$

where $map[r,c,b]$ is 1 if $y[r,c,b] > 0$ and 0, otherwise, and

$$\beta[b] = max(|z[r,c,b]|) \tag{34}$$

$\forall r, c$.

The matrix is now an image. For each pair of synthetical images constructed in the way mentioned, the M.A.D. transform was applied. For this stage, it was helpful to use the following programs: *fhist* [2], *gclm* [2] [14] and *maf* [5]. For each pair of images of the data-set, the union of the maps indicating positive values for each co-occurrence matrix was generated; the M.A.D. Transform was only applied to the pixels included in that map. Tables 1 and 2 show the results obtained for the sets of original images. Tables 3 and 4 show the results obtained for the sets of illumination-corrected images[1].

Each single cell value in Tables 1 and 3 is the average absolute correlation value between the normalized co-occurrence matrix of the $b-th$ color bands and the $m-th$ MAD components of all pairs of images of the $l-th$ lesion of the $p-th$ patient. It is computed in the following way:

$$E[\rho_{p,l,b,m}] = \frac{1}{(n_s - \delta t)n_c^2} \sum_{s=1}^{n_s-\delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} a_{p,l,s,i,\delta_t,j,b,m} \tag{35}$$

where

$$a_{p,l,s,i,\delta_t,j,b,m} = \frac{1}{2}(|\rho[X_{p,l,s,i,b}, MAD_{p,l,s,i,\delta t,j,b}, m]| + |\rho[X_{p,l,s+\delta t,j,b}, MAD_{p,l,s,i,\delta t,j,b}, m]|) \tag{36}$$

and ρ is the correlation coefficient, δt is the time increment (in this case, δt is one week). The indexes i and j indicate the capture number within a session. $MAD_{p,l,s,i,\delta t,j,b,m}$ equals to $MAD[X_{p,l,s,i,b}, X_{p,l,s+\delta t,j,b}, m]$.

Each single cell value in Tables 2 and 4 is the standard deviation of the averaged absolute correlation values between the normalized co-occurrence matrix of the $b-th$ color bands and the $m-th$ MAD components of all pairs of images of the $l-th$ lesion of the $p-th$ patient. It is computed in the following way:

$$S[\rho_{p,l,b,m}] = \sqrt{\frac{1}{(n_s - \delta t)n_c^2} \sum_{s=1}^{n_s-\delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} (a^2_{p,l,s,i,\delta_t,j,b,m}) - E[\rho_{p,l,b,m}]^2} \tag{37}$$

where the symbols have the same meaning as before.

[1] The original HIPS formatted images were used in combination with the TIFF formatted illumination function to produce the illumination-corrected images in HIPS format. These images were equalized and then the co-occurrence matrix was computed.

Table 1. Average Absolute Correlation Values per Lesion of *RGB* skin data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.163069	0.181529	0.052689
(1,B)	0.093642	0.216019	0.109057
(1,C)	0.150723	0.199478	0.137738
(2,A)	0.117731	0.151841	0.187481
(2,B)	0.094710	0.230283	0.130987
(2,C)	0.111509	0.133244	0.182223
(3,A)	0.038838	0.195560	0.163730
(3,B)	0.032641	0.206640	0.203355
(3,C)	0.043867	0.228722	0.198393
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.149579	0.083112	0.188200
(1,B)	0.236058	0.085506	0.155400
(1,C)	0.261095	0.131136	0.178398
(2,A)	0.238930	0.149425	0.080667
(2,B)	0.247828	0.052757	0.160879
(2,C)	0.233855	0.098236	0.082241
(3,A)	0.252858	0.062821	0.119769
(3,B)	0.280739	0.109337	0.099062
(3,C)	0.285794	0.086110	0.127319
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.167162	0.170654	0.168428
(1,B)	0.184625	0.193093	0.195325
(1,C)	0.226380	0.242602	0.240919
(2,A)	0.244500	0.255134	0.260686
(2,B)	0.192198	0.197892	0.200079
(2,C)	0.279305	0.296285	0.289731
(3,A)	0.165526	0.181658	0.180395
(3,B)	0.204145	0.216882	0.217735
(3,C)	0.192946	0.202501	0.202845

Table 2. Standard Deviation of the Absolute Correlation Values per Lesion of the *RGB* skin data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.025971	0.016633	0.030963
(1,B)	0.028369	0.019309	0.026687
(1,C)	0.031737	0.014496	0.019820
(2,A)	0.059834	0.030295	0.014737
(2,B)	0.040232	0.010890	0.029503
(2,C)	0.070107	0.042084	0.076832
(3,A)	0.012120	0.010609	0.013386
(3,B)	0.017749	0.009709	0.013645
(3,C)	0.029494	0.014647	0.023083
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.028772	0.023656	0.017764
(1,B)	0.026801	0.043394	0.016797
(1,C)	0.022621	0.023489	0.020474
(2,A)	0.033550	0.029944	0.027297
(2,B)	0.018773	0.023604	0.023727
(2,C)	0.059343	0.032142	0.047481
(3,A)	0.011745	0.019359	0.008171
(3,B)	0.017558	0.016535	0.030054
(3,C)	0.022625	0.032087	0.036531
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.013872	0.012831	0.012198
(1,B)	0.024374	0.025512	0.028243
(1,C)	0.011900	0.014789	0.014719
(2,A)	0.031265	0.031427	0.034579
(2,B)	0.014287	0.013550	0.014184
(2,C)	0.038647	0.043020	0.036207
(3,A)	0.021784	0.021555	0.022279
(3,B)	0.021174	0.023636	0.023704
(3,C)	0.012659	0.014226	0.016471

Table 3. Average Absolute Correlation Values per Lesion of the illumination-corrected *RGB* skin data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.170079	0.176821	0.073168
(1,B)	0.183584	0.193015	0.093375
(1,C)	0.208084	0.204445	0.076428
(2,A)	0.099910	0.214296	0.151165
(2,B)	0.150016	0.196696	0.171897
(2,C)	0.070576	0.187925	0.195604
(3,A)	0.113855	0.222418	0.142154
(3,B)	0.024691	0.208501	0.202745
(3,C)	0.090282	0.221817	0.190266
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.154035	0.074054	0.201950
(1,B)	0.203793	0.137015	0.198929
(1,C)	0.186940	0.094268	0.245279
(2,A)	0.249320	0.059774	0.159474
(2,B)	0.229558	0.132791	0.145819
(2,C)	0.282807	0.088674	0.073255
(3,A)	0.255834	0.040157	0.182085
(3,B)	0.276600	0.100579	0.105682
(3,C)	0.285825	0.080852	0.142917
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.177938	0.183928	0.178287
(1,B)	0.214027	0.220947	0.222161
(1,C)	0.200914	0.208052	0.203250
(2,A)	0.181045	0.188425	0.187791
(2,B)	0.202083	0.209892	0.211410
(2,C)	0.220413	0.235589	0.234932
(3,A)	0.179750	0.188971	0.186960
(3,B)	0.208101	0.217542	0.217578
(3,C)	0.198283	0.208827	0.208337

Table 4. Standard Deviation of the Absolute Correlation Values per Lesion of the illumination-corrected RGB skin data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.039767	0.018222	0.036756
(1,B)	0.043357	0.034397	0.047152
(1,C)	0.017927	0.021666	0.017569
(2,A)	0.032620	0.019728	0.024057
(2,B)	0.088681	0.040340	0.058499
(2,C)	0.045544	0.024256	0.043097
(3,A)	0.027636	0.008405	0.017570
(3,B)	0.021223	0.014558	0.014621
(3,C)	0.033461	0.023677	0.024106
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.045084	0.031534	0.017648
(1,B)	0.045152	0.059842	0.043288
(1,C)	0.024923	0.029271	0.006875
(2,A)	0.023577	0.043846	0.027413
(2,B)	0.075262	0.052789	0.051839
(2,C)	0.033527	0.032352	0.030808
(3,A)	0.017790	0.018396	0.011767
(3,B)	0.018079	0.021996	0.018503
(3,C)	0.021804	0.046565	0.038868
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.014500	0.015754	0.014663
(1,B)	0.028974	0.029582	0.035410
(1,C)	0.006570	0.005675	0.006290
(2,A)	0.009182	0.010115	0.010683
(2,B)	0.009957	0.011238	0.011484
(2,C)	0.020532	0.024252	0.025135
(3,A)	0.007396	0.007803	0.009210
(3,B)	0.030657	0.033255	0.033002
(3,C)	0.015375	0.017395	0.019219

Looking at the upper part of Table 1, it seems that for some cases, most of the contribution in the textural change is given by the red and the green bands (cases $(1, A)$ and $(1, C)$), and in others, by the green and the blue bands (the remaining cases). On the other hand, after illumination correction, the values reported in Table 3 indicate that for all the cases of the patient 1, most of the contribution in the textural change is given by the red and the green band and, for instance, the contribution of the red band in the textural change of the case $(3, A)$ is quite significant.

There is an aspect that was not taken into account during the design of the experiment. This was the lesion to skin ratio. As was mentioned in Section 2.1 of [9], the size between lesions [2] has a significant variance. In other words, the proportion of the area covered by the lesion with respect to the area covered by the skin is quite variable. On the other hand, there are cases for which the skin is completely covered by very small lesions, such that the whole classified region indicating skin could be considered part of the class lesion, and the second stage of the classification scheme, is not necessary. This means, that it may be necessary to redesign the experiment in order to compare more homogeneous data.

The same procedure as before was repeated for the cases $(1, A)$, $(1, B)$, $(1, C)$ and $(2, B)$, now using thematic maps indicating lesions. This means, that the normal skin was excluded. For patient 1, the thematic maps indicating lesions generated after illumination correction of the original images were used (see Figures 22 (case $(1, A)$), 23 (case $(1, B)$) and 24 (case $(1, c)$) in [10]). The objects were selected using an automated algorithm developed that assumes that the object of largest size in the thematic map is the desired lesion. The outputs are presented in Section A. For patient 2, the thematic maps used were produced with user-interaction assuming circular shape of the lesion. The user was required to provide for each single image, one single point, which was assumed to be the lesion center. A radius of that circle was also required. It was assumed that for a given lesion, all the shapes had the same radius. This was done for the lesion B (see Figure in Section A). Lesion C was excluded because it is partically covered by hair and the manual selection of one small homogeneous region turns to be difficult. For patient 3, the procedure was not applied, because it can be considered that the whole region indicating skin, is covered by lesion.

As illustration examples, some co-occurrence matrixes computed with data of illumination corrected images are shown in Appendix B. For the cases $(1, A)$, $(1, B)$, $(1, C)$ and $(2, B)$, the co-occurrence matrixes were computed using the data indicated in Appendix A. To compute the co-occurrence matrixes of the cases $(2, C)$, $(3, B)$ and $(3, C)$ the corresponding data sets used can be seen in Figures 6 to 8 of [9] respectively. For the cases $(2, A)$ and $(3, A)$ an artificial map for the down-sampled version of the original images was generated, indicating

[2] This means the size between the lesions of the different cases: $(1, A)$, $(1, B)$, $(1, C)$, $(2, A)$, $(2, B)$, $(2, C)$, $(3, A)$, $(3, B)$ and $(3, C)$.

Table 5. Average Absolute Correlation Values per Lesion of *RGB* lesion data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.228414	0.241037	0.249670
(1,B)	0.087517	0.291845	0.257182
(1,C)	0.021994	0.444337	0.468673
(2,B)	0.021226	0.524634	0.387158
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.299963	0.154635	0.242142
(1,B)	0.334106	0.108684	0.195718
(1,C)	0.101731	0.469909	0.454413
(2,B)	0.061194	0.400099	0.518758
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.268627	0.300383	0.275886
(1,B)	0.270752	0.272509	0.266351
(1,C)	0.449892	0.046755	0.039726
(2,B)	0.399107	0.017429	0.019905

Table 6. Standard Deviation of the Absolute Correlation Values per Lesion of the *RGB* lesion data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.075189	0.051723	0.120361
(1,B)	0.050564	0.057280	0.092927
(1,C)	0.010095	0.050346	0.045895
(2,B)	0.012165	0.078557	0.100383
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.068989	0.070599	0.110723
(1,B)	0.091141	0.094951	0.078275
(1,C)	0.022922	0.034339	0.042043
(2,B)	0.014210	0.087458	0.069111
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.015704	0.025307	0.022419
(1,B)	0.077885	0.095661	0.094163
(1,C)	0.024233	0.025523	0.019999
(2,B)	0.014610	0.009277	0.008870

Table 7. Average Absolute Correlation Values per Lesion of the illumination-corrected *RGB* lesion data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.156864	0.251143	0.306729
(1,B)	0.134858	0.280693	0.287114
(1,C)	0.020100	0.438892	0.477563
(2,B)	0.024732	0.531542	0.375503
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.356152	0.151465	0.151860
(1,B)	0.326464	0.178192	0.196218
(1,C)	0.103229	0.475666	0.448373
(2,B)	0.064307	0.391578	0.524883
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.263711	0.298394	0.279176
(1,B)	0.292042	0.290461	0.282628
(1,C)	0.449129	0.048744	0.041427
(2,B)	0.400400	0.015945	0.019321

Table 8. Standard Deviation of the Absolute Correlation Values per Lesion of the illumination-corrected *RGB* lesion data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.073514	0.043622	0.104251
(1,B)	0.084878	0.075100	0.108610
(1,C)	0.012400	0.049313	0.043281
(2,B)	0.011909	0.085662	0.100957
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.069711	0.055424	0.111003
(1,B)	0.094358	0.099449	0.092184
(1,C)	0.022383	0.033330	0.039751
(2,B)	0.017439	0.090638	0.071341
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.013473	0.022410	0.021080
(1,B)	0.068711	0.085562	0.081845
(1,C)	0.022226	0.023979	0.019838
(2,B)	0.013120	0.008677	0.009668

the whole image as being lesion. The contribution of each single color band to the first M.A.D. component of these examples can be seen in Table 9.

According to the added results (see upper part of Tables 5 and 7), it is surprising to see that while in the medical area the redness is the variable used for measuring change in lesions with psoriasis, it is mostly the green if not the blue band which contributes most significantly to the first M.A.D. component. These results have to be interpreted with caution, because of the small size of the data-set used.

6 Back-projection to the original data

To back-project the output of previous Section to the original data the following aspects have to be taken into account:

- For each pixel in the image and for each single band value, co-occurrences were computed with neighbors in eight directions. The output of the M.A.D. Transform is referred to these directions. There is not a one-to-one correspondence that allows to back-project the output of the M.A.D. Transform to the original data. Therefore, some constraints have to be set.
- The M.A.D. Transform is a difference of canonical variables, which come from an eigen-decomposition. Eigenvectors give a direction, but not a sense. The constraint for the M.A.D. Transform is that the eigenvectors are chosen so that the correlation between the canonical variables is positive ([11], page 85). However, an aspect not considered here, but that could be explored further, is if there are real cases for which the a negative correlation between canonical variables could be more interesting to analyze; for instance, a region that in a given time is normal and in a next time is ill, and viceversa.
- Again, the M.A.D. Transform is a difference of canonical variables. For the case of the co-occurrence matrix, each one of these variables are a new co-occurrence matrix corresponding to a new (unknown) texture.

To back-project the data from the M.A.D. Transform to the original images, the following criterion was applied. As it was mentioned, the M.A.D. components are computed with the co-occurrence matrix of each single band of two equalized images. For each pixel in each equalized image, co-occurrence pairs in eight directions were formed for each single band value. Each pair is used as index in a weighting function given by the absolute value of the first M.A.D. component. This was done, because only the magnitude of the change was considered of interest. For a given pixel, the M.A.D. absolute values corresponding to each co-occurrence pair formed by the pixel with its neighbors pixel in 8-connectivity [3], are summed and saved in a synthetical image. The histogram of this image is finally divided in two parts by a user set threshold. The segmentation output

[3] 3 bands by 8 pairs gives 24 weights

obtained is composed by a region indicating more textural change and region indicating less textural change.

The output of the back-projection scheme proposed is shown for two examples: the lesions $(1, A)$ and $(1, B)$. In both cases, the illumination corrected images with the lesion region equalized were used. The output of the M.A.D. Transform was back-projected to the original data in the way previously described. Note in each case that for each image of the pair, the segmented regions showing more and less textural change are approximately corresponding. This gives the evidence that it is possible to avoid the alignment and registration of images applying the M.A.D. transform to multi-spectral descriptors that are invariant to rotations and translations.

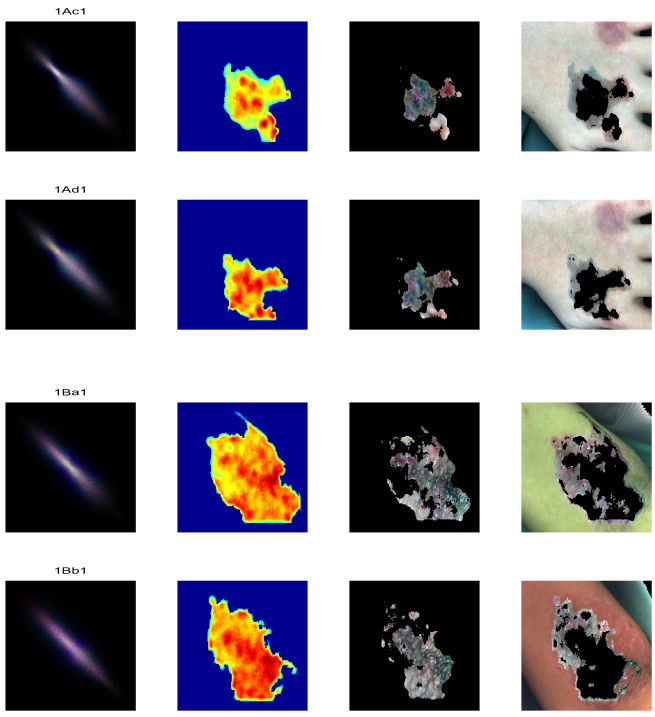


Fig. 1. From the M.A.D. Transform to the Original Data.
From left to right: co-occurrence matrix, back-projection synthetical map, regions with more textural change and regions with less textural change. From top to bottom: the first two rows are data for the case $(1, A)$ (the first images of the third and fourth session) and the last two rows, for the case $(1, B)$ (the first images of the first and second session).

However, it has to said that, in general, the output of the back-projection approach was not satisfactory. Each step and constraints set for the back-projection should be revised. The first M.A.D. component is the difference of canonical variables of the co-occurrence matrixes, that, in fact, correspond to a new textural version of the original images. For the back-projection, a combination of the original data and the first M.A.D. component was used, what could have been an unappropriate mixture of information. As it was mentioned in [12], the simple difference of the spectral bands, turns to be difficult to show the change in all the bands simultaneously. Thus, may be, for simplicity, in order to obtain a more suitable back-projection scheme, to start with a single difference of one pair of co-occurrence matrixes could be more a more fortunate way to follow.

7 Conclusions

The application of the Multi-variate Alteration Detection Transform to the normalized co-occurrence matrix of lesions with psoriasis patterns is a suitable approach for detecting texture changes in time that avoids the interaction with the user and the lesion pattern registration. The most significant contribution in the texture change of lesions with psoriasis is given -for the pilot data-set- by the green and blue bands. There is evidence that it is possible to back-project the M.A.D. Transform output to the original data, such that corresponding regions showing more and less textural change can be delineated. However, in order to be able to generalize the approach, the optimization of each single stage of the whole procedure has to be considered.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A Selected Lesions

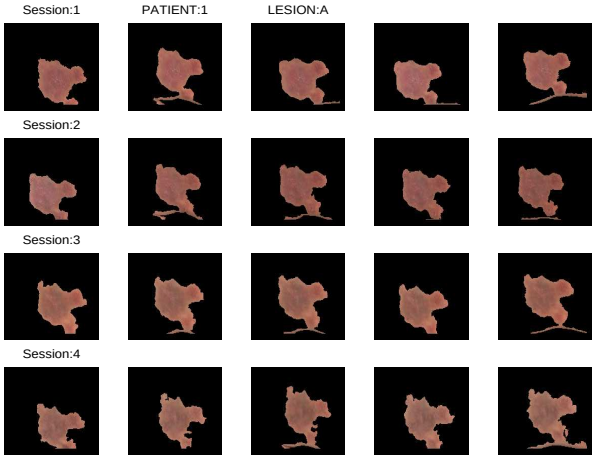


Fig. 2. Regions indicating the selected lesion A of patient 1 in the illumination-corrected images.

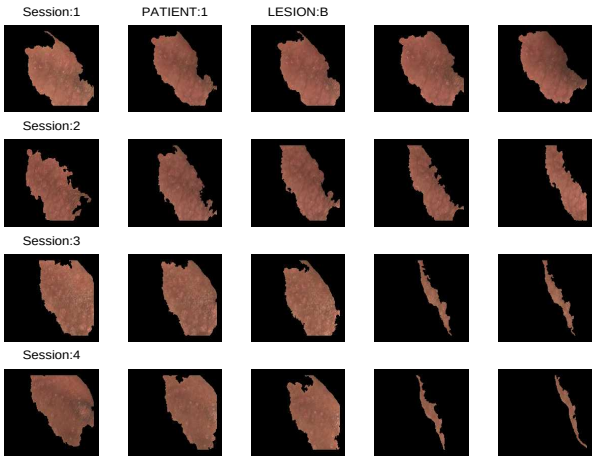


Fig. 3. Regions indicating the selected lesion B of patient 1 in the illumination-corrected images.

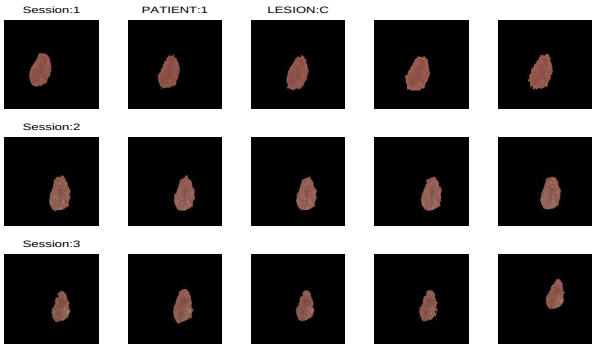


Fig. 4. Regions indicating the selected lesion C of patient 1 in the illumination-corrected images.

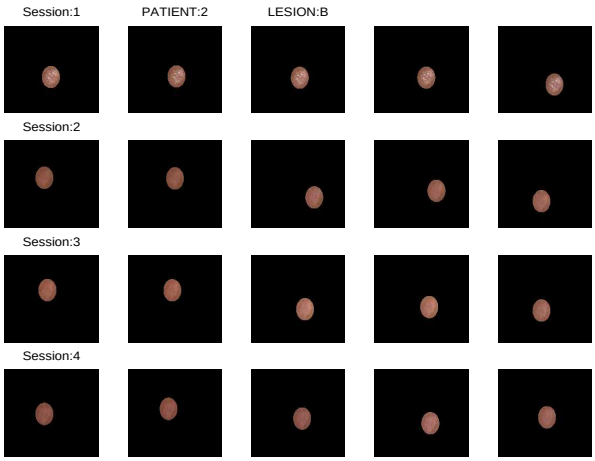


Fig. 5. Regions indicating the selected lesion B of patient 2 in the illumination-corrected images.

B Co-occurrence matrix of selected pairs of images and their MAD Transform

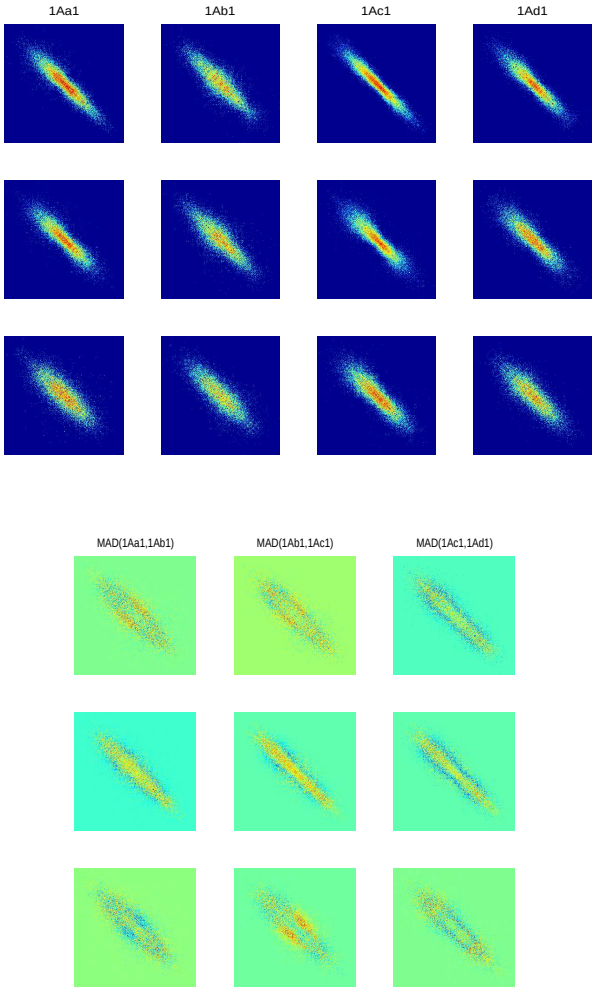


Fig. 6. Co-occurrence matrixes and M.A.D. Components for the case $(1, A)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 1, Lesion A); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

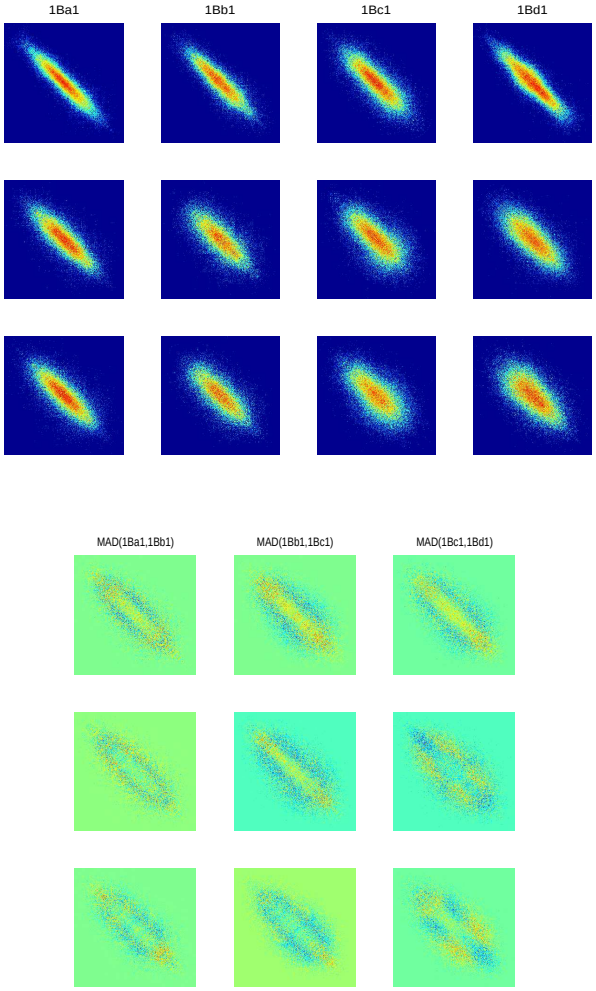


Fig. 7. Co-occurrence matrixes and M.A.D. Components for the case $(1, B)$.
The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 1, Lesion B); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

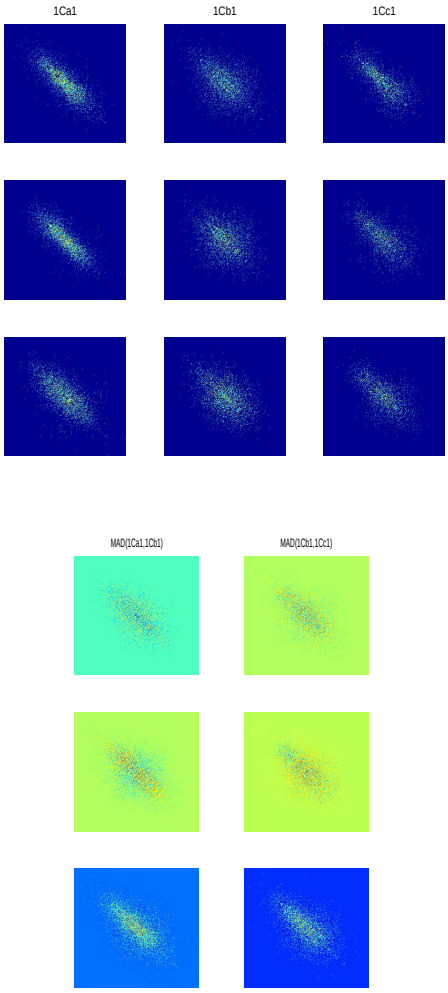


Fig. 8. Co-occurrence matrixes and M.A.D. Components for the case $(1, C)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 1, Lesion C); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

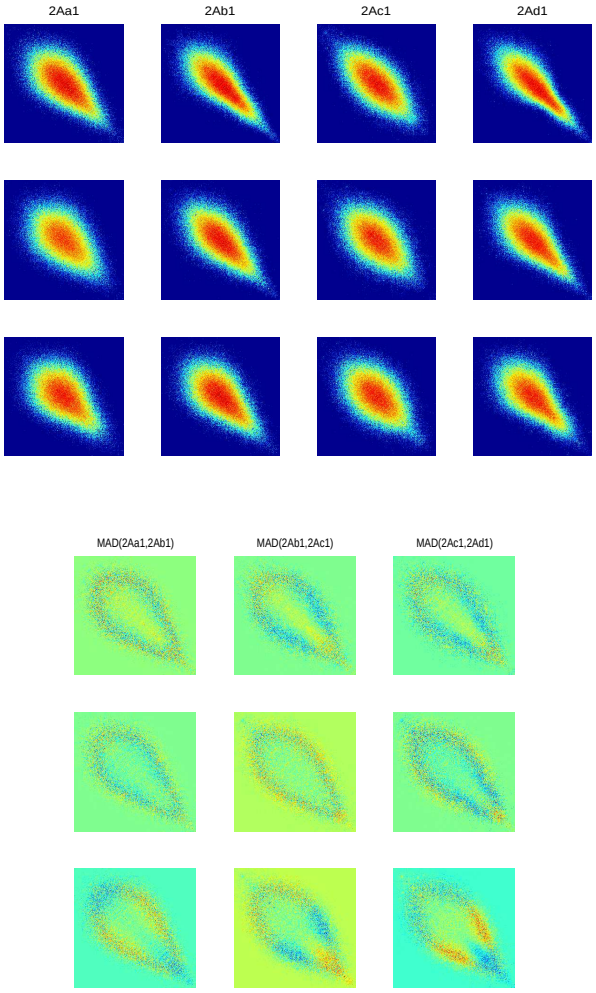


Fig. 9. Co-occurrence matrixes and M.A.D. Components for the case $(2, A)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 2, Lesion A); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

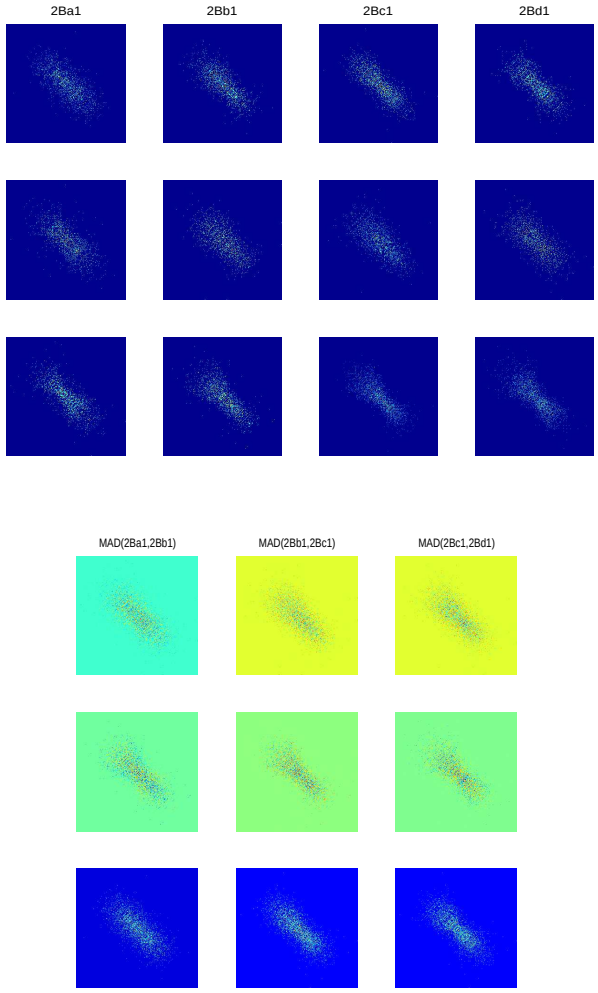


Fig. 10. Co-occurrence matrixes and M.A.D. Components for the case $(2, B)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 2, Lesion B); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

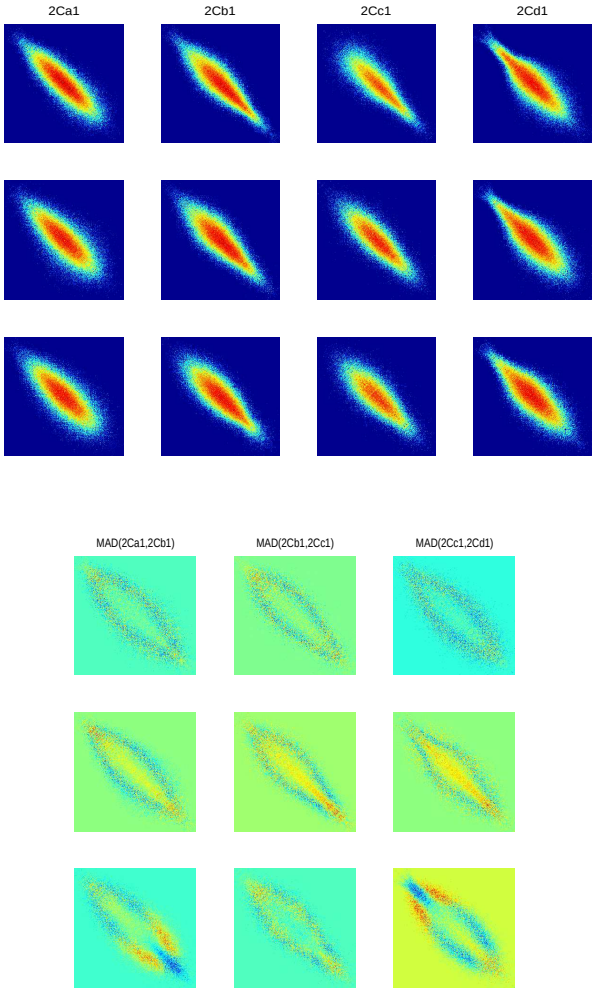


Fig. 11. Co-occurrence matrixes and M.A.D. Components for the case $(2, C)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 2, Lesion C); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

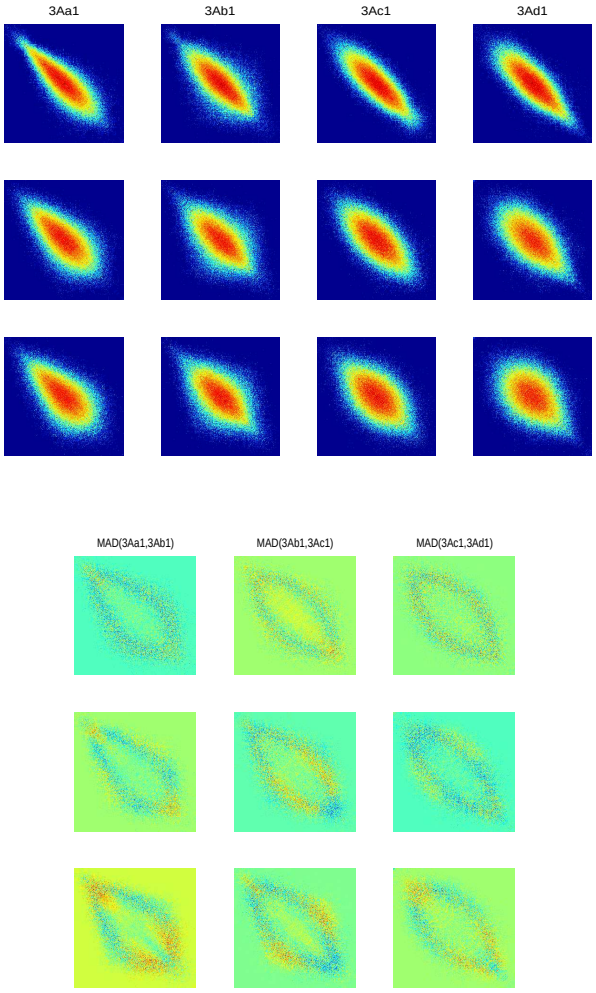


Fig. 12. Co-occurrence matrixes and M.A.D. Components for the case $(3, A)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 3, Lesion A); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

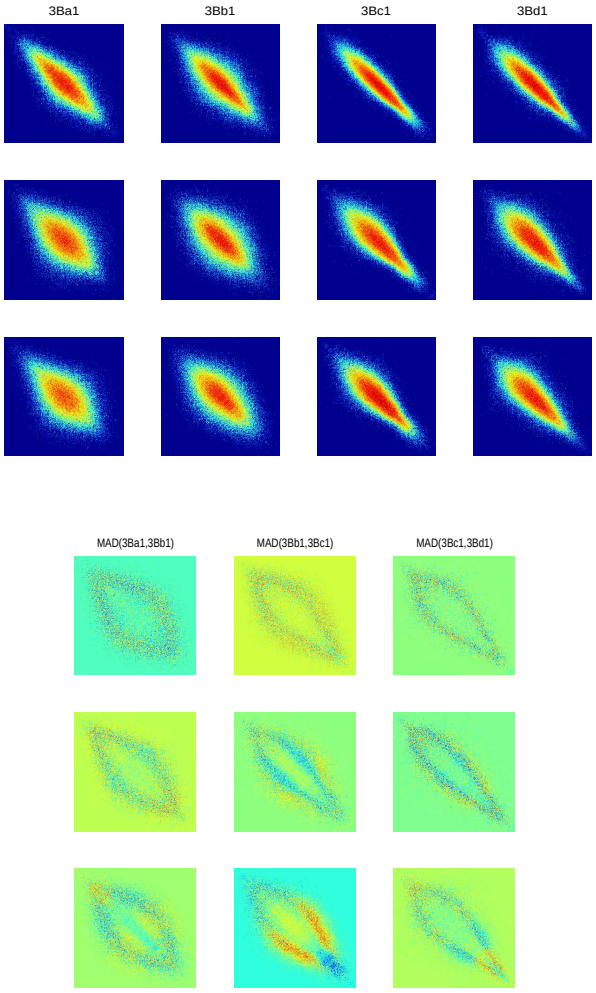


Fig. 13. Co-occurrence matrixes and M.A.D. Components for the case $(3, B)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 3, Lesion B); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

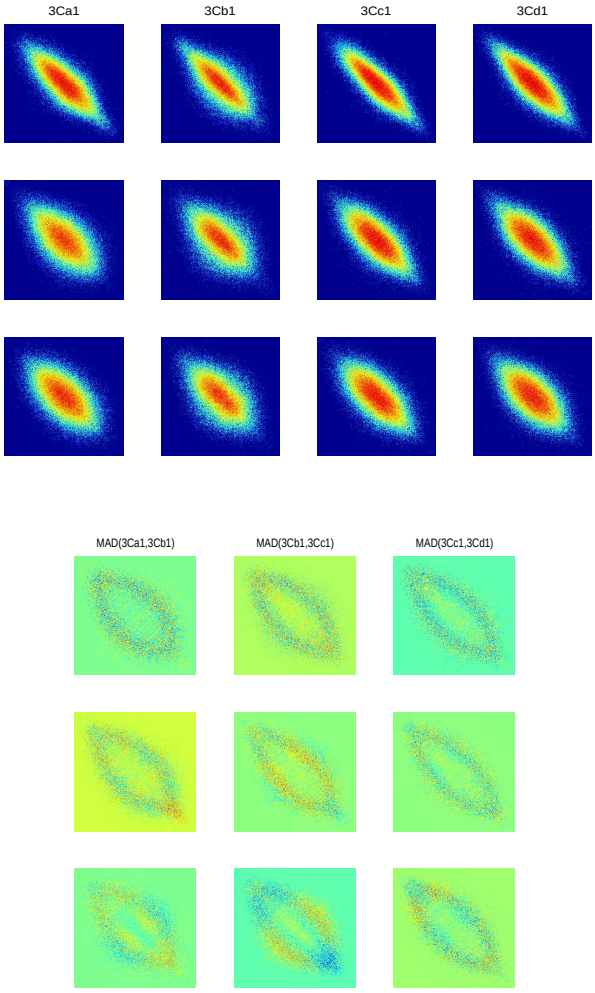


Fig. 14. Co-occurrence matrixes and M.A.D. Components for the case $(3, C)$. The upper part shows the co-occurrence matrixes of the lesion region of selected illumination-corrected images of (Patient 3, Lesion C); from top to bottom, the rows correspond to the red, green and blue band respectively. The lower part shows the MAD components of pairs of consecutive co-occurrence matrixes; from top to bottom, the row number is associated with the MAD component number.

Table 9. Correlation values between the Color Bands and the respective first MAD component for the images of previous Figures

LESION	t_1	t_2	$RED(t_1)$	$GREEN(t_1)$	$BLUE(t_1)$	$RED(t_2)$	$GREEN(t_2)$	$BLUE(t_2)$
1A	a1	b1	-0.043395	-0.281037	0.368620	0.076882	0.296416	-0.397385
1A	b1	c1	0.157975	0.259383	-0.401074	-0.009873	-0.281179	0.280133
1A	c1	d1	0.092474	-0.275296	0.272046	0.360614	-0.083418	-0.280742
1B	a1	b1	0.305761	-0.230549	0.054734	-0.129127	-0.198322	0.343047
1B	b1	c1	0.088509	-0.316239	0.258445	0.320800	-0.068653	-0.275978
1B	c1	d1	0.312352	-0.252936	0.015724	0.163600	-0.290688	0.204880
1C	a1	b1	0.001901	-0.378092	0.510217	-0.069119	0.390545	-0.540887
1C	b1	c1	-0.009486	-0.449951	0.490224	0.001886	0.456618	-0.472794
2A	a1	b1	-0.043036	0.239584	-0.185652	0.216014	-0.191849	0.097690
2A	b1	c1	0.324745	-0.083721	-0.018579	-0.001702	0.211669	-0.210786
2A	c1	d1	0.042016	-0.227695	0.187835	0.266919	-0.140662	0.097644
2B	a1	b1	-0.007383	0.543504	-0.394928	0.019575	0.638896	-0.266538
2B	b1	c1	0.001548	0.637515	-0.242365	-0.047060	-0.669888	0.171656
2B	c1	d1	-0.028495	0.500740	-0.460677	-0.056241	-0.664664	0.193865
2C	a1	b1	-0.015721	-0.194092	0.227168	0.194830	0.101938	-0.189855
2C	b1	c1	0.038893	0.194974	-0.199133	0.048887	0.194801	-0.246159
2C	c1	d1	0.071718	0.195087	-0.221466	0.048802	-0.219207	0.138725
3A	a1	b1	0.083018	-0.214939	0.163499	0.117317	-0.231709	0.145831
3A	b1	c1	0.087001	-0.224819	0.167746	0.143994	-0.213472	0.115006
3A	c1	d1	0.100927	-0.222693	0.149637	-0.087361	0.226875	-0.149954
3B	a1	b1	0.029576	-0.227130	0.222873	-0.001799	0.225008	-0.224343
3B	b1	c1	-0.023228	0.256221	-0.183855	-0.011463	-0.186459	0.174127
3B	c1	d1	0.004763	-0.193600	0.189768	0.011727	0.192159	-0.204803
3C	a1	b1	0.095635	0.177159	-0.248092	-0.110164	0.262144	-0.176593
3C	b1	c1	0.062883	-0.255678	0.199650	0.105009	-0.224865	0.158335
3C	c1	d1	0.073072	-0.232236	0.174764	0.133484	-0.244558	0.131380

References

1. M. Augusteijn and L. Clemens. A performance evaluation of texture measures for image classification and segmentation using the cascade-correlation architecture. *IEEE International Conference on Neural Networks, IEEE World Congress on Computational Intelligence*, 7:4300–4305, 1994.
2. J. Carstensen. *Description and Simulation of visual texture.* PhD thesis, I.M.S.O.R., Technical University of Denmark, Lyngby, April 1992.
3. R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.
4. K. Hilger. *Exploratory Analysis of Multivariate Data.* Ph.d. thesis, Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby, November 2001.
5. R. Larsen. Maf and other transformations applied to remote sensing. Master's thesis, Institute for Mathematical stastistics and Operations Research (IMSOR), Technical University of Denmark, DTU, 1991. pp. xii+130+205.
6. J. Lira and L. Frulla. An automated region growing algorithm for segmentation of texture regions in sar images. *International Journal of Remote Sensing*, 19(18):3595–3606, 1998.
7. J. Lira and G. Maletti. A supervised classifier for multispectral and textured images based on an automated region growing algorithm. *European Space Agency Publications*, SP-434:153–158, 1998.
8. J. Liu, K. Bowyer, D. Goldgof, and S. Sarkar. A comparative study of texture measures for human skin treatment. *International Conference on Information, Communications and Signal Processing*, 1:170–174, September 1997. ICICS'97.
9. G. Maletti and B. Ersbøll. A hierarchical classification scheme of psoriasis images. Technical Report 6, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
10. G. Maletti and B. Ersbøll. Illumination correction in psoriasis lesions images. Technical Report 7, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
11. A. Nielsen. *Analysis of Regularly and Irregularly sampled spatial, multivariate, and muti-temporal data.* PhD thesis, Institute of Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark, October 1994. No. 6.
12. A. Nielsen, K. Conradsen, and J. Simpson. Multivariate alteration detection (m.a.d.) and m.a.f. postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Environ.*, 64:1–19, 1998.
13. T. Ojala, M. Peitikäinen, and D. Harwood. A comparative study of textre measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.
14. S. Peckinpaugh. An improved method for computing gray-level coocurrence matrix based texture measures. *Graphical Models and Image Processing*, 53(6):574–580, November 1991.
15. L. Wang. Vector choice in the texture spectrum approach. *International Journal of Remote Sensing*, 15(18):3823–3829, 1994.
16. L. Wang and D. He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910, 1990.
17. J. Weszka, Ch. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4):269–285, April 1976.

4.2.5 A Combined Alignment and Registration scheme of Psoriasis Lesions Images

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 9
Pages: 1 – 42
Reference: [33]

A Combined Alignment and Registration scheme of Psoriasis Lesion Images

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. A two-stage registration scheme of psoriasis lesion patterns is proposed. In the first stage, global rotation and translation effects of assumed equally scaled psoriasis lesion patterns are removed. In the second stage, only local translation effects are removed. In both stages a novel algorithm for extreme value detection is applied. The output of each stage is evaluated with the M.A.D. transform and the correlation coefficient. Results are shown for real examples.

1 Introduction

We use a set of segmented psoriasis lesion patterns from *RGB* images taken at the Gentofte Hospital, Denmark, during pilot sessions with three invited patients. For each patient, three lesions were followed once a week for at least three weeks. In each session, five images of each lesion were taken. Following the notation used in previous work ([13], [11], [12], [14]), the group of images corresponding to a given patient and lesion is here called "case (patient, lesion)". Part of the mentioned segmented patterns[1] was generated by a two-stage hierarchical classification scheme [11], and another part, was produced by user-interaction[2]. This was done so, in order to exclude shadow effects in the captured objects [12]. One case[3] was excluded, because corresponding points between sessions were hard to find, by visual assessment.

Once a spacial pattern has been segmented, morphological descriptors like area, perimeter, compactness, perimeter roughness, eccentricity, shape and texture can be estimated ([9], page 340). In order to register two spatial patterns, it is required that the descriptors are sensitive to the transformations (rotations, translations and/or changes of scale) whose effect we want remove trough a distance minimization process. This minimal distance between two objects is

[1] For the cases $(1, A)$, $(1, B)$ and $(1, C)$ see Figures 2 to 4 in [14]. For the cases $(3, B)$ and $(3, C)$, see Figures 7 and 8 respectively in [11]

[2] See Figure 5 in [11] for the case $(2, B)$. Case $(2, A)$ was generated in the same way.

[3] The case $(2, C)$ was hair all over.

assumed to be referred to the position, angle and scale for which the patterns are matching. When this distance is Euclidean and the minimization is done over rotations and scale of equally centered objects then we have the full Procrustes distance [3]; when the same is only done over rotations, then we have the partial Procrustes distance. Procrustes Analysis is a useful approach for aligning two landmark registered shapes [8]. For the case that these shapes are not landmark registered, the approach could be extended by setting some constraints, for instance, with regard to the topology given by the connectivity between the landmarks. A list of references to landmark-free approaches can be found in [3], page 305.

It could be nice to include textural information [6], [10], [1], [7], [20], [19] during the registration process. Previous work that combines shape with textural information was published in [18]. In the present work, it is assumed that the shape of the object is invariant to scale and that, for each pattern, the pixels locations are landmarks that have to be matched. It is assumed that the objects of largest size in the thematic map indicating lesions are corresponding. However, they could not have the same size. Therefore, the most convenient set of pixels that will define the set of landmarks to be registered has to be found through a distance minimization process.

The basic approach could be to compare the pixel values in their original form. However, there are many ways to transform the data and present them to a distance function. A review of linear and non linear decomposition approaches can be found in [8]. Since the red band of each single image is the most correlated with the first principal component of the image [13], it is from this band the pattern values are compared.

This work is composed of the following sections. First, a review of measures of distances and similarities is given. Secondly, a general algorithm to obtain the minimal distance between two patterns is proposed. In many cases, the cost of the estimation of the values of a function is expensive. Therefore, to obtain one of its extreme values (for instance, its minimum) could be carried out on a small representative data-set. Then, a section detailing a scheme that uses this algorithm to carry out the initial alignment of patterns follows. The initial alignment implies removing global translation and rotation effects of two assumed equally scaled patterns. The output of the initial alignment scheme is used by an internal local registration scheme described afterwards. In this case, it is assumed that the local rotation effects in the interior of the patterns are not significant, and only local translation effects in the interior of the patterns are removed.

2 Distances and Similarities

A mathematical representation of how close or similar two objects or data-sets or groups of observations are, is given by the distance. The selection of the measure of distance used changes the results. A short review of definitions follows from [15].

Definition 1. *Given two points $\boldsymbol{x}$ and $\boldsymbol{y}$, a real-valued function $d(\boldsymbol{x}, \boldsymbol{y})$ is a distance function if it has the following properties:*

- *symmetry:* $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$;
- *non-negativity:* $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$;
- *identification mark:* $d(\boldsymbol{x}, \boldsymbol{x}) = 0$

Definition 2. *A metric is a distance for which also the following properties are satisfied:*

- *definiteness:* $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ *if and only if* $\boldsymbol{x} = \boldsymbol{y}$;
- *triangle inequality:* $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$.

Definition 3. *A measure of similarity $s(\boldsymbol{x}, \boldsymbol{y})$ is such that it satisfies the following properties:*

- $s(\boldsymbol{x}, \boldsymbol{y}) = s(\boldsymbol{y}, \boldsymbol{x})$;
- $s(\boldsymbol{x}, \boldsymbol{y}) > 0$
- $s(\boldsymbol{x}, \boldsymbol{y})$ *increases as the similarity between* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *increases.*

Some frequently used measures of distances between two data-sets $\boldsymbol{x}$ and $\boldsymbol{y}$ are [21], [15]:

- Distance of Minkowsky

$$d(\boldsymbol{x}, \boldsymbol{y}) = (\sum_{i=1}^{m} |x_i - y_i|^r)^{\frac{1}{r}} \tag{1}$$

- Euclidean Distance

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \tag{2}$$

- Distance of Manhattan

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{m} |x_i - y_i| \tag{3}$$

- Distance of Camberra

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{m} \frac{|x_i - y_i|}{|x_i + y_i|}} \tag{4}$$

– Distance of Chebychev

$$d(\boldsymbol{x},\boldsymbol{y}) = max_{i=1}^{m}|x_i - y_i| \tag{5}$$

– Quadratic Distance

$$d(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^T Q(\boldsymbol{x}-\boldsymbol{y}) \tag{6}$$

where Q is a positive definite $m \times m$ matrix of weights.

– Distance of Mahalanobis

$$d(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}-\boldsymbol{y})^T V^{-1}(\boldsymbol{x}-\boldsymbol{y}) \tag{7}$$

where V is the covariance of $\boldsymbol{x}$ and $\boldsymbol{y}$.

– One minus Correlation

$$d(\boldsymbol{x},\boldsymbol{y}) = 1 - |\frac{\sum_{i=1}^{m}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{m}(x_i-\overline{x})^2 \sum_{i=1}^{m}(y_i-\overline{y})^2}}| \tag{8}$$

– Scalar Product [9]

$$d(\boldsymbol{x},\boldsymbol{y}) =< \boldsymbol{x},\boldsymbol{y} > \tag{9}$$

which can be standardized in different ways, like range, standard deviation or number of elements. Some more distance measures can be found in [17]. Similarity measures can be transformed into dissimilarity measures by means of tricks like [5]:

$$s(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{1+d(\boldsymbol{x},\boldsymbol{y})} \tag{10}$$

or

$$s(\boldsymbol{x},\boldsymbol{y}) = c - d(\boldsymbol{x},\boldsymbol{y}) \tag{11}$$

with c a constant.

In order to evaluate, for our registration problem, the appropriateness in using one of the measures of distances previously listed, the following aspects are considered:

– Feature space properties: results of clustering depend on the choice of the measure of distance.
– Set elements correspondence: the way the elements of two sets are ordered affects the measure of distance between the sets.
– Computational time: the number of mathematical operations required to obtain each given measure of distance is counted.

With respect to the feature space, Duda and Hart ([4], page 213) pointed out that the selection of the Euclidean distance (See Equation 2) implies that the feature space is isotropic: the clusters will be invariant to translations or rotations; however, they will not be invariant to linear transformations in general or to transformations that distort the distance relationship. In order to achieve invariance the data can be normalized prior to clustering. Normalization implies the introduction of information that gives the procedure meaning. Some normalization suggestions given in [4] are, for instance:

- invariance to displacements and scale changes: translate and scale the axes so that all the features have zero mean and unit variance. This prevents certain features from dominating distance calculations. This approach is appropriate if the spread of the values is due to normal random variation. The problem with our data set is that, for instance, changes in the class variances could be of interest to detect lesion changes in time. For instance, when scaling appears in the lesion, the variance of the distribution changes. This means, that a normalization by variance of the original data is not the most appropriate decision. Furthermore, changes in the redness of the lesions, which imply mean class displacements along the time, are also of interest for the dermatologists.
- invariance to rotation: rotate the axes so that they coincide with the sample covariance matrix (transformation to principal components). This can be preceded or followed by normalization for scale.

Many other normalized distance or similarity functions could be used to normalize the data.

With respect to the pixel correspondence, the following is considered. When data sets are compared, the order given to the elements of each single set (i.e. the correspondence established between pairs of same indexed elements, one of each set) plays a relevant role. Note, for instance, in the equations of distances given, that the index i establishes the correspondence between the $i-th$ element of the set $\boldsymbol{x}$ and the $i-th$ element of the set $\boldsymbol{y}$. For our case, by placing a rotated and translated version of a pattern over another pattern, and computing a measure of distance, location correspondence of pixels could be assumed. On the other hand, as it was mentioned before, spatial registration of patterns requires descriptors sensitive to the transformations that are expected to be removed. However, for prediction purposes, it could be nice to use descriptors invariant to the transformations that need to be removed in order to achieve registration of patterns, like the normalized histogram or the co-occurrence matrix (See [14]).

Results obtained regarding to the computational time can be seen in Table 1. As it can be deduced, Equation 3 has the smallest computational cost.

3 The Extreme Value Detection Algorithm

The Extreme Value Detection Algorithm (E.V.D.A.) is an iterative process that assumes that the expected minimum (extreme value) of the distance function between two objects is near the center of masses of the domains of those objects. In the first iteration, a mapping template centered on the center of mass of the domain of one of the objects is used to define the subset of the domain of the distance function for which their values will be estimated. This mapping template includes many points of the neighborhood of the center of mass, and only

Table 1. Number of operations required for the computation of each measure of distance between two $d-$dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of m elements, with $d = 1$.

Distance	+	−	∗	/	||	^	if	Order
Minkowsky	m	m		1	m	m+1		4m+2
Euclidean	m	m				m+1		3m+1
Manhattan	m	m			m			3m
Camberra	2m	m		m	2m			6m
Chebychev		m			m		2m	4m
Quadratic	2m	m	2m					5m
Mahalanobis	3m	2m	2+m	3				6m+5
Correlation	5m	4m	m+1	1		2m+2		12m+4

few that are far away of the center. The extreme value of the calculated values of the distance function is obtained, and the same mapping template is centered on that first guess of the extreme value. This step is repeated until the guessed extreme value does not change anymore.

Formally, the Extreme Value Detection Algorithm (E.V.D.A.) is an algorithm that iteratively approximates the extreme value f_{ext} (minimum or maximum) of a smooth and discrete function F defined over a subset L_F of a lattice $L = \{[r, c] | 1 \leq r \leq M, 1 \leq c \leq N\}$, such that the location of f_{ext} is $[r, c]_{ext}$ and f_{ext} is unique. For each iteration, a mapping template is centered on the position of the actual guess $\hat{[r, c]}_{ext}$ of the location of that extreme value. This template is Bolduc and Levine's [2] three parameters retinal mapping model $M(r_f, \omega, \alpha)$, where:

- r_f is the foveal radius,
- ω is the receptive field overlap factor, and
- α is the receptive field size-to-eccentricity ratio, (the size of the receptive field is the diameter of the receptive field circle, and the eccentricity of a receptive field is, in our case, the radial distance from the center of the receptive field to the actual guess $\hat{[r, c]}_{ext}$ of the location the extreme value).

The initial actual guess of the location of the extreme value can be any element of L_F (the use of the center of mass of the domain is suggested). Following Bolduc and Levine's model, receptive field centers are arranged along rays originated at $\hat{[r, c]}_{ext}$, and along circles also placed at the same location (See Eq. 1-3 of [2] for details of the derivation of the positioning parameters θ and k, where θ is the angle between adjacent rays and k is the ratio of radii of adjacent circles). The circle of radius r_{nr} is such that it is the circle of maximum radius for which at least one point belongs to L_F. The function F is computed for each center of receptive fields placed on the rings and for each element belonging to the fovea. The location of f_{ext} guessed for the next iteration is given by the location of the extreme value $\hat{f}_{ext}$ of F computed until now. If more than one element of L_F can be the actual guess, then one of them is randomly selected. The algorithm stops when the values of $\hat{[r, c]}_{ext}$ between two consecutive iterations do not change

anymore. The convergence is guaranteed.

E.V.D.A. has a maximal processing time of $O(n_{r_f} + nr * (round(\frac{2*\pi}{\theta})))$ per iteration, where n_{r_f} is the number of elements of the fovea, nr is the number of rings and $round((\frac{2*\pi}{\theta}))$ is the number of receptive fields placed on each ring. It has a fast convergence because, for each iteration, a local neighborhood of the estimated minimum value composed of the elements of the fovea is analyzed in detail, at the same time other parts of the whole domain are taken into account by sampling the centers of receptive fields placed over rings of increasing radius at a logarithmic scale.

Some examples of mapping templates are shown in Figure 3.

The E.V.D.A. has been tested using Wilson's Model [22] parameters ($\omega = 0.5$ and $\alpha = 0.25$) on lattices of 100 by 100 elements. The foveal radius was increased from 1 to 50. For each foveal radius, the algorithm was tested 20 times. The initial guess of the location of the extreme value was randomly set. The minimum values of the following functions were found:

Table 2. Tested functions

Nr.	Function	Domain
1	$Z = (X^2 + Y^2)^{-\frac{1}{2}}$	$L_Z = \{[r, c] \mid -99 \leq r \leq 0, -99 \leq c \leq 0\}$
2	$Z = (X^2 + Y^2)$	$L_Z = \{[r, c] \mid -49 \leq r \leq 50, -49 \leq c \leq 50\}$
3	$Z = 1 - \frac{1}{(X^2+Y^2+1)}$	$L_Z = \{[r, c] \mid -49 \leq r \leq 50, -49 \leq c \leq 50\}$
4	$Z = 1 - \frac{1}{(X^2+Y^2+1)}$	$L_Z = \{[r, c] \mid -89 \leq r \leq 10, -69 \leq c \leq 30\}$

For each function and foveal radius, the number of estimated function values and the average number of iterations is shown in Fig. 2. Note that for all the testing functions, being the convergence guaranteed, the percentage of evaluated points increases unnecessarily when the foveal radius increases. The cases for which the function is not smooth, contextual information could be useful.

4 Initial alignment

Some definitions to be used in the present section are first introduced. Afterwards, an algorithm to carry out the initial alignment is proposed.

Given a set of segmented patterns of a given lesion, in the initial alignment stage, global translations and rotations effects between pairs of patterns are removed. For comparison purposes, it is assumed that the dispersion of the elements of the data-sets is of interest, while the mean differences are not. Some

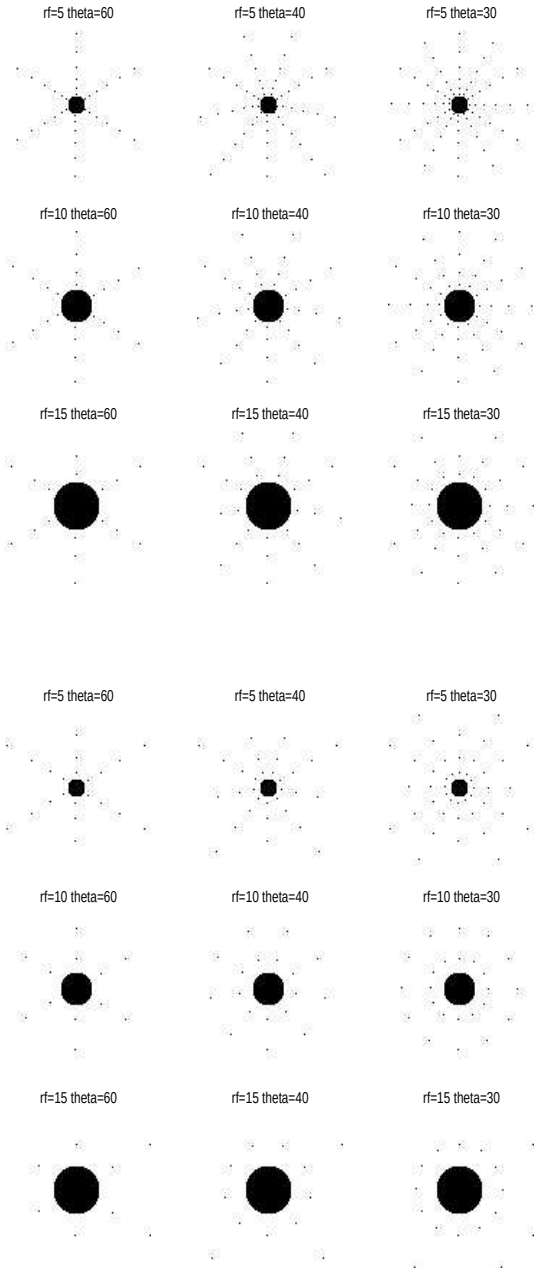


Fig. 1. Mapping templates
They are shown on a domain of 100 by 100 pixels of size with different foveal radius ($r_f = 5, 10, 15$ pixels) and different angles between adjacent rays of ($\theta = 60, 40, 30$ degrees) for the positioning parameter $k = 0.4$ (first three rows) and $k = 1$ (last three rows).

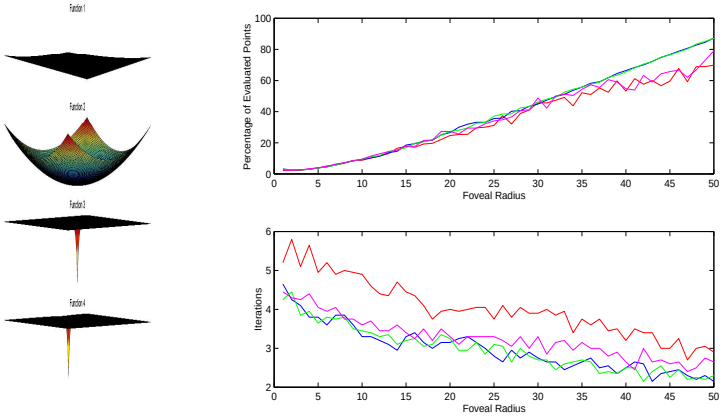


Fig. 2. Tested functions and results of the Extreme Value Detection Algorithm for foveal radius varying from 1 to 50.

definitions follow:

Definition 4. *Let two windows A and B of size $n_a = n_b$ with domains L_A and L_B such that $\exists \boldsymbol{t} \wedge \hat{\theta}$ for which:*

$$L_B = Rot[Trans[L_A, \boldsymbol{t}], \hat{\theta}] \tag{12}$$

where Rot and $Trans$ are the rotation and translation function respectively defined in a 2-D discrete space. Rot and $Trans$ are commutative.

For each window, we compute a texture descriptor, and then take a measure of distance between pairs of windows. The measure of distance used is Manhattan distance (See Equation 3). The texture descriptor is the volume under the surface formed by the normalized values of the pixels of the window. In this way, the quantization level information is removed, while the texture is preserved.

Definition 5. *The difference of volume $dV_{(A),(B)}$ between the windows A and B is, in general, defined as:*

$$dV_{(A),(B)} = dV_{(B),(A)} = \frac{1}{2} \sum_{i=1}^{n_{A \cap B}} |w_{(A)}[i] - w_{(B)}[i]| \tag{13}$$

where, in the present work,

$$w_{(J)}[i] = \frac{x_{(J)}[i]}{\sum_{i=1}^{n_J} x_{(J)}[i]} \tag{14}$$

$x_{(J)}[i]$ is the value of the pixel located at the $i-th$ position of the $J-th$ window of n_J elements. Note that $0 \leq dV \leq 1$.

During the initial alignment, it is assumed, for a given lesion, that Equation 12 is true for the intersection of the translated and rotated versions of the pair of pattern domains, for which the distance (see Equation 13) between the corresponding pixel values is minimal. It is also assumed that the objects to be aligned have the same scale. However, as mentioned before, the input of the initial alignment stage is a set of segmented patterns (this means, the overlaps between segmented objects indicating lesion and the original data), which may not have the same area. (See, for instance, Figures 2 to 4 in [14]). The most important result produced by this stage is the specification of the number of landmarks to be registered in a next stage, this means, that the area of the objects is finally forced to be constant. Since some lesion patterns are occluded in the original images, it is left to the user to decide if the number of landmarks is given by the area of the intersection of the whole set or of a subset of the aligned domains (objects in thematic maps indicating lesion).

The algorithm that carries out the initial pattern alignment is detailed here. First, the n largest sized objects belonging to each of the thematic maps obtained in the previous stage are found. Misclassified pixels forming holes within regions indicating lesions are previously assigned to the class lesion[4]. Correspondence by decreasing object area is assumed. From now, it is assumed that the overlay of the object with the pixel values is a window. For each given window S and a given reference window R, each of the rotations $Rot(R, \hat{\theta})$ is centered on the position of each pixel of S, and the angle $\hat{\theta}_{opt}$ and location $[i, j] \in L$ for which the difference of volume is minimal is found. L is the lattice on which the images including the patterns R and S are defined. In practice, pattern objects are rotated in a range varying from -45 to 45 degrees. For each iteration, this range is divided into five intervals. Neighboring intervals containing the minimum of the evaluating function dV (Equation 13) define the range for the next iteration. This minimum is computed using the Extreme Value Detection Algorithm (E.V.D.A.) detailed in Section 3. The algorithm stops when the minimum between two iterations differ with less than a user-provided threshold β.

The output of the initial alignment stage is back-projected to the original data and a new set of images containing the aligned shapes is obtained. These images can afterwards be registered and realigned.

[4] The function *bwfill* of Matlab Version 6.1.0.450 (R12.1) is used.

4.1 Results and Discussion

The following parameters were set for the initial alignment of the lesions:

- The threshold β for the minimal angle was 1 degree.
- One object per thematic map was tracked within and between sessions related to the same lesion.
- For all the cases, the mapping template used was $M(0.3, 0.5, 0.25)$. The first parameter in M is a percentage of the smallest side of the smallest rectangle that includes the lesion. The last parameters correspond to Wilsons's Model.
- The lesion pattern of the first image of each sequence was the reference window S.
- The initial alignment with original red band data is presented here.

The Algorithm E.V.D.A. defines searching areas starting from the center of mass of the pattern domain, because it is expected that the maximum overlap between a pair of pattern domains will be near their center of masses.

The number of computations of values of the minimizing function F decreases along the iterations, because values of previous iterations can be preserved.

The set of aligned lesion patterns can be seen in Section A. For all the cases, 10 rows and columns were removed from the borders. For the cases $(1, A)$, $(1, B)$, $(1, C)$, $(3, B)$ and $(3, C)$ the classification outputs were used; this means that the image data were down-sampled with a factor of 4 [5]. For the cases $(2, A)$ and $(2, B)$ no down-sampling was done.

The Extreme Value Detection Algorithm used on average 3 iterations. Clearly, the use of a template providing the same amount of information for each ring, while the spacing between rings is given at a logarithmic scale, is advantageous for the fast detection of the emergence of an extreme value in the sampled domain. On the other hand, looking for the extreme value in the same location as the center of the domain of the reference window is a reasonable criteria, that helps to reduce the convergence time. This is added to the fact that the sampling rate is increased in the foveal region. However, it has to be mentioned that some extended tests have shown that if the initial guess of the position of the extreme value was too far away from the real value, the algorithm stopped in a local extreme value.

In general, it can be said that, by visual examination, the results are satisfactory. Looking through some of the alignment results, it is obvious that the illumination problem is greater than expected. See, for instance, how much the shadow location varies in the case $(3, A)$. In fact, as it was reported in [12], the shadows do affect the classification results. However, a comparison of the alignment results before and after illumination correction, which is shown in

[5] One pixel every four columns and rows was selected.

Appendix B, is given only as an illustrative example [6]. Correlation values before and after illumination correction are compared in Table 3. Details can be found in Table 9.

Table 3. Average correlation and standard deviation of a randomly selected aligned lesion pattern of the lesion A of the patient 2 with the remaining aligned lesion patterns of the same patient and lesion.

(2,A)	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B
Before	0.5650	0.1195	0.3819	0.1103	0.4824	0.0755
After	0.3173	0.0568	0.2292	0.0489	0.1919	0.0470

Even when the illumination was corrected, the initial alignment results did not improve. May be the measure of distance used gets a bit confused in large data-sets. Note in the fourth and fifth capture of the first session of the case $(3, B)$, that the E.V.D.A. clearly stopped in a local minima, and it was not able to align the data correctly. This could imply that the measure of distance is less sensitive to changes when the data set is too large. This alignment problem is, in fact, not trivial. The intersection of the pattern domains was used to define the region for which the measure of distance is computed. When the pattern domains are (almost) overlapping, the area of this region is maximal. We expect a minimal difference of volume near the center of masses of the domains. For this reason, the initial alignment of the pattern corresponding to the first patient, for which the pattern domains were very similar, worked very well. However, for the most of the remaining cases, the patterns do not have a common shape: some lesion borders are given by the image border. We could think of different scale-shapes, or what is more precise, after an ideal alignment, the intersection of the domains is far from being equal to the union of the domains. Therefore, the E.V.D.A. finds a minimum when the overlap of the patterns domains is near a maximal and not where the patterns have a maximal overlap. To find the minimal intersection sample size prior to the alignment could probably improve the results.

On the other hand, it is mentioned only as a comment, that the initial alignment output using the $|B - G|$ lesion data was satisfactory for the cases $(1, A)$, $(1, B)$, $(1, C)$ and $(3, C)$. However, particularly for the cases $(3, A)$ and $(3, B)$ it was not possible to align the whole data set in an acceptable way. May be due to the same reasons as explained before.

[6] Due the high contrast between the lesion and the rest, in both cases, the $|B - G|$ band. The images where down-sampled a factor of 16: one pixel every 16 columns and rows was selected.

For a given lesion, the expected correlation value of a randomly selected aligned lesion pattern with any other remaining aligned pattern of the same set (case) is given in Table 4. The values of each single μ cell were computed according to the following equation:

$$\mu_{p,l,b} = \frac{1}{n_s n_c} \sum_{i=1}^{n_s} \sum_{j=1}^{n_c} E[corr(\mathsf{X}_{p,l,i,j,b}, \mathsf{X}_{p,l,:,:,b})] \tag{15}$$

where $E[corr(\mathsf{X}_{p,l,i,j,b}, \mathsf{X}_{p,l,:,:,b})]$ is defined as:

$$E[corr(X_{p,l,s,c,b}, X_{p,l,:,:,b})] = \frac{1}{n_s n_c - 1}(\sum_{i=1}^{n_s} \sum_{j=1}^{n_c} corr(X_{p,l,s,c,b}, X_{p,l,i,j,b}) - 1) \tag{16}$$

and p is the index for patient, l, for lesion, s, for session, i, for capture, b for color band. The symbol n_c is the number of captures of the $s-th$ session. The session s takes values in {'a','b','c','d'}, where the order in the sequence indicates the session week. The symbol n_s is the number of sessions. The color band b takes values in {'R','G','B'}.

The values of each single σ cell were computed according to the following Equation:

$$\sigma_{p,l,b} = \sqrt{\frac{1}{n_s n_c} \sum_{i=1}^{n_s} \sum_{j=1}^{n_c} (E[corr(\mathsf{X}_{p,l,i,j,b}, \mathsf{X}_{p,l,:,:,b})] - \mu_{p,l,b})^2}. \tag{17}$$

where the sub-indices have the same meaning as before.

Table 4. Average correlation and standard deviation of a randomly selected aligned lesion pattern with the remaining aligned lesion patterns of the same patient and lesion, per color band, patient and lesion.

(Patient, Lesion	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B	$\mu_{\|B-G\|}$	$\sigma_{\|B-G\|}$
(1,A)	0.6918	0.0468	0.5770	0.0538	0.3679	0.0595	0.6941	0.0323
(1,B)	0.6696	0.0932	0.5241	0.0944	0.4413	0.0843	0.4517	0.1305
(1,C)	0.2534	0.0806	0.1651	0.0642	0.1138	0.0561	0.4211	0.0585
(2,A)	0.5162	0.0606	0.3561	0.0589	0.5720	0.0588	0.6486	0.0367
(2,B)	0.6681	0.0227	0.5736	0.0331	0.7260	0.0219	0.4854	0.0432
(3,C)	0.2550	0.0288	0.0757	0.0350	0.1289	0.0202	0.3524	0.0672

5 Combined registration and alignment

According to the experience of the dermatologist that lesions with psoriasis do not significantly change their shape and size along the time: the major change

happens within the lesions instead. After the initial alignment it is assumed that the aligned patterns have the same size and shape. Until now, for a given lesion, elements of pairs of aligned patterns located at the same position were considered corresponding points. This hypothesis is now refined: we assume that there is some error produced by small internal displacements. These displacements are assumed to be small translations only.

In the first part of the present Section, a novel registration scheme is proposed. This scheme uses the E.V.D.A. (See Section 3) and thin plate spline interpolations, the theory of which was included for completeness. A subsection of results and discussion follows.

5.1 A registration scheme

In general, in order to define the correspondence, two m-dimensional sets of landmarks T and Y, called the source and the target respectively have to be specified [3]. In the present scheme, $m = 1$ and, for each given image taken during a given session, the source T is given by the locations of pixels belonging to the aligned pattern corresponding to the first image of the first session, and the target Y should be given by their corresponding but unknown locations of pixels. In order to reduce the computational time, for each pattern, only a subset of T and Y is considered at first. Pixels placed on a grid composed of square cells with a user-defined number of points form the subset of T for which their corresponding points will be found. The correspondence between the source subset and the target subset is established as follows. For each pixel belonging to the grid, a user-defined circular neighborhood is considered. At the same location in the second image, another user-defined circular neighborhood[7] - larger than the previous one[8]- is taken into account. See an example in Figure 3. The difference of volume values dV of all[9] the translations of the first window over the second window are considered. The position of the pixels belonging to the second image whose neighborhood has the minimal difference of volume compared to the window of reference the pattern is saved in a new grid.

Afterwards, two different approaches are considered:

– Approach 1: The window sizes defining the searching areas in the target image are reduced along the iterations. The process is repeated, but a target pixel is moved to a new position only if the difference of volume between the source window and the new guess of the estimated corresponding target window is smaller than the difference of volume between the source window

[7] These window sizes are established by visual inspection of the image containing the error between pairs of images taken during the same session.
[8] Computed as a function of the spacing between the points on the grid
[9] In practice, only the points defined by the mapping template used in E.C.V.A. were evaluated

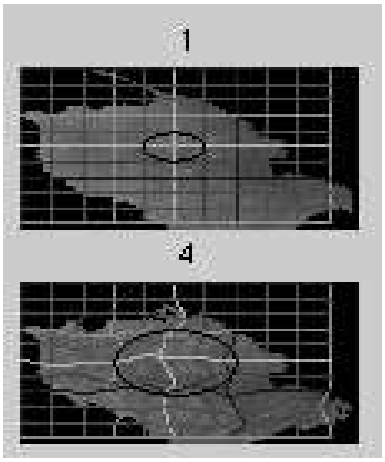


Fig. 3. Pair of corresponding grids for two segmented lesions of the case $(1, B)$ .On the right, the user-defined grid with one example neighborhood delineated in black. On the left, the corresponding grid with the evaluating neighborhood of the example on the right.

and the actual guess of the estimated corresponding window. The user is repeated for a user-provided number of iterations.

- Approach 2: The average grid is computed. The set of distance vectors from the grid of the reference pattern image to the average grid is used to translate all the grids to a new position. This process is repeated for a user-provided number of iterations.

Then, to generate the whole target set Y for each single image, thin plate splines are applied to the rest of the values of the source T and the rest of the new lesion image can be generated (See Section 5.2).

At the end of this stage, it is expected that the pattern correlation within and between sessions increases. Also, when comparing two patterns belonging to the same session, randomly Gaussian noise should be observed in their difference.

5.2 Pairs of thin-plate splines

This topic is described following Dryden and Mardia [3]. Conceptually, the thin-plate spline is the most natural interpolating function for data in two dimensions, because it minimizes the amount of bending in transforming between two configurations.

Let the set of 2-dimensional landmarks be $\mathsf{T} = [\boldsymbol{t}_1\boldsymbol{t}_2...\boldsymbol{t}_k]^T$, on the first figure mapped exactly into $\mathsf{Y} = [\boldsymbol{y}_1\boldsymbol{y}_2...\boldsymbol{y}_k]$ on the second figure, with $m = 2$. There

are $2k$ interpolation constraints:

$$(\boldsymbol{y}_j)_r = \phi_r(\boldsymbol{t}_j), r = 1, 2, j = 1, ..., k \tag{18}$$

and $\phi = (\phi_1(\boldsymbol{t}_j), \phi_2(\boldsymbol{t}_j))^T$ for the 2-dimensional deformation.

A set of k pairs of thin-plate splines is given by k bivariate functions

$$\phi[T] = (\phi_1[T], \phi_2[T])^T \tag{19}$$

where

$$Y_{kx2} = [S^T_{kxk}\mathbf{1}_{kx1}\, T_{kx2}][W^T_{2xk}\boldsymbol{c}_{2x1}A_{2x2}]^T \tag{20}$$

where $S_{i,j} = \sigma(t_i - t_j)$ and $\sigma(h) = ||h||^2 log(||h||)$ if $||h|| > 0$ and 0 in the contrary case. Obviously, $S = S^T$. In order to define the bending energy, six more constraints are introduced:

$$[\mathbf{1}^T_{1xk}0_{1x1}\mathbf{0}_{1x2}][W^T_{2xk}\boldsymbol{c}_{2x1}A_{2x2}]^T = \mathbf{0}_{1x2} \tag{21}$$

$$[T^T_{2xk}\mathbf{0}_{2x1}0_{2x2}][W^T_{2xk}\boldsymbol{c}_{2x1}A_{2x2}]^T = 0_{2x2}. \tag{22}$$

Pairs of thin-plate splines which satisfy the constraints of Equation 21 and 22 are called natural thin-plate splines. Let the matrix symmetric positive definite matrix $\Gamma_{(k+3)x(k+3)}$ be defined with the left side matrix of Equations 20, 21 and 22 as follows:

$$\Gamma = \begin{pmatrix} S^T & 1 & T \\ 1 & 0 & 0 \\ T^T & 0 & 0 \end{pmatrix} \tag{23}$$

The inverse of γ exists if S^{-1} exists. Writing the partition of γ^{-1} as:

$$\begin{pmatrix} \Gamma^{11}_{kxk} & \Gamma^{12} \\ \Gamma^{21} & \Gamma^{22} \end{pmatrix} \tag{24}$$

we have that $W_{kx2} = \Gamma^{11}_{kxk}Y_{kx2}$ and $\Gamma^{21}_{1xk}Y_{kx2} = [c^T_{1x2}A^T_{2x2}]$

5.3 Results and Discussion

The following parameters were set to produce the outputs:

- The number of points in the grid was set to a multiple of 10. The minimum side of the smallest rectangle that includes the lesion was divided in ten pieces; the size of each piece defined the size of square cells forming the grid.
- The radii of the circular window sizes used for the registration were a function of the separation between points in the grid. For the source image, the radius was 1 times the separation between points in the grid; for the target image, it was 2.

- The mapping template was $M(0.30, 0.5, 0.25)$.
- The user-defined number of iterations was 10.
- The red band values were used for both approaches.
- For each case, the lesion pattern of the first image of the first session was used as the reference window.

Initially, occluded lesions were handled in the same way as the others. However, this considerably affected the results. After removing them from the sets to be registered, the algorithm was again applied and the results were more satisfactory.

Clearly, the separation between points in the grid is crucial during the registration stage, because the size of the source and target window depends on it: if the window size for the target image is too small, the error in the registration is not randomly distributed within the lesion and some structure appears. Too large window sizes, in some cases, only increased the computational time. Thus, they have to be selected after visual inspection of the difference between pairs of initially aligned lesions belonging to the same session. Setting these parameters per lesion could improve the results.

The registration outputs of the second approach are presented in Section C. Each single cell value in these tables was computed using Equation 16. In correspondence with Table 4, but using the correlation values of registration outputs, Tables 5 and 6 were generated.

Table 5. Average correlation and standard deviation per color band, patient and lesion of a randomly selected registered (with Approach 1) lesion pattern with the remaining patterns of the same patient and lesion.

(Patient, Lesion	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B	$\mu_{\lvert B-G\rvert}$	$\sigma_{\lvert B-G\rvert}$
(1,A)	0.6917	0.0458	0.5670	0.0438	0.3682	0.0557	0.6582	0.0319
(1,B)	0.6772	0.0724	0.5364	0.0769	0.4605	0.0697	0.4086	0.1129
(1,C)	0.2768	0.0723	0.1902	0.0582	0.1313	0.0452	0.4150	0.0586
(2,A)	0.5451	0.0635	0.3686	0.0629	0.5779	0.0619	0.6031	0.0535
(2,B)	0.6624	0.0235	0.5737	0.0328	0.7097	0.0195	0.4107	0.0367
(3,C)	0.2811	0.0396	0.0821	0.0292	0.1410	0.0225	0.3246	0.0515

Averages per color band of the values of the Tables included in Section C are shown in Tables 7 and 8. In Table 7, each single cell value is the average per color band of only the correlation values within sessions (i.e. rows (a, a), (b, b), (c, c) and (d, d)). In Table 8, each single cell value is the average per color band of all the correlation values within and between sessions (this means, from rows (a, a), (a, b) ... to (c, d), (d, d)). Note that the second registration approach has a better performance than the first registration approach. However, the improvement in

Table 6. Average correlation and standard deviation per color band, patient and lesion of a randomly selected registered (with Approach 2) lesion pattern with the remaining patterns of the same patient and lesion.

(Patient, Lesion	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B	$\mu_{\lvert B-G\rvert}$	$\sigma_{\lvert B-G\rvert}$
(1,A)	0.6853	0.0447	0.5689	0.0505	0.3666	0.0586	0.6728	0.0345
(1,B)	0.6819	0.0790	0.5380	0.0814	0.4608	0.0743	0.4486	0.1209
(1,C)	0.2915	0.0841	0.1993	0.0720	0.1361	0.0586	0.4325	0.0586
(2,A)	0.5781	0.0573	0.4222	0.0553	0.6204	0.0533	0.6499	0.0406
(2,B)	0.6643	0.0258	0.5720	0.0394	0.7147	0.0252	0.4642	0.0369
(3,C)	0.2865	0.0382	0.0999	0.0331	0.1476	0.0203	0.3675	0.0604

the correlation values, compared with the output of the initial alignment stage, is not significant.

Table 7. Average correlation values within sessions per color band of the outputs of the first and second stage of the combined alignment and registration scheme.

Stage	RED	GREEN	BLUE	|B-G|
Initial Alignment	0.7288	0.6377	0.6432	0.7040
Registration (Approach 1)	0.7197	0.6235	0.6329	0.6751
Registration (Approach 2)	0.7412	0.6507	0.6537	0.7412

Table 8. Average correlation values within and between sessions per color band of the outputs of the first and second stage of the combined alignment and registration scheme.

Stage	RED	GREEN	BLUE	|B-G|
Initial alignment	0.5938	0.4935	0.5001	0.5780
Registration (Approach 1)	0.5998	0.4937	0.4915	0.5408
Registration (Approach 2)	0.6114	0.5021	0.5043	0.5741

Regarding the distribution of the noise by comparing pairs of images within a given session, the first session of the registered examples is presented in Appendix E. For the case $(1, A)$, the kind of structure that the M.A.D. Transform [16] detects on the lesion center after the alignment tends to disappear after the registration, however a small structure appears after registration on the top left side of the lesion. For the case $(1, B)$, there is a set of three points that appear on the first absolute M.A.D. component of the first and second capture, that tends to disappear after registration, but at the same time, it tends to appear on the first absolute M.A.D. component of the second and third capture, while it was

not there after the initial alignment. For the case $(1, C)$, structure tends to be noise after registration, but not significantly. Note also that near the borders, the results do not improve much; this is due to the fact that the input to the registration algorithm does not include skin data near the lesion. For the case $(2, A)$ it could be said that the noise tends to be more randomly distributed after registration, the same for the case $(2, B)$. Particularly, for the case $(2, B)$ it has to be mentioned that, by visual assessment, the registration within sessions of images not belonging to the first session look seriously corrupted due to the presence of hair partially covering the lesion. For the case $(3, C)$ it seems that a kind of structure appear after registration, in spite of having increased the correlation within sessions (See Table 14).

Conclusions

The application of the Extreme Value Detection Algorithm during the initial alignment has shown to produce satisfactory results. Results analysis indicates that large variability in the data sets sizes influences the detection of the extreme value.

Two different approaches for pattern registration were proposed. Results indicate that using an average grid (approach 2) produces better registration results than using the first pattern of a given sequence as reference (approach 1) in the registration process. Compared to the outputs of the initial alignment scheme, it has been shown that the second approach used in the second stage of the present scheme improves the correlation between the registered patterns. However, this improvement is not significant. The registration of aligned patterns within and between sessions is clearly not an easy problem.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A The Initial Alignment Stage Outputs

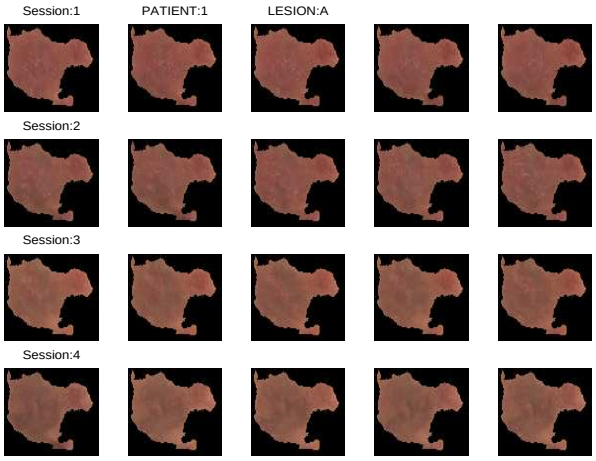


Fig. 4. Initial alignment of lesion patterns for the case (Patient 1, Lesion A).

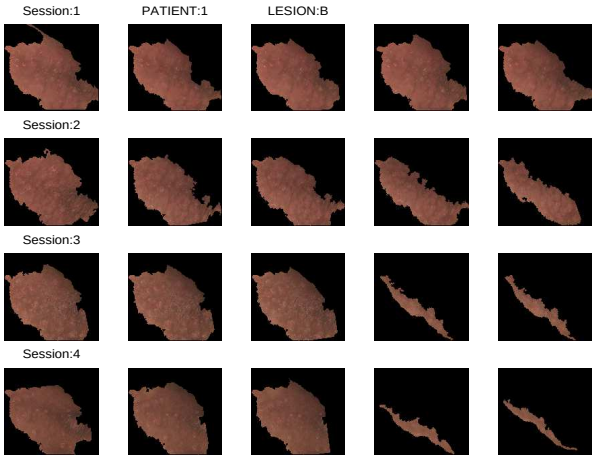


Fig. 5. Initial alignment of lesion patterns for the case (Patient 1, Lesion B).

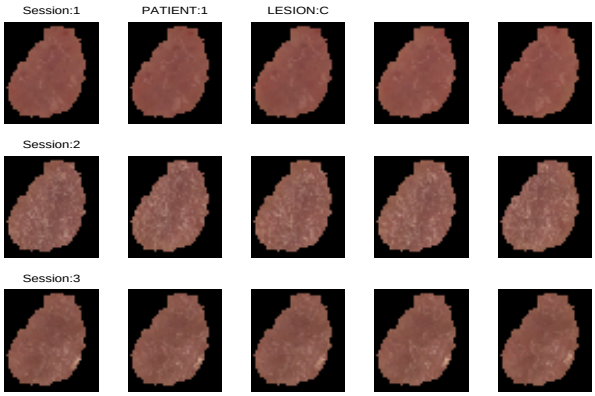


Fig. 6. Initial alignment of lesion patterns for the case (Patient 1, Lesion C).

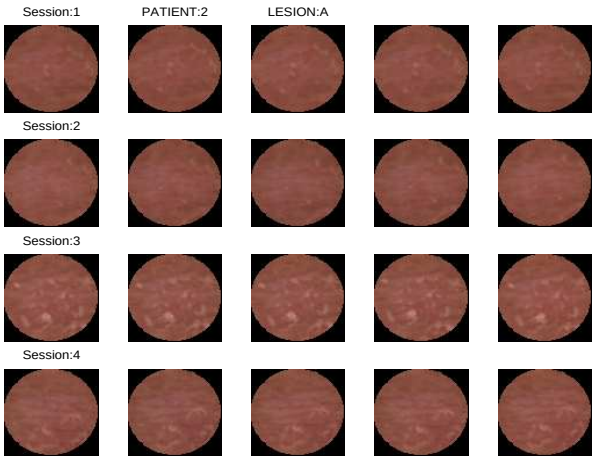


Fig. 7. Initial alignment of lesion patterns for the case (Patient 2, Lesion A).

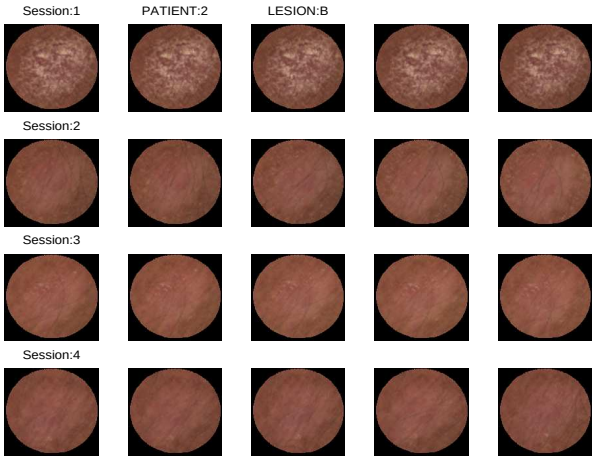


Fig. 8. Initial alignment of lesion patterns for the case (Patient 2, Lesion B).

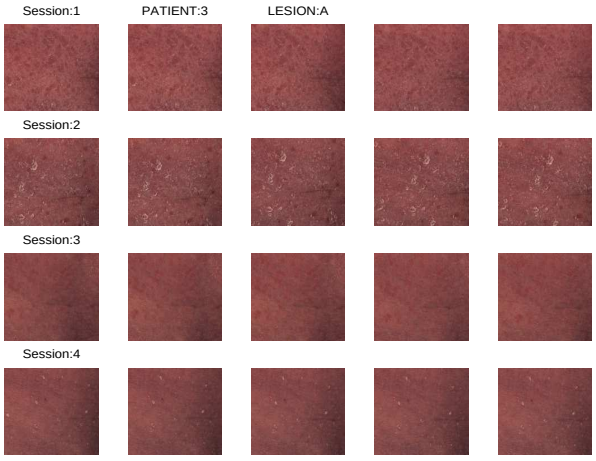


Fig. 9. Initial alignment of lesion patterns for the case (Patient 3, Lesion A).

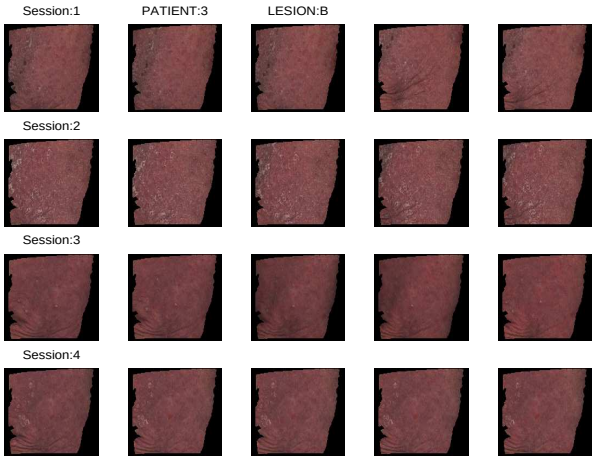


Fig. 10. Initial alignment of lesion patterns for the case (Patient 3, Lesion B).

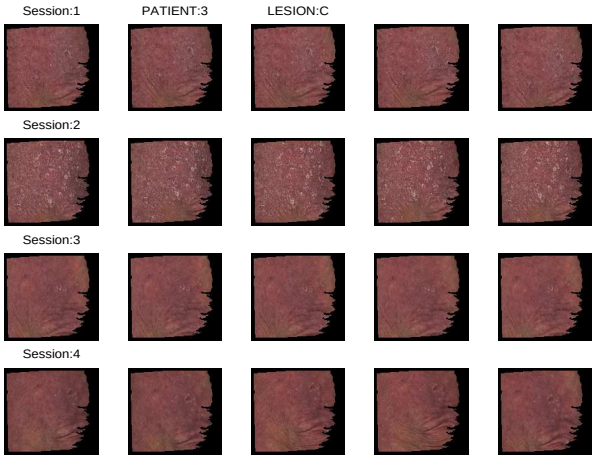


Fig. 11. Initial alignment of lesion patterns for the case (Patient 3, Lesion C).

B Illumination Correction: an example.

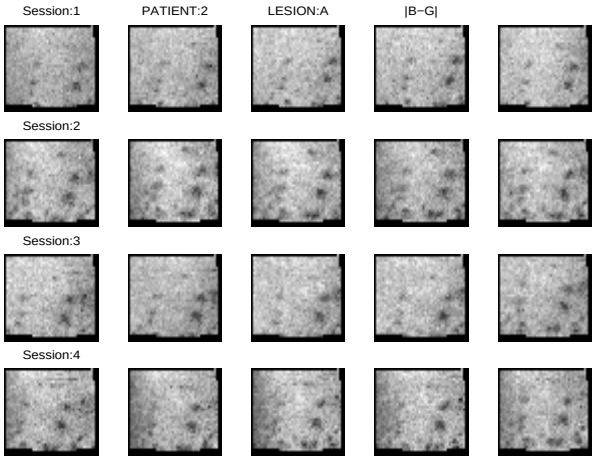


Fig. 12. Initial alignment of lesions for not illumination corrected images of (Patient 2, Lesion A).

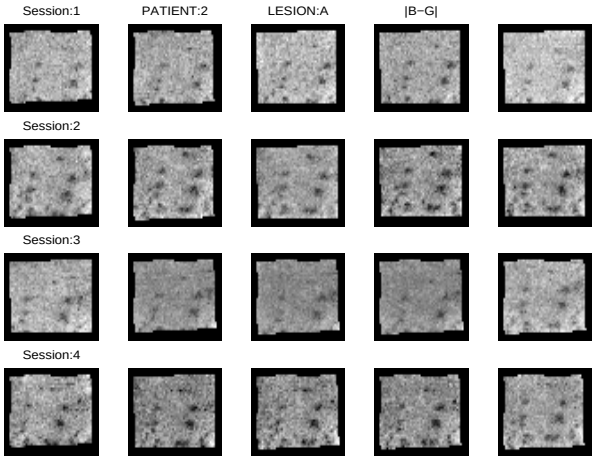


Fig. 13. Initial alignment of lesions for illumination corrected images of (Patient 2, Lesion A).

Table 9. Within and between sessions average correlation and standard deviation of randomly selecting an aligned lesion as reference image for the registration per color band for the case (Patient 2, lesion A).

Before	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B
(a,a)	0.6807	0.1722	0.5122	0.2510	0.5945	0.2114
(a,b)	0.6114	0.0792	0.3953	0.0667	0.4696	0.0544
(a,c)	0.4123	0.0678	0.2339	0.0470	0.3856	0.0442
(a,d)	0.6231	0.0885	0.4071	0.0795	0.5023	0.0630
(b,b)	0.8538	0.0851	0.6993	0.1660	0.7093	0.1566
(b,c)	0.3705	0.0930	0.2052	0.0625	0.3597	0.0548
(b,d)	0.8313	0.0456	0.6579	0.0726	0.6649	0.0533
(c,c)	0.5420	0.2715	0.4259	0.3230	0.5309	0.2694
(c,d)	0.3623	0.0926	0.1963	0.0689	0.3831	0.0557
(d,d)	0.8896	0.0677	0.7763	0.1262	0.7674	0.1284
After	μ_R	σ_R	μ_G	σ_G	μ_B	σ_B
(a,a)	0.5015	0.2620	0.4086	0.3048	0.4218	0.2966
(a,b)	0.2607	0.0819	0.1782	0.0687	0.1491	0.0685
(a,c)	0.2810	0.1361	0.2136	0.1071	0.2031	0.1059
(a,d)	0.2535	0.0763	0.1758	0.0565	0.1562	0.0538
(b,b)	0.4755	0.2707	0.4010	0.3036	0.3429	0.3332
(b,c)	0.2837	0.1233	0.2179	0.0899	0.1712	0.0904
(b,d)	0.3484	0.1030	0.2475	0.0721	0.1756	0.0723
(c,c)	0.5849	0.2579	0.5039	0.2913	0.4614	0.3127
(c,d)	0.3281	0.1006	0.2343	0.0863	0.2001	0.0805
(d,d)	0.5510	0.2403	0.4355	0.2881	0.3798	0.3145

C Registration

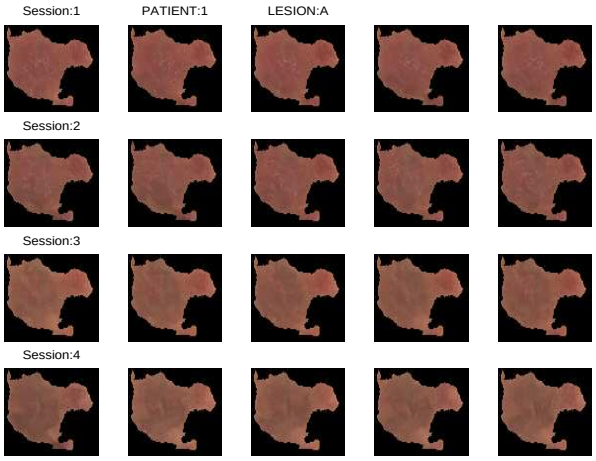


Fig. 14. Registered pattern for the case (Patient 1, Lesion A).

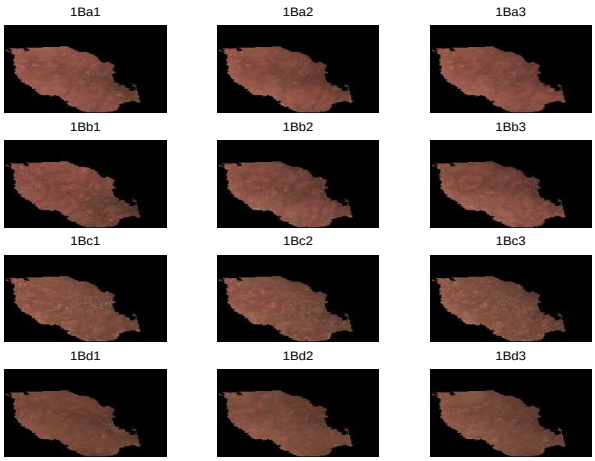


Fig. 15. Registered pattern for the case (Patient 1, Lesion B).

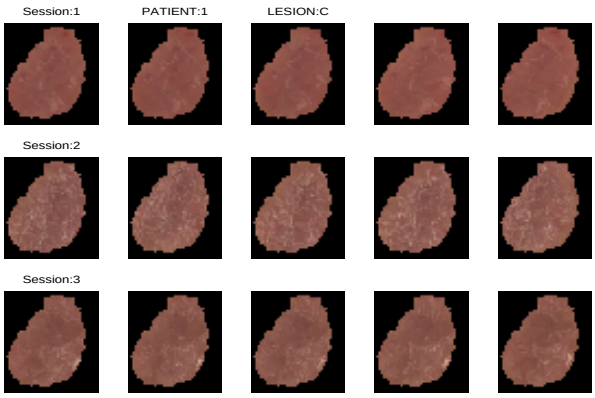


Fig. 16. Registered pattern for the case (Patient 1, Lesion C).

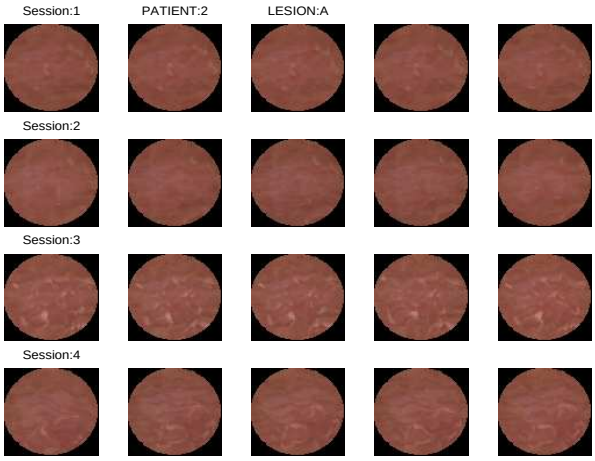


Fig. 17. Registered pattern for the case (Patient 2, Lesion A).

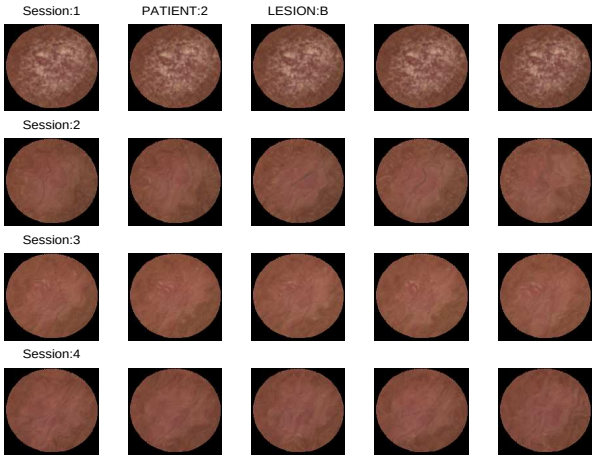


Fig. 18. Registered pattern for the case (Patient 2, Lesion B).

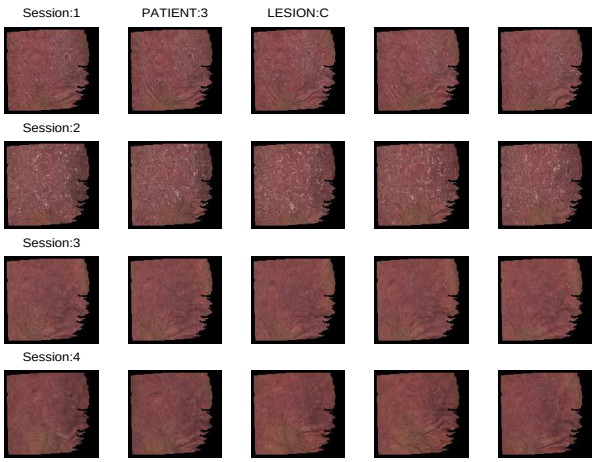


Fig. 19. Registered pattern for the case (Patient 3, Lesion C).

D Result Evaluation

Table 10. Within and between sessions average correlation values of randomly selecting an aligned lesion as reference image for the registration per color band.

Sessions	(1,A)	(1,B)[10]	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.8305	0.8785	0.5349	0.8612	0.9605	0.5010
(a,b)	0.6642	0.6850	0.1342	0.6632	0.5556	0.2969
(a,c)	0.5949	0.7029	0.2424	0.4274	0.6500	0.1010
(a,d)	0.6055	0.7247		0.3704	0.5728	0.2619
(b,b)	0.8186	0.6231	0.3872	0.8314	0.7772	0.5089
(b,c)	0.6415	0.6096	0.1497	0.4730	0.6796	0.0723
(b,d)	0.5888	0.6129		0.4438	0.6543	0.3045
(c,c)	0.9421	0.8213	0.7540	0.7376	0.8376	0.7408
(c,d)	0.8183	0.6835		0.3676	0.6850	0.1876
(d,d)	0.8973	0.7943		0.7260	0.7847	0.4772
GREEN						
(a,a)	0.7908	0.7915	0.4754	0.7958	0.9609	0.2962
(a,b)	0.5011	0.5695	0.0620	0.4698	0.4892	0.0509
(a,c)	0.5917	0.5108	0.1245	0.2391	0.5475	-0.0291
(a,d)	0.5300	0.5956		0.2302	0.4893	0.0077
(b,b)	0.6839	0.5944	0.3617	0.7490	0.6794	0.3118
(b,c)	0.5128	0.4492	0.0421	0.2439	0.5803	-0.0463
(b,d)	0.4150	0.5050		0.2926	0.5270	0.0962
(c,c)	0.8539	0.6684	0.6921	0.6683	0.7708	0.6197
(c,d)	0.6632	0.4967		0.1903	0.5698	0.0914
(d,d)	0.8133	0.7120		0.6683	0.7022	0.3814
BLUE						
(a,a)	0.6664	0.7584	0.4385	0.8829	0.9692	0.3748
(a,b)	0.3283	0.5201	0.0347	0.7156	0.6348	0.1470
(a,c)	0.3114	0.4066	-0.0106	0.4841	0.6962	0.0296
(a,d)	0.2267	0.4946		0.4555	0.6469	0.0890
(b,b)	0.6092	0.5605	0.3456	0.8528	0.7927	0.3505
(b,c)	0.2874	0.3651	0.0160	0.5204	0.7468	0.0655
(b,d)	0.1183	0.4215		0.5192	0.7149	0.1555
(c,c)	0.7839	0.6090	0.6920	0.7593	0.8653	0.5334
(c,d)	0.5173	0.4063		0.4233	0.7524	0.0944
(d,d)	0.7538	0.6495		0.7626	0.8239	0.3388
|B-G|						
(a,a)	0.8757	0.7140	0.6136	0.8598	0.8423	0.6066
(a,b)	0.7071	0.4475	0.4054	0.7181	0.3133	0.3197
(a,c)	0.7234	0.5563	0.4604	0.6792	0.3059	0.3719
(a,d)	0.6361	0.4538		0.4971	0.3368	0.4287
(b,b)	0.8156	0.5339	0.4991	0.7951	0.6708	0.5283
(b,c)	0.6941	0.4281	0.3274	0.6511	0.5717	0.0902
(b,d)	0.6052	0.3616		0.5642	0.5380	0.2546
(c,c)	0.8232	0.6938	0.6379	0.7986	0.7470	0.7904
(c,d)	0.6730	0.4662		0.5778	0.5499	0.3575
(d,d)	0.7582	0.5891		0.8307	0.6875	0.5860

Table 11. Within and between sessions standard deviation correlation values of randomly selecting an aligned lesion as reference image for the registration per color band.

Sessions	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.1077	0.1067	0.2948	0.0821	0.0241	0.2542
(a,b)	0.0478	0.1849	0.0766	0.0341	0.0238	0.0665
(a,c)	0.1074	0.0633	0.0385	0.0179	0.0255	0.0317
(a,d)	0.1142	0.0889		0.0253	0.0233	0.0925
(b,b)	0.0994	0.3662	0.4070	0.1050	0.1306	0.2461
(b,c)	0.0750	0.1142	0.0520	0.0240	0.0407	0.0209
(b,d)	0.0801	0.1816		0.0180	0.0265	0.1064
(c,c)	0.0303	0.1396	0.1464	0.1937	0.0919	0.1745
(c,d)	0.0400	0.0676		0.0479	0.0294	0.1081
(d,d)	0.0845	0.1612		0.2042	0.1221	0.2985
GREEN						
(a,a)	0.1151	0.1689	0.3293	0.1345	0.0249	0.3535
(a,b)	0.0337	0.1825	0.0511	0.0526	0.0122	0.0294
(a,c)	0.0697	0.0769	0.0376	0.0259	0.0285	0.0128
(a,d)	0.0929	0.0953		0.0322	0.0261	0.0489
(b,b)	0.1608	0.3508	0.4100	0.1632	0.1882	0.3444
(b,c)	0.0393	0.1180	0.0516	0.0368	0.0398	0.0142
(b,d)	0.0580	0.1752		0.0145	0.0342	0.0650
(c,c)	0.0754	0.2435	0.1838	0.2631	0.1327	0.2504
(c,d)	0.0456	0.0849		0.0522	0.0365	0.0910
(d,d)	0.1246	0.2135		0.2623	0.1651	0.3341
BLUE						
(a,a)	0.1774	0.1889	0.3422	0.0749	0.0193	0.3137
(a,b)	0.0508	0.1681	0.0450	0.0395	0.0157	0.0262
(a,c)	0.0898	0.0775	0.0250	0.0276	0.0223	0.0197
(a,d)	0.1322	0.0824		0.0334	0.0261	0.0412
(b,b)	0.2006	0.3679	0.4304	0.0991	0.1214	0.3250
(b,c)	0.0357	0.1028	0.0406	0.0302	0.0282	0.0121
(b,d)	0.0654	0.1333		0.0182	0.0237	0.0514
(c,c)	0.1112	0.2858	0.1833	0.1911	0.0783	0.3022
(c,d)	0.0497	0.0822		0.0527	0.0229	0.0272
(d,d)	0.1513	0.2553		0.1852	0.0964	0.3476
|B-G|						
(a,a)	0.0641	0.2177	0.2457	0.0767	0.0806	0.2062
(a,b)	0.0279	0.2555	0.1691	0.0467	0.0345	0.0881
(a,c)	0.0196	0.0771	0.0819	0.0217	0.0237	0.0548
(a,d)	0.0296	0.1434		0.0306	0.0185	0.0957
(b,b)	0.0951	0.3882	0.3351	0.1172	0.1811	0.2422
(b,c)	0.0359	0.2225	0.0755	0.0383	0.0398	0.0558
(b,d)	0.0537	0.2142		0.0190	0.0321	0.1128
(c,c)	0.0897	0.2252	0.1933	0.1177	0.1297	0.1154
(c,d)	0.0213	0.1245		0.0475	0.0565	0.1500
(d,d)	0.1313	0.3032		0.0963	0.1625	0.2351

Table 12. Average Correlation Values between pairs of Sessions per Patient, Lesion and Color Band of the outputs of the Registration Approach 1

Sessions	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.8176	0.8753	0.5962	0.8446	0.9246	0.5469
(a,b)	0.6584	0.6875	0.1892	0.7030	0.5979	0.3246
(a,c)	0.5902	0.6904	0.2724	0.4994	0.6634	0.1512
(a,d)	0.6294	0.7384		0.4391	0.6213	0.2980
(b,b)	0.8030	0.6537	0.3545	0.8044	0.7555	0.5378
(b,c)	0.6222	0.6123	0.1968	0.5354	0.6509	0.1002
(b,d)	0.6062	0.6318		0.4699	0.6366	0.3388
(c,c)	0.9289	0.8168	0.6575	0.6460	0.7953	0.6500
(c,d)	0.8251	0.6864		0.4061	0.6597	0.2088
(d,d)	0.9011	0.8256		0.6855	0.7330	0.4948
GREEN						
(a,a)	0.7641	0.8139	0.5423	0.7589	0.9176	0.3445
(a,b)	0.5118	0.5832	0.1110	0.4900	0.5501	0.0473
(a,c)	0.5650	0.5065	0.1710	0.2717	0.5489	-0.0177
(a,d)	0.5196	0.6106		0.2887	0.5466	0.0189
(b,b)	0.6900	0.6184	0.3260	0.7038	0.6565	0.3322
(b,c)	0.4957	0.4549	0.0921	0.3140	0.5474	-0.0269
(b,d)	0.4229	0.5244		0.3157	0.5333	0.1131
(c,c)	0.8308	0.6739	0.5809	0.5229	0.7155	0.5174
(c,d)	0.6480	0.4956		0.2217	0.5578	0.0941
(d,d)	0.8074	0.7438		0.6134	0.6625	0.3957
BLUE						
(a,a)	0.6284	0.7868	0.5087	0.8732	0.9374	0.4220
(a,b)	0.3351	0.5340	0.0754	0.7275	0.6651	0.1530
(a,c)	0.2945	0.4119	0.0318	0.5228	0.6964	0.0583
(a,d)	0.2450	0.5326		0.4749	0.6768	0.1045
(b,b)	0.5957	0.5862	0.2989	0.8195	0.7614	0.3665
(b,c)	0.2634	0.3680	0.0509	0.5747	0.7027	0.0913
(b,d)	0.1555	0.4477		0.5137	0.6914	0.1690
(c,c)	0.7542	0.6188	0.5790	0.6702	0.8150	0.4418
(c,d)	0.5389	0.4102		0.4376	0.7150	0.1052
(d,d)	0.7531	0.6871		0.7184	0.7793	0.3499
|B-G|						
(a,a)	0.8638	0.7247	0.6234	0.8574	0.7736	0.5611
(a,b)	0.7012	0.4244	0.4127	0.7097	0.2698	0.2867
(a,c)	0.6794	0.4944	0.4469	0.6754	0.2608	0.3194
(a,d)	0.6005	0.3927		0.4274	0.2702	0.3604
(b,b)	0.8005	0.5111	0.4937	0.7636	0.5829	0.5148
(b,c)	0.6335	0.3810	0.3303	0.6355	0.4647	0.0873
(b,d)	0.5608	0.3064		0.4561	0.4486	0.2876
(c,c)	0.7854	0.6755	0.5892	0.7572	0.6817	0.7298
(c,d)	0.6309	0.4144		0.4947	0.4749	0.3389
(d,d)	0.7418	0.5884		0.7919	0.6264	0.5672

Table 13. Standard Deviation of the Correlation Values between pairs of Sessions per Patient, Lesion and Color Band of the outputs of the Registration Approach 1

Sessions	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.1072	0.0980	0.2416	0.0799	0.0457	0.2467
(a,b)	0.0443	0.1484	0.0890	0.0228	0.0220	0.0656
(a,c)	0.0991	0.0500	0.0540	0.0358	0.0266	0.0288
(a,d)	0.0915	0.0610		0.0565	0.0266	0.0973
(b,b)	0.1074	0.3228	0.3550	0.1019	0.1306	0.2319
(b,c)	0.0744	0.0787	0.0691	0.0344	0.0581	0.0315
(b,d)	0.0723	0.1465		0.0408	0.0389	0.1115
(c,c)	0.0375	0.1379	0.1795	0.1874	0.1078	0.1862
(c,d)	0.0281	0.0612		0.0442	0.0332	0.1048
(d,d)	0.0706	0.1301		0.1670	0.1421	0.2855
GREEN						
(a,a)	0.1262	0.1388	0.2704	0.1251	0.0514	0.3468
(a,b)	0.0285	0.1364	0.0687	0.0404	0.0236	0.0369
(a,c)	0.0650	0.0705	0.0359	0.0440	0.0223	0.0131
(a,d)	0.0716	0.0625		0.0698	0.0265	0.0617
(b,b)	0.1575	0.3197	0.3560	0.1530	0.1847	0.3342
(b,c)	0.0357	0.0931	0.0680	0.0491	0.0402	0.0200
(b,d)	0.0555	0.1439		0.0552	0.0296	0.0735
(c,c)	0.0865	0.2377	0.2195	0.2563	0.1485	0.2499
(c,d)	0.0351	0.0760		0.0509	0.0326	0.0774
(d,d)	0.1206	0.1868		0.2096	0.1754	0.3218
BLUE						
(a,a)	0.1941	0.1549	0.2821	0.0662	0.0384	0.3041
(a,b)	0.0401	0.1142	0.0667	0.0269	0.0168	0.3041
(a,c)	0.0944	0.0674	0.0539	0.0319	0.0168	0.0232
(a,d)	0.1069	0.0562		0.0563	0.0322	0.0490
(b,b)	0.2066	0.3340	0.3763	0.0941	0.1269	0.3170
(b,c)	0.0418	0.0725	0.0442	0.0350	0.0309	0.0178
(b,d)	0.0639	0.1046		0.0526	0.0194	0.0569
(c,c)	0.1251	0.2753	0.2200	0.1803	0.0961	0.2875
(c,d)	0.0380	0.0721		0.0424	0.0273	0.0186
(d,d)	0.1431	0.2253		0.1512	0.1149	0.3356
|B-G|						
(a,a)	0.0699	0.2020	0.2336	0.0721	0.1202	0.2393
(a,b)	0.0247	0.2405	0.1730	0.0355	0.0403	0.0829
(a,c)	0.0221	0.0762	0.0656	0.0187	0.0373	0.0620
(a,d)	0.0317	0.0917		0.0276	0.0380	0.0807
(b,b)	0.1012	0.4065	0.3426	0.1223	0.2168	0.0807
(b,c)	0.0305	0.2088	0.0737	0.0318	0.0396	0.0541
(b,d)	0.0533	0.1721		0.0336	0.0596	0.1229
(c,c)	0.1094	0.2375	0.2095	0.1265	0.1633	0.1229
(c,d)	0.0262	0.0889		0.0609	0.0574	0.1412
(d,d)	0.1435	0.2987		0.1131	0.1903	0.2487

Table 14. Average Correlation Values between pairs of Sessions per Patient, Lesion and Color Band of the outputs of the Registration Approach 2

Sessions	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.8250	0.8922	0.6495	0.9003	0.9636	0.5667
(a,b)	0.6719	0.6866	0.1831	0.7215	0.5699	0.3209
(a,c)	0.5981	0.6911	0.2977	0.5147	0.6606	0.1492
(a,d)	0.5972	0.7488		0.4643	0.6060	0.3074
(b,b)	0.8092	0.6412	0.3612	0.8457	0.7605	0.5333
(b,c)	0.6299	0.5996	0.1686	0.5429	0.6568	0.0907
(b,d)	0.5760	0.6348		0.4982	0.6300	0.3319
(c,c)	0.9290	0.8543	0.7393	0.7548	0.8154	0.6925
(c,d)	0.8145	0.6987		0.4422	0.6717	0.2241
(d,d)	0.8778	0.8276		0.7182	0.7676	0.5144
GREEN						
(a,a)	0.7785	0.8244	0.5844	0.8599	0.9650	0.3761
(a,b)	0.5124	0.5774	0.1089	0.5268	0.5209	0.0600
(a,c)	0.5947	0.5096	0.1745	0.3343	0.5596	0.0004
(a,d)	0.5198	0.6188		0.3088	0.5361	0.0391
(b,b)	0.6795	0.6061	0.3288	0.7683	0.6610	0.3309
(b,c)	0.5028	0.4351	0.0595	0.3299	0.5444	-0.0326
(b,d)	0.4031	0.5178		0.3501	0.5068	0.1139
(c,c)	0.8271	0.7162	0.6751	0.6879	0.7275	0.5767
(c,d)	0.6526	0.5130		0.2730	0.5586	0.1269
(d,d)	0.7918	0.7331		0.6548	0.6887	0.4191
BLUE						
(a,a)	0.6528	0.7969	0.5471	0.9189	0.9726	0.4414
(a,b)	0.3419	0.5264	0.0621	0.7517	0.6416	0.1600
(a,c)	0.3245	0.4076	0.0343	0.5558	0.6988	0.0524
(a,d)	0.2304	0.5305		0.5258	0.6704	0.1160
(b,b)	0.5966	0.5750	0.3050	0.8648	0.7701	0.3651
(b,c)	0.2723	0.3508	0.0164	0.5802	0.7090	0.0757
(b,d)	0.1142	0.4407		0.5583	0.6890	0.1680
(c,c)	0.7514	0.6714	0.6658	0.7739	0.8348	0.4935
(c,d)	0.5358	0.4292		0.4855	0.7314	0.1178
(d,d)	0.7331	0.6783		0.7582	0.8063	0.3631
|B-G|						
(a,a)	0.8701	0.7391	0.6490	0.8715	0.8455	0.6351
(a,b)	0.6960	0.4479	0.4141	0.7355	0.3003	0.3535
(a,c)	0.7084	0.5506	0.4755	0.6811	0.3018	0.3826
(a,d)	0.6240	0.4588		0.5097	0.3445	0.4076
(b,b)	0.8021	0.5310	0.4797	0.7971	0.6276	0.5532
(b,c)	0.6640	0.4185	0.3429	0.6593	0.5243	0.1153
(b,d)	0.5886	0.3559		0.5416	0.4965	0.3061
(c,c)	0.7955	0.7039	0.6391	0.7956	0.7214	0.7548
(c,d)	0.6291	0.4492		0.5739	0.5254	0.3608
(d,d)	0.7384	0.5778		0.8129	0.6750	0.5915

Table 15. Standard Deviation of the Correlation Values between pairs of Sessions per Patient, Lesion and Color Band of the outputs of the Registration Approach 2

Sessions	(1,A)	(1,B)	(1,C)	(2,A)	(2,B)	(3,C)
RED						
(a,a)	0.1085	0.0879	0.2196	0.0525	0.0207	0.2369
(a,b)	0.0450	0.1529	0.0729	0.0199	0.0149	0.0742
(a,c)	0.1049	0.0549	0.0429	0.0125	0.0242	0.0365
(a,d)	0.1201	0.0668		0.0276	0.0280	0.0938
(b,b)	0.1046	0.3420	0.3705	0.0850	0.1410	0.2349
(b,c)	0.0817	0.0915	0.0539	0.0182	0.0465	0.0264
(b,d)	0.0946	0.1492		0.0266	0.0466	0.1017
(c,c)	0.0371	0.1114	0.1419	0.1375	0.0995	0.1641
(c,d)	0.0435	0.0662		0.0254	0.0299	0.1052
(d,d)	0.0986	0.1320		0.1804	0.1293	0.2712
GREEN						
(a,a)	0.1204	0.1352	0.2655	0.0782	0.0202	0.3337
(a,b)	0.0338	0.1489	0.0617	0.0437	0.0128	0.0383
(a,c)	0.0706	0.0674	0.0260	0.0227	0.0284	0.0138
(a,d)	0.0954	0.0808		0.0386	0.0329	0.0637
(b,b)	0.1629	0.3404	0.3722	0.1288	0.1871	0.3354
(b,c)	0.0348	0.0971	0.0659	0.0288	0.0439	0.0170
(b,d)	0.0682	0.1429		0.0343	0.0463	0.0670
(c,c)	0.0887	0.2079	0.1767	0.1797	0.1457	0.2215
(c,d)	0.0548	0.0784		0.0361	0.0288	0.0797
(d,d)	0.1390	0.1966		0.2288	0.1672	0.3089
BLUE						
(a,a)	0.1828	0.1516	0.2785	0.0429	0.0156	0.2986
(a,b)	0.0517	0.1311	0.0660	0.0279	0.0212	0.0287
(a,c)	0.0958	0.0651	0.0332	0.0156	0.0214	0.0245
(a,d)	0.1356	0.0753		0.0319	0.0285	0.0471
(b,b)	0.2050	0.3543	0.3936	0.0757	0.1284	0.3186
(b,c)	0.0351	0.0797	0.0473	0.0251	0.0387	0.0168
(b,d)	0.0828	0.1154		0.0353	0.0423	0.0485
(c,c)	0.1267	0.2383	0.1821	0.1324	0.0882	0.2629
(c,d)	0.0554	0.0790		0.0282	0.0228	0.0280
(d,d)	0.1656	0.2361		0.1543	0.1040	0.3298
|B-G|						
(a,a)	0.0668	0.1926	0.2155	0.0658	0.0785	0.1927
(a,b)	0.0334	0.2376	0.1468	0.0412	0.0717	0.0917
(a,c)	0.0190	0.0684	0.0695	0.0242	0.0411	0.0514
(a,d)	0.0314	0.1222		0.0238	0.0224	0.0783
(b,b)	0.1031	0.3907	0.3391	0.1076	0.1990	0.2360
(b,c)	0.0308	0.2195	0.0693	0.0299	0.0388	0.0703
(b,d)	0.0535	0.2001		0.0128	0.0476	0.1310
(c,c)	0.1037	0.2167	0.1870	0.1094	0.1427	0.1275
(c,d)	0.0348	0.1167		0.0452	0.0498	0.1400
(d,d)	0.1423	0.3082		0.1083	0.1653	0.2322

E First M.A.D. Component of pairs of registered patterns of the same session

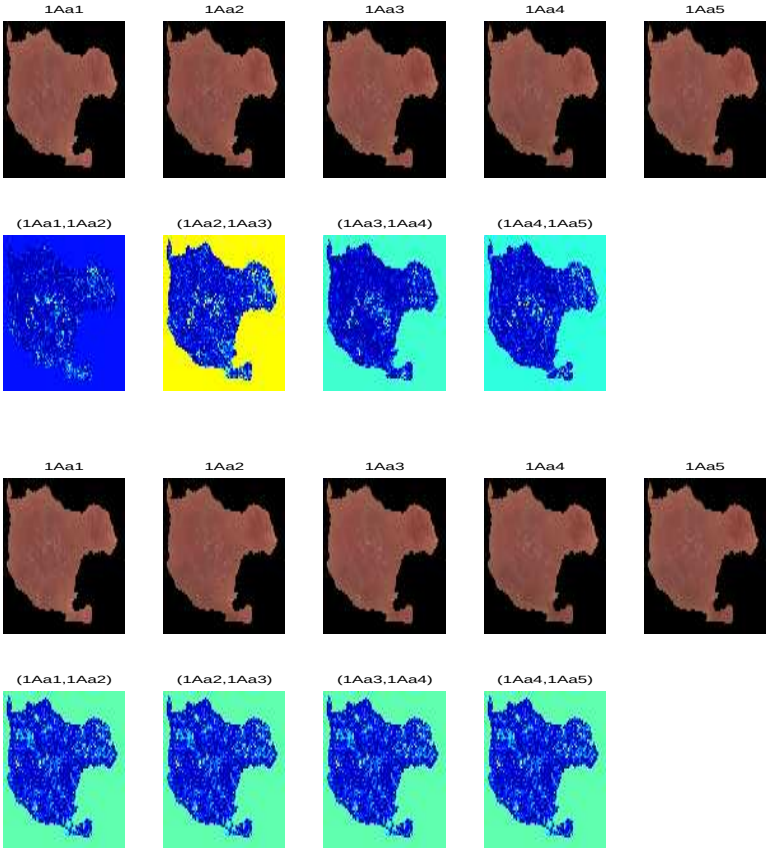


Fig. 20. Images of the first session of the case $(1, A)$ before (top) and after (bottom) registration, and respective first M.A.D. components

References

1. M. Amadasum and R. King. Some texture measures for image segmentation. *Canadian Journal of Electrical and Computer Engineering*, 15(1):18–21, 1990.
2. M. Bolduc and M. Levine. A real-time foveated sensor with overlapping receptive fields. *Real-Time Imaging*, (3):195–212, 1997.

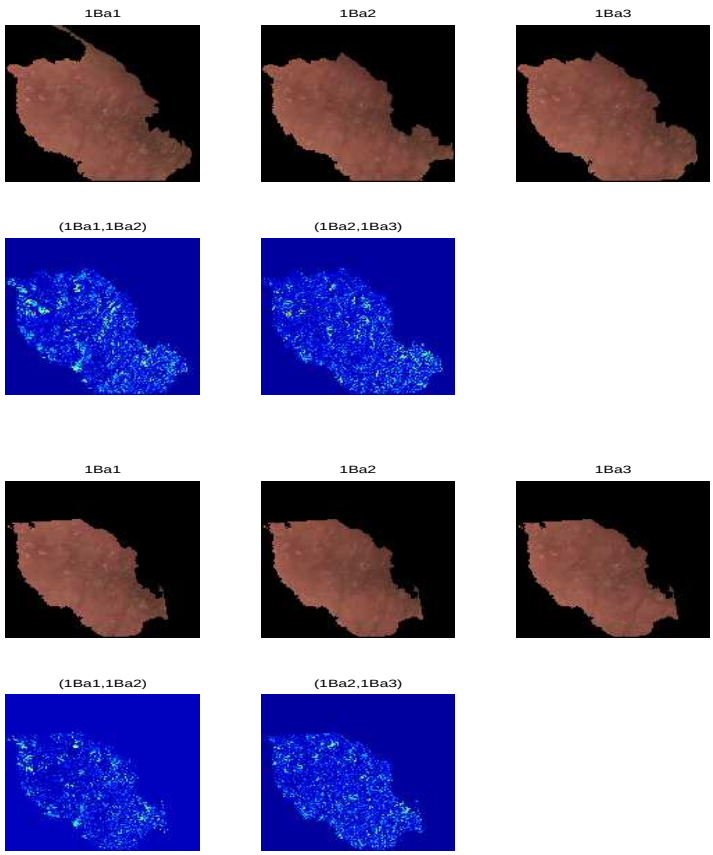


Fig. 21. Images of the first session of the case $(1, C)$ before (top) and after (bottom) registration, and respective first M.A.D. components

3. I. Dryden and K. Mardia. *Statistical Shape Analysis*, volume 1. John Wiley and Sons Ltd., West Sussex, England, 1998.
4. R. Duda and P. Hart. *Pattern Classification and Scene Analysis.* ISBN 0-471-22361-1. John Wiley & Sons, California, 1973.
5. A. Gordon. *Classification. Monographs on Applied Probability and Statistics.* ISBN 0-412-22850-5. Chapman and Hall and Methuen, Inc., London, 1981.
6. R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.
7. D. He and L. Wang. Texture features based on texture spectrum. *Pattern Recognition*, 24(5):391–399, 1991.
8. K. Hilger. *Exploratory Analysis of Multivariate Data.* Ph.d. thesis, Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby, November 2001.

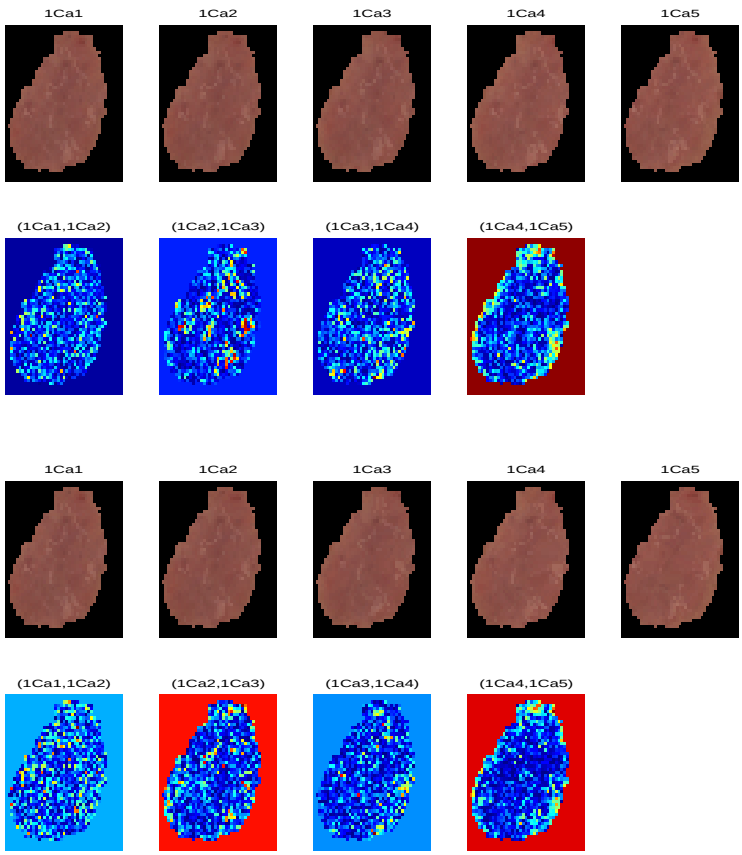


Fig. 22. Images of the first session of the case $(1, C)$ before (top) and after (bottom) registration, and respective first M.A.D. components

9. J. Lira. *Introducción al Tratamiento Digital de Imágenes.* Instituto Politécnico Nacional, Universidad Nacional Autónoma de México, Fondo de Cultura Económica., México D.F., 1 edition, September 2002.
10. J. Lira and G. Maletti. A supervised classifier for multispectral and textured images based on an automated region growing algorithm. *European Space Agency Publications*, SP-434:153–158, 1998.
11. G. Maletti and B. Ersbøll. A hierarchical classification scheme of psoriasis images. Technical Report 6, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
12. G. Maletti and B. Ersbøll. Illumination correction in psoriasis lesions images. Technical Report 7, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
13. G. Maletti and B. Ersbøll. Principal component analysis of psoriasis lesions images. Technical Report 5, Department of Informatics and Mathematical Modelling.

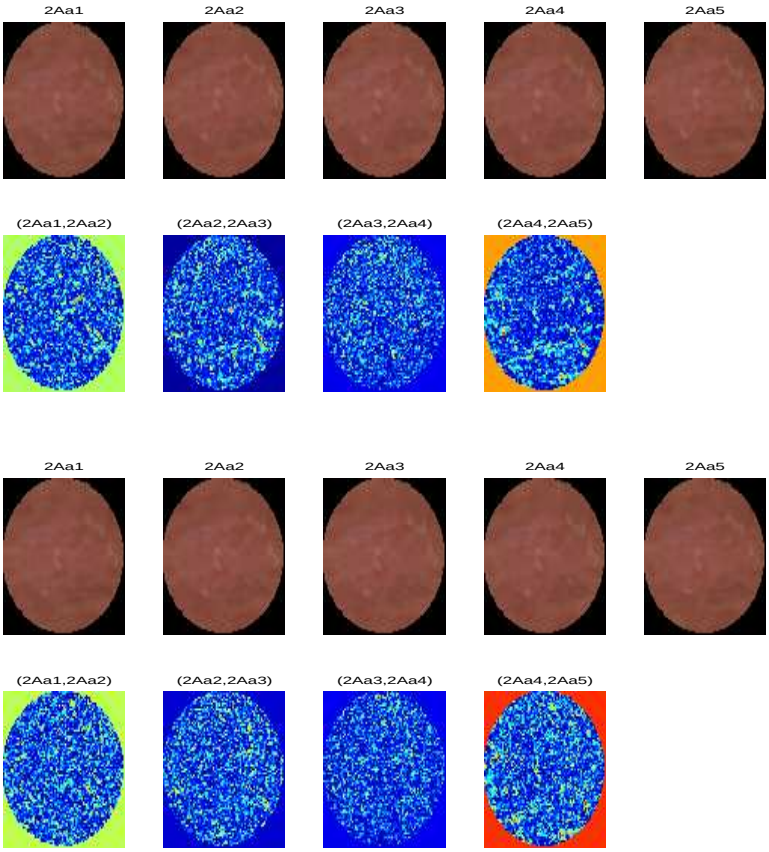


Fig. 23. Images of the first session of the case $(2, A)$ before (top) and after (bottom) registration, and respective first M.A.D. components

Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

14. G. Maletti and B. Ersbøll. Texture alteration detection in bitemporal images of lesions with psoriasis. Technical Report 8, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
15. K. Mardia, J. Kent, and J. Bibby. *Multivariate Analysis*. Academic Press, Inc., San Diego, CA92101, 10 edition, 1979.
16. A. Nielsen, K. Conradsen, and J. Simpson. Multivariate alteration detection (m.a.d.) and m.a.f. postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Environ.*, 64:1–19, 1998.
17. M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Brooks Cole Publishing Company, Pacific Grove, CA 93950, USA., 2nd edition, 1999.

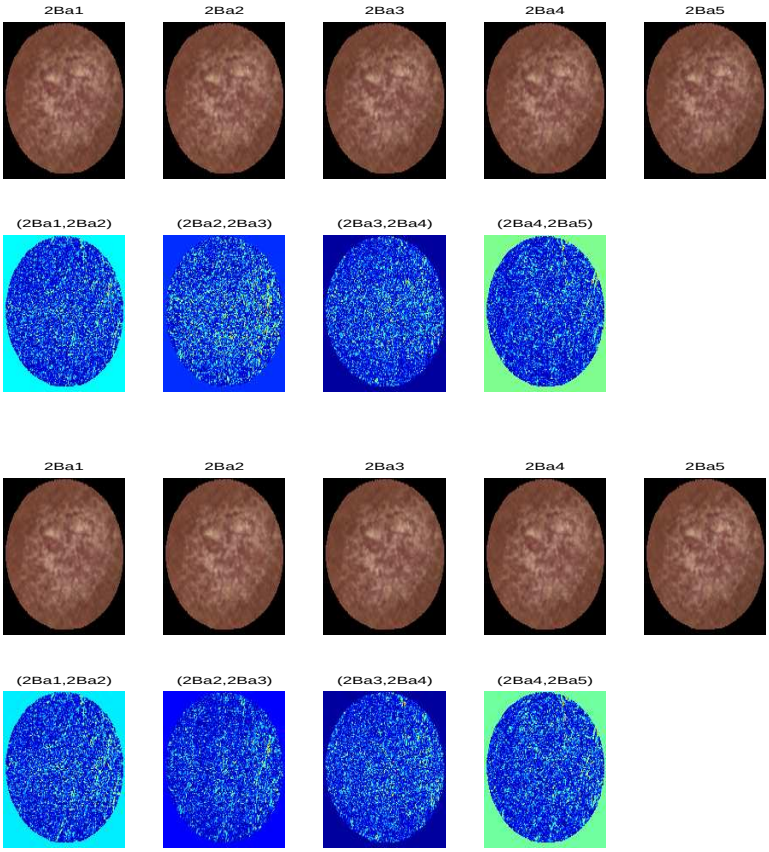


Fig. 24. Images of the first session of the case $(2, B)$ before (top) and after (bottom) registration, and respective first M.A.D. components

18. M. Stegmann. Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2000.
19. L. Wang. Vector choice in the texture spectrum approach. *International Journal of Remote Sensing*, 15(18):3823–3829, 1994.
20. L. Wang and D. He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910, 1990.
21. R. Wilson and T. Martinez. Improved heterogeneous distance function. *Journal of Artificial Intelligence Research*, (6):1–34, 1997.
22. S. Wilson. On the retino-cortical mapping. *International Journal of Man-Machine Studies*, 18(4):361–389, 1983.

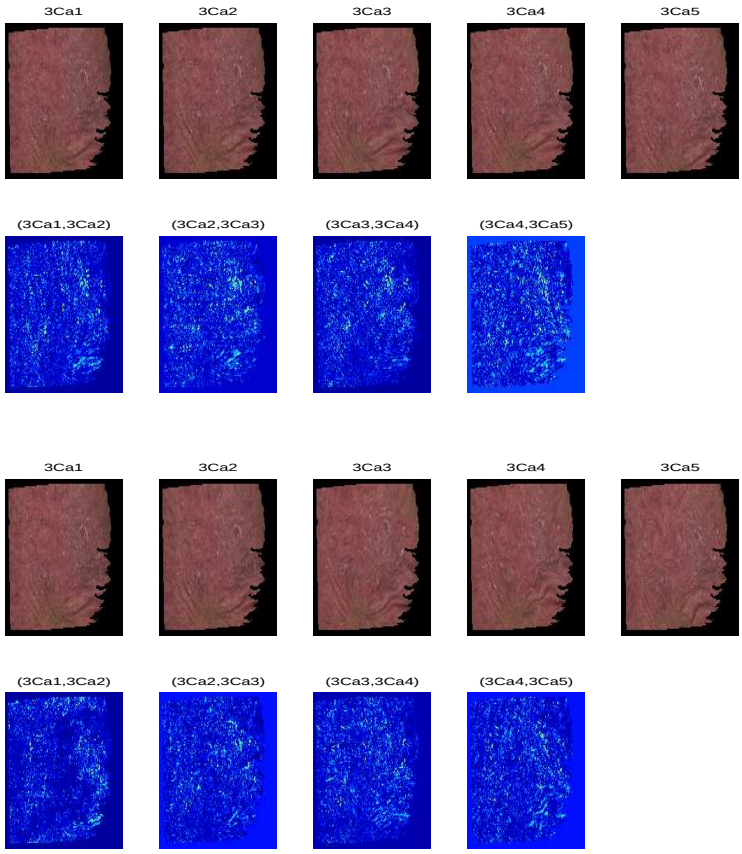


Fig. 25. Images of the first session of the case $(3, C)$ before (top) and after (bottom) registration, and respective first M.A.D. components

4.2.6 Change Detection in Registered Psoriasis Lesions Image Patterns

Authors: G. Maletti and B. Ersbøll

Technical Report: Internal I.M.M.
Number: 10
Pages: 1 – 14
Reference: [32]

Change Detection in Registered Psoriasis Lesions Image Patterns

Gabriela Maletti and Bjarne Ersbøll

Department of Informatics and Mathematical Modelling,
Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark
{gmm, be}@imm.du.dk
http://www.imm.dtu.dk/image

Abstract. The M.A.D. Transform is applied to pairs of registered psoriasis lesion patterns within and between weekly sessions. Color band contribution to the first M.A.D. component is analyzed.

1 Introduction

We use a set of registered psoriasis lesion patterns segmented from a set of images taken at the Gentofte Hospital, Denmark during four pilot sessions with three invited patients. On those occasions, three lesions per patient were captured five times during each session, along four weeks. The images were labelled with four characters, indicating patient (1,2,3), lesion (A, B, C), session (a, b, c, d) and capture (1, 2, 3, 4, 5) respectively. From a subset of these images, lesion patterns were segmented and registered within and between sessions using the schemes reported in [4] and [3] respectively.

To evaluate lesion changes between sessions, physicians make scores of a four-variable set: redness, scaling, thickness and body area covered by the lesions. Manual scoring depends strongly on the dermatologist, who can show criteria variations along sessions due to the huge amount of patients and lesions observed during each working day.

The main objective of the present work is to analyze changes in time [8], [10], [1] from a statistical point of view of the registered data sets mentioned. A suitable statistical approach for analyzing changes in time in a multi-variable fashion is the Multi-variate Alteration Detection Transform (M.A.D.) [9]. It transforms two sets of multi-variate observations into a difference between two linear combinations of the original variables explaining maximal change in all variables simultaneously. The magnitude of the lesion changes in time could be shown using the absolute value of the first M.A.D. component. Alternatively, the color band contribution to the change could be calculated.

The present work is composed of the following sections. First, a short description of the registration algorithm is given. Secondly, general aspects of the

M.A.D. Transform are described. Then, an experiment to compute the color band contribution in the lesions change in time is designed. A section of results and discussion follows. Results for each single case can be found in the Appendix.

2 Registration Scheme

For completeness, the registration scheme used to align the patterns is briefly described. Details can be found in [3]. The registration scheme is composed of two stages. In the first stage, global rotation and translation effects of assumed segmented patterns of equally scaled psoriasis lesions are removed. In the second stage, it is assumed that there are no local rotation effects and, therefore, only local translation effects are removed. In both stages, the Extreme Value Detection Algorithm (E.V.D.A.) is applied to compute the minimal distance between pairs of translated and rotated patterns.

E.V.D.A. is an iterative process that expects the translation effects between two objects to be matched are removed when their center of masses match. In the first iteration, a mapping template centered on the center of mass of the domain of one of the objects is used to define the subset of the domain of the distance function for which the values are going to be estimated. This mapping template includes many points of the neighborhood of the center of mass, and only few that are far away from the center. The extreme value of the calculated values of the distance function is obtained, and the same mapping template is centered on that first guess of the extreme value. This step is repeated until the guessed extreme value does not change anymore.

3 M.A.D. Transform

Following [8], let X_1 and X_2 be two sets of Gaussian distributed variables of dimension m_1 and m_2 respectively, with $m_1 \leq m_2$, $E\{X_1\} = E\{X_2\} = 0$ transformed with the coefficients from a standard canonical correlation analysis $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in

$$\begin{aligned} Y_{1i} &= \boldsymbol{a}_{1i}^T X_1, i = 1, ..., m_1 \\ Y_{2i} &= \boldsymbol{a}_{2i}^T X_2, i = 1, ..., m_2, \end{aligned} \tag{1}$$

positively correlated and with unit variance, such that the variance of their difference is maximized.

$$Var[\boldsymbol{a}_1^T X_1 - \boldsymbol{a}_2^T X_2] = \tag{2}$$

$$Var[\boldsymbol{a}_1^T X_1] + Var[\boldsymbol{a}_2^T X_2] - 2Cov[\boldsymbol{a}_1^T X_1, \boldsymbol{a}_2^T X_2] \tag{3}$$

$$= 2(1 - Corr[\boldsymbol{a}_1^T X_1, \boldsymbol{a}_2^T X_2]) \tag{4}$$

The M.A.D. Transform consists of the variates to get when the corresponding canonical variates are subtracted in reverse order, which means that the m_1^{th}

difference shows maximum variance among such variables, and the $(m_1 - j)^{th}$ difference shows maximum variance subject to the constraint that this difference is not correlated with the previous j ones. The M.A.D. variates are invariant to linear and affine scaling, which is not the case of the Principal Components. The dispersion matrix of the M.A.D. variates is

$$D[a_1^T X_1 - a_2^T X_2] = 2(I - R) \tag{5}$$

where I is the $m_1 \times m_1$ unit matrix and R is the $m_1 \times m_1$ matrix containing the sorted canonical correlations on the diagonal and zeros off the diagonal.

4 Results and Discussion

The contribution of each single color band in showing the lesion change in a period of one week is analyzed here. Results are confronted with the original image data. Lesion changes between sessions are also described using the absolute value of the first M.A.D. component.

Following with the notation used in previous work ([6], [4], [5], [7], [3]), the groups of images corresponding to a given patient and lesion are here called "case (patient, lesion)". The cases registered in [3] are $(1, A)$, $(1, B)$, $(1, C)$, $(2, A)$, $(2, B)$ and $(3, C)$, which can be seen in the Appendix C of the referred work.

The set of 108 registered images produced by the registration scheme (see Appendix C in [3]) allowed the construction of 377 pairs[1] of registered lesions with the same time increment of one week between them. For each image of a given patient and lesion, each one of the images of the next session were associated. For each pair of synthetical images constructed in the way mentioned, the M.A.D. transform was applied [2]. The M.A.D. Transform was only applied to the pixels belonging to the lesion. Tables 1 and 2 show the results obtained for the sets of registered images.

Each single cell value in Table 1 is the average absolute correlation value between the $b-th$ color bands and the $m-th$ MAD components of all pairs of images of the $l-th$ lesion of the $p-th$ patient. It is computed in the following way:

$$E[\rho_{p,l,b,m}] = \frac{1}{(n_s - \delta t) n_c^2} \sum_{s=1}^{n_s - \delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} a_{p,l,s,i,\delta_t,j,b,m} \tag{6}$$

[1] For each lesion the number of pairs is $(1, A) = 25 \times 3$; $(1, B) = 9 \times 3$; $(1, C) = 25 \times 2$; $(2, A) = 25 \times 3$; $(2, B) = 25 \times 3$ and $(2, C) = 25 \times 3$ respectively.

where

$$a_{p,l,s,i,\delta_t,j,b,m} = \frac{1}{2}(|\rho[X_{p,l,s,i,b}, MAD_{p,l,s,i,\delta t,j,b}, m]| + |\rho[X_{p,l,s+\delta t,j,b}, MAD_{p,l,s,i,\delta t,j,b}, m]|) \tag{7}$$

and ρ is the correlation coefficient, δt is the time increment (in this case, δt is one week). The indexes i and j indicate the capture number within a session. $MAD_{p,l,s,i,\delta t,j,b,m}$ equals to $MAD[X_{p,l,s,i,b}, X_{p,l,s+\delta t,j,b}, m]$.

Each single cell value in Table 2 is the standard deviation of the averaged absolute correlation values between the normalized co-occurrence matrix of the $b-th$ color bands and the $m-th$ MAD components of all pairs of images of the $l-th$ lesion of the $p-th$ patient. It is computed in the following way:

$$S[\rho_{p,l,b,m}] = \sqrt{\frac{1}{(n_s - \delta t)n_c^2} \sum_{s=1}^{n_s-\delta t} \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} (a^2_{p,l,s,i,\delta_t,j,b,m}) - E[\rho_{p,l,b,m}]^2} \tag{8}$$

where the symbols have the same meaning as before.

Table 1. Average Absolute Correlation Values per Lesion of *RGB* registered lesion data with their respective M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.239941	0.312891	0.479555
(1,B)	0.194602	0.342770	0.404496
(1,C)	0.502525	0.582041	0.602087
(2,A)	0.273979	0.444085	0.298466
(2,B)	0.232567	0.325108	0.154075
(3,C)	0.399863	0.536893	0.529572
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.071068	0.267727	0.141211
(1,B)	0.086706	0.122383	0.123006
(1,C)	0.211631	0.176254	0.224510
(2,A)	0.185631	0.198860	0.123149
(2,B)	0.123701	0.157495	0.116539
(3,C)	0.314972	0.258697	0.225947
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.246138	0.150544	0.097587
(1,B)	0.324035	0.278036	0.255027
(1,C)	0.254546	0.171539	0.071888
(2,A)	0.246606	0.161208	0.262824
(2,B)	0.302258	0.267418	0.318841
(3,C)	0.212062	0.127252	0.186059

Excluding from Table 1 the case for which the registered lesion patterns are partially covered by hair (Patient 2, Lesion B), we can observe that, in general,

Table 2. Standard Deviation of the Absolute Correlation Values per Lesion of the *RGB* registered lesion data with the M.A.D. Components

(Patient,Lesion)	(R,MAD1)	(G,MAD1)	(B,MAD1)
(1,A)	0.082760	0.088492	0.076644
(1,B)	0.062977	0.088879	0.101755
(1,C)	0.071854	0.051745	0.044919
(2,A)	0.145003	0.130561	0.081781
(2,B)	0.043526	0.078260	0.060818
(3,C)	0.120938	0.129143	0.099749
(Patient,Lesion)	(R,MAD2)	(G,MAD2)	(B,MAD2)
(1,A)	0.047430	0.051808	0.072253
(1,B)	0.042134	0.102374	0.107343
(1,C)	0.088767	0.095334	0.078513
(2,A)	0.103476	0.125162	0.076504
(2,B)	0.055127	0.103429	0.061193
(3,C)	0.167413	0.165183	0.135105
(Patient,Lesion)	(R,MAD3)	(G,MAD3)	(B,MAD3)
(1,A)	0.021716	0.022367	0.038063
(1,B)	0.074971	0.062505	0.061780
(1,C)	0.052423	0.057971	0.050960
(2,A)	0.023292	0.017698	0.027054
(2,B)	0.027727	0.028042	0.025397
(3,C)	0.069320	0.051744	0.049821

most of the contribution in the first M.A.D. component is given either by the green or the blue band. These results can be verified, by visual examination of the registered lesions image patterns decomposed in its *RGB* bands included in Appendix B.

Some outputs produced during the experiment mentioned before are shown in Appendix A. For each case, registered image patterns, corresponding to the first image of each session, and the absolute first M.A.D. components of pairs of those patterns with a time separation of one week can be seen. In order to show the magnitude of the change, the absolute value of the first M.A.D. component has been computed. Light colors in the figures corresponding to the absolute value of the first M.A.D. components imply more change in the lesion; dark colors imply the opposite situation. For instance, for the case $(1, A)$, in the upper part of the figure we have four registered patterns, corresponding to the first, second, third and fourth week. In the lower part of the same figure, we have from the left to the right, the absolute value of the first M.A.D. component from the first to the second week, from the second to the third week, and, from the third to the fourth week respectively. The figures corresponding to the remaining cases are presented in the same way. Some comments regarding to visual examination of the results follow.

For the case $(1, A)$, it can be seen that most of the change from the first to the second week occurs in the central part of the lesion. Afterwards, from the second to the third week, it spreads out. From the third to the fourth week, most of the change occurs near the borders. The same behavior can be described for the case $(1, B)$ and for the first three weeks of the case $(1, C)$. In the upper part of the first M.A.D. component from the first to the second week of the case $(1, B)$ a large change is indicated. Looking at the original images, it seems that this aspect could be indicating different illumination conditions [5] between the captures, more than lesion changes. The reader is reminded that the mentioned cases do correspond to the same patient. For the case $(2, A)$ an interesting peak can be observed in the absolute value of the first M.A.D. component from the first to the second week. This is not easily appreciable in the original images. Clearly the M.A.D. transform is able to show a change that cannot be seen in the original images. For the case $(2, B)$, it can again be observed that the change starts in the center of the lesion and in the next week it spreads out. It was mentioned in [3], that the registration algorithm gets a bit confused in the presence of hair. Hair displacements between sessions can not easily be seen in the absolute value first M.A.D. component. This may be because of the quality of the registration output itself. For the case $(3, C)$, the presence of scaling in the second week is clearly indicated by large absolute values of the first M.A.D. component, from the first to the second week, and, from the second to the third week. However, peaks appearing in the absolute value of the first M.A.D. component from the third to the fourth week are clearly skin displacements attained after folding the elbow. In general, it could be said that the absolute value of the first M.A.D. component shows in a very clear manner, the appearing and disappearing of scaling.

An aspect that was not considered here, because it is not the objective of the present work, is that the M.A.D. transform, as it was conceived in [9], requires positive correlation of the canonical variables. The cases for which the canonical variables are not positively correlated[2], could be interesting to analyze. This is referred to the meaning of the output of the difference of negatively correlated canonical variables. If the M.A.D. Transform is extended to all the orientations of the two sets of eigenvectors by flipping the eigenvectors values of the original variables, a set of four possible solutions is produced.

5 Conclusions

The small number of lesions evaluated is not enough to make general conclusions about the disease behavior. However, for the cases analyzed the following can be mentioned. Looking at the absolute value of the first M.A.D. component

[2] In the psoriasis lesions data set there cases for which the correlation between pairs of images is negative. See for instance, in Tables 6, 8 and 10 of [3] the values corresponding to the case $(3, C)$.

of pairs of registered lesion patterns separated by a period of one week, it can be observed that in the beginning, most of the change occurs within the lesion, while in the next week it spreads out and is distributed on the whole lesion. In some cases it has been observed, that in the lasts sessions, the change occurs near the lesion borders.

Regarding the contribution of each single color band to the first M.A.D. component, it has been shown that, in general, it is either the green or the blue band, and not the red band which contributes most in showing the change.

Acknowledgments

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number *STVF* 56-00-0123) for supporting the present work. To the dermatologists Lone Skov and Bo Bang of the Gentofte Hospital of Denmark and to the anonymous patients, for their collaboration during the image acquisition sessions.

Appendix

A Registered Images and First M.A.D. Components

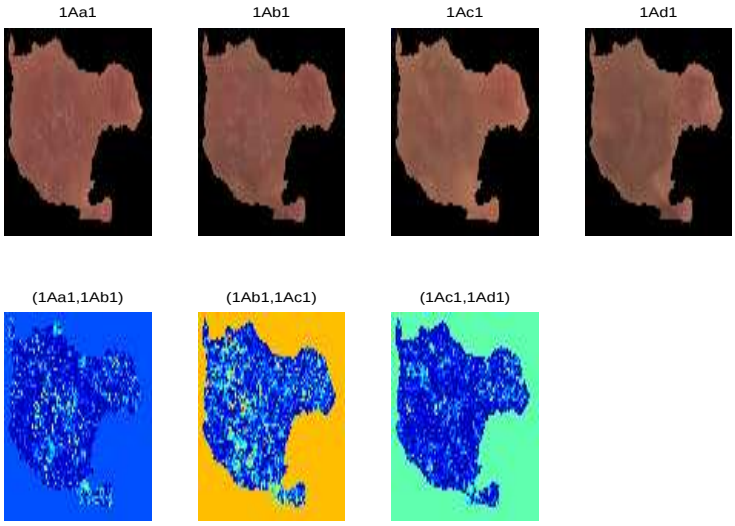


Fig. 1. Registered Images for the case $(1, A)$ and the First M.A.D. components per week.

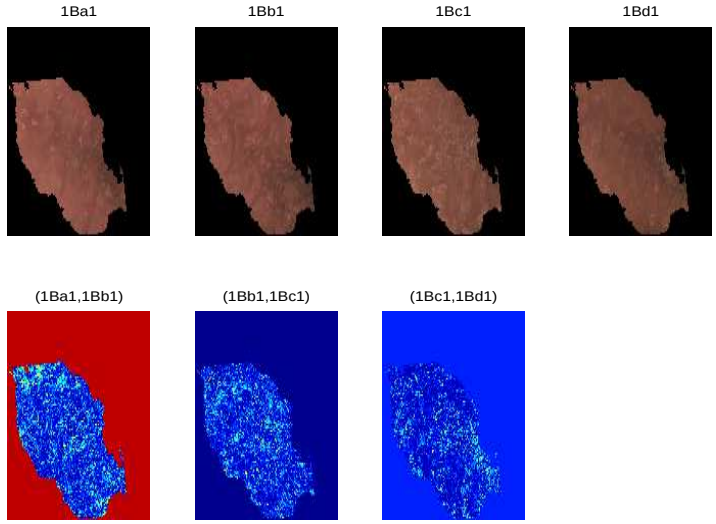


Fig. 2. Registered Images for the case $(1, B)$ and the First M.A.D. components per week.

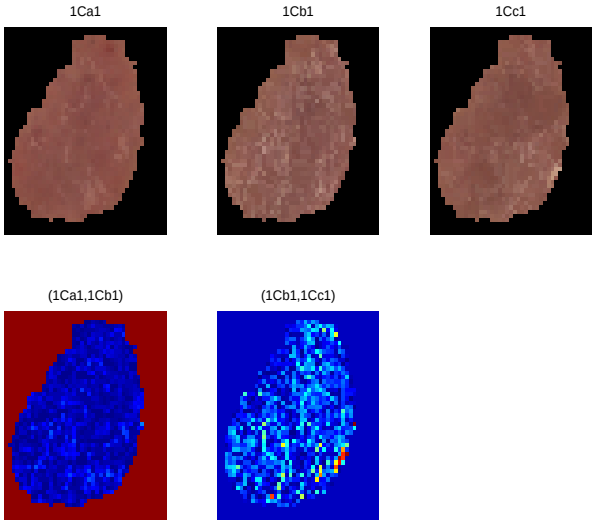


Fig. 3. Registered Images for the case $(1, C)$ and the First M.A.D. components per week.

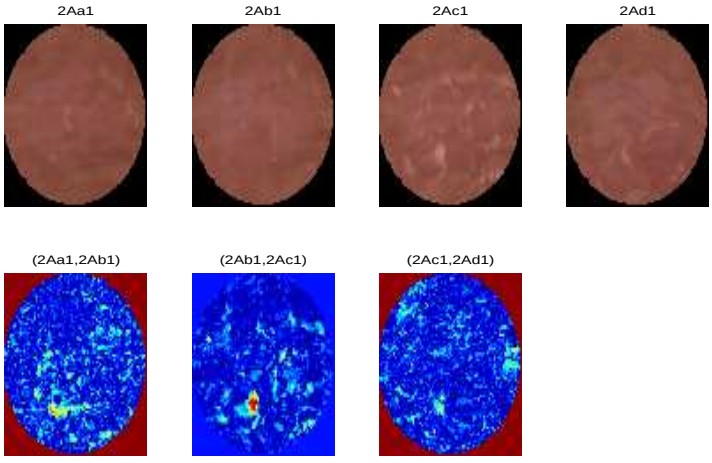


Fig. 4. Registered Images for the case $(2, A)$ and the First M.A.D. components per week.

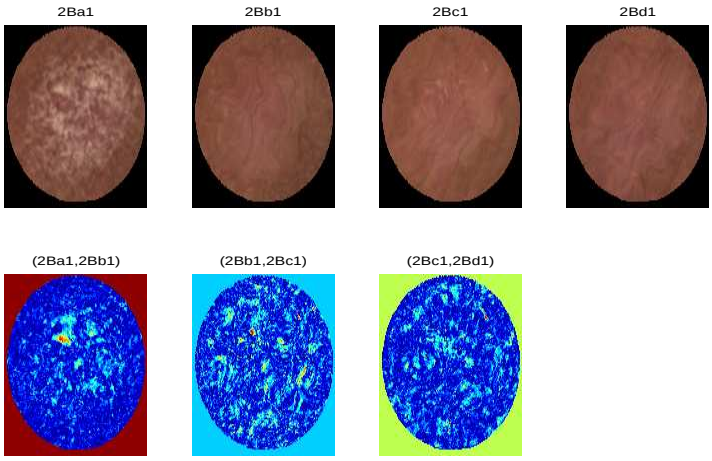


Fig. 5. Registered Images for the case $(2, B)$ and the First M.A.D. components per week.

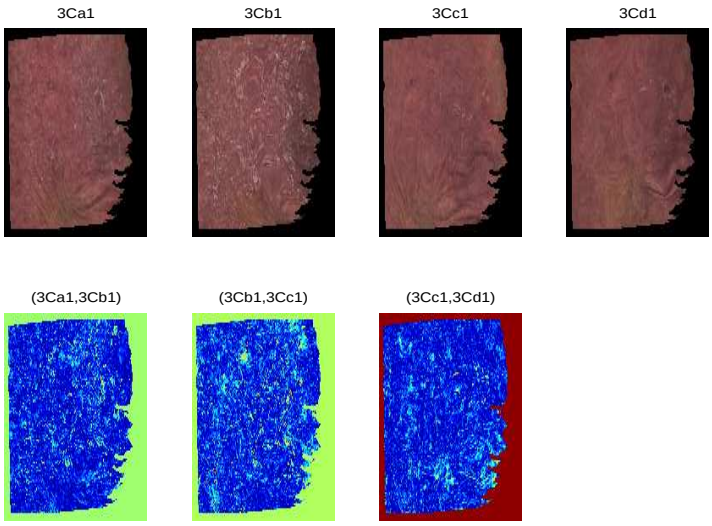


Fig. 6. Registered Images for the case $(3, C)$ and the First M.A.D. components per week.

B *RGB* bands of the registered images

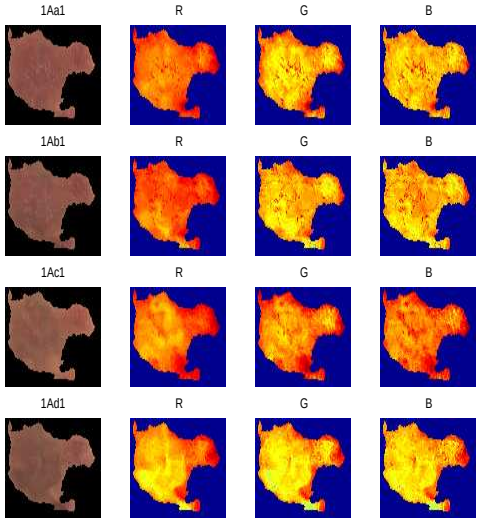


Fig. 7. *RGB* bands of registered images of the case $(1, A)$

References

1. B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–204, 1981.
2. R. Larsen. Maf and other transformations applied to remote sensing. Master's thesis, Institute for Mathematical stastistics and Operations Research (IMSOR), Technical University of Denmark, DTU, 1991. pp. xii+130+205.
3. G. Maletti and B. Ersbøll. A combined alignment and registration scheme of psoriasis lesions images. Technical Report 9, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
4. G. Maletti and B. Ersbøll. A hierarchical classification scheme of psoriasis images. Technical Report 6, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
5. G. Maletti and B. Ersbøll. Illumination correction in psoriasis lesions images. Technical Report 7, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
6. G. Maletti and B. Ersbøll. Principal component analysis of psoriasis lesions images. Technical Report 5, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

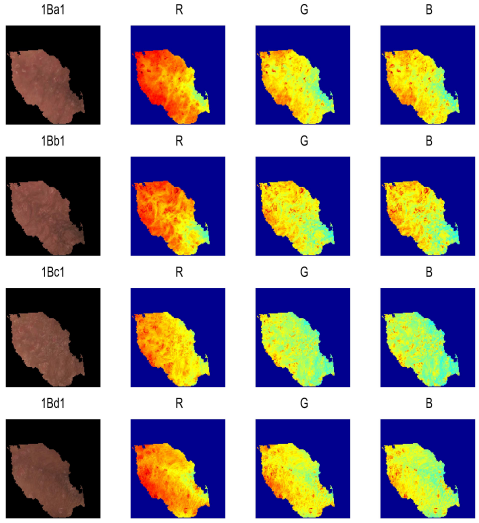


Fig. 8. RGB bands of registered images of the case $(1, B)$

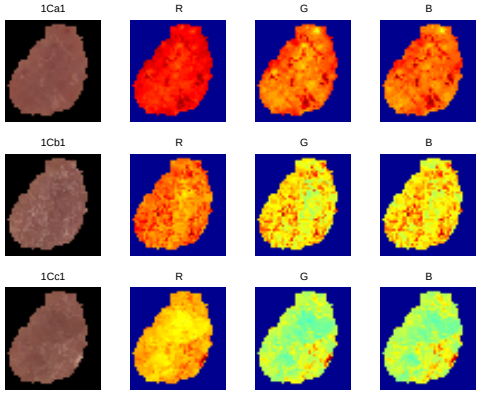


Fig. 9. RGB bands of registered images of the case $(1, C)$

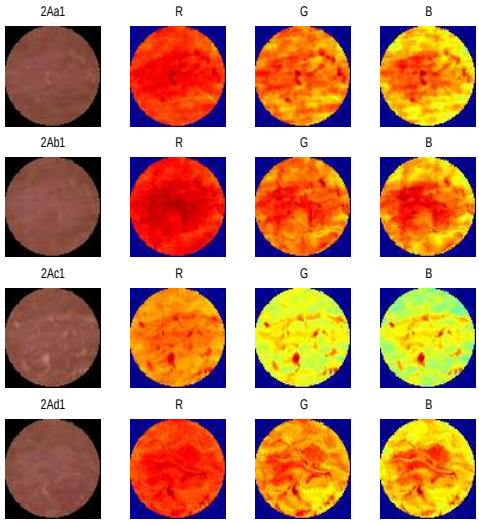


Fig. 10. RGB bands of registered images of the case $(2, A)$

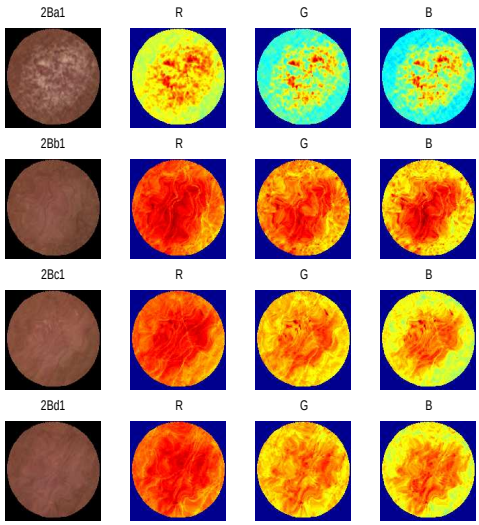


Fig. 11. RGB bands of registered images of the case $(2, B)$

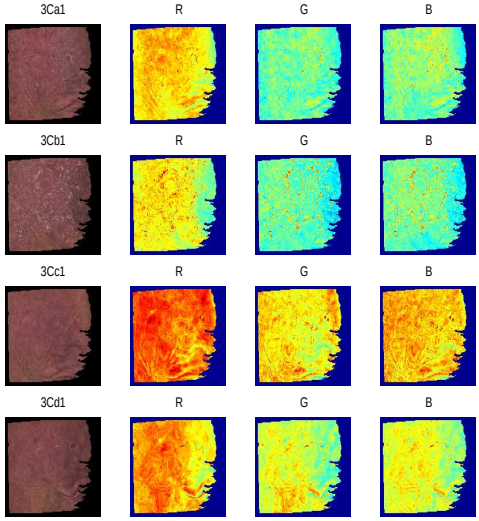


Fig. 12. *RGB* bands of registered images of the case $(3, C)$

7. G. Maletti and B. Ersbøll. Texture alteration detection in bitemporal images of lesions with psoriasis. Technical Report 8, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.
8. A. Nielsen. *Analysis of Regularly and Irregularly sampled spatial, multivariate, and muti-temporal data*. PhD thesis, Institute of Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark, October 1994. No. 6.
9. A. Nielsen, K. Conradsen, and J. Simpson. Multivariate alteration detection (m.a.d.) and m.a.f. postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Environ.*, 64:1–19, 1998.
10. M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Brooks Cole Publishing Company, Pacific Grove, CA 93950, USA., 2nd edition, 1999.

Appendix A

Appendix A

A.1 Set of Original Images

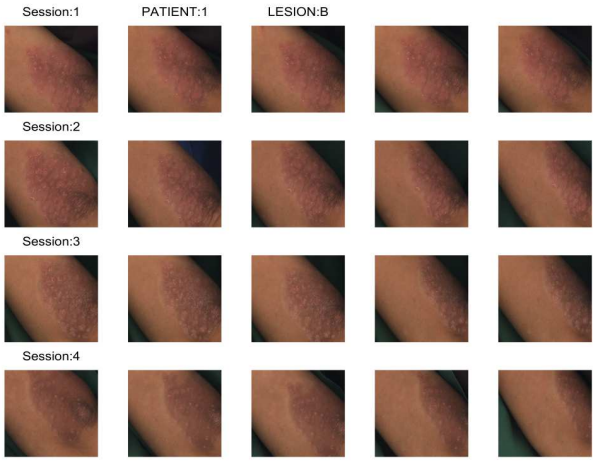


Figure A.1: Set of Original Images of (Patient 1, Lesion B).

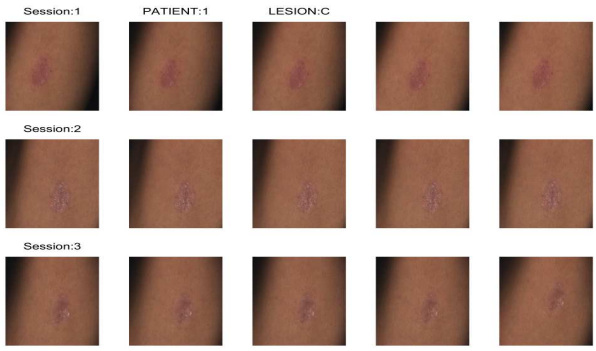


Figure A.2: Set of Original Images of (Patient 1, Lesion C).

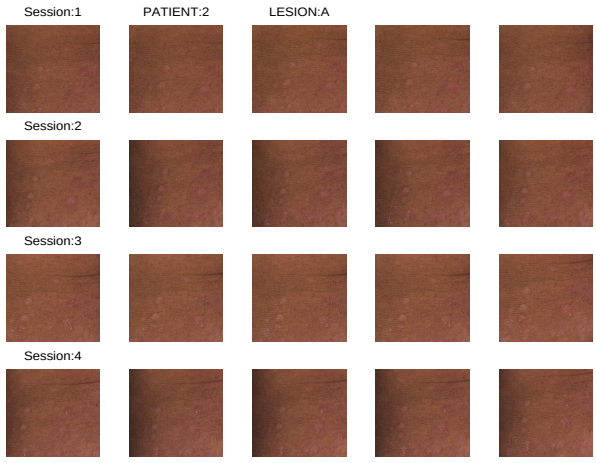


Figure A.3: Set of Original Images of (Patient 2, Lesion A).

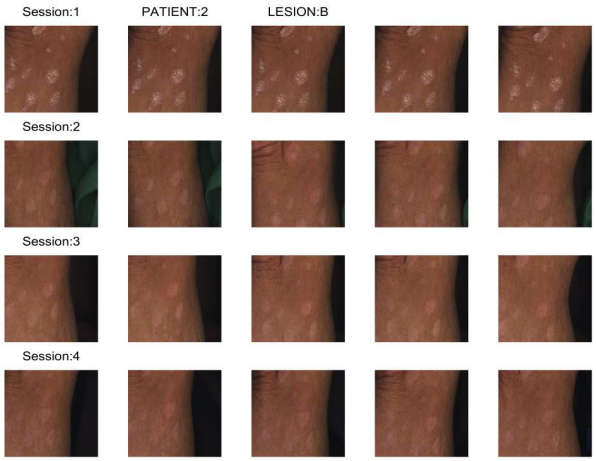


Figure A.4: Set of Original Images of (Patient 2, Lesion B).

Figure A.5: Set of Original Images of (Patient 2, Lesion C).

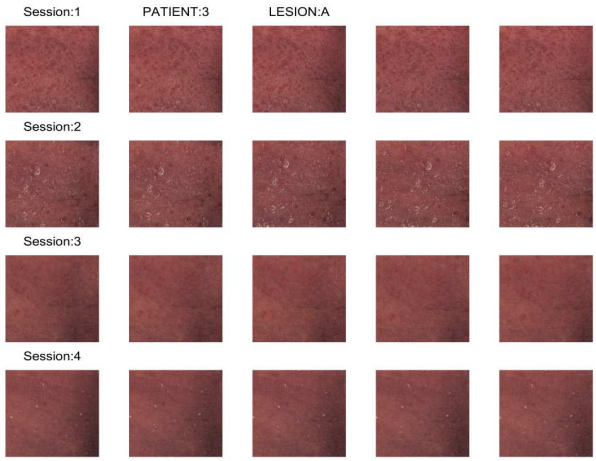


Figure A.6: Set of Original Images of (Patient 3, Lesion A).

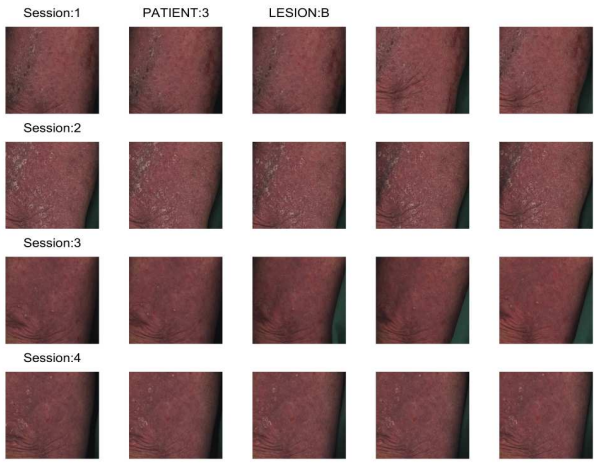


Figure A.7: Set of Original Images of (Patient 3, Lesion B).

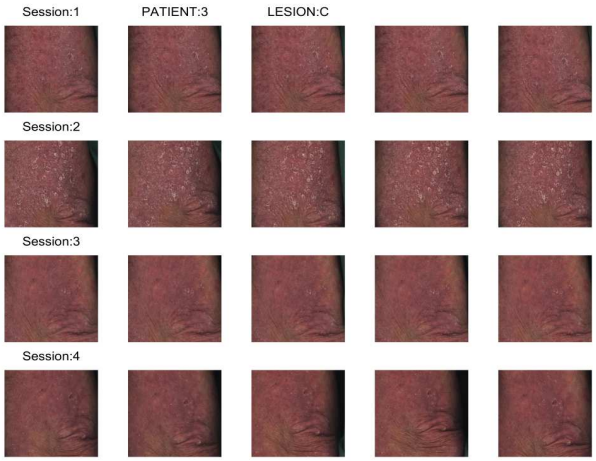


Figure A.8: Set of Original Images of (Patient 3, Lesion C).

Bibliography

[1] Videometer. http:www.videometer.dk.

[2] J. Benedetto and M. Frazier. *Wavelets. Mathematics and Applications*. Studies in Advanced Mathematics. CRC Press, Inc., Washington University in St. Louis, 1994.

[3] M. Biehl and A. Mietzner. Statistical mechanics of unsupervised structure recognition. *J. Phys. A.: Math. Gen.*, 27:1885–1897, 1994.

[4] A. Buhot and M. Gordon. Phase transitions in optimal unsupervised learning. *Physical Review E*, 57(3):3326–3333, 1998.

[5] K. Cave. The featuregate model of visual selection. *Psychological Research*, 62:182–194, 1999.

[6] D. Chung and G. Sapiro. Segmenting skin lesions with partial differential equation based image processing algorithms. *IEEE International Conference on Image Processing*, 3(24):404–407, 2000.

[7] W. Clarke, P. Lachenbruch, and B. Broffitt. How non-normality affects the quadratic discriminant function. *Communications in Statistics*, A8:1285–1301, 1979.

[8] University Bonn Computer Vision Group, Computer Science. Collection of microtextures. http://www-dbv.cs.uni-bonn.de/image/browse.html .

[9] K. Conradsen. *En Introduktion til Statistik. Bind 2.* IMSOR DTH., Lyngby, 1984.

[10] C. Craver-Lemley, M. Arterberry, and A. Reeves. Illusory illusory conjunctions: The conjoining of features of visual and imagined stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4):1036–1049, 1999.

[11] S. Donoho and L. Rendell. Rerepresenting and restructuring domain theories: A constructive induction approach. *Journal of Artificial Intelligence Research*, (2):411–446, April 1995.

[12] I. Dryden and K. Mardia. *Statistical Shape Analysis*, volume 1. John Wiley and Sons Ltd., West Sussex, England, 1998.

[13] J. Fürnkranz. Integrative windowing. *Journal of Artificial Intelligence Research*, 8:129–164, 1998.

[14] G. Hance, S. Umbaugh, R. Moss, and W. Stoecker. Unsupervised color image segmentation. *IEEE Engineering in Medicine and Biology*, 15(1):104–111, 1996.

[15] R. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.

[16] J. Hegdé and F. Felleman. The popout in some conjunction searches is due to perceptual grouping. *Cognitive Neuroscience*, 10:143–148, 1999.

[17] J. Hertz, A. Krogh, and G. Thorbergsson. Phase transitions in simple learning. *Journal of Physics A (Mathematical and General)*, 22(12):2133–2150, 1989.

[18] K. Hilger. *Exploratory Analysis of Multivariate Data.* Ph.d. thesis, Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby, November 2001.

[19] M. Hodgson. What size window for image classification? a cognitive perspective. *Photogrammetric Engineering and Remote Sensing*, 64(8):797–807, 1998.

[20] T. Horowitz and J. Wolfe. Visual search has no memory. *Letters to Nature*, 394(6693):575–577, 1998.

[21] H. Ioannidou. Statistical interpretation of the planck constant and the uncertainty relation. *International Journal of Theoretical Physics*, 22(12):1129–1139, 1983.

[22] Y. Kareev. Through a narrow window: working memory capacity and the detection of covariation. *Cognition*, 56:263–269, 1995.

[23] T. Kirkpatrick and D. Thirumalai. Comparison between dynamical theories and metastable states in regular and glassy mean-field spin models with underlying first-order-like phase transitions. *Physical Review A.*, 37(11):4439–4448, 1988.

[24] R. Larsen. Maf and other transformations applied to remote sensing. Master's thesis, Institute for Mathematical stastistics and Operations Research (IMSOR), Technical University of Denmark, DTU, 1991. pp. xii+130+205.

[25] J. Lira. *Introducción al Tratamiento Digital de Imágenes.* Instituto Politécnico Nacional, Universidad Nacional Autónoma de México, Fondo de Cultura Económica., México D.F., 1 edition, September 2002.

[26] J. Lira and L. Frulla. An automated region growing algorithm for segmentation of texture regions in sar images. *International Journal of Remote Sensing*, 19(18):3595–3606, 1998.

[27] J. Lira and G. Maletti. A supervised classifier for multispectral and textured images based on an automated region growing algorithm. *European Space Agency Publications*, SP-434:153–158, 1998.

[28] J. Lira and G. Maletti. A supervised contextual classification scheme based on an automated region growing algorithm. *Computers and Geosciences*, 28(8):951–959, 2002.

[29] G. Logan. The code theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, 103(4):603–649, 1996.

[30] A. Maintz and A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[31] G. Maletti. A supervised contextual classifier for multi-spectral and textural images based on an automated region growing algorithm. Master's thesis, National Autonomous University of Mexico., August 1998. Adviser: Jorge Lira Chávez.

[32] G. Maletti and B. Ersbøll. Change detection in registered psoriasis lesions image patterns. Technical Report 10, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[33] G. Maletti and B. Ersbøll. A combined alignment and registration scheme of psoriasis lesions images. Technical Report 9, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[34] G. Maletti and B. Ersbøll. A hierarchical classification scheme of psoriasis images. Technical Report 6, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[35] G. Maletti and B. Ersbøll. Illumination correction in psoriasis lesions images. Technical Report 7, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[36] G. Maletti and B. Ersbøll. Principal component analysis of psoriasis lesions images. Technical Report 5, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[37] G. Maletti and B. Ersbøll. Texture alteration detection in bitemporal images of lesions with psoriasis. Technical Report 8, Department of Informatics and Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark., March 2003.

[38] G. Maletti, B. Ersbøll, and K. Conradsen. A contextual classifier that only requires one prototype pixel for each class. *IEEE Transactions on Nuclear Science*, 49(3):700–706, June 2002.

[39] G. Maletti, B. Ersbøll, and K. Conradsen. Iterative extended mean shift algorithm. 4^{th} *International Conference on Computer Vision, Pattern Recognition & Image Processing, 2002 (C.V.P.R.I.P. 2002), part of the* 6^{th} *Joint Conference on Information Sciences (J.C.I.S.)*, pages 793–796, 8–14 March 2002. Durham, North Carolina, USA.

[40] G. Maletti, B. Ersbøll, and K. Conradsen. Mean shift detection using active learning in dermatological images. *The International Society of Optical Engineering International Symposium Medical Imaging 2002 (M.I.), Conference on Image Processing*, pages 1790–1797, 23–28 February 2002. San Diego, California, USA.

[41] G. Maletti, B. Ersbøll, K. Conradsen, and J. Lira. An initial training set generation scheme. *Proceedings of the* 11^{th} *Scandinavian Conference on Image Analysis.*, pages 706–714, 2001.

[42] M. McGuire and H. Stone. Techniques for multiresolution image registration in the presence of occlusions. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1476–1479, May 2000.

[43] M. Møller. Supervised learning on large redundant training sets. *International Journal of Neural Systems*, 4(1):15–25, 1993.

[44] A. Nielsen. *Analysis of Regularly and Irregularly sampled spatial, multivariate, and muti-temporal data*. PhD thesis, Institute of Mathematical Modelling. Technical University of Denmark., Kgs. Lyngby. Denmark, October 1994. No. 6.

[45] A. Nielsen, K. Conradsen, and J. Simpson. Multivariate alteration detection (m.a.d.) and m.a.f. postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Environ.*, 64:1–19, 1998.

[46] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* Mc. Graw Hill, Series in Systems Science, 1965.

[47] J. Röning and M. Riech. Registration of nevi in successive skin images for early detection of melanoma. 14^{th} *International Conference on Pattern Recognition*, 1:352–357, 1998.

[48] G. Rosenfield and K. Fitzpatrick-Lins. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2):223–227, February 1986.

[49] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Skin detection in video under changing illumination conditions. 15^{th} *International Conference on Pattern Recognition*, 1:839–842, 2000.

[50] J. Sowa. Peircean fundations for a theory of context. *Lecture Notes in Computer Science*, 1257:41–64, 1997.

[51] M. Stegmann. Active appearance models: Theory, extensions and cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2000.

[52] S. Stehman. Estimating standard errors of accuracy assessment statistics under cluster sampling. *Remote Sensing of Environment*, 60:258–269, 1997.

[53] J. Theeuwes and F. Kooi. Parallel search for a conjunction of contast polarity. *Vision Res.*, 34(22):3013–3016, 1994.

[54] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[55] A. Treisman and H. Schimdt. Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14:107–141, 1982.

[56] G. Vetter, M. Stadler, and J. Haynes. Phase transitions in learning. *The Journal of Mind and Behavior*, 8(2, 3):335 [233]–350 [248], 1997.

[57] Y. Wang and T. Adali. *Signal Processing for Magnetic Resonance Imaging and Spectroscopy. Stocastic Model Based Image Analysis*, chapter 14, pages 1–34. 2000. (Topics in Biomedical Imaging Course Notes.).

[58] T. Watkin and J. Nadal. Optimal unsupervised learning. *J. Phys. A: Math. Gen.*, 27m:1899–1915, 1994.

[59] J. Wolfe. Guided search 2.0: The upgrade. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, pages 1295–1299, 1993.