COVARIANCE SHRINKING IN ACTIVE SHAPE MODELS WITH APPLICATION TO GYRAL LABELING OF THE CEREBRAL CORTEX

Zhen Yang,
Image Analysis and Communications Laboratory, Dept. of Electrical and Computer Engineering, The Johns Hopkins University, zyang11@jhu.edu

Aaron Carass, and
Image Analysis and Communications Laboratory, Dept. of Electrical and Computer Engineering, The Johns Hopkins University, aaron_carass@jhu.edu

Jerry. L. Prince
Image Analysis and Communications Laboratory, Dept. of Electrical and Computer Engineering, The Johns Hopkins University, prince@jhu.edu

Abstract

Active shape models (ASMs) have been widely used in segmentation tasks in medical image analysis. Complex structures and a limited number of training samples can, however, result in the failure to capture the complete range of shape variations. Various modifications to the point distribution model (PDM) have been proposed to increase the flexibility of the model. Still model parameters are often determined empirically without respect to the underlying data structure. We explore shrinkage covariance estimation in building a PDM by combining the sample covariance matrix with a target covariance matrix estimated from a low-dimensional constrained model. Instead of using a global shrinkage intensity, we apply a spatially varying shrinkage intensity field to better adapt to the spatially varying characteristic of a complex shape. The parameters of the constrained model and the amount of shrinkage are determined in a data-driven fashion, so that the resulting distribution is optimized in representing the underlying data. The PDM, which we call SC-PDM, shows an increased flexibility in fitting new shapes and at the same time, is robust to noise. We demonstrate the effectiveness of using SC-PDM to label gyral regions on the human cerebral cortex.

Keywords

Active shape model; point distribution model; covariance shrinking; cerebral cortex; gyral labeling

1. INTRODUCTION

Active shape models (ASMs) [1] have been widely used in medical image segmentation. A known difficulty is that the subspace dimensionality of a point distribution model (PDM) is limited by the training sample size, resulting in an over-constrained shape when segmenting complex structures with high inter-subject variations–see Fig. 1 for examples on the human cerebral cortex. Various approaches have been proposed to bridge the gap between the PDM subspace and the real distribution in the high-dimensional shape space. These have included introducing synthetic variations on the training samples [2], mixing the sample covariance with an artificial covariance matrix [3], relaxing the model constraint at the fitting stage [4, 5], partitioning the coordinates into local blocks in spatial or spatial-frequency domain [6, 7]. A common problem of the above work, is that model parameters are determined

empirically, and are not optimized with respect to the underlying data structure. Also, the balance between flexibility and robustness has not been evaluated in past work.

Many of the above modifications to the PDM can be viewed as covariance regularizations. For example, combining sample covariance with a constructed covariance matrix is a type of shrinkage estimation of the covariance [8, 9]. The partition approach [6] enforces independence (zeros in the covariance matrix) between local groups of variables. Crimi et al. [10] discussed several covariance regularization techniques in building PDMs and proposed a Bayesian approach using a Gaussian prior. In this work, we explore theshrinkage covariance regularization for improving the PDM, and discuss how the model parameters can be optimized in representing the underlying data.

The idea behind covariance shrinking is to combine the estimation of an unrestricted high-dimensional model, here the sample covariance $\mathbf{S}$, with the estimation from a lower dimensional restricted sub-model, $\mathbf{T}$, so that a new estimate $\mathbf{S}^* = \lambda\mathbf{T} + (1 - \lambda)\mathbf{S}$ can outperform both $\mathbf{S}$ and $\mathbf{T}$ in terms of accuracy and statistical efficiency. $\mathbf{T}$ is often called the shrinkage target and $\lambda \in [0, 1]$ is called the *shrinkage intensity*, which can be determined analytically by minimizing the mean squared error (MSE) [8, 9]. In this paper, we first design a low-dimensional covariance model according to the prior knowledge of the shape and estimate the model parameters. Instead of using a global shrinkage intensity, we compute a spatially varying shrinkage intensity field to adapt to the spatially varying nature of a complex shape. The resulting PDM, called the shrinkage covariance PDM (SC-PDM) has an increased flexibility in fitting new shapes and at the same time is robust to noise. We apply it to labeling of gyral regions on the human cerebral cortex.

2. SHRINKAGE COVARIANCE ESTIMATION

2.1. Shrinkage target

Let y_i^k be the k-th coordinate of the i-th landmark point. Let $s(y_i^k, y_j^l)$ be the element in sample covariance $\mathbf{S}$ that corresponds to the covariance between y_i^k and y_j^l. As in [9], the covariance matrix is parametrized in terms of variances and correlations, i.e., $s(y_i^k, y_j^l) = r(y_i^k, y_j^l) \sqrt{s\left(y_i^k, y_i^k\right) s\left(y_j^l, y_j^l\right)}$. The low-dimensional constrained model $\mathbf{T}(\cdot)$ for the covariance matrix should conform to the presumed low-dimensional structure in the data set, in order to avoid introducing significant bias to the estimator. Based on the observation that points close to each other are highly correlated, we model the correlation between y_i^k and y_j^l as a decreasing function of d_{ij}, the distance between the i-th and the j-th landmark point. The element $t(y_i^k, y_j^l)$ of the target covariance model $\mathbf{T}(\cdot)$ has the following form,

$$t\left(y_i^k, y_j^l\right) = \sqrt{s\left(y_i^k, y_i^k\right) s\left(y_j^l, y_j^l\right)} \frac{r\left(y_i^k, y_i^l\right) + r\left(y_j^k, y_j^l\right)}{2} \varphi\left(d_{ij}; \sigma_d\right), \quad (1)$$

where $\varphi(d_{ij}; \sigma_d)$ is the decreasing function of d_{ij} with parameter σ_d, which is Gaussian in our approach. We note that d_{ij} can be any distance measure, e.g., Euclidean or geodesic distance. Since the sample variance and correlation can be computed directly from the training data, the only parameter to be estimated for the constrained model is σ_d, which we compute by maximum likelihood.

2.2. Spatially varying shrinkage intensity

The optimal shrinkage intensity can be determined analytically. Ledoit and Wolf [8] derived a simple theorem for choosing the that minimizes the mean squared error without knowing the underlying distributions. Schafer [9] proposed, for small sample inference, to compute by replacing all expectations, variances and covariances by their unbiased sample counterparts,

$$\lambda = \frac{\Sigma_{i=1}^{p} \left[\widehat{\mathrm{Var}} \left(s_i \right) - \widehat{\mathrm{Cov}} \left(t_i, s_i \right) \right]}{\Sigma_{i=1}^{p} \left(t_i - s_i \right) 2}. \quad (2)$$

Accordingly, decreases as the variance of the sample covariance decreases, so with increasing sample size the influence of the target **T** diminishes. also depends on the correlation between **S** and **T**. If both are positively correlated then the weight put on the shrinkage target decreases. The denominator in Eqn. (2) protects the shrinkage estimate **S*** from a misspecified target **T**. In our approach, instead of using a global shrinkage intensity, we apply a shrinkage intensity field that has different values associated with each entry of the covariance matrix,

$$\lambda \left(y_i^k, y_j^l \right) = \frac{\Sigma_{i'j'} k_{ij,i'j'} \left[\widehat{\mathrm{Var}} \left(y_{i'}^k, y_{j'}^l \right) - \widehat{\mathrm{Cov}} \left(y_{i'}^k, y_{j'}^l \right) \right]}{\Sigma_{i'j'} k_{ij,i'j'} \left[t \left(y_{i'}^k, y_{j'}^l \right) - s \left(y_{i'}^k, y_{j'}^l \right) \right]^2}. \quad (3)$$

Here $k_{ij,i'j'} = (d_{ii'}; \ _d)(d_{jj'}; \ _d)$ is a distance based kernel function constructed by the same decreasing function $(d_{ij}; \ _d)$ as in Eqn. (1). In Eqn. (3), as compared to Eqn. (2), we see that each term in the numerator/denominator is weighted by the kernel function $k_{ij,i \ j}$, so that the amount of shrinkage at each entry of the covariance is computed from the statistics of spatially adjacent points. The shrinkage covariance is then computed entry-wise as a weighted sum of the corresponding entry of the sample covariance and that of the target covariance. With the spatially varying shrinkage intensity field, one can better model the correlations between a local group of points and adapt to the spatially varying nature of a complex shape.

3. EXPERIMENTAL RESULTS

3.1. Model flexibility and robustness

We tested the proposed SC-PDM in approximating shapes with known point correspondences, for both noise-free and noisy cases, and compared with 1) the original PDM, 2) a PDM combined with a smoothness model (SMT-PDM) [3], and 3) a PDM shrunk to the proposed constrained model with a global shrinkage intensity (GSC-PDM). The shape being modeled, as shown in Fig. 2 is a network of sulcal curves that forms the boundaries of gyral regions of the left hemisphere of the human cerebral cortex, mapped to the unit sphere. Sulcal curves are manually delineated and uniformly sampled to get the homologous landmark points. Twenty subjects are used for training the PDMs and another twenty subjects are used for testing. For the noisy case–shown in Fig. 2–block noise with random positions, intervals, andmagnitudes are added to the landmark coordinates to simulate errors in corresponding point detection. We approximate the shape with different numbers of principal shape modes with coefficients truncated by three times the singular value associated with that mode and and compute the mean approximation errors (MAE) over all test shapes. For SMT-PDM and GSC-PDM, we test the model with different shrinkage intensity.

As shown in Fig. 3, the original PDM has the largest approximation error in both the noise-free and noisy cases due to its limited number of eigenmodes. SMT-PDM reaches the least approximation error with some shrinkage intensity in both cases when using a large numbers of shape modes. However, as shown in Fig. 2, the low approximation error is a result of over-fitting. Compared to SMT-PDM, the error of GSC-PDM and SC-PDM decreases much faster as the shape mode number increases, and the approximation error does not increase much from the noise-free to the noisy case. This implies that the target covariance estimated from the proposed constrained model is a better approximation of the underlying true covariance compared to the smoothness model. SC-PDM has a lower approximation error than GSC-PDM, which verifies the effectiveness of the spatially varying shrinkage intensity field.

3.2. Real segmentation task

Gyral labeling is a challenging task because of the large inter-subject variability in brain morphometry. We applied a prior constrained point set registration method to identify the sulcal curve network on a subject cortical surface and label the gyral regions [11]. As the iteration increases, the shape is approximated by all shape modes of SC-PDM with increasing coefficient threshold. To reduce the shape variations introduced by the surfaces of different subjects, all the subject surfaces were mapped to a unit-sphere, and the PDM building and point set registration were performed on the unit-sphere. Fifty subjects were obtained from the Open Access Series of Imaging Studies (OASIS) database [12] and manually delineated by a trained expert. Twenty subjects were used as training samples, and the remaining subjects were used for testing. Fig. 1 shows the detected sulcal curves and labeled gyral regions on one subject. Fig. 4 shows the average dice measure for each gyral region. The overall average dice measure (area weighted) is 0.85 and the average dice coefficient of all regions is above 0.7. Compared to our previous results in [11], although the overall dice is about the same, the results on the low performance structures —e.g. the frontal pole, temporal pole, and the divisions of the cingulate sulcus—are improved. These structures have few sulcal features to follow and can only be estimated from prior knowledge. Thus SC-PDM conveys the constraints from the shape prior, while also capturing the large inter-subject variations.

4. CONCLUSION

We presented a shrinkage covariance based PDM to address the over-constrained problem of a traditional ASM. In our approach, we fit a constrained covariance model to the training data and apply a spatially varying shrinkage intensity to combine the sample covariance and the constrained target covariance. Since the parameters of the constrained model and the amount of shrinkage are determined in a data-driven fashion, the resulting distribution is optimized in representing the underlying data. Experimental results demonstrate increased accuracy and robustness of the resulting SC-PDM in fitting shapes. The results on real data show the potential to use SC-PDM for segmenting complex structures—gyral regions on the cerebral cortex—with limited (20) training samples. Future work will focus on the form of the lower dimensional model and its parameter estimation.

REFERENCES

[1]. Cootes TF, Taylor CJ, Cooper DH, Graham J, et al. Active shape models-their training and application. Computer vision and image understanding. 1995; 61(1):38–59.

[2]. Cootes TF, Taylor CJ. Combining point distribution models with shape models based on finite element analysis. Image and Vision Computing. 1995; 13(5):403–409.

[3]. Wang Y, Staib LH. Boundary finding with prior shape and smoothness models. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2000; 22(7):738–743.

[4]. Shen D, Herskovits EH, Davatzikos C. An adaptive-focus statistical shape model for segmentation and shape modeling of 3-d brain structures. Medical Imaging, IEEE Transactions on. 2001; 20(4):257–270.

[5]. Tao X, Prince JL, Davatzikos C. Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain. IEEE Transactions on Medical Imaging. 2002; 21(5):513–524. [PubMed: 12071622]

[6]. Davatzikos C, Tao X, Shen D. Hierarchical active shape models, using the wavelet transform. Medical Imaging, IEEE Transactions on. 2003; 22(3):414–423.

[7]. Amberg, M.; Lüthi, M.; Vetter, T. Local regression based statistical model fitting; Proceedings of the 32nd DAGM conference on Pattern recognition; 2010; Springer. p. 452-461.

[8]. Ledoit O, Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance. 2003; 10(5):603–621.

[9]. Schäfer J, Strimmer K, et al. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical applications in genetics and molecular biology. 2005; 4(1)

[10]. Crimi A, Sporring J, Bruijne M, et al. Prior knowledge regularization in statistical medical image tasks. Probabilistic Models for Medical Image Analysis. 2009

[11]. Yang, Z.; Carass, A.; Chen, C.; Prince, JL. SPIE Medical Imaging. International Society for Optics and Photonics; 2012. Simultaneous cortical surface labeling and sulcal curve extraction; p. 831414-831414.
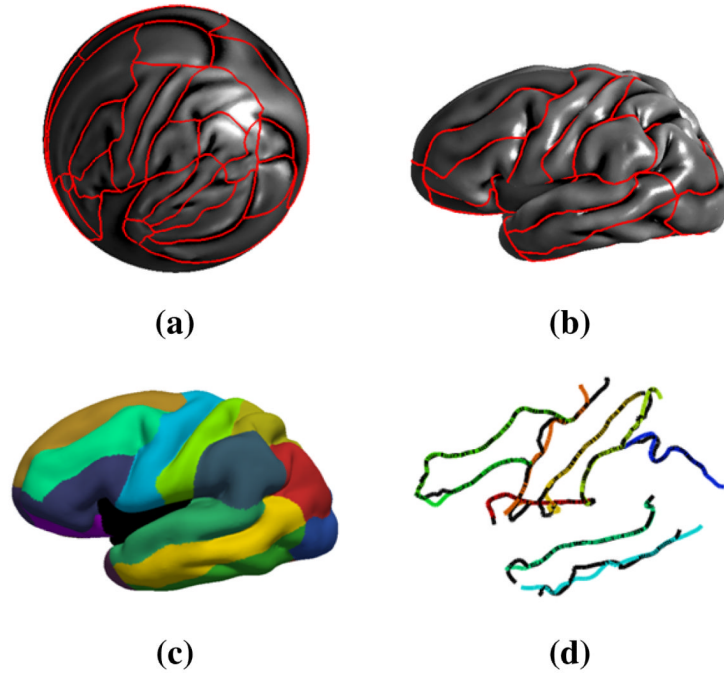
[12]. Open access series of imaging studies (OASIS). http://www.oasis-brains.org/

Fig. 1.
Example SC-PDM result for gyral labeling. (**a**) A sulcal curve network matched to a subject feature map, shown mapped to a sphere. (**b**) The same sulcal curve network shown on the corresponding partially inflated cortical surface. (**c**) Labeled gyral regions generated from the sulcal curve network. (**d**) Major sulcal curves in the curve network (colored) overlayed with manual delineations (black). See Sec. 3 for complete details.
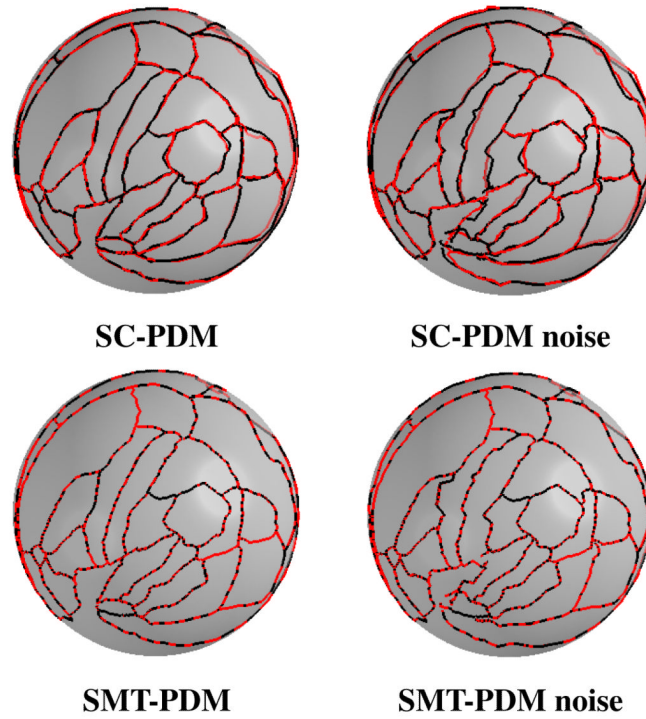
Fig. 2.
Example shape approximation results from using all shape modes with coefficients truncated by three times the singular value associated with that mode. The black curve shows the test shape, while red denotes the approximated shape.
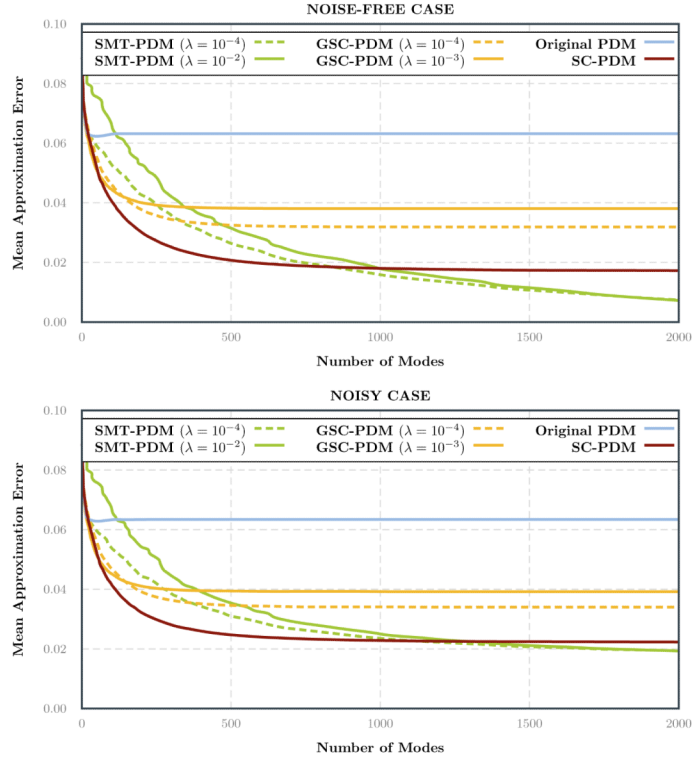
Fig. 3.
The top figure shows the mean approximation error *vs.* the number of modes for the noise-free case. The bottom figure shows the mean approximation error *vs.* the number of modes for a case with noise added (see text for details of the noise). The PDMs shown are the global shrinkage intensity model (GSC-PDM) with two values of $\lambda = 10^{-3}$ and 10^{-4}, a smoothness model (SMT-PDM) [3] with two values of $\lambda = 10^{-2}$ and 10^{-4}, the original PDM, and the shrinkage covariance model (SC-PDM).
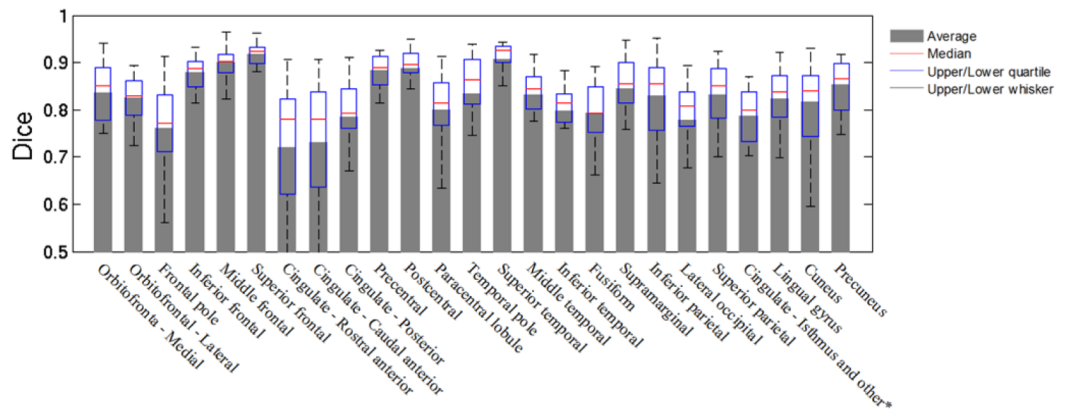
Fig. 4.
Statistics on the dice coefficients for each gyral region over all subjects.