Coarse-to-Fine Statistical Shape Model by Bayesian Inference

Ran He, Stan Li, Zhen Lei, and ShengCai Liao

Institute of Automation, Chinese Academy of Sciences, Beijing, China
rhe@nlpr.ia.ac.cn

Abstract. In this paper, we take a predefined geometry shape as a constraint for accurate shape alignment. A shape model is divided in two parts: fixed shape and active shape. The fixed shape is a user-predefined simple shape with only a few landmarks which can be easily and accurately located by machine or human. The active one is composed of many landmarks with complex shape contour. When searching an active shape, pose parameter is calculated by the fixed shape. Bayesian inference is introduced to make the whole shape more robust to local noise generated by the active shape, which leads to a compensation factor and a smooth factor for a coarse-to-fine shape search. This method provides a simple and stable means for online and offline shape analysis. Experiments on cheek and face contour demonstrate the effectiveness of our proposed approach.

Keywords: Active shape model, Bayesian inference, statistical image analysis, segmentation.

1 Introduction

Shape analysis is an important area in computer vision. A common task of shape analysis is to recover both pose parameters and low-dimensional representation of the underlying shape from an observed image. Applications of shape analysis spread from medical image processing, face recognition, object tracking and etc.

After the pioneering work on active shape model (ASM) put forward by Cootes and Taylor [1,2], various shape models have been developed for shape analysis, which mainly focus on two parts: (1) statistic framework to estimate the shape and pose parameters and (2) optimal features to accurately model appearance around landmarks. For parameter estimation, Zhou, Gu, and Zhang [3] propose a Bayesian tangent shape model to estimate parameters more accurately by Bayesian inference. Liang et al. [4] adopt Markov network to find an optimal shape which is regularized by the PCA based shape prior through a constrained regularization algorithm. Li and Ito [5] use AdaBoosted histogram classifiers to model local appearances and optimize shape parameters. Thomax Brox et al. [6] integrated 3D shape knowledge into a variational model for pose estimation and image segmentation. For optimal features, van Ginneken et al. [7] propose a non-linear ASM with Optimal Features (OF-ASM), which allows distributions of multi-modal intensities and uses a k-nearest neighbors classifier for local textures classification. Federico Sukno et al. [8] further develop

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 54–64, 2007.

this non-linear appearance model, incorporating a reduced set of differential invariant features as local image descriptors. A Cascade structure containing multiple ASMs is introduced in [9] to make location of landmarks more accurate and robust. However, these methods will lose their efficiency when dealing with complicated geometry of shapes or large texture variations.

Can we utilize some accurate information to simplify ASM algorithm and make shape parameters estimation more robust? For example, we can utilize face detection algorithm to detect the coordinates of eyes and mouth or manually label these coordinates when we want to find a facial contour for further analysis. In this paper, the problem of shape analysis is addressed from three aspects. Firstly, we present geometry constrained active shape model (GCASM) and divide it in two parts: fixed shape and active shape. The fixed shape is a user-predefined shape with only a few points and lines. Those points could be easily and accurately located by machine or human. The active one is a user's desired shape and is composed of many landmarks with a complex contour. It will be located automatically with the help of the fixed shape. Secondly, Bayesian inference is introduced to make parameter estimation more robust to local noise generated by the active shape, which leads to a compensation factor and a smooth factor to perform a coarse-to-fine shape search. Thirdly, optimal features are selected as local image descriptors. Since the pose parameters can be calculated by the fixed shape, classifiers are trained for each landmark without scarifying performance.

The rest of the paper is organized as follows: In Section 2, we begin with a brief review of ASM. Section 3 describes our proposed algorithm and Bayesian inference. Experimental results are provided in Section 4. Finally, we draw the conclusions in Section 5.

2 Active Shape Models

This section briefly reviews the ASM segmentation scheme. We follow the description and notation of [2]. An object is described by points, referred as landmark points. The landmark points are (manually) determined in a set of *N* training images. From these collections of landmark points, a point distribution model (PDM) [10] is constructed as follows. The landmark points (x_1, y_1, … , x_n, y_n) are stacked in shape vectors.

$$x = (x_1, y_1, ..., x_n, y_n)^T . \tag{1}$$

Principal component analysis (PCA) is applied to the shape vectors *x* by computing the mean shape, covariance and eigensystem of the covariance matrix.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad S = \frac{N}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T . \tag{2}$$

The eigenvectors corresponding to the *k* largest eigenvalues λ_j are retained in a matrix $\Phi = (\phi_1 \mid \phi_2 \mid \cdots \mid \phi_k)$. A shape can now be approximated by

$$x \approx \bar{x} + \Phi b . \tag{3}$$

Where b is a vector of k elements containing the shape parameters, computed by

$$b = \Phi^T (x - \bar{x}) . \tag{4}$$

When fitting the model to a set of points, the values of b are constrained to lie within a range

$$| b_j | \leq \pm c\sqrt{\lambda_j} . \tag{5}$$

where c usually has a value between two and three.

Before PCA is applied, the shapes can be aligned by translating, rotating and scaling so as to minimize the sum of squared distances between the landmark points. We can express the initial estimate x of a shape as a scaled, rotated and translated version of original shape

$$x = M(s,\theta)[x] + t . \tag{6}$$

Where $M(s,\theta)$ and t are pose parameters (See [1] for details). Procrustes analysis [11] and EM algorithm [3] are often used to estimate the pose parameters and align the shapes. This transformation and its inverse are applied both before and after projection of the shape model. The alignment procedure makes the shape model independent of the size, position, and orientation of the objects.

3 Coarse-to-Fine Statistical Shape Model

3.1 Geometry Constrained Statistical Shape Model

To make use of the user-predefined information, we extend PDM to two parts: active shape and fixed shape. The active shape is a collection of landmarks to describe an object in the basic PDM. It is composed of many points with a complex contour. The fixed shape is a predefined simple shape accurately marked by user or machine. It is composed of several connected lines between these points which can be easily and accurately marked by machine or human. Considering there are tremendous points in a line, we present a line with several equidistant points. Thus the extended PDM is constructed as follows. The landmarks (x_1, y_1, … , x_m, y_m) are stacked in active shape vectors, and landmarks (x_{m+1}, y_{m+1}, … , x_n, y_n) are stacked in fixed shape vectors.

$$x = (x_1, y_1, ..., x_m, y_m, x_{m+1}, y_{m+1}, ..., x_n, y_n)^T . \tag{7}$$

As in PDM, a shape can now be approximated by

$$x \approx \bar{x} + \Phi b . \tag{8}$$

When aligning shapes during training, the pose parameters of a shape (scaling, rotation and translation) are estimated by the fixed shape. An obvious reason is that the fixed shape is simpler and more accurate than the active one.

Taking cheek contour as an example, the active shape is composed of landmarks in a cheek contour and the fixed shape is composed of 13 landmarks derived from three manual labeled points: left eye center, right eye center and mouth center. Five

landmarks are added equidistantly between two eyes center to represent horizontal connected line. And five landmarks are inserted equidistantly in the vertical line passing the mouth center and perpendicular to the horizontal line. (See left graph of Fig.1 for details) During training, two shapes are aligned according to the points between two eyes only. Each item of b reflects a specific variation along the corresponding principle component (PC) axis. Shape variation along first three PCs is shown in right graph of Fig.1. The interpretation of these PCs is straight forward. The first PC describes left-right head rotations. The second PC accounts for face variation in vertical direction: long or short. And the third one explains a human face fat or thin.

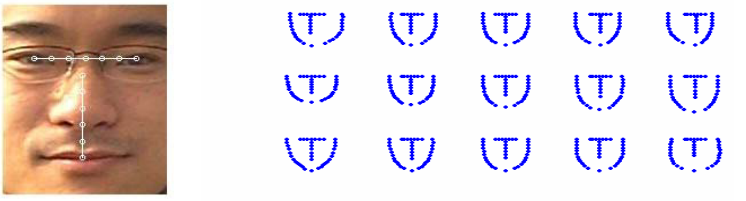

Fig. 1. The fixed shape and shapes reconstructed by the first three PCs. The thirteen white circles in left image are points of the fixed shape. In right image, the middle one in each row is the mean shape.

3.2 Bayesian Inference

When directly calculating shape parameter b by formula (4), there is an offset between the reconstructed fixed shape and the given fixed shape. But the fixed shape is supposed to be accurate. This noise comes from reconstruction error of the active shape. Inspired by paper [3], we associate PCA with a probabilistic explanation. An isotropic Gaussian noise item is added to both fixed and active shape; thereby we can compute the poster of model parameters. The model can be written as:

$$y = \bar{x} + \Phi b + \varepsilon \,. \tag{9}$$

$$y - \bar{x} - \Phi b = \varepsilon \,. \tag{10}$$

Where the shape parameter b is a n-dimensional vector distributed as multivariate Gaussian $N(0,\Lambda)$ and $\Lambda = diag(\lambda_1,...,\lambda_k)$. ε denotes an isotropic noise on the whole shape. It is a n-dimensional random vector which is independent with b and distributes as

$$p(\varepsilon) \sim \exp\{-\| \varepsilon \|^2 / 2(\sqrt{\rho})^2\} \,. \tag{11}$$

$$\rho = \sum_{i=1}^{n} \alpha_i \| y_i^{old} - y_i \|^2 \,. \tag{12}$$

Where y^{old} is the shape estimated in the last iteration and y is an observed shape in the current iteration. a_i is classification confidence related to a classifier used in locating a

landmark. When a_i is 0, which implies that classifier can perfectly predict shape's boundary; when a_i is 1, which means classifier fails to predict the boundary.

Combing (10) and (11) we obtain the likelihood of model parameters:

$$P(b \mid y) = const P(y \mid b) P(b) \sim \exp(-\frac{1}{2}[(y - \bar{x} - \Phi b)^T (y - \bar{x} - \Phi b) / \rho + b^T \Lambda^{-1} b]) \quad (13)$$

Let $\frac{\partial}{\partial b}(\ln P(b \mid y)) = 0$, we get:

$$b_j = (\lambda_j \setminus (\lambda_j + \rho)) \phi_j^T (y - \bar{x}) . \quad (14)$$

Combining (4), we obtain:

$$b_j = (\lambda_j \setminus (\lambda_j + \rho)) b_j . \quad (15)$$

It is obvious that value of b_j will become smaller after updating of (15) ($\rho \geq 0$). This updating will slow down search speed. Hence, a compensation factor p_1 is introduced to make shape variation along eigenvectors corresponding to large eigenvalues more aggressive (see formula 18). If p_1 is equal to $(\lambda_{\max} + \rho) / \lambda_{\max}$, we get

$$b_j = ((\lambda_{\max} + \rho) \setminus \lambda_{\max}) \times (\lambda_j \setminus (\lambda_j + \rho)) \times b_j . \quad (16)$$

Formula (16) shows that the parameter b_j corresponding to a larger eigenvalue will receive a small punishment. And the parameter b_j corresponding to a small eigenvalue will become smaller after updating. Moreover, we expect a smooth shape contour and neglect details in the first several iterations. A smooth factor p_2 (see formula 18) is introduced to further punish the parameter b_j. It is noticed that ρ is smaller than the largest eigenvalue and will become smaller. The p_2 regularizes the parameters by enlarging the punishment. As in Fig.2, the reconstructed shape's contour by Bayesian inference is smoother than the one by PCA in regions pointed by the black arrows. Although the PCA reconstruction can remove some noise, the reconstructed shape is still unstable when the image is noisy. Formula (18) makes the parameter estimation more robust to local noise.

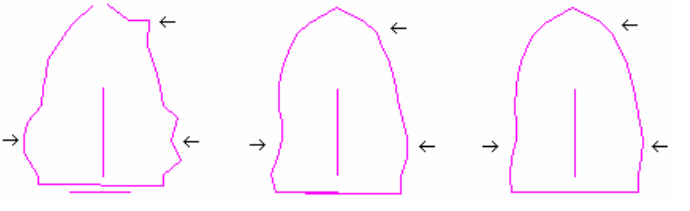

Fig. 2. Shapes reconstructed from PCA and Bayesian Inference. Left shape is mean shape after desired movements; middle shape is reconstructed by PCA; right shape is reconstructed by Bayesian Inference. The black arrows highlight the regions to be compared.

3.3 Optimal Features

Recently, optimal features are applied in ASMs and have drawn more and more attentions. [5, 6, 7] Experimental results show that optimal features can make shape segmentation more accurate. But a main drawback of optimal features method is that it takes ASMs more time to find the desired landmarks due to extract optimal features in each iteration. An efficient speed-up strategy is to select a subset of the available features for all landmarks. [6, 7] It is clear that textures around different landmarks are different. It is impossible for a single subset of optimal features to describe various textures around all the landmarks.

In GCASM, the pose parameters of scale, rotation and translation can be calculated by the fixed shape. All landmarks can be categorized into several groups, for each of which we select the same discriminate features. When search a shape, the image is divided into several areas according to the categories. For each area, the same optimal features are extracted to determine movement. Optimal features are features reported in both paper [6] and [7]. Fig.3 shows classification results for each landmark. The Mean classification accuracy is 76.67%. We can learn about that landmarks near jaw and two ears have low classification accuracy, and the landmarks near cheek have high classification accuracy. Considering this classification error, we introduce Bayesian Inference and a_i of formula (12) to make shape estimation more robust.

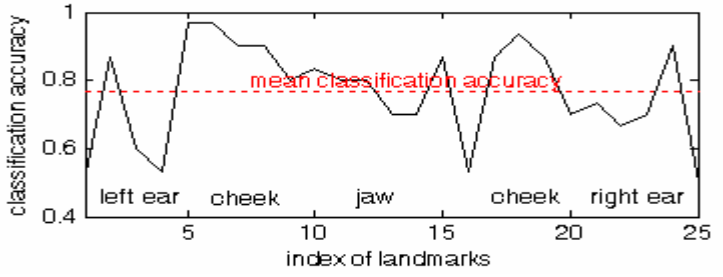

Fig. 3. Classification results for each cheek landmark. Classification accuracy stands for a classifier's ability to classify whether a point near the landmark is in or outside of the shape. The points around the indices of 4 and 22 are close to ears and the points around the index of 13 are close to jaw.

3.4 Coarse-to-Fine Shape Search

During image search, main differences between GCASM and ASM lie in twofold. One is that since the pose parameters of GCASM have been calculated by the fixed shape, we needn't to think about the pose variation during iterative updating procedure. The other is that the fixed shape is predefined accurately in GCASM. After reconstruction from the shape parameters, the noise will make the reconstructed fixed shape leave away from the given fixed shape. Because the fixed shape is supposed to be accurate, it should be realigned to the initial points. The iterative updating procedure of GCASM and ASM are shown in Fig.4. We use formula (17) to calculate shape parameter $b=[b_1,\ldots,b_k]^T$ and normalize b by formula (18).

$$b_j = \Phi_j^T (y - \bar{x}) . \tag{17}$$

$$b_j = (p_1 \lambda_j / (\lambda_j + p_2 \rho)) b_j \tag{18}$$

Where $1 \le p_1 \le (\lambda_{\max} + p_2 \rho) / \lambda_{\max}$, $p_2 \ge 1$.

We call the parameter p_1 compensation factor which makes shape variation in a more aggressive way. The parameter p_2 is a smooth factor which gives a penalty to the shape parameter when shape has a large variation. The compensation factor and smoother factor give more emphasis on shape parameters corresponding to large eigenvalues. This can adjust a shape along major PCs and neglect shape's local detail in initial several iterations. When the algorithm converges ($\rho \to 0$), $p_1 \lambda_j / (\lambda_j + p_2 \rho)$ is equal to 1. Hence, the compensation factor and smoother factor lead a coarse-to-fine shape searching. Here, we simply set $p_1 = (\lambda_{\max} + p_2 \rho) / \lambda_{\max}$, $\alpha_i = 1$ and $p_2 = 4$.

Obviously, the formula (18) can also be used in ASMs to normalize the shape parameter.

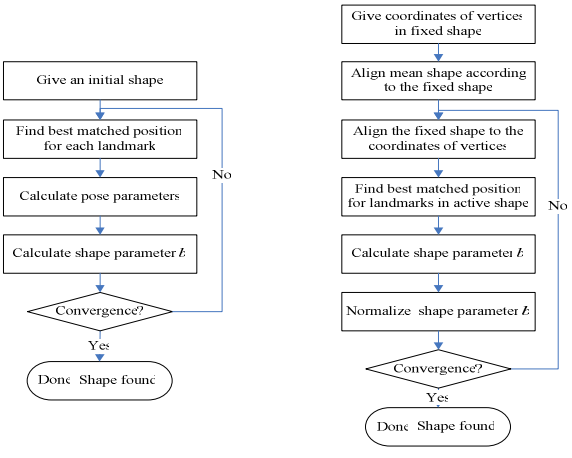

Fig. 4. Updating rules of ASM and GCASM. The left block diagram is the basic ASM's updating rule and the right block diagram is GCASM updating rule.

4 Experiments

In this section, our proposed method is tested on two experiments: cheek contour search and facial contour search. A total of 100 face images are randomly taken from the XM2VTS face database. [12] Each image is aligned by coordinates of two eyes. The average distance between two eyes is 80 pixels. Three points of the fixed shape including two eyes and mouth center are manually labeled. The fixed shape takes a shape of letter 'T'. Hamarneh's ASM source code [13] is taken as the standard ASM without modification. Optimal features are collected from features reported in both paper [7] and [8]. The number of optimal features is reduced by sequential feature

selection [14]. In this work, all the points near the landmarks are classified by linear regression to predict whether they lie in or out of a shape.

4.1 Experiments on Cheek Contour

A designed task to directly search a cheek contour without eyes, brow, mouth, and nose is presented to validate our method. A total of 25 cheek landmarks are labeled manually on each image. The PCA thresholds are set to 99% for every ASMs. The fixed shape is composed of points between two eyes and mouth. As in Fig.3, it is difficult to locate points around landmarks near ears and jaw. When a contour shape is simple and textures around landmarks are complex, the whole shape will be dragged off from the right position if there are several inaccurate points. It is clear that the cheek shape can be accurately located with the help of the fixed shape.

Fig. 5. Comparison of different algorithms' cheek searching results: Shapes in first column are results of ASM searching; Shapes in second column are results of simple OF-ASM; Shapes in third column are results of the basic GCASM; Shapes in fourth column are results of GCASM with optimal features; Shapes in fifth column are results of GCASM with optimal features and Bayesian inference

As in Fig.5, first two columns are the searching results of ASM and OF-ASM. It is clear that the searching results miss desired position because of local noise. Several inaccurate landmarks will drag the shape from desired position. It also illustrates that optimal features can model contour appearance more accurately.

As illustrated in the last three columns in Fig.5, searching results are well trapped in a local area when the fixed shape is introduced. Because the fixed shape is accurate without noise, reconstructed shape will fall into a local area around the fixed shape even if some landmarks are inaccurate. Every landmark will find a local best matched point instead of a global one. Comparing the third and fourth column, we can learn about that optimal features can locate landmarks more accurately. But optimal

features couldn't keep local contour detail very well. There is still some noise in searching results. Looking at the fifth column of Fig.5, it is clear that borders of the shapes become smoother. The Bayesian inference can further improve the accuracy.

4.2 Experiments on Facial Contour

A total of 96 face landmarks are labeled manually on each image. The PCA thresholds are set to 95% for every ASMs. Three landmarks are inserted into two eyes to present horizontal connected line. And three landmarks are inserted between mouth and horizontal line to present the vertical line. For the sake of simplicity, optimal features don't used in this subsection. The results are shown in table 1.

Table 1. Comparison results of traditional ASM and our method without optimal features

	Face	F.S.O	Cheek Contour
ASM	7.74	6.45	11.4
Our algorithm	4.68	4.41	5.47
Improvement	39.5%	31.6%	52.0%

Where F.S.O. means five sense organs. Location error is measured in pixel. It is clear that our algorithm is much more accurate than ASM.

Fig. 6. Comparison results of ASM and GCASM with Bayesian inference. The first row is ASM results, and the second row is our results.

Fig.6 shows a set of searching results of basic ASM and GCASM with Bayesian inference. In the case, there are wrinkles and shadings on the facial contour or other facial sub-parts. It is clear that our method can recover the shape from local noise. A direct reason is that the shape variation is restricted in a local area when combining accurate information in ASM. The Bayesian inference holds the whole shape and smoothes the shape border.

5 Conclusion

This work focuses on an interesting topic how to combine some accurate informationgiven by user or machine to further improve shape alignment accuracy. The PDM is extended by adding a fixed shape which is generated from given information. After PCA reconstruction, local noise of the active shape will make the whole shape unsmooth. Hence Bayesian inference is proposed to further normalize parameters of the extended PDM. Both compensate factor and smooth factor lead a coarse-to-fine shape adjustment. Comparisons of our algorithm and the ASM algorithms demonstrate the effectiveness and efficiency.

Acknowledgements

This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Supporting Platform Project #2006BAK08B06, National 863 Program Projects #2006AA01Z192 and #2006AA01Z193, Chinese Academy of Sciences 100 people project, and the Authen-Metric Collaboration Foundation.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D., Graham, J.: Active shape models-Their training and application. Comput. Vis. Image Understanding 61(1), 38–59 (1995)
2. Cootes, T.F., Taylor, C.J.: Statistical models of appearance for computer vision, Wolfson Image Anal. Unit, Univ. Manchester, Manchester, U.K., Tech. Rep (1999)
3. Zhou, Y., Gu, L., Zhang, H.-J.: Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference. In: IEEE Conf. on Computer Vision and Pattern Recognition, Madison, WI (June 2003)
4. Liang, L., Wen, F., Xu, Y.Q., Tang, X., Shum, H.Y.: Accurate Face Alignment using Shape Constrained Markov Network. In: Proc. CVPR (2006)
5. Li, Y.Z., Ito, W.: Shape parameter optimization for Adaboosted active shape model. In: ICCV, pp. 259–265 (2005)
6. Brox, T., Rosenhahn, B., Weickert, J.: Three-Dimensional Shape Knowledge for Joint Image Segmentation and Pose Estimation. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) Pattern Recognition. LNCS, vol. 3663, pp. 109–116. Springer, Heidelberg (2005)
7. Ginneken, B.V., Frangi, A.F., Staal, J.J., ter Har Romeny, B.M., Viergever, M.A.: Active shape model segmentation with optimal features. IEEE Transactions on Medical Imaging 21(8), 924–933 (2002)
8. Sukno, F., Ordas, S., Butakoff, C., Cruz, S., Frangi, A.F.: Active shape models with invariant optimal features IOF-ASMs. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 365–375. Springer, Heidelberg (2005)
9. Zhang, S., Wu1, L.F., Wang, Y.: Cascade MR-ASM for Locating Facial Feature Points. The 2nd International Conference on Biometrics (2007)
10. Dryden, I., Mardia, K.V.: The Statistical Analysis of Shape. Wiley, London, U.K (1998)

11. Goodall, C.: Procrustes methods in the statistical analysis of shapes. J.Roy. Statist. 53(2), 285–339 (1991)
12. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: Proc. AVBPA, pp. 72–77 (1999)
13. Hamarneh, G.: Active Shape Models with Multi-resolution, http://www.cs.sfu.ca/~hamarneh/software/asm/index.html
14. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classiers. Pattern Recognition, 25–41 (2000)