# MULTI-SET MULTI-TEMPORAL CANONICAL ANALYSIS OF PSORIASIS IMAGES

$D.D.Gomez^1, G.Maletti^2$

$A.A.Nielsen^1, B.Ersboll^1$

[1]Technical University of Denmark
Informatics and Mathematical Modelling
DK-2800 Kgs. Lyngby, Denmark

[2]The Royal Veterinary
and Agricultural University
DK-2630 Taastrup, Denmark

## Abstract

*Nowadays, the medical tracking of dermatological diseases is imprecise, mainly due to the lack of suitable objective methods to evaluate the lesion. The severity of the disease is currently scored by doctors merely by means of visual examination. In this work, multi-set canonical correlation analysis over registered images is proposed to track the evolution of the disease automatically. This method transforms the original images into sets of variables that exhibit decreasing degree of similarity, based on correlation measures. Due to this property, these new variables are more suitable to detect where changes occur. An experiment with 5 different time series collected from psoriasis patients during 4 different sessions is conducted. The analysis of the obtained results points out some patterns that can be used both to interpret and summarize the evolution of the lesion and to achieve a better image registration.*

*keywords: multiset canonical correlation analysis, image registration, psoriasis*

## 1. INTRODUCTION

One of the main problems in the treatment of dermatological diseases is the difficulty to track the evolution of the disease. Doctors receive the visit of their patients several times to control its evolution. However, due to the fact there are not objective methods to summarize the lesion, doctors make scorings and take notes which they can remember the actual condition of the patient in a next visit. The drawbacks of this doctors dependency notes can be easily found. Different doctors can have different perceptions of the same patient. This implies that if one patient has to be treated by another doctor, the notes of the previous doctor are useless for the second and the case history is lost.

With this work, multiset canonical correlation [1] is proposed as a tool to detect where the changes are produced. This statistical method has been applied successfully in the analysis of multi temporal remote sensing data[2]. With this work , its applicability to dermatological diseases is analyzed. The study is carried out patients with psoriasis[3]. A disease that which consists of red thick areas with silver scales that affect about between 1% and 2% of the population of the United States and United Kingdom.

Results of the study show many interesting features that convey the method as tool to detect the changes in dermatological diseases. Furthermore, the results points out to that the study can also be extended to improve the registration.

## 2 MULTISET CANONICAL CORRELATION ANALYSIS

Multiset Canonical Correlation Analysis (MCCA)[1] emerged as a natural extension of the theory of canonical correlation analysis developed by Hotelling[4]. Given n sets of random variables $X_1, X_2, ..., X_n$, with dimensions $m_1, m_2, ...., m_n$ ( $m_1 \leq m_2 \leq .... \leq m_n$), initially, MCCA searches for variables $U^T = [U_1, U_2, ..., U_n]$ given by:

$$\begin{aligned} U_1 &= a_1^T X_1, V\{U_1\} = a_1^T \Sigma_{11} a_1 \\ U_2 &= a_2^T X_2, V\{U_2\} = a_2^T \Sigma_{22} a_2 \\ &\cdots \quad \cdots \quad \cdots \\ U_n &= a_n^T X_n, V\{U_n\} = a_n^T \Sigma_{nn} a_n \end{aligned} \tag{1}$$

with dispersion matrix:

$$\mathbf{\Sigma_U} = \begin{bmatrix} a_1^T \Sigma_{11} a_1 & a_1^T \Sigma_{12} a_2 & \dots & a_1^T \Sigma_{1n} a_n \\ a_2^T \Sigma_{21} a_1 & a_2^T \Sigma_{22} a_2 & \dots & a_2^T \Sigma_{2n} a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T \Sigma_{n1} a_1 & a_n^T \Sigma_{n2} a_2 & \dots & a_n^T \Sigma_{nn} a_n \end{bmatrix}$$

where $\Sigma_{ij} = Cov(X_i, X_j)$ and $Cov(U_i, U_j) = a_i^T \Sigma_{ij} a_j$. These variables $U_1, ..., U_n$ are chosen to exhibit a high measure of similarity between them in terms of covariances. The measure of similarity is, generally, determined by optimizing one of the next criteria based on $\Sigma_U$:

1. To maximize the sum of its elements:
   $max\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_i^T\Sigma_{ij}a_j\right)\quad(SUMCOR)$[1]
2. To maximize sum of squared element:
   $max\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i^T\Sigma_{ij}a_j)^2\right)\quad(SSQCOR)$
3. To maximize its largest eigenvalue. (MAXVAR)
4. To minimize its smallest eigenvalue. (MINVAR)
5. To minimize determinant:
   $min\,(det\Sigma_U)\quad(GENVAR)$

When $n = 2$, if the correlation matrix is used instead of the covariance matrix, these measures are equivalent to the traditional canonical correlation analysis.
Usually, a constraint is applied to these measurement with the goal of getting more interpretability. Two constraints widely used are given by:

1. the projection vectors are unit vectors within each set:
   $a_i^T a_i = 1$ *(AA)*
2. the variance of the new variables is equal to 1
   $V(U_i) = a_i^T\Sigma_{ii}a_i = 1$ *(ACA)*

The variables obtained by optimizing one of the previous criteria[5] are called the *first stage* or *first order canonical variables*. Kettering[1], suggested an extension of this first order set to higher order canonical sets. If $U^1, ..., U^{s-1}$, are the first s-1 canonical set, one simple way to obtain the $U^s$ canonical set is to optimize the criteria described previously subject to the restriction

$$corr\{U_j^i, U_j^s\} = 0 \quad (i = 1, ..., s-1; j = 1, ..., m).$$

In this work, the most interesting features are found in the first order canonical set. If we have two images that are more similar their difference in the first canonical set is minimal. The higher order canonical sets will only be displayed for a better perspective.

# 3. EXPERIMENTAL RESULTS

With the goal of showing the suitability of using MCCA to detect changes in dermatological diseases, a study is performed. The data, obtained in collaboration with the dermatological department of Gentofte Hospital in Denmark, consists on 5 temporal series of psoriasis images. These series were collected from three different patients in 4 sessions at different times. In each session, 5 images were taken with small displacements. The series are labelled with 1A, 1B, 1C, 2A, 3C[2], where the number indicates the patient and the letter makes reference at the area where the lesion is found. Due to difficulties with the schedule of patient number one, the last session of area C could not be collected. In order to be sure that the images are suitable for the study, Videometer lab, a high technology equipment for image acquisition[3][6] was utilized. This equipment assures to maintain the light and geometrical conditions during the different sessions allowing that images taken in different times can be compared.

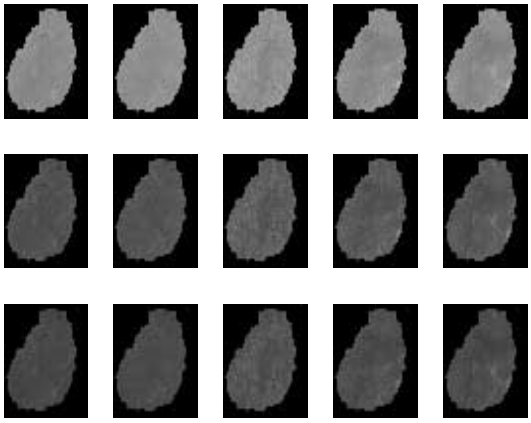

Figure 1: Aligned and registered images of the series 1C. Rows 1-3 : sessions 1-3

With the objective of establishing a correspondence between the pixels in the different images, which allows to conduct the study properly, the series were aligned and registered[7]. It was observed that the last session of the series 1B presented a poor alignment. This was due to the fact that the region covered by the images in this session differs considerably from the previous sessions. The mentioned session was excluded from the analysis to avoid the negative effects that this misalignment may cause in the results. This makes that the data used in the study consist of three series of four sessions and two series of three sessions.

The series 1C is chosen to illustrate the results. Figure 1 shows this series after the alignment and the registration. The results obtained with the other series will be display numerically with the purpose of generalization.

Each of the images is discomposed in its three chromatic bands to discover how the final results are affected by them. Thus each session is composed of 15 variables(5 images x 3bands). MCCA was applied repeatedly to the data, using each of the 5 optimization criteria with the constraints *AA* and *ACA*.

Figure 2 shows the canonical variables sets for the case 1C where sumcor and *AA* are elected as optimization criterium and constraints. It can be observed that the high degree of similarity over the different sessions is in the first

[1]The names has been chosen according to Kettering, 1971
[2]This nomenclature is chosen to have consistent with previous works[7]
[3]www.videometer.com

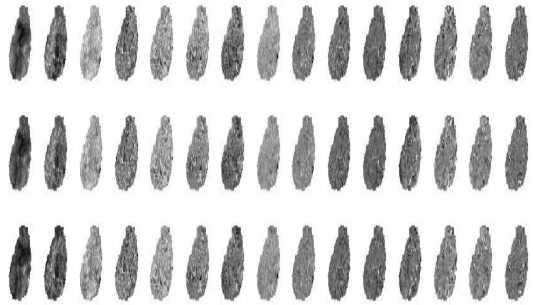

Figure 2: Canonical correlation sets for series 1C. Columns 1-15: canonical sets order 1 - canonical sets order 15. Rows 1-3: Session 1-3.

| opt.Crit / AA | (a,b) | (a,c) | (b,c) |
|---|---|---|---|
| SUMCOR | 0.953428 | 0.8673 | 0.924249 |
| SSQCOR | 0.952967 | 0.865278 | 0.922850 |
| MAXVAR | 0.953138 | 0.865857 | 0.923038 |
| MINVAR | 0.852435 | 0.747123 | 0.931232 |
| GENVAR | 0.704900 | 0.634624 | 0.828202 |
| opt.Crit / ACA | (a,b) | (a,c) | (b,c) |
| SUMCOR | 0.981693 | 0.971973 | 0.981918 |
| SSQCOR | 0.981694 | 0.971970 | 0.981919 |
| MAXVAR | 0.981693 | 0.971972 | 0.981918 |
| MINVAR | 0.949110 | 0.847497 | 0.953976 |
| GENVAR | 0.981832 | 0.971481 | 0.982050 |

Table 1: Correlations between the variables of the first order canonical set of the series 1C and for each optimization criteria and both constraints

order canonical set. The reason that this optimization criterium with this constraint is chosen for display is that exhibits many interesting features.

With the aim of having a deeper understanding of the behavior of the disease, the correlations within the sets of first canonical components were computed for all the different possibilities under study.

Table 1 shows the values of the correlations between the variables of the first canonical set for the series 1C. This table reveals a few interesting facts. Firstly, it is noticed a high correlation between the first canonical components when the criteria SUMCOR, SSQCOR and MAXVAR are used. Therefore these three criteria are preferred to detect the changes rather than MINVAR or GENVAR. The high correlation of these three methods will allow to detect where the images have less similarity and therefore where the disease exhibits more changes. It is also observed that the result of these three criteria are almost identical. So due to its simplicity and interpretability SUMCOR is preferred. Also to notice that the smallest correlation is found between the

| AA | 1A | 1B | 1C | 2A | 3C |
|---|---|---|---|---|---|
| SUMCOR | 0.8989 | 0.9817 | 0.9150 | 0.9158 | 0.7868 |
| SSQCOR | 0.8989 | 0.9817 | 0.9137 | 0.9157 | 0.7840 |
| MAXVAR | 0.8989 | 0.9817 | 0.9140 | 0.9158 | 0.7855 |
| MINVAR | 0.3151 | 0.4133 | 0.8436 | 0.4002 | 0.4722 |
| GENVAR | 0.4272 | 0.8803 | 0.7226 | 0.4043 | 0.4751 |
| ACA | 1A | 1B | 1C | 2A | 3C |
| SUMCOR | 0.9215 | 0.9855 | 0.9785 | 0.9271 | 0.8455 |
| SSQCOR | 0.9215 | 0.9855 | 0.9785 | 0.9271 | 0.8454 |
| MAXVAR | 0.9215 | 0.9855 | 0.9785 | 0.9271 | 0.8455 |
| MINVAR | 0.7681 | 0.9813 | 0.9169 | 0.4251 | 0.4991 |
| GENVAR | 0.9029 | 0.9854 | 0.9785 | 0.9147 | 0.5558 |

Table 2: Average Absolute Correlation Values for each optimization criteria and constraint

first and the last session. This makes sense because it is supposed the highest change occurs between the beginning and the end of the treatment and therefore the images are less similar.

Although the constraint *ACA* maximizes the correlation [2], the obtained correlations are not much more different to the obtained with the option *AA*. This implies that, in terms of the correlation, none of the constraint should be preferred.

General results are reported for all the cases in Table 2. Each single cell is the average absolute value of correlation within the first canonical sets of a given series. These values are computed according to:

$$\bar{\rho} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mid Corr\{Y_{i1}, Y_{j1}\} \mid$$

where n is the number of sessions and $Corr\{Y_{i1}, Y_{j1}\}$ is the correlation between the firsts canonical variate corresponding to the i-th and j-th session respectively.

These results support the previous hypothesis to use the optimization given by sumcor.

Although it outside the scope of the objective of this research, briefly to point out that the fact of the high correlations (considerably higher than the correlation between the original bands's [7] ), it converts this first order set as a suitable feature to obtain better registration [8].

Looking to establish if one of the constrain *AA* or *ACA* is preferable, the contribution of each single band is analyzed.

Table 3 shows these contributions averaged by band for our example case 1C.

It can be noticed that when the constraint *AA* is used the options we are interested gave the bands the same weight. In Table 4, it can be observed that the same pattern happens for all the series. So according to the results, the mean of the original bands could be used as a good approach to detect the changes.

| Opt.Crit:AA | $\mu_R, \sigma_R^2$ | $\mu_G, \sigma_G^2$ | $\mu_B, \sigma_B^2$ |
|---|---|---|---|
| SUMCOR | 0.201, 0.010 | 0.233, 0.032 | 0.222, 0.024 |
| SSQCOR | 0.202, 0.009 | 0.234, 0.031 | 0.224, 0.023 |
| MAXVAR | 0.202, 0.009 | 0.234, 0.031 | 0.224, 0.023 |
| MINVAR | 0.128, 0.065 | -0.154, 0.087 | 0.066, 0.030 |
| GENVAR | 0.160, 0.079 | -0.150, 0.110 | -0.011, 0.001 |
| Opt.Crit:ACA | $\mu_R, \sigma_R^2$ | $\mu_G, \sigma_G^2$ | $\mu_B, \sigma_B^2$ |
| SUMCOR | -0.398, 2.407 | 1.826, 30.72 | 0.215, 18.44 |
| SSQCOR | -0.396, 2.407 | 1.826, 30.74 | 0.213, 18.43 |
| MAXVAR | -0.397, 2.405 | 1.826, 30.72 | 0.214, 18.44 |
| MINVAR | -0.044, 5.893 | 1.629, 33.50 | -0.541, 13.00 |
| GENVAR | -0.195, 2.453 | 1.816, 33.02 | 0.025, 16.93 |

Table 3: For series 1C, mean and variance of the first canonical variate weights color band values for the different optimization criterion and constraints

| Case | $\mu_R, \sigma_R^2$ | $\mu_G, \sigma_G^2$ | $\mu_B, \sigma_B^2$ |
|---|---|---|---|
| 1A | 0.2363, 0.0001 | 0.2335, 0.0001 | 0.2342, 0.0002 |
| 1B | 0.2372, 0.0024 | 0.3058, 0.0012 | 0.3118, 0.0002 |
| 1C | 0.2011, 0.0101 | 0.2335, 0.0324 | 0.2226, 0.0245 |
| 2A | 0.2526, 0.0001 | 0.2600, 0.0002 | 0.2540, 0.0003 |
| 3A | 0.2048, 0.0030 | 0.2153, 0.0018 | 0.2238, 0.0011 |

Table 4: For all the cases, mean and variance of the corresponding first canonical variate weights color band values for the optimization criterion SUMCOR and *AA*

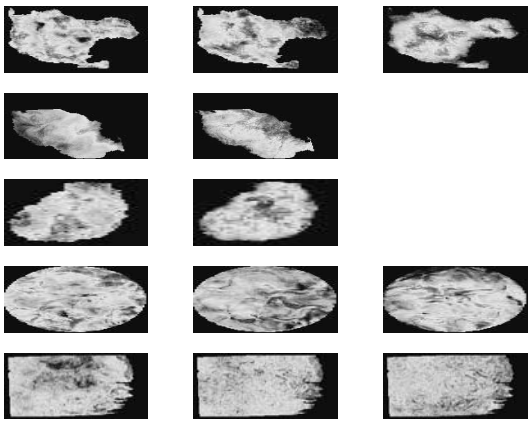

Figure 3: Changes in the sessions. column i: different between the first order canonical variable of the session i+1 and the first order canonical variable of the session i. Row 1: series 1A, Row 2: series 1B, Row 3: series 1C, Row 4: series 2A, Row 1: series 3C

# 4. SUMMARY AND CONCLUSIONS

In this paper a method to detect changes in dermatological diseases based on MCCA has been proposed. The method has been successfully applied to 5 different series of psoriasis images. The method shows where the changes are produced. In the special case of the psoriasis, it has been showed that at the beginning of the treatment, the changes happened mainly in the border of the disease with the normal skin and when the treatment is applied this changes are produced inside the lesion. Furthermore, the spectral combination that enhances the similarity of the image in different sessions in the psoriasis can be expressed through the mean of the trichromatic bands. It has also been noticed the high correlation between the variables in the first canonical set. This points out the possibility of improve the registration using these variables instead of the original variables.

# Acknowledgments

To express our gratitude to the dermatologist doctors Lone Skov and Bo Bang who made possible the development of this work.

# References

[1] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, Vol. 58, pp. 433-451, 1971.

[2] A. A. Nielsen, "Multiset Canonical Correlations Analysis and Multispectral, Truly Multitemporal remote Sensing Data," *IEEE transactions on image processing*, Vol. 11, pp. 293-305, 2002.

[3] C. Camisa, *Handbook of psoriasis,* Blackwell Science, 1998.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28, pp. 321-377, 1936.

[5] A. A. Nielsen, "Analysis of regularly and irregularly sampled spatial, multivariate and multitemporal data," *Ph.D. dissertation,*, Inst. Math.Statist.Oper.Res.,Tech.Univ.Denmark, Lyngby, Denmark, 1992.

[6] D. Delgado, B. Ersbøll and J.M Carstensen, "Building an Image-Based System to automatically Score psoriasis," *Proceedings* $13^{th}$ *Scandinavian Conference* pp. 557-564, 2003.

[7] G. Maletti and B. Ersbøll, "A combined aligment a registration scheme of psoriasis lesions images," *International Journal in Information Sciences*, accepted, 2003.

[8] J. Kim and J. A. Fessler, "Image Registration using Robust Correlation," *2002 IEEE International symposium on Biomedical Imaging*, pp. 353-356 , 2002.