Dense Iterative Contextual Pixel Classification using Kriging

Melanie Ganz

In medical applications, segmentation has become an ever more important task. One of the competitive schemes to perform such segmentation is by means of pixel classification. Simple pixel-based classification schemes can be improved by incorporating contextual label information. Various methods have been proposed to this end, e.g., iterative contextual pixel classification, iterated conditional modes, and other approaches related to Markov random fields. A problem of these methods, however, is their computational complexity, especially when dealing with high-resolution images in which relatively long range interactions may play a role. We propose a new method based on Kriging that makes it possible to include such long range interactions, while keeping the computations manageable when dealing with large medical images.

Dynamic Warping for Multiple Multivariate Sequences

Dennis Levin Herzog

Abstract: The alignment of a pair of sequences through dynamic time warping is a general approach, e.g., to compare sequences which are similar but vary in their dynamics. The alignment of multiple sequences is necessary in order to allow, e.g., accurate interpolation of human movements for model animation.
However, dynamic time warping algorithms allow only the alignment of two sequences. The extension to multiple sequences is too time consuming, computationally. Thus a rather poor approach is generally applied, where all sequences of a set of sequences are aligned one-by-one through pair-wise dynamic warping to a single (appropriate) sequence of the set. However, it is neither clear whether such an appropriate sequence exist nor how to find it.
Here, we present an Hidden-Markov-Model (HMM) based approach which overcomes this dilemma. The alignment of the whole sequence set is computed based on an HMM trained for the set. The approach scales (computationally) with the number of sequences; and handles all sequences equally. We use a special structured, and constrained HMM in order to avoid problems in the training phase of the model and in order to archive results that are comparable to those by dynamic time warping for a single sequence pair.

Object Reconstruction from Linear Cross Sections

Ojaswa Sharma

Non-destructive testing techniques are used extensively in a wide variety of fields. Acoustics is used as one of such techniques in ultrasound, echo-sounding and other fields. Acoustic waves (arranged in a specific geometry) penetrate the subject and present the inside structural details. The return signals are generally obtained as linear cross sections embedded in a higher dimensional space (2D or 3D). In order to completely understand the structure of the subject, it is important to accurately reconstruct the space between the acoustic beams. Here we present an algorithm to extract useful signal mass from the raw acoustic signals and then reconstruct the missing information between successive beams using homotopy continuation method. We show results of

smooth reconstruction in both 2D and 3D that are guaranteed to be at least C^1 continuous and topologically plausible.

Structured Light 3D Tracking System for Measuring of Movements in PET Brain Imaging

Oline V. Olesen

Patient movements are deteriorating for the image quality especially for high resolution PET scanners. A new proposal of a 3D head tracking system for high resolution PET brain imaging is setup and demonstrated. A prototype of a tracking system based on structured light with a DLP projector and CCD camera is setup on a model of the HRRT PET scanner. Methods to reconstruct a 3D point cloud of simple surfaces based on phase-shifting techniques are demonstrated. The captured images are calibrated for the non-linear projector output. The results are convincing and a first step toward a fully automated tracking system for measuring of head movements in PET imaging.

Tracing the structure of manifolds by the Auto Diffusion function

Katarzyna Gebal

The eigenvectors of the Laplace-Beltrami operator were recently used to analyze the structure of the Riemmanian manifolds. Instead of the eigenvectors directly, we propose the Auto Diffusion function (ADF) which can be expressed by the eigensolutions of the Laplace-Beltrami operator in a way that has a simple physical interpretation related to the diffusion process. If ink is injected at the point of the manifold the ADF describes how much of the ink remains at this point after some time t of the diffusion process occurring within the manifold. For a given shape, the extrema of the ADF tend to lie on the tips of features (i.e. protrusions), and its level sets and gradients tend to encircle or align with the features. ADF is controlled by a single time parameter which can be interpreted as feature scale, so features at different scales will be traced when using different t. This analysis method allows us to find feature points, to extract skeletons and to segment the manifolds.

Mesh Segmentation Using Laplacian Eigenvectors and Gaussian Mixtures

Avinash Sharma

In this work we propose a new mesh segmentation algorithm based on Laplacian embedding. We analyse the geometric properties of the Laplacian eigenvectors and we devise a practical method that combines spectral clustering with Gaussian mixtures. We attempt to characterize the projection of the mesh onto each one of its eigenvectors based on both the discrete nodal domain theory and the principal component interpretation of the eigenvectors. We devise an unsupervised probabilistic method, based on Gaussian mixture with an optimal BIC criterion, to reveal the structure of each eigenvector.

Based on this structure, we select a subset of eigenvectors among the set of eigenvectors associated with the smallest non-null eigenvalues and we embed the mesh in the isometric space spanned by this selection. The final mesh segmentation is performed via unsupervised classification based on learning a Gaussian mixture model of the embedded mesh.

Height and Tilt Geometric Texture

Vedrana Andersen

We propose a new intrinsic representation of geometric texture over triangle meshes. Our approach extends the conventional height field texture representation by incorporating displacements in the tangential direction in the form of a normal tilt. This texture representation offers a good practical compromise between functionality and simplicity: it can efficiently handle and process geometric texture too complex to be represented as a height field without having recourse to a full blown mesh editing algorithms. The height-and-tilt representation proposed here is fully intrinsic to the mesh. Texture editing and animation, such as bending or waving, can be intuitively controlled and added back to an arbitrary base mesh. We also provide simple methods for texture extraction and transfer using our height-and-field representation.

A groupwise mutual information metric for cost efficient selection of a suitable reference in cardiac computational atlas construction

Corné Hoogendoorn

Computational atlases based on nonrigid registration have found much use in the medical imaging community. To avoid bias to any single element of the training set, there are two main approaches: using a (random) subject to serve as an initial reference and posteriorly removing bias, and a true groupwise registration with a constraint of zero average transformation. Major drawbacks are the possible selection of an outlier on one side, potentially leading to problems with registration performance, and a final approximation of the average image in which the structure of interest is not at all like the population average, or on the other hand a prohibitive computational load.
We propose an inexpensive means of reference selection based on a groupwise correspondence measure which avoids the selection of outliers. Thus, it improves tractability of reference selection and robustness of automated atlas construction.
Applying the groupwise correspondence measure results in a measure of information that each subject carries about the remaining set of subjects. This creates one-dimensional scores reflecting centrality of the subjects based on image intensities. The subject manifold itself is dependent on the chosen representation, either as shapes (meshes, parameterizable surfaces or m-Reps) or as deformation fields, but we believe the intensity-based centrality relates sufficiently well to centrality on the subject manifold. Based on this, the subject with greatest intensity-based centrality is chosen as the initial reference for atlas construction.

Analysis of gait using a treadmill and a Time-of-flight camera

Rasmus R. Jensen

We present a system that analyzes human gait using a tread-mill and a Time-of-flight camera. The camera provides spatial data with local intensity measures of the scene, and data are collected over several gait cycles. These data are then used to model and analyze the gait. For each frame the spatial data and the intensity image are used to fit an articulated model to the data using a Markov random field. To solve occlusion issues the model movement is smoothened providing the missing data for the occluded parts. The created model is then cut into cycles, which are matched and through Fourier fitting a cyclic model is created. The output data are: Speed, Cadence, Step length and Range-of-motion. The described output parameters are computed with no user interaction using a setup with no requirements to neither background nor subject clothing.

Simulating Human Motion in Real-Time

Morten Engell-Nørregård

Human motion simulation in real time has two, conflicting main goals. One is the realism and precision. The other is the speed. It is not possible two reach both goals to perfection but we can find usable compromises. Previous work has focused on either one or the other, or has sacrificed generality in favor of speed or precison. We will focus on improving both objectives, while retaining sufficient generality. I will outline the avenues we have chosen to try and obtain this goal.

Human motion estimation

Søren Hauberg

In this paper, we present a novel approach to three dimensional human motion estimation from monocular video data. We employ a particle filter to perform the motion estimation. The novelty of the method lies in the choice of state space for the particle filter. Using a non-linear inverse kinematics solver allows us to perform the filtering in end-effector space. This effectively reduces the dimensionality of the state space while still allowing for the estimation of a large set of motions. Preliminary experiments with the strategy show good results compared to a full-pose tracker.

Bicycle Chain Shape Models

Stefan Sommer

In this paper we introduce landmark-based pre-shapes which allow mixing of anatomical landmarks and pseudo-landmarks, constraining consecutive pseudo-landmarks to satisfy planar equidistance relations. This defines naturally a structure of Riemannian manifold on these preshapes, with a natural action of the group of planar rotations. Orbits define the shapes. We develop a Geodesic Generalized Procrustes Analysis procedure for a sample set on such a preshape spaces and use it to compute Principal Geodesic Analysis. We demonstrate it on an elementary synthetic example as well on a dataset of manually annotated vertebra shapes from X-ray. We re-landmark them consistently and show that PGA captures the variability of the dataset better than its linear counterpart, PCA.

Fractal Dimension and Its Properties: In Application to Mammographic Breast Cancer Risk Assessment

Gopal Karemore1

Structural texture measures are used to address the aspect of breast cancer risk assessment in screening mammograms. The current study investigates whether texture properties characterized by local Fractal Dimension (FD) and Lacunarity contribute to asses breast cancer risk. It also introduces one of the novel fractal properties called succolarity. FD represents the complexity while the Lacunarity characterize the gappiness of a fractal. Our cross-sectional case-control study includes mammograms of 50 patients diagnosed with breast cancer in the subsequent 2-4 years and 50 matched controls. The longitudinal double blind placebo controlled HRT study includes 39 placebo and 36 HRT treated volunteers for two years. ROIs with same dimension (250*150 pixels) were created behind the nipple region on these radiographs. Box counting method was used to calculate the fractal dimension (FD) and the Lacunarity. Paired t-test and Pearson correlation coefficient were calculated. It was found that there were no differences between cancer and control group for FD (P=0.8) and Lacunarity (P=0.8) in cross-sectional study whereas earlier published heterogeneity examination of radiographs (BC-HER) breast cancer risk score separated groups (p=0.002). In the longitudinal study, FD decreased significantly (P<0.05) in the HRT treated population while Lacunarity remained insignificant (P=0.2). FD is negatively correlated to Lacunarity (-0.74, P<0.001), BIRADS (-0.34, P<0.001) and Percentage Density (-0.41, P<0.001). FD is invariant to the mammographic texture change from control to cancer population but marginally varying in HRT treated population. This study yields no evidence that lacunarity or FD are suitable surrogate markers of mammographic heterogeneity as they neither pick up breast cancer risk, nor show good sensitivity to HRT.

Classification in medical images using adaptive metric KNN

Chen Chen, Francois Lauze, Konstantin Chernoff, Gopal Karemore, and Mads Nielsen

Image Group, Department of Computer Science, University of Copenhagen, Denmark

ABSTRACT

This paper presents a new approach to perform classification in medical images using adaptive metric $k-$nearest neighbors (KNN) algorithm. A distance function is needed in order to identify the k nearest neighbors in feature space, and the standard Euclidean distance is commonly used.[?] Instead of using the standard Euclidean distance, we propose to use the Mahalanobis distance metric so that the structure of samples is better represented. The covariance matrix of Mahalanobis distance can be estimated in different ways. In this paper, Mahalanobis distance metrics based on three different covariance matrices are estimated for our proposed adaptive metric KNN: the empirical covariance matrix based on the data set itself, the theoretical covariance matrix based on Brownian Image Model (BIM),[?] and a novel optimized covariance matrix obtained by minimizing a smooth energy function. In order to validate this approach, a set of leave-one-out experiments have been performed on cardiovascular disease (CVD) data and mammogram data. The results show that the proposed adaptive metrics improve on the standard Euclidean one, especially for CVD data, where the empirical, theoretical and even preliminary optimized metric KNN classifiers have better areas under ROC curve (AUC) of 0.9137, 0.9023 and 0.8902 respectively, as compared to 0.8270 from the standard Euclidean one.

chen@diku.dk

Learning Graphical Model Structure with Bayesian Sparse Linear Factor Models

Ricardo Henao, Ole Winther
rhenao@binf.ku.dk, owi@imm.dtu.dk

July 24, 2009

Abstract

In recent years, sparse linear models have been a very active field of research in machine learning community and it is also well known that learning the structure of graphical models, specifically directed acyclic graphs (DAG), is rather a difficult task to solve. In this work we present an approach to learn DAG structure based on some recent developments to Bayesian sparse factor models. Assuming that the observed variables can be ordered in such way they can be represented as a DAG and that the value of each variable is a linear combination of values already taken by previous variables plus a driving signal, we can write a data vector $\mathbf{x}$ with d variables as $\mathbf{x} = \mathbf{PAPx} + \mathbf{z}$, where $\mathbf{A}$ is a strictly lower triangular weight matrix, $\mathbf{P}$ is a permutation matrix encoding the correct order of the variables and $\mathbf{z}$ is the driving signal. If $\mathbf{A}$ is square we can rewrite the problem as $\mathbf{x} = \mathbf{Bz} = \mathbf{P}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{Pz}$ and we end up with a linear factor model with two restrictions, (i) $\mathbf{B}$ must be permutable to a triangular form since $(\mathbf{I} - \mathbf{A})^{-1}$ is triangular and (ii) $\mathbf{z}$ must be non-Gaussian independent variables to ensure identifiability (up to scaling and permutations of columns, $\mathbf{P}_c$). We have developed a three-step algorithm which uses Bayesian slap and spike sparsity priors to estimate an ensemble of factor models, permutations and DAG models. It avoids combination of discrete (permutation) and continuous estimation (parameters) which are prone to local minima. In order to estimate the factor model we specify a Bayesian model where $\mathbf{B}$ has a slap and spike mixture prior to allow for sparsity, $\mathbf{z}$ has a Laplace distribution and the inference process is carried out by Gibbs sampling on the complete model hierarchy which includes a residual term. The factor model is invariant to $\mathbf{P}$ and $\mathbf{P}_c$ but we can make a stochastic search for $\mathbf{P}$ and $\mathbf{P}_c$ within the Gibbs sampling by accepting new permutations matrices according log likelihood ratios (Metropolis-Hastings) for $\mathbf{PBP}_c$ *masked to be lower triangular*. This produces a list of candidate orderings that can be used in the DAG estimation step by specifying a similar Bayesian model on $\mathbf{x} = \widehat{\mathbf{P}}\mathbf{A}\widehat{\mathbf{P}}\mathbf{x} + \mathbf{z}$, were $\widehat{\mathbf{P}}$ is a candidate ordering and $\mathbf{A}$ is strictly lower triangular with slap and spike priors. The final outcome of the algorithm is an ensemble of factor and DAG models. Model selection among these are performed using a test log likelihood. Results are presented on artificial and real data (flow cytometry measurements).We compare our model with the ICA based related method LINGAM (Linear Non-Gaussian Acyclic Model for Causal Discovery), showing that we can achieve better results with less number of examples and also that our model scales better with the number of variables.

Quantifying glossiness of yoghurt

Flemming Møller

Danisco A/S, department of Physical Food Science, Brabrand, Denmark
DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

The sensory quality of yoghurt can be altered when changing the milk composition or processing conditions. Part of the sensory quality may be assessed visually. It is described how a non-contact method for quantifying surface gloss and grains in yoghurt can be made. It was found that the standard deviation of the entire image evaluated at different scales in a Gaussian Image Pyramid was a measure for graininess of yoghurt. This methodology is used to predict glossinesand to evaluate effect of yoghurt composition.

Keywords: yoghurt, visual quality, light reflection, image quantification

1 INTRODUCTION

Graininess is a visual defect that can occur in yoghurt. It is particles or lumps of different sizes randomly distributed in the yoghurt gel. Terms like 'specks' or 'grains' have been used to describe this defect. Although the presence of these specks or grains does not affect the nutritional quality of the yoghurt, it is an aesthetic issue and therefore the problem is of concern for the dairy industry.

In literature there are several examples where yoghurt grains have been quantified after water dilution by image analysis or light scattering [1-2]. In this study light reflection from the yoghurt surface is recorded, analysed and related to production parameters and sensory data.

2 MATERIALS AND METHODS

2.1 Yoghurt production

24 different stirred yoghurt samples were produced, differing in composition (i.e. concentration of whey protein, fat and starter culter). All samples were evaluated by a trained sensory panel and by the equipment described in the next section.

2.1 Set-up

The method presented is an imitation of how visual inspection is typically done in the food industry, i.e. an inspector looks at the surface and records how the light is reflected. The physical set-up comprises light from 6 LEDs and a standard grayscale camera. Figure 1 illustrates how light from one LED is reflected from a surface and detected by a camera. Smooth and very shiny yoghurt acts as a perfect mirror, resulting in a sharp image of the LED whereas grainy yoghurt is a "poor mirror" and hence results in a very diffuse image, see Fig. 2.
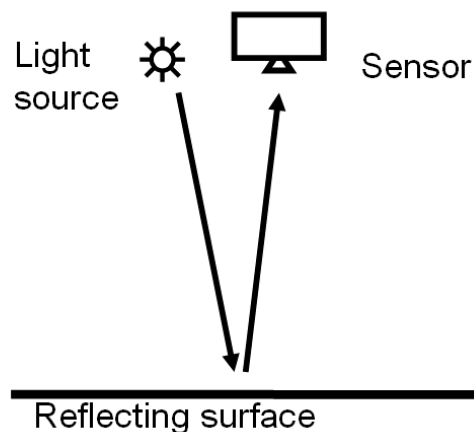

Figure 1 Drawing of the experimental set-up. The light source is LEDs and the sensor a standard camera.

Shiny yoghurt has reflection from one or very few points (Fig. 2, left). The grainy product results in an image with reflection from many smaller points (Fig. 2, right). A commercial version of the instrument is available from www.videometer.com
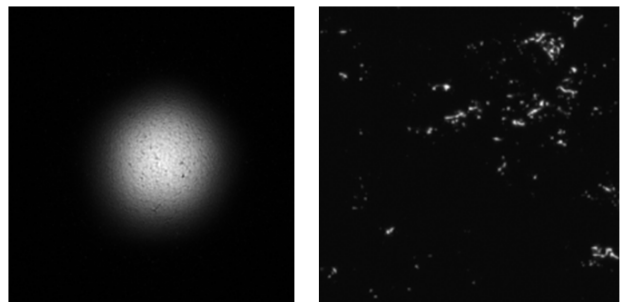

Figure 2 Typical reflection images. A smooth product with no grains and good reflection (left) and a very grainy yoghurt (right).

2.2 Image analysis

Different approaches for quantification of surface reflection images were evaluated. One of the most robust methods for predicting yoghurt gloss turned

out to be a Gaussian pyramid [4] and GLCM correlation values [5]. The original image was filtered with a Gaussian smoothing filter and subsampled to create the next level (or scale) in the pyramid. This may continue until a number of image scales have been constructed. A number of statistical and GLCM parameters were calculated at each pyramid level. Figure 3 shows how different fat levels of yoghurt influences the GLCM correlation parameter as a function of scale.
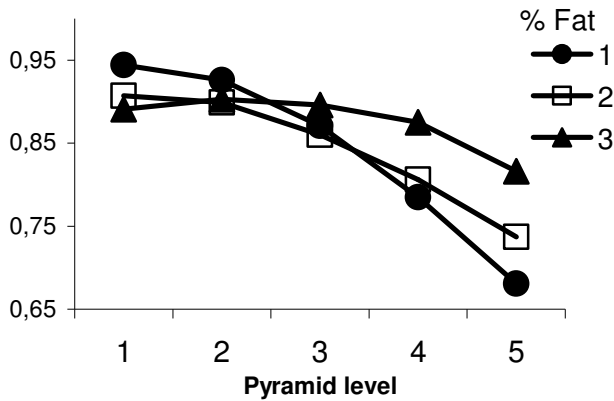

Figure 3 Standard deviation as a function of image pyramid scale. Figure 2 shows the corresponding smooth and grainy yoghurt images.

2.3 Data analysis

Partial Least Squares regression (PLS1) was made in Unscrambler 9.2 (CAMO PROCESS AS, Norway). Full cross-validation (leave one out) was used. By the minimising RMSEP was the correlation parameter found to give the best prediction of yoghurt glossiness. 50-50 MANOVA was used to test for any significance of the production parameters [6].

3 RESULTS

Data from reflection images showed a high correlation to sensory parameters such as glossy, see Fig. 4.

50-50 MANOVA indicated that several of the production parameters had a significant effect on the surface reflection of the evaluated stirred yoghurts, see Tab. 1.
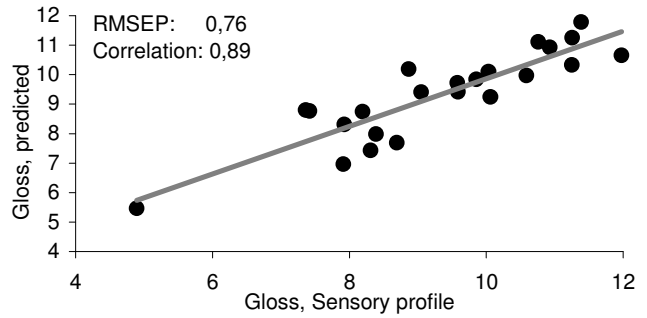

Figure 4 The predicted versus measure values for 'gritty' based on a PLS1-model using normalized standard deviation values from an Gaussian Image pyramid (r^2=0,89).

Table 1 The effect of composition parameters on the surface reflection.

Parameter	p-value
Whey protein Conc.	0.001369
Fat	0.001406
Starter	0.128756
WPC * Fat	0.677213
WPC * Starter	0.859970
Fat * Starter	0.555556

REFERENCES

[1] Sodini, I. et al.: Effect of Milk Base and Starter Culture on Acidification, Texture, and Probiotic Cell Counts in Fermented Milk Processing, Journal of Dairy Science, 85, pp. 2479-2488 (2002)

[2] Kailasapathy, K. & Supriadi, D.: Effect of partially replacing skim milk powder with whey protein concentrate on the sensory qualities of lactose hydrolysed acidophilus yoghurt, Milchwissenschaft, 53, pp. 385-389 (1998)

[3] Nayar, SK et al.: Surface Reflection: Physical and Geometrical Perspectives, tech. report CMU-RI-TR-89-07, Robotics Institute, Carnegie Mellon University, March, 1989

[4] Adelson, EH el al.: Pyramid methods in image processing, RCA Engineer , 29-6, pp. 33-41 (1984).

[5] Haralick et al.: Textural Features for Image Classification, IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, pp. 610-621 (1973)

[6] Langsrud, Ø.: 50-50 multivariate analysis of variance for collinear responses. J. R. Stat. Soc. Ser. D 51:305–317. (2002)

Non-Parametric Image Regression in a Log-Euclidean Framework

Christof Seiler[1], Xavier Pennec[2], Mauricio Reyes[1]

[1]University of Bern,
Institute for Surgical Technology and Biomechanics,
Stauffacherstrasse 78, 3014 Bern, Switzerland

[2]INRIA, Sophia Antipolis, France

July 24, 2009

1 Introduction

In this abstract a non-parametric approach to anatomical image regression based on one prediction variable is presented. We predict CT images of the femur given the age of a patient by calculating a weighted average of 193 femur CT images. The calculations are done in a Log-Euclidean framework to preserve topology and prevent folding. This implies that the transformations, in that case vector fields, are diffeomorphic even after statistical computations. In [2] the authors showed kernel regression formulated with Fréchet weighted means to take into account the non-Euclidean nature of medical images and applied it to images of the brain. In contrast, in this work we formulate kernel regression in a Euclidean way in a Log-Euclidean framework. This simplifies and speeds-up the process significantly and still considers the non-Euclidean nature of the manifold-valued data. To find the optimal kernel bandwidth we perform cross-validation.

2 Methods

To setup correspondences between anatomical images, a set of images are registered to a reference. We use the novel symmetric diffeomorphic registration approach described in [5]. What is new in this registration framework is the efficient optimization in the log-domain. As a consequence, the results of the registration are so-called velocity fields. Velocity fields can be looked at as generators for diffeomorphic deformations. Applying the Log-Euclidean framework [1] on these fields allows us to compute statistics, e.g. averages, and still preserve diffeomorphism. In the Log-Euclidean framework, velocity fields are reg-

1

ular elements in a vector space; this allows us to use simple Euclidean statistics instead of complicated non-linear techniques, which we needed when working in the space of diffeomorphic transformations. To map resulting velocity fields into diffeomorphic transformations the exponential is calculated. To go from diffeomorphic transformation back to velocity fields, a logarithmic mapping is performed. For a detailed survey of the methodology we refer to [4]. Our kernel regression function is

$$\hat{m}_\sigma(x) = \exp\left(\frac{\sum_{i=1}^{N} K_\sigma(x - x_i)v_i}{\sum_{i=1}^{N} K_\sigma(x - x_i)}\right) \tag{1}$$

where K_σ, exp and v_i are Gaussian kernel function with σ bandwidth, mapping from velocity fields to diffeomorphic deformations and ith velocity field, respectively.

We use the kernel regression method to predict images based on prediction variables, in this case the age of a person. The quality of kernel regression methods strongly depends on the selection of bandwidth parameters. To select a bandwidth parameter we apply cross-validation with penalty functions. The penalty and corresponding weighting functions penalize very small bandwidth values. Bandwidth values equal to zero are not interesting because they are just a nearest neighbor interpolation of the data. In our case we solve the following minimization problem:

$$\hat{\sigma} = \operatorname*{argmin}_{\sigma \in \mathbb{R}} \sum_{i=1}^{N} ||\log(\hat{m}_\sigma(x_i)) - v_i||^2 \Xi(W_{\sigma,i}(x_i)), \tag{2}$$

where N, log, σ, Ξ are total number of images, mapping from diffeomorphic deformations to velocity fields, bandwidth and penalty function, respectively, and

$$W_{\sigma,i}(x_i) = \frac{K(0)}{\sum_{j=1}^{N} K(\sigma^{-1}(x_i - x_j))}, \tag{3}$$

is the weighting function. For details we refer to [3].

3 Results

In Fig. 1, the cross-validation results for selecting the optimal bandwidth σ are shown. We calculated the cost function with five different penalty functions: (Generalized) cross-validation, Shibata's model selector, Akaike's information criterion, Akaike's finite prediction error and Rice's T. The residual sum of squares (RSS) correspond to the cost function being minimized in (2).

In Fig. 2, two regressed images using $\sigma = 3.0$, show differences between a 50 and 80 years old bone.
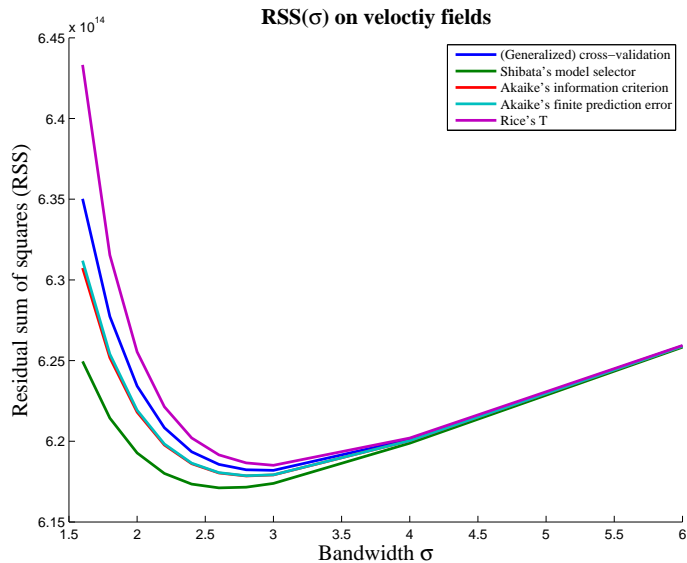
Figure 1: Cross validation graph for bandwidth selection. Five penalty functions were calculated: (Generalized) cross-validation, Shibata's model selector, Akaike's information criterion, Akaike's finite prediction error and Rice's T. The residual sum of squares (RSS) corresponds to the cost function being minimized in (2).

4 Discussion

Comparing the 50 and 80 year old bone indicates some minor changes with respect to age. Firstly, the caput collum diaphysis angle decreases. Secondly, the cortical shell decreases in volume.

With only one predicting variable for the regression we have a reasonable amount of samples to make predictions. Increasing the dimension of the predictors will introduce the need of a much bigger sampling size. So, it seems that for higher dimensional predictions, e.g. predicting images based on age, weight and height, the problem needs to be tackled with a parametric approach. We are currently working on a parametric regression model.

References

[1] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A Log-Euclidean Framework for Statistics on Diffeomorphisms. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, pages 924–931, 2006.

Figure 2: Age regression: (a) 50 years and (b) 80 years old bone.

[2] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi. Population Shape Regression from Random Design Data. In *IEEE 11th International Conference on Computer Vision - ICCV*, pages 1–7, 2007.

[3] Wolfgang Härdle. *Applied Nonparametric Regression (Econometric Society Monographs)*. Cambridge University Press, January 1992.

[4] Xavier Pennec. Statistical Computing on Manifolds: From Riemannian Geometry to Computational Anatomy. In *Emerging Trends in Visual Computing*, volume 5416 of *LNCS*, pages 347–386. Springer, 2008.

[5] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, pages 754–761, 2008.

Creating point correspondence for UCLP children undergoing surgery

S. S. Thorup[1,2], T. A. Darvann[2], N. V. Hermann[2,3], P. Larsen[2], R. R. Paulsen[1,2], Alex A. Kane[4], Daniel Govier[4], Lun-Jou Lo[5], S. Kreiborg[2,3], R. Larsen[1]

1) DTU Informatics, Technical University of Denmark, Lyngby, Denmark
2) 3D Craniofacial Image Research Laboratory, (School of Dentistry, University of Copenhagen; Copenhagen University Hospital Rigshospitalet; DTU Informatics), Copenhagen, Denmark
3) Pediatric Dentistry and Clinical Genetics, School of Dentistry, University of Copenhagen, Copenhagen, Denmark
4) Department of Plastic and Reconstructive Surgery, Washington University School of Medicine, St. Louis, MO, USA
5) Department of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital, Taipei, Taiwan

Introduction and Aim

Cleft Lip and/or Palate (CLP) is the most common congenital craniofacial malformation and is caused by lack of fusion of facial processes in the young fetus leading to clefting that may be either uni- or bilateral. CLP is treated by surgical closure of the lip and palate soon after birth. However, a major problem in habilitation is still asymmetry of the constructed nose and upper lip. The aim of the study was to investigate registration problems occurring when registering subjects with Unilateral Cleft Lip and Palate (UCLP) as this may be problematic due to the large morphological changes from before to after surgery. Successful, automatic, non-rigid, volumetric registration would allow population studies of head shape as well as detailed evaluation of surgical outcome.

Material

The material consisted of 3-dimensional CT scans of 23 Taiwanese infants with UCLP. The infants were scanned before lip repair at the age of 3 months, and again after lip repair at the age of 12 months. In UCLP, the cleft can occur on either the left or right side of the face, and in order to increase the effective sample size *n*, the right sided clefts were mirrored prior to analysis.

Method

Figure 1 illustrates problems occurring in the cleft region when non-rigid image registration (NRIR) is used for registration between before- and after-surgery volumes. The optimal transformation F needed is indicated schematically in the lower part of Figure 2 but is not possible due to too large local deformations. Instead, it is proposed to bridge the large change between the before- and after-surgery-atlases by a non-rigid deformation field *T* (upper part of Figure 2) determined by thin-plate-splines (TPS) using manually placed landmarks. The method is motivated by the observation that a point distribution model containing landmarks from both ages is distinctly bi-modal, making it probable that *T* will be a good initial guess for the deformation field for *any* UCLP lip surgery.

Algorithm 1 (Figure 1):
1) Register an individual (before surgery) to the before-surgery-atlas using non-rigid image registration. The result is a deformed individual resembling the atlas; deformation field *A*.
2) Deform the result from 1) to the after-surgery-atlas using landmark-based (Figure 2) TPS deformation; deformation field *T*.
3) Register the result from 2) to the individual after surgery using non-rigid image registration; deformation field *B*.
4) Compute the final deformation field *F* for the individual by summing the three computed deformation fields:
$$F = A + T + B.$$

(a) Original before-surgery atlas

(b) After-surgery-atlas deformed to resemble the before-surgery-atlas.

(c) Original after-surgery atlas

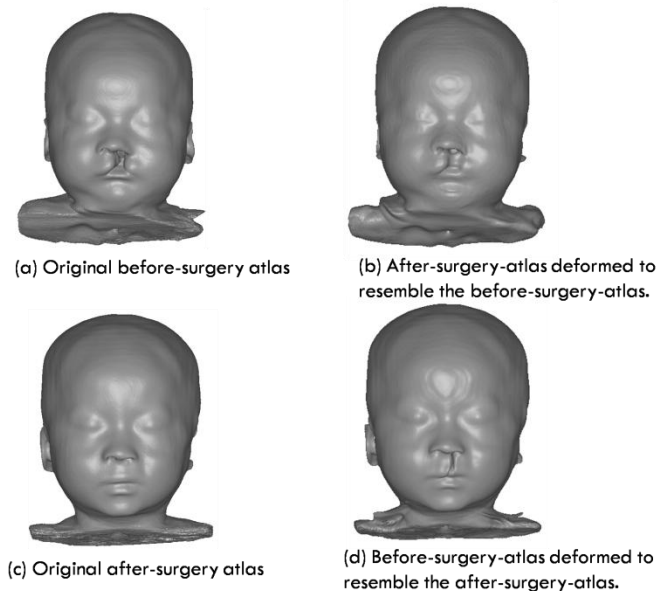(d) Before-surgery-atlas deformed to resemble the after-surgery-atlas.

Figure 1 - Before- and after lip surgery atlases (n = 23) (a and c, respectively), and the problems encountered in the cleft region when registering the atlases to each other using NRIR (b and d). Similar problems occur when registering single individuals.

3 months | 12 months

atlas → (T) → TPS deformed atlas

(A) ↑   (B) ↓
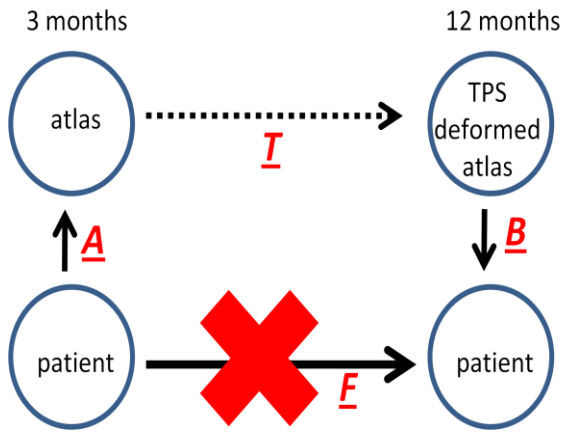
patient → ✗ (F) → patient

Figure 2 - All arrows indicate NRIR, except the dashed arrow which represents the TPS. The proof-of-principle is to create correspondence between the 3-month and 12-month atlases. Future work aims at involving all arrows except the one with the cross to achieve the best performance.

Results

A proof-of-principle was obtained by applying the deformation field T to the before-surgery-atlas (Figure 3a) and subsequently deforming the result (Figure 3c) to the after-surgery atlas (Figure 1c). A validation was carried out by computing the difference (closest distances) between the result (Figure 3c) and the after-surgery-atlas (Figure 1c).
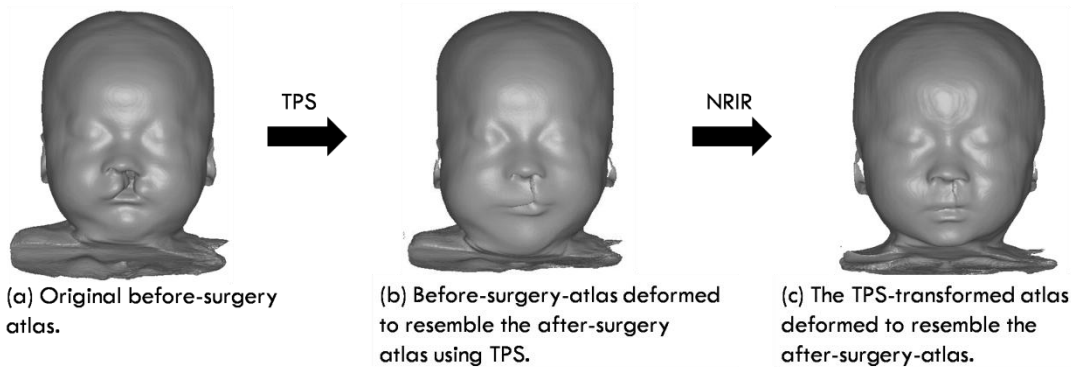

(a) Original before-surgery atlas.

TPS →

(b) Before-surgery-atlas deformed to resemble the after-surgery atlas using TPS.

NRIR →

(c) The TPS-transformed atlas deformed to resemble the after-surgery-atlas.

Figure 3 - The TPS brings the cleft to an almost closure ((a) to (b)), bringing the anatomy into the capture range of the NRIR. The overall head shape is also affected by the TPS (b). Adding non-rigid image registration (NRIR) afterwards improves the cleft closure, as well as the overall head shape (c).

Discussion and Conclusion

The results indicate that the use of the proposed method (Algorithm 1) will enable automatic determination of the deformation field due to lip surgery and growth in (unseen) individuals. Closest point differences (errors) between the original atlas (Figure 1c) and the TPS-transformed (Figure 3c) were on average small (0.1 and 0.3 mm for soft tissue facial surface and bony skull surface, respectively). Relatively large errors were seen close to the cleft in the bony skull surface, but this may be explained by the fact that only soft tissue landmarks were used for the TPS deformation.

A similar validation study was carried out on the deformed surfaces from Figure 1 i.e. a) was compared to b) and c) compared to d) by closest difference. Errors were on average larger than the ones found when using Algorithm 1 (For Figure 1 (a) - (b) the errors were 1.36 and 1.19 mm and for (c) - (d) the errors were 0.23 and 0.52 mm for soft tissue facial surface and bony skull surface, respectively). This indicates that Algorithm 1 improves the registration.

Future work will include validating the method using single patients in a leave-one-out fashion. Hence, all transformations in Figure 2 will be evaluated.

References

[Ólafsdóttir et al., 2007]   Ólafsdóttir H, Darvann TA, Hermann NV, Oubel E, Erbøll BK, Frangi AF, Larsen P, Perlyn CA, Morriss-Kay GM, Kreiborg S. *Computational mouse atlases and their application to automatic assessment of craniofacial dysmorphology caused by the Crouzon mutation Fgfr2*. J Anat. 2007; 211(1): 37-52.

[Thorup, 2008]   Thorup SS. *Quantication of craniofacial growth in mice with craniofacial dysmorphology caused by the Crouzon mutation Fgfr2^{C342Y}*. Master's thesis. Technical University of Denmark, 2008.

Sparse but emotional decomposition of lyrics

Michael Kai Petersen, Morten Mørup, and Lars Kai Hansen

Technical University of Denmark, DTU Informatics,
Building 321, DK.2800, Kgs.Lyngby, Denmark
{mkp,mm,lkh}@imm.dtu.dk
http://www.imm.dtu.dk

Abstract. Both low-level semantics of song texts and our emotional responses can be encoded in words. A cognitive model might therefore be constructed to bottom-up define term vector distances between lyrics and affective adjectives, which top-down constrain the latent semantics according to neurophysiological dimensions that capture how we perceive the emotional context of songs. Projecting the lyrics and adjectives as vectors into a semantic space using LSA their cosine similarities can be mapped as emotions over time. And subsequently applying a three-way Tucker tensor decomposition of the derived matrices to potentially identify similarities across thousands of songs, a number of time series and affective components emerge which appear in general to form dramatic curvatures determining the structure of the lyrics.

1   Introduction

Whether it being a soundscape of structured peaks or tiny black characters lined up across a page, we rely on syntax for parsing sequences of symbols, which based on hierarchically nested structures allow us to express and share the meaning contained within the lyrics of a song or a melodic phrase. As both low-level semantics of textual fragments and our affective responses can be encoded in words, a cognitive model might therefore be constructed to bottom-up define term vector distances between lyrics and affective adjectives using LSA [1], which top-down constrain the latent semantics according to neurophysiological dimensions that capture how we perceive the emotional context of songs. Meaning, that the affective adjectives function not only as markers in the semantic space, but also represent points in an emotional plane, framed by the psychological axes of valence and arousal that have neural correlates pertaining to two distinct networks in the brain [2]. That is, the brain applies an analysis-by-synthesis approach, which infers structure from bottom-up processing of statistical regularities, that are continuously compared against stored patterns of top-down labeled gestalts [3]. And cognitively speaking our emotional reactions can in this sense be thought of as top-down labels that we consciously assign to what is perceived [4].

2   Related work

During the past decade advances in neuroimaging technologies enabling studies of brain activity have established that musical structure to a larger extent than previously thought is being processed in 'language' areas of the brain [5]. Specifically related to songs, fMRI brain imaging experiments show that neural processes involved in perception and action when covertly humming the melody or rehearsing the song text activate overlapping areas in the brain. This indicates that core elements of lyrical music appear to be treated in a fashion similar to those of language [6], which might again be supported by the electrophysical evidence of language and music competing for the same neural resources when processing syntax and semantics [7]. Recent studies of the interaction between phonology and melody indicate that "wovels sing whereas consonants speak", meaning that wovels and melodic intervals may have similar functionalities related to the generative structure of syntax in language and music, whereas consonants seem rather related to lexical distinction crucial for learning words [8]. Projecting term vector distances onto a structured representation of affective adjectives may not previously have been implemented to model emotional context of media, although a similar approach has recently been applied in fMRI neuroimaging studies, where instead distances between nouns and verbs were used to predict what parts of the brain would be activated in response to words belonging to different semantic categories [9]. These patterns were in turn similarly inferred from the statistics of word co-occurrences in a large natural language corpus, although here based on the Google data set containing one trillion n-gram counts of word sequences. Likewise the affective adjectives used as semantic markers in the proposed model appear to have neural correlates, as intonation cues processed in the auditory cortices associated with the feelings of anger, sadness, relief or joy, have been identified as distinct patterns of activation in recent fMRI brain imaging studies [10].

3   Emotional tensors

Having previously applied our LSA analysis-by-synthesis model to a small sample of lyrics [11], defining the cosine similarity between individual lines of lyrics and each the affective adjectives:

happy, funny, sexy
romantic, soft, mellow, cool
angry, aggressive
dark, melancholy, sad

these terms could be interpreted as being distributed across an emotional plane framed by the psychological dimensions of valence and arousal. Here the dimension of valence describes how pleasant something is perceived, along an axis going from positive to negative associated with words like 'happy' or 'sad', while

arousal captures the amount of involvement ranging from passive states like 'mellow' and 'sad' to active aspects of excitation as reflected in 'angry' or 'happy'.

To probe whether the semantic contours previously identified in a small sample are commonly found in lyrics, we in this paper move beyond the LSA second-order analysis of emotions over time and apply a higher-order PCA principal component analysis. In this case a three-mode factor analysis to derive similarities across 8457 lyrics selected from LyricWiki. Unfolding the two-dimensional LSA into a 3-way array, it can be decomposed into a core tensor and three matrices in the model proposed by Tucker [12]:

$$x_{ijk} \approx \mathcal{G} \times \mathbf{A} \times \mathbf{B} \times \mathbf{C} = \sum_{lmn}^{LMN} g_{lmn} a_{il} b_{jm} c_{kn} \tag{1}$$

where $\mathcal{G}$ becomes a core array that defines the strengths by which the J_N columns or vector loadings of the $\mathbf{A}^{Time \times L}(\geq 0)$, $\mathbf{B}^{Emotions \times M}$ (unconstrained), and $\mathbf{C}^{Songs \times N}(\geq 0)$ matrices interact. Or in other words the model will relate all potential linear interactions between vectors representing the three modes. Here the variables L, M and N correspond to the number of components or columns in the orthogonal factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, which could in turn be interpreted as principal components in each of the three modes [13].

Assuring that the model retains only the minimal number of components necessary for representing the data, the Tucker tensor decomposition is fitted using a sparse regression algorithm to prune excess components. A hierarchical Bayesian approach, ARD automatic relevance determination is applied to determine the amount of sparsity imposed on the core array as well as the loadings. Enabling that the relevance of different features can be determined based on hyperparameters, which define a range of variation for the underlying parameters. And to provide a sparse representation, these are modeled as the width of an exponential and Laplace prior distribution for parameters, which are non-negativity constrained and unconstrained respectively assigned to the loadings and core [14]. Assuming a higher degree of signal than noise in the data, a signal to noise ratio of 1dB has been chosen in this implementation of the sparse Tucker model. In order to optimize the likelihood function, the ARD Tucker model is applied ten times to the LSA matrices and the decomposition achieving the lowest negative logarithmic probability value based on 500 iterations is selected to provide the best fit.

4   Discussion

Applying the selected sparse ARD Tucker model to the LSA matrices (Fig.1), results in six time-series components when the lyrics are resampled to a fixed length of 32 lines. Illustrating the dramatic shape over time in a song, the six curves are in turn linked to ten combinations of emotional components retrieved from the analysis. Here the square black to white matrix, with quadrants of increasing size signifying negative or positive correlation respectively, illustrates

which affective adjectives are combined in the ten groups of emotions. The ten emotional mixtures derived from the tensor decomposition are in turn associated with only three groups of lyrics, as the sparse regression algorithm and automatic relevance determination will remove less significant components in order to assure the most parsimonious representation. While six time-series components are identified, only the third and sixth corresponding to the red and yellow curves are strongly correlated with specific groups of emotions and lyrics. These 3-way correlations between time series dramatic shapes and specific emotions identified in the three groups of songs are illustrated as saturated white squares in the three greyscale core arrays at the bottom of the figure (Fig.1).

In the leftmost greyscale core array, the vertical saturated axis signifies the dramatic curves that are coupled with the emotions along the horizontal axis in the first group of songs. Here the third time-series component or red dramatic curve, consisting of two peaks culminating halfway into the song and then subsequently fading out, is associated with the 'dark' texture characteristic of the seventh group of emotions. This predominantly 'dark' texture can clearly be made out when zooming into one of the top 20 lyrics most representative of the first group of songs, "End of the night" by The Doors. A sustained activation of 'dark' components stand out in the tenth row of the matrix, contrasted with equally strong triggering of 'soft' elements (Fig.2). Supporting that the mixture of 'dark' and 'soft' components seems to be a significant texture found in lyrics in general, similar to the 'dark' and 'soft' peaks previously identified in songs [11]. The red time series curve is also found in the third group of songs, where it stands out saturated in the third row of the rightmost core array, now coupled with 'happy, soft' and 'sad' feelings, which make up the the sixth column of emotions in the core array. And to a lesser extent associated with a 'funny' and 'cool' combination of affective components, corresponding to the ninth column in the core array (Fig.1). Saturated contrasts between positive and negative valence come out as illustrated in one of the top 20 samples, the Queen song "Funny how love is" (Fig.2), where the 'happy' and 'sad' contrasts literally appear as strongly saturated lines framing the very top and bottom rows, as well as the combination of 'funny' and to a lesser extent 'angry' peaks interleaved in rows 2 and 8 of the matrix. Indicating that the simultaneous triggering of 'happy' and 'sad' previously observed in a few samples, appears to extend beyond the individual peaks of songs and rather generalize into emotional building blocks that are frequently encountered in lyrics. In contrast the second group of songs is characterized by an ascending dramatic curve, corresponding to the sixth time-series component or yellow curve, building up forming a couple of tops before reaching a climax towards the very end of the lyrics (Fig.1). Standing out saturated in the sixth row of the core array, this curve is more narrowly restricted to the 'happy, soft' and 'sad' emotions, without the 'funny-angry' contrasts, as illustrated by one the top 20 lyrics in the second group of songs, the Bon Jovi song "Not fade away" (Fig.2), where primarily 'happy' components appear synchronized with 'sad' aspects, sustained throughout the song and culminating with an activation of 'soft-dark' textures at the very end.

5   Conclusion

Applying a Tucker 3-way tensor decomposition to LSA emotions over time matrices of 8457 lyrics, the results indicate that two time series components characterize the lyrics in general. Either a dramatic curve which reaches an emotional climax early in the song before fading out, associated with either 'dark' textures or 'happy-sad' plus 'funny-angry' contrasts. Or reversely a curve that builds up to form an ascending line culminating at the very end based on mainly 'happy-sad' contrasts. Suggesting that the contrasting pairs of affective components previously identified as semantic contour in a small sample of lyrics [11], appear as emotional building blocks to be combined with temporal structures that are in general found in lyrics.

References

1. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. Harshman, R.: *Indexing by latent semantic analysis*, Journal of the American Society for Information Science, 41, p.391–407, (1990)
2. Posner, J., Russell, J.A., Gerber, A. et al. : *The neurophysiological bases of emotion: an fMRI study of the affective circumplex using emotion-denoting words*, Human Brain Mapping, 50, p.883–895, (2009)
3. Borenstein, E., Ullman, S. : *Combined top-down / bottom-up segmentation*, IEEE Transactions on pattern anlysis and machine intelligence, 30:12, p.2109–2125, (2008)
4. Barrett, L.F. : *Solving the emotion paradox: categorization and the experience of emotion*, Personality and social psychology review,10:1, p.20-46, (2006)
5. Patel, A.D.: *Music, language and the brain*, Oxford University Press (2009)
6. Callan, D.E., Tsytsarev, V., Hanakawa, T. et al: *Song and speech: Brain regions involved with perception and covert production*, NeuroImage, 31:3, pp. 1327–1342, (2006)
7. Koelsch,S. *Neural substrates of processing syntax and semantics in music*, Current Opinion in Neurobiology, 15, pp. 207–212, (2005)
8. Kolinsky,R., Lidji, P., Peretz, I., Besson, M. Morais, J:. *Processing interactions between phonology and melody: Vowels sing but consonants speak*, Cognition, 112, pp. 1–20, (2009)
9. Mitchel T.M. Shikareva, S.V., Carlson, A. et al.: *Predicting human brain activity associated with the meanings of nouns*, Science, Volume 320:30, p.1191-1195, (2008)
10. Ethofer, T., Van De Ville, D. and Scherer, K. Vuilleumier, P. : *Decoding of emotional information in voice-sensitive cortices*, Current Biology, Volume 19:12, (2009)
11. Petersen, M.K., Hansen, L.K., Butkus, A: *Semantic contours in tracks based on emotional tags*, In: Kronland-Martinet, R. et al. (Eds.): Genesis of meaning of sound and music, LNCS 5493, pp. 45–66, Springer (2009)
12. Tucker, L.R.: *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31:3, pp. 279-311, 1966
13. Kolda, T.C., Bader, B.W.: *Tensor decompositions and applications*, SIAM review, june, 2008
14. Mørup, M., Hansen, L.K.: *Automatic relevance determination for multi-way models*, Journal of Chemometrics,DOI 10.1002/cem.1223, 2009
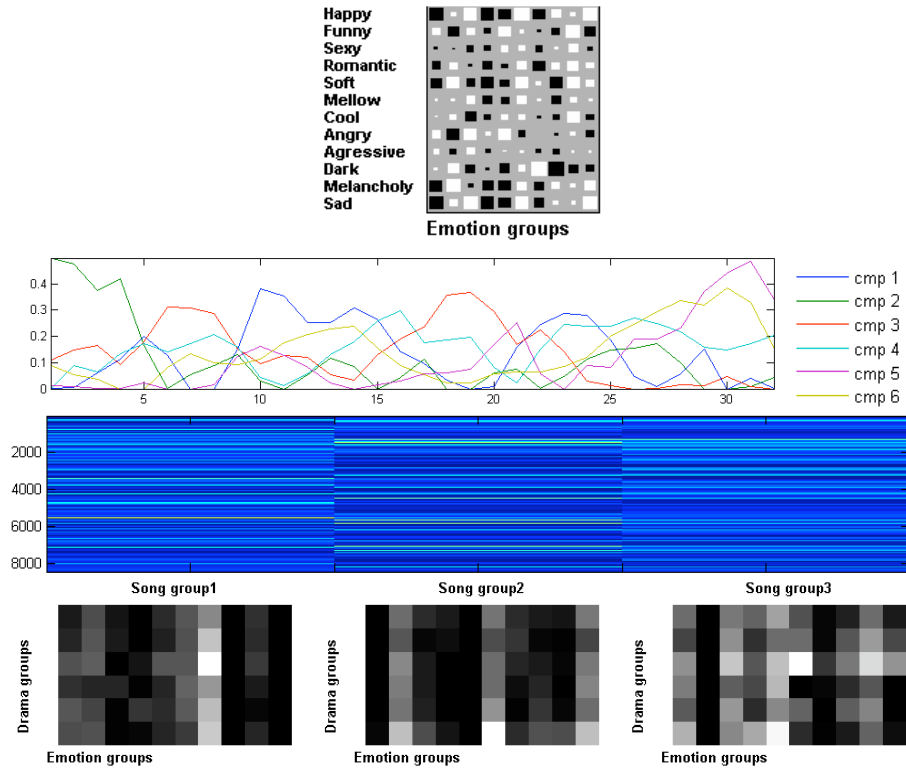
Fig. 1. TUCKER 3-way tensor decomposition of the LSA lyrics emotions over time matrices across 8457 songs, fitted to 10 combinations of emotions related to 6 time-series components represented within 3 groups of songs - based on a sparse regression algorithm pruning excess components and a hierarchical Bayesian ARD automatic relevance determination approach applied to determine the amount of sparsity.
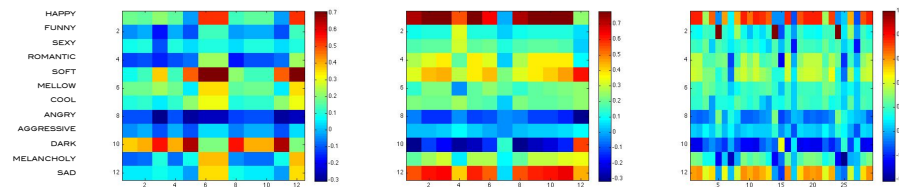

Fig. 2. Song groups 1-3 samples LSA emotions over time: "End of the night" characterized by the 'soft-dark' components that appear saturated in row 5 and 10 of the matrix, "Not fade away" characterized by mainly 'happy-sad' contrasts (rows 1 and 12), "Funny how love is" characterized by 'happy-sad' (rows 1 and 12) and 'funny-angry' contrasts (rows 2 and 8),