Multimodal Prior Appearance Models based on Regional Clustering of Intensity Profiles

François Chung and Hervé Delingette

Asclepios Research Team, INRIA Sophia-Antipolis, France
`francois.chung@inria.fr`

Abstract. Model-based image segmentation requires prior information about the appearance of a structure in the image. Instead of relying on Principal Component Analysis such as in Statistical Appearance Models, we propose a method based on a regional clustering of intensity profiles that does not rely on an accurate pointwise registration. Our method is built upon the Expectation-Maximization algorithm with regularized covariance matrices and includes spatial regularization. The number of appearance regions is determined by a novel model order selection criterion. The prior is described on a reference mesh where each vertex has a probability to belong to several intensity profile classes.

1 Introduction

Intensity profiles were among the first image representations used to describe appearance for segmentation purposes. Cootes used intensity profiles to build Statistical Appearance Models [1]. They are sampled in training images and both mean profile and its principal modes of variation are extracted for each landmark. Intensity and gradient profiles were used to optimize image forces of deformable models [2]. The idea was to better discriminate organ contours in images by comparing intensity profiles, using two generic models and checking the similarity variation with the normalized cross-correlation. A thorough study on intensity profiles can be found in [3].

Several issues may be raised with statistical appearance methods based on Principal Component Analysis (PCA). First, they require an accurate point-wise registration as the statistical analysis of appearance is performed at each point. Defining homologous points for 3D structures is difficult and therefore registering those points accurately is still considered challenging. A second limitation common to most appearance models (*e.g.* Active Appearance Models) is that they are *monomodal*, *i.e.* they rely on the hypothesis that the probability density function is well described by a single Gaussian distribution. This hypothesis is often violated by the presence of pathologies but also by the fact that shape is not necessarily correlated with appearance (see for instance livers of Fig. 2 where top regions corresponding to the lungs vary in size). Instead of having one mode with large covariance, it is preferable for image segmentation or image detection purposes to have several modes with lower covariance.

In this paper, we describe how to build a Multimodal Prior Appearance Model from a training set of P meshes. Intensity profile classes are estimated for each mesh and not for each point (*i.e.* without the need for any registration). Registration between subjects is only used to estimate the posterior probabilities on a reference mesh. Furthermore, the proposed method is fully automated and one single threshold $\mathcal{J}$ controls the number of classes. Finally, we introduce new regularization strategies of covariance matrices in the Expectation-Maximization algorithm (EM) and the *OSI* index to determine the optimal number of classes.

2 Building Multimodal Prior Appearance Models

2.1 Principles

In Fig. 1, we overview our automated method to create a Multimodal Prior Appearance Model. The input is a set of P meshes $\mathcal{M}_p$ corresponding to the segmentation of the same structure in different images. The meshes may have different number of vertices, or even different topologies. At each vertex i of $\mathcal{M}_p$, we extract M regularly sampled intensities to build an intensity profile of dimension M along the normal direction, noted $\mathbf{x}_i^p$. This profile can extend inward, outward or both sides, depending on the application. Note that intensity profiles act as feature vectors that could be replaced by any other local or global features such as isophote curvature, texture descriptors, oriented filters, etc. Changing the feature vector would only change the regularization of the covariance matrices (section 2.2).
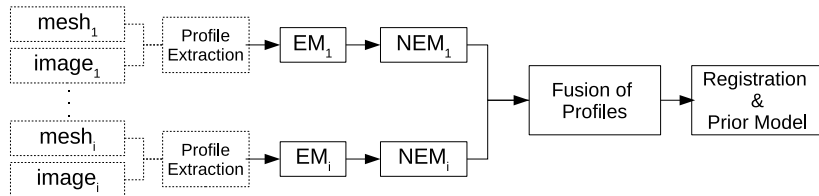

Fig. 1. Proposed pipeline for Multimodal Prior Appearance Model construction.

For each mesh, we propose to automatically cluster the profiles using an EM classification. The number of classes, a hyperparameter, is selected in an automatic fashion through a model order selection based on a new criterion (section 2.3). Classification is improved by performing spatial regularization of the posterior probabilities (section 2.4). The creation of the prior is done in two steps. First, all intensity profile classes from the P subjects are compared and classes corresponding to the same tissues are possibly merged. Finally, all P meshes are registered to the same reference mesh $\mathcal{M}^\star$ and a unique prior model is created. Each vertex i of $\mathcal{M}^\star$ is given a probability $\tilde{\gamma}_i^m$ to belong to a reference class m (section 2.5).

2.2 EM classification of intensity profiles

The classification for a given mesh $\mathcal{M}_p$ is formulated in the context of a probability density estimation using Gaussian Mixture Models. The approach is semi-parametric because the number of classes $\mathcal{K}_p$ is an unknown parameter. In the remainder, intensity profile class is denoted by *mode* and we write μ_k^p, $\mathbf{\Sigma}_k^p$, π_k^p and $^p\gamma_i^k$ for the mean, covariance matrix, mixing coefficient and posterior probabilities of class k from mesh $\mathcal{M}_p$.

Initialization EM is initialized with the Fuzzy C-Means algorithm (FCM). After convergence, FCM cluster centers and data membership values are used to initialize the EM mode means μ_k^p, posterior probabilities $^p\gamma_i^k$, mixing coefficients π_k^p and covariance matrices $\mathbf{\Sigma}_k^p$. FCM are themselves initialized with random cluster centers.

Coping with missing data Since segmented structures may be close to the image boundaries, profiles are likely to be incomplete. In order to use those incomplete profiles in the classification, a valid dimension range $M_i \leq M$ is determined for each profile i. In the E-step, the term $\exp((\mathbf{x}_i^p)^T(\mathbf{\Sigma}_k^p)^{-1}(\mathbf{x}_i^p))$ is only computed for the valid dimension range and multiplied by M/M_i. In the M-step, the mode means and covariance matrices are normalized by the total number of valid values for each index u of the profile ($1 \leq u \leq M$). FCM have also been extended in similar fashion to cope with incomplete data.

Covariance matrix regularization The EM log-likelihood maximization may lead to local maxima or degeneracy. For profiles of dimension M (typically $M \approx 10$), degeneracy of covariance matrices $\mathbf{\Sigma}_k^p$ may occur due to the coarse sampling (typically $N \approx 4000$) of this high dimensional space. We propose 3 distinct methods to regularize $\mathbf{\Sigma}_k^p$ based on a regularization parameter h ($0 \leq h \leq 1$). The first one is based on *Spectral Regularization*. The covariance matrix is diagonalized $\mathbf{\Sigma}_k^p = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$ and the h lowest eigenvalues are set to 1% of the highest eigenvalue, thus leading to a new diagonal matrix $\hat{\mathbf{\Lambda}}_h$. The inverse is then computed as $^h(\mathbf{\Sigma}_k^p)^{-1} = \mathbf{P}\hat{\mathbf{\Lambda}}_h^{-1}\mathbf{P}^T$. This can be seen as performing PCA and filtering the matrix by discarding high frequencies. In a second approach, *Diagonal Regularization*, the covariance matrix is regularized towards a diagonal matrix controlled by parameter h. The u,v element of matrix $^h\mathbf{\Sigma}_k^p$: $^h(\mathbf{\Sigma}_k^p)_{u,v} = (1-h)(\mathbf{\Sigma}_k^p)_{u,v} + h\,\delta_{u,v}\,(\mathbf{\Sigma}_k^p)_{u,v}$ ($\delta_{u,v}$ is the Kronecker symbol). The higher h, the more diagonal dominant the covariance matrix. This approach has been used in climate modeling to cope with missing values [4]. The last approach, *Constant Regularization*, regularizes the covariance matrix towards a constant matrix $\mathbf{I}d\,\mathrm{tr}(\mathbf{\Sigma}_k^p)/M$ controlled by parameter h : $^h(\mathbf{\Sigma}_k^p) = (1-h)\,(\mathbf{\Sigma}_k^p) + h\,\mathbf{I}d\,\mathrm{tr}(\mathbf{\Sigma}_k^p)/M$. With a high value of h, the covariance matrix converges towards a diagonal matrix with the same variance. The choice of a covariance matrix regularization method depends on the nature of the data. We have tested the three methods on profiles and we found that *Diagonal* or *Constant Regularization* method with $h = 0.9$ leads to the most intuitive classification results.
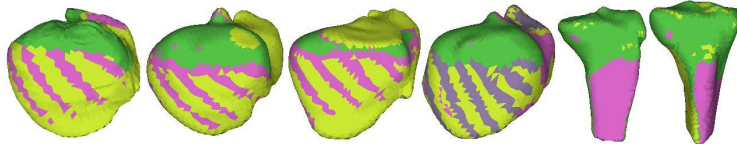
Fig. 2. EM classification of outward profiles performed on 4 livers and 2 tibias.

Table 1. Selection of the optimal number of EM modes for outward and inward profiles.

		Outward profiles					Inward profiles				
Regularization	h	OSI	FVQ	AIC	AIC_c	BIC	OSI	FVQ	AIC	AIC_c	BIC
Diagonal	0.9	4	2	3	3	2	2	2	4	4	2
Diagonal	1.0	3	2	3	3	2	2	2	4	4	2
Constant	0.9	4	2	5	5	2	2	2	6	6	2
Constant	1.0	3	2	3	3	2	2	2	5	5	2

2.3 Model Order Selection

The objective of model order selection is to find the number of modes that best represents the data without any under or overfitting. To this end, we propose to estimate a criterion measuring the quality of EM classification for a given number of modes $\mathcal{K}$. We then keep the number of modes $\mathcal{K}_{opt}$ that maximizes (or minimizes) that criterion. Several criteria based on information theory have been proposed in the literature such as the Akaike Information Criterion (AIC) and the second-order AIC (AIC_c). The often preferred EM criterion is the Bayesian Information Criterion (BIC) [5]. We also investigated FCM criteria such as cluster validity indices [6] and the Fuzzy Vector Quantization [7].

In addition, we propose a new non parametric model order selection criterion called *Overlap Separation Index* (OSI) inspired by Kim *et al.* [6]. This criterion is solely based on the posterior probabilities $^p\gamma_i^k$ and penalizes the overlap between modes while encouraging their separation. More precisely, the criterion is computed as $OSI = \frac{C_1}{C_2}$. The first term C_1 sums the amount of overlap $2\,^p\gamma_i^s/(^p\gamma_i^r + ^p\gamma_i^s)$ between the best two modes r and s for profile i (*i.e.* modes with the highest $^p\gamma_i^k$). The second term C_2 is the minimum separation between any pair of modes. The separation between pair of modes r and s is computed as the sum of $2\,^p\gamma_i^s/(^p\gamma_i^r + ^p\gamma_i^s)$ for all profile i being classified to mode r and $2\,^p\gamma_i^r/(^p\gamma_i^r + ^p\gamma_i^s)$ for all profile i being classified to mode s.

Table 1 shows the performance of the different model order selection criteria with varying regularization methods (spectral method has been discarded here). The number of modes being tested varies between 2 and 10, which takes around 15 minutes for a mesh with few thousand points. For outward profiles, the expected number of modes is at least 3 (air, bones, soft tissue) while for inward profiles 2 modes are expected (parenchyma and non-parenchyma). Based on Table 1 and further analysis on 6 other liver meshes, we found that *OSI* criterion gave the most consistent results with a limited sensitivity to regularization methods and h parameter.

2.4 Spatial Regularization

EM does not take into account the neighborhood information of profiles. This leads to non smooth posterior probability maps $^p\gamma_i^k$, which impairs the fusion of appearance regions. To account for the connectivity between profiles, we use the Neighborhood EM algorithm (NEM) [8] since it nicely extends EM and leads to efficient computation (compared to Markov Random Field). NEM is an alternate optimization of the L functional :

$$\mathcal{L}(^p\gamma_i^k, \pi_k^p, \mu_k^p, \mathbf{\Sigma}_k^p) = L(^p\gamma_i^k, \pi_k^p, \mu_k^p, \mathbf{\Sigma}_k^p) + \beta \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{N} {}^p\gamma_i^k \ {}^p\gamma_j^k \ v_{ij} \tag{1}$$

$$L(^p\gamma_i^k, \pi_k^p, \mu_k^p, \mathbf{\Sigma}_k^p) = \sum_{k=1}^{K} \sum_{i=1}^{N} {}^p\gamma_i^k \ \log(\pi_k^p G(\mathbf{x}_i^p | \mu_k^p, \mathbf{\Sigma}_k^p)) - \sum_{k=1}^{K} \sum_{i=1}^{N} {}^p\gamma_i^k \log(^p\gamma_i^k)$$

where $G(\mathbf{x}_i^p | \mu_k^p, \mathbf{\Sigma}_k^p))$ is the Gaussian probability density function.

The former term L leads to the classical EM [9] while the latter is a spatial regularization term controlled by β. The neighborhood parameter v_{ij} sets the amount of smoothing and is non-zero only if profile i is neighbor to profile j. $\mathcal{L}$ functional is minimized with an alternate optimization leading to a modified E-step where the posterior probabilities are iteratively estimated until a fixed point value is reached [8]. In our setup, profiles are extracted from 2-simplex meshes for which each vertex has only 3 neighbors [10]. Thus, v_{ij} has only three non-zero values, which substantially speeds-up the computation. In practice, less than 5 iterations are necessary to obtain stable posterior probabilities.

As neighborhood parameter v_{ij}, we choose the correlation coefficient between neighboring profiles i and j. With this choice, the spatial regularization of posterior probabilities is stronger between similar neighboring profiles, similarly to anisotropic diffusion in image processing. This prevents the blurring of tissue modes that would have occurred with a constant v_{ij} value.

The choice of the β parameter is an important issue and we set this parameter automatically by using an heuristic proposed by Dang [11]. It consists in using NEM with increasing values of β, and then detecting the β value above which the log-likelihood $\mathcal{L}$ sharply decreases (see left of Fig. 3). Indeed, too much spatial regularization leads to a significantly worse profile classification captured by the $\mathcal{L}$ functional. The proposed approach is a fully automatic way to spatially regularize posterior probabilities.

2.5 Fusion of modes from P meshes

The objective of this section is to merge the modes from the profile classification performed on P meshes into a single Multimodal Prior Appearance Model. This is done in two steps. The first step consists in comparing and merging the appearance regions extracted on the same structure for P different subjects. Profiles of each mesh $\mathcal{M}_p$ have been classified and lead to $\mathcal{K}_p$ modes ($\mathcal{K}_p$ may

Fig. 3. $\mathcal{L}$ w.r.t β (left), liver before (middle) and after (right) NEM regularization.

vary among meshes, *e.g.* due to the occurrence of pathologies). In order to have a meaningful comparison, an intensity normalization is required (*e.g.* to cope with the different nature or settings of the imaging systems). This may be done by histogram normalization or many other approaches proposed in the literature. With CT images, we found best not to perform any normalization.

In order to merge similar regions, we measure the similarity between any pair of modes $(\mu_k^p, \mathbf{\Sigma}_k^p), (\mu_l^q, \mathbf{\Sigma}_l^q)$ for $p \neq q$ by using the Jaccard index (ratio of the intersection of two sets over their union) of the region spanned by the mean and standard deviation $\mu_k^p \pm \sqrt{\sigma_k^p}$ where σ_k^p is the diagonal of the covariance matrix (see Fig. 4). A threshold $\mathcal{J}$ between 0 and 1 is used to decide whether two modes k and l are equivalent. Thus, we create a graph where nodes represent the modes and arcs link the modes found to be equivalent. The number of connected components of this graph is the number of independent modes $\mathcal{K}$. For connected components with only one node (*i.e.* without equivalence), modes are directly included in the prior with a new index m. For connected components having r equivalent nodes, we compute the mean of the new mode as the weighted sum of profile means μ_r^p with the weight $\sum_i^p \gamma_i^r$ while covariance matrices are recomputed. This computation leads to $\mathcal{K}$ independent modes $(\tilde{\mu}_m, \tilde{\mathbf{\Sigma}}_m)$ and an equivalence table $\eta(p, m)$ establishing the new index m of mode k.

An alternative to this first step could be to perform an EM classification of all profiles for all P subjects with model order selection to find the optimal number of modes. This approach would lead to a more time consuming task, which would need to be performed each time a new dataset is added. Instead, we prefer to achieve a separate clustering of each dataset followed by a merging of all modes. Another advantage of our approach is that it is not biased by the variation of mesh resolution between datasets.

The second step provides a geometric embedding for the independent modes. To this end, we register non-rigidly all P meshes $\mathcal{M}^p$ towards the same reference shape with a coarse-to-fine deformable surface approach where each mesh is registered towards a binary image with globally-constrained deformations [10]. After defining a reference mesh $\mathcal{M}^\star$ on the reference shape, each posterior probability $^p\gamma_i^k$ is resampled on $\mathcal{M}^\star$ using a closest point approach. Finally, for each vertex i of $\mathcal{M}^\star$, we compute the posterior probability $\tilde{\gamma}_i^m$ by summing and normalizing the posterior probabilities associated to each mode : $\tilde{\gamma}_i^{\eta(p,m)} += {}^p\gamma_i^m$.
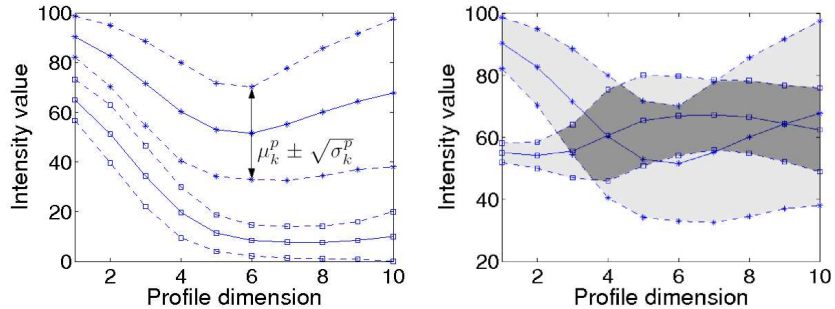
Fig. 4. Left: μ_k^p (solid lines) and $\sqrt{\sigma_k^p}$ (dashed lines) of 2 modes (asterisk and square) after EM classification of profiles (σ_k^p is the diagonal of $\mathbf{\Sigma}_k^p$). Right: Similarity between pairs of modes defined as the ratio of the intersection (dark gray) over the union (light gray) of their variance surface.

In practice, this approach leads to sparse probabilities where only a few modes have non-zero posterior probabilities (as opposed to performing an E-step based on the mode means and covariances).

3   Results

We tested our method on 7 livers segmented from CT images and 4 tibias (cropped at knee level) segmented from MR images (see Fig. 2). For both structures, outward profiles (10 samples extracted every mm) were generated from meshes with ≈ 4000 vertices. As said before, EM classification using *Diagonal* or *Constant Regularization* with $h = 0.9$ leads to the most intuitive results. The optimal number of modes were estimated with the *OSI* criterion. NEM was launched ≈ 10 times to find the optimal β (see left of Fig. 3). For the livers, an initial total number of 24 modes leads to 14 new modes after the merging of profiles with $\mathcal{J} = 0.6$. With a lower $J = 0.5$, the number of modes goes down to 6, thus providing a simple way to taylor the complexity of the prior. For tibias, an initial total number of 11 modes leads to 6 new modes with $\mathcal{J} = 0.5$. For both structures, the Multimodal Prior Appearance Model has been built from these new modes (see Fig. 5).

4   Conclusion

We propose in this paper a method that builds a Multimodal Prior Appearance Model from the regional clustering of intensity profiles. The main advantage of our approach is that modes are built without requiring an accurate pointwise registration. Another advantage is that a meaningful prior may be built with very few datasets (in fact one dataset suffices), which makes it well suited for a bootstrapping strategy. Furthermore, the prior is multimodal therefore able
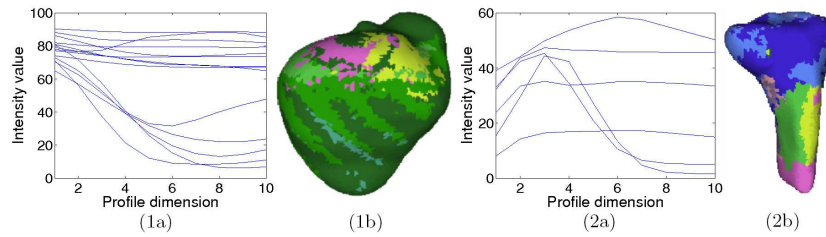
Fig. 5. Multimodal Prior Appearance Model for the livers (1) and tibias (2). For both structures, plot of the new mode means (a) and the new classification (b).

to cope with large variation of appearances including pathologies. We have also introduced the *OSI* index and included spatial anisotropic regularization of EM classification. Future work will focus on the use of this prior in image segmentation.

References

1. Cootes, T., Taylor, C.: Using grey-level models to improve active shape model search. In: Proceedings of ICPRE. Volume 1. (1994) 63–67
2. Gilles, B.: Anatomical and kinematical modelling of the musculoskeletal system from mri. Phd Thesis, University of Geneva (August 2007)
3. Heimann, T.: Statistical Shape Models for 3D Medical Image Segmentation. VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG (2009)
4. Schneider, T.: Analysis of incomplete climate data. J. Clim. **14**(5) (2001) 853–871
5. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2) (1978) 461–464
6. Kim, D.W., Lee, K.H., Lee, D.: On cluster validity index for estimation of the optimal number of fuzzy clusters. Pattern Recognition **37** (2004) 2009–2025
7. Saha, S., Bandyopadhyay, S.: A new cluster validity index based on fuzzy granulation-degranulation criterion. 15th ADCOM 2007 (2007) 353–358
8. Ambroise, C., Dang, M., Govaert, G.: Clustering of spatial data by the em algorithm. Quantitative Geology and Geostatistics **9** (1997) 493–504
9. Hathaway, R.J.: Another interpretation of the em algorithm for mixture distributions. Statistics and probability letters **4**(2) (1986) 53–56
10. Montagnat, J., Delingette, H.: Globally constrained deformable models for 3D object reconstruction. Signal Processing **71**(2) (1998) 173–186
11. Dang, M., Govaert, G.: Spatial fuzzy clustering using em and markov random fields. In: Int. Journal of System Research and Information Science. (1998) 183–202