Journal of Southeast University (English Edition) Vol. 21, No. 3, pp. 263 – 267 Sept. 2005 ISSN 1003—7985

Optimal shape space and searching in the active shape model

He Lianghua[1,2] Zou Cairong[2] Zhao Li[2] Hu Die[2]

([1] Research Center of Learning Science, Southeast University, Nanjing 210096, China)

([2] Department of Radio Engineering, Southeast University, Nanjing 210096, China)

Abstract: A novel idea, called the optimal shape subspace (OSS) is first proposed for optimizing active shape model (ASM) search. It is constructed from the principal shape subspace and the principal shape variance subspace. It allows the reconstructed shape to vary more than that reconstructed in the standard ASM shape space, hence it is more expressive in representing shapes in real life. Then a cost function is developed, based on a study on the search process. An optimal searching method using the feedback information provided by the evaluation cost is proposed to improve the performance of ASM alignment. Experimental results show that the proposed OSS can offer the maximum shape variation with reserving the principal information and a unique local optimal shape is acquired after optimal searching. The combination of OSS and optimal searching can improve the ASM performance greatly.

Key words: active shape model; shape subspace; search subspace; principal component analysis

Accurate alignment of faces is very important for extraction of good facial features for success of applications such as face recognition, expression analysis and face animation. The active shape model (ASM)[1] proposed by Cootes et al. is one successful shape model for object localization. In it, the local appearance model, which represents the local statistics around each landmark, efficiently finds the best candidate point for each landmark in searching the image. The solution space is constrained by the properly trained global shape model. Based on the accurate modeling of the local features, ASM obtains good results in shape localization.

Usually, the dimensionality of the principal shape model is chosen to explain as high as 95% to 98% of variation in the tangent shape space (TSS) so that the ASM model can approximate any shape in TSS accurately. The underlying assumption is that if the reconstruction error is small, the overall error of the ASM search will also be small. This underlying assumption was taken for granted by previous works without any justification. However, our analysis of optimal shape subspace and optimal search processing for ASM shows that this is inadequate.

In standard ASM search, each point is first searched in image shape subspace (ISS) according to the profiles perpendicular to the object contour around it. Then shape mode constraint is performed to adjust the search result. These two steps are done alternately and independently. If no evaluation is made to the search result and no direction constraint is given to the search process, the search will never stop and the result will be very sensitive to the patch noise. Several evaluation methods are proposed[1-4], in particular that by Huang et al.[3] can be used in a new image search, but it needs to build in the model of active appearance model (AAM). As for shape model constraints, no discussion about it appears in previous works.

In this paper, the optimal shape subspace (OSS) is proposed after analyzing the properties of ASM search subspace. The idea is as follows: To minimize the error between the search result and the input, we should construct a search subspace which not only constrains the search in the principal shape space, but also allows the shape to vary as much as possible. The OSS allows good variations of shape information with minimal dimension. On the other hand, a simple evaluation method based on the point searching cost value in ISS is presented to measure the quality of the ASM search, to produce reliable search results. We also propose a method to constrain the search action according to the shape adjust in OSS. Finally, combining the evaluation and the constraints, we get an optimal searching method with the information from both ISS and OSS.

1 Optimal Shape Subspace in ASM

1.1 Traditional ASM modeling

To train the ASM model, shapes in ISS should first be annotated in the image domain, then these shapes in ISS are aligned into those in TSS. When a

Received 2005-03-28.

Foundation item: 21st Century Education Revitalization Project (No. 301703201).

Biographies: He Lianghua (1977—), male, graduate; Zou Cairong (corresponding author), male, doctor, professor, cairong@ seu. edu. cn.

training set of the tangent shapes in TSS is given, the basis functions for the KLT are obtained by solving the eigenvalue problem $\boldsymbol{\Lambda} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi}$, where $\boldsymbol{\Sigma}$ is the covariance matrix, $\boldsymbol{\Phi}$ is the eigenvector matrix of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ is the corresponding diagonal matrix of eigenvalues. After being trained by principal component analysis (PCA), the ASM model can be written as $\boldsymbol{x} = \bar{\boldsymbol{x}} + \boldsymbol{\Phi}_t\boldsymbol{b}$, where $\bar{\boldsymbol{x}}$ is the mean tangent shape vector, $\boldsymbol{\Phi}_t = \{\phi_1 \mid \phi_2 \mid \dots \mid \phi_t\}$ is a submatrix of $\boldsymbol{\Phi}$ containing the principal eigenvectors corresponding to the largest eigenvalues (sorted so that $\lambda_i \geqslant \lambda_{i+1}$), and $\boldsymbol{b}$ is a vector of shape parameters. For a given shape, its shape parameter is given by $\boldsymbol{b} = \boldsymbol{\Phi}_t^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}})$. The number of eigenvectors(t) to retain can be chosen in several ways described detailed in Ref. [5]. But they are all based on the reconstruction error which Zhao et al. [4] thought is not enough. They proposed a P-TSS shape subspace and got some better results.

1.2 Optimal shape subspace

We can see from above that the TSS in standard ASM and Zhao et al. 's[4] is constructed by only considering the principal shapes and discarding all the other components. This is to some extent inadequate since every model to be described has much variance and we cannot express them all in a given model space. Thus, we should consider the shape variance as well as the shape itself. As an optimal TSS, we think that it should have such properties. First, it can allow the shape it reconstructs to vary as much as possible. When reconstructing a shape in it, we should have the minimal reconstruction error. Secondly, its dimensions should be as few as possible. Not only can it reduce the computation, but also the constraint when reconstruction is minor. Thirdly, which is the most important, the total changes between the search result and its reconstruction must be as small as possible, since we believe, to some degree, that the points after the search are the best candidates in that iteration. The aim of reconstruction is to adjust the points which deviate too much, but not to reconstruct a new shape. We call the shape subspace which satisfies such three properties the optimal shape subspace.

For the first one, the larger t is, the more shapes this subspace can reconstruct. While for the second, the fewer the better. Considering the first and the second together, we decompose OSS into two parts: the principal shape subspace (PSSS) which controls the shape in searching and the principal variance shape subspace (PVSS) which controls the shape variance in searching. They are both the subspace of $\boldsymbol{\Phi}$ with orthogonality: OSS = PSSS $\oplus$ PVSS.

When constructing, the PSSS shares the same meaning of TSS and can be constructed like $\boldsymbol{\Phi}_t$ in standard ASM. In this paper, we adopt the rule of Zhao et al. [4], that is, about 72% proportion of the variance exhibited in the training data is selected. But for PVSS, the construction suffers some difficulties. We should find one parameter to stand for the shape variance for every eigenvector. Suppose that every element of a given eigenvector is the same and its projection coefficient changes, the whole shape will only transform in shift, but will not change in itself (see the last row in Fig. 1). That is, only the variance of an eigenvector itself controls the shape transformation. Thus, to select the eigenvectors with the largest variance to construct PVSS is somewhat reasonable.

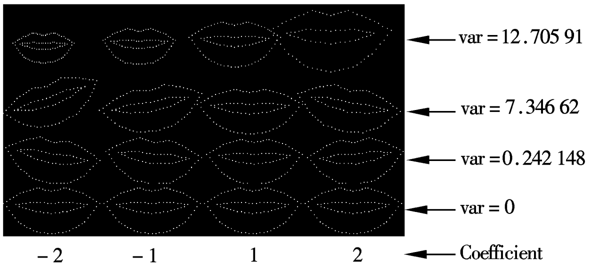

Fig. 1 Relationship between the eigenvector's variance and its deformability

In total, the proposed algorithm to construct the OSS is as follows:

① Select the eigenvectors with 72% proportion of the variance in the training data.

② Get the correlation matrix of the left eigenvectors.

③ Get the variance for the left eigenvectors.

④ Select the largest variance eigenvectors with the least correlation coefficient. The number can be chosen in many ways, one can set the threshold of the total number or the value of the variance. In this paper, we select the number semi-automatically, that is, we manually select the principal deformations in all the eigenvector candidates after printing.

There are two main improvements in constructing OSS. First, the major shape and shape variance are considered. So not only can it reconstruct the shape similar to the training with the least reconstruction error, but it can also approximate the shape which has as much difference from the training as much as possible. Secondly, the basis to control shape variance is selected after removing correlation, which would make the space more compressive.

2 Optimal Search Processing in ASM

After the OSS is constructed, the next task is searching. Usually, searching is performed in ISS and TSS alternately. No evaluation is given to the temporary search result and no discussion is made about the relation between the ISS and TSS when searching. The underlying assumption is that the result after iteration in ISS and TSS is completely right. Obviously this is not always right. Thus, we cannot get a stable result. It can be oscillatory when most of the points are in the right place in searching or the reverse. If we can evaluate the search result right away, we will know the resulting situation and decide whether it is right, so that we can select the best candidates to make further iteration. Point-to-point error is used by most researchers to evaluate the searching result shape[1,2,4]. But it cannot be used because we do not know the true position when searching. In Ref. [3], a novel ASM + AAM evaluation method is proposed after training with the Adaboost classifier, but it is somewhat complicated. So what can be used reasonably and easily?

Let us return to the search process. We know that there are two parts in ASM, one is OSS to adjust the shape after searching, the other is ISS in which every point is searched. So after iteration, we have two kinds of parameters: $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ in OSS and $(\beta_1, \beta_2, \ldots, \beta_M)$ in ISS. N and M are the dimension of OSS and the point number of the model. β_i is the projection error in ISS of the i-th point. Its value shows the possibility that the point i is the right i-th point. If it is large, the possibility will be small and maybe the search result will be wrong. Homoplastically, for the whole model, if the total error $\sum \beta_i$ is small, it dedicates more points with a large possibility that they are in the right place. The experimental results (see Fig. 2) show that around the right shape, the larger the total error, the further from the right shape the searching shape is. In other words, it means that this total error value can evaluate the search result. So, in this paper, $\sum \beta_i$ is used as the cost value to evaluate the search result.

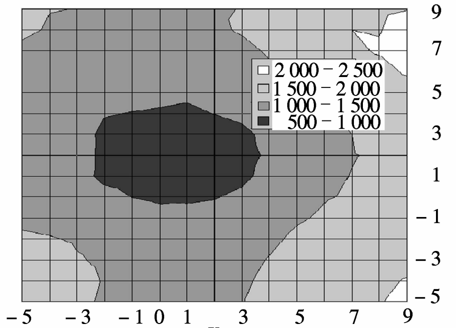

Fig. 2 Model cost value in X and Y shift

α_i is the projection coefficient in the OSS. Its value denotes the shape transformation along the i-th basis. In PSSS, every basis is for main shapes. That the projection coefficient is larger or smaller than the threshold indicates that the search result did not like the shape any more. That is, most of the points are in the wrong places. Maybe the initial scale is too large or too small and maybe the shift of X or Y is too much. The later shifting can be omitted because of the detection before. So the initial scale is the key point. Now we can conclude that if the α_i in PSSS is out of range, we can believe to some extent that the reason is because of the initial shape scale. This is very useful in forecasting the search direction in the next iteration. In PVSS, if α_i is out of range, the searching result will transform too much in the i-th rotation direction. This can be brought out by wrong searching completely or by some points searching correctly and others incorrectly. Just like in PSSS, the possibility of the whole shape incorrectly searching is small. So it is mainly because of unbalanced searching. A logical reason for this unbalance is that some points are near the right place, and some are far from the right place, which can appear when the initial shape intersects with the right. This is the second conclusion.

After iteration, we get the out of range coefficients $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n_1}, \alpha_{21}, \alpha_{22}, \ldots, \alpha_{2n_2}$ and the total cost value $C = \sum_{i=1}^{M} \beta_i$. Let n_1 be the number in PSSS and n_2 be the number in PVSS. Then we can forecast the distortion between the right shape and the search result in scale and shift in this way:

$$\Delta_S = X_{\text{old}} \prod_{i=1}^{n_2} \left(\frac{\alpha_{2i}}{T_i} - 1 \right) \tag{1}$$

$$\Delta_O = \delta \prod_{i=1}^{n_1} \left(\frac{a_{1i}}{T_i} - 1 \right) \tag{2}$$

where threshold T_i and coefficient δ are constant values. But the former threshold T_i should be selected very carefully, since it not only contains the value information, but also contains the scale or shift direction information. The final initial shape position for the next iteration we used is

$$X_{\text{new}} = X_{\text{old}} \left(1 + \prod_{i=1}^{n_1} \frac{a_{1i}}{T_i} \right) + \lambda \prod_{j=1}^{n_2} \frac{a_{2j}}{T_j} \tag{3}$$

The initiate shape before searching is more similar to the right shape after preprocessing in such a way. Considering the performance robustness, we search the shape around X_{new} (± 5 pixels in X and Y direction) through a few steps circulation in scale and shift and choose the one with the least cost value C as X'_{new}.

In total, the method for the optimal searching is as follows:

① Initiate the shape with the detection algorithm[6] for the first time;

② Calculate $C_{\text{old}} = \sum_{i=1}^{M} \beta_{\text{old}\, i}$;

③ Search as for the traditional ASM;

④ Get the parameters of $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1n_1}, \alpha_{21}, \alpha_{22}, \ldots, \alpha_{2n_2}$, and the total cost value $C_{\text{new}} = \sum_{i=1}^{M} \beta_{\text{new}\, i}$;

⑤ Compute X_{new} from Eq. (3) and get the best X'_{new} after searching;

⑥ Repeat ③ to ⑤, until $|C_{\text{new}} - C_{\text{old}}| < T$.

This algorithm has three major improvements over traditional ASM searching methods. First, the ISS and OSS are considered at the same time when searching. They restrict each other with a strong relation, whereas in the traditional method, they are used separately. Secondly, it provides a new evaluation method for the search shape and a preprocessing set for the new shape being searched, so we can make a more controllable search. Thirdly, the final result is determinate, not oscillatory any more. This is very important because we do not know exactly how many searches are enough for a new image in the traditional method, but now, there is no such question.

3 Experiments and Results

The database used consists of 1 406 face images from the FERET[6], the AR[5] databases and other collections. 87 landmarks are labeled on each face. We randomly selected 703 images as the training images and the others as the testing images. Multi-resolution search is used, using four levels with resolution of 1/8, 1/4, 1/2, 1 of the original image in each dimension. At most 10 iterations are run at each level. The ASM uses profile models of 11 pixels long (five points on either side) and searches two pixels either side.

3.1 Optimal shape subspace

On each testing image, we make two kinds of tests with different ways to initialize the start mean shape. In the first method, we initialize it after face detection. In the second, we initialize the starting mean shape with displacements from the true position by ±10 pixels in both X and Y. Point location accuracy is used to give the evaluation for every search subspace in two categories. The comparison results are shown in Fig. 3. The X coordinate is point-to-point error and the Y coordinate is the percentage of the samples whose point-to-point errors are less than the given point-to-point errors value. We can see that in the first situation, the search error is mainly around five pixels for OSS, much better than the other two methods with nearly the same error of about 20 pixels. In the second experiment, although the difference among three methods is very small, searching with OSS is still the best.

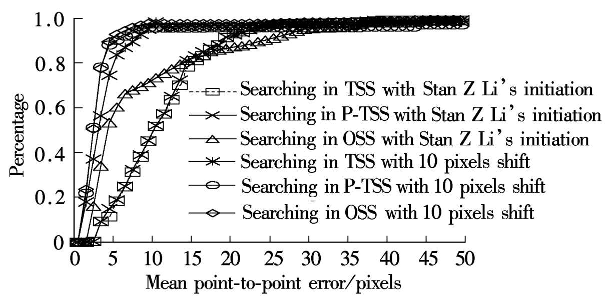

Fig. 3 Comparison of different searching subspaces with different initiating mean shapes

3.2 Optimal searching

In this experiment, optimal searching, the traditional ASM searching and Zhao et al.'s searching are compared using the initialization of a face detection method[7]. Like the experiment in testing OSS, point location accuracy is also computed to evaluate their performance. The result is shown in Fig. 4. It can be seen that optimal searching has about 50 percent samples whose point-to-point error is less than 5 pixels, obviously better than the other two. Both optimal searching and Zhao et al.'s method are better than the traditional searching algorithm.

Lastly, the performance of OSS + optimal searching, which we call optimal ASM, is tested in the same way. The results are shown in Fig. 5. From the figure we can see that by joining OSS with optimal searching, the result is overwhelmingly better.

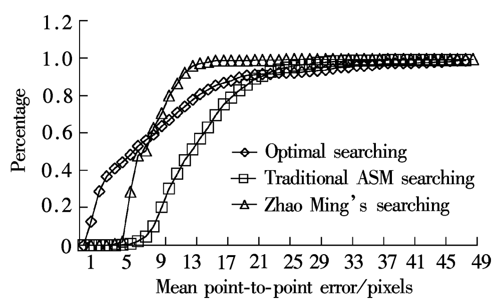

Fig. 4 Comparison of different search methods

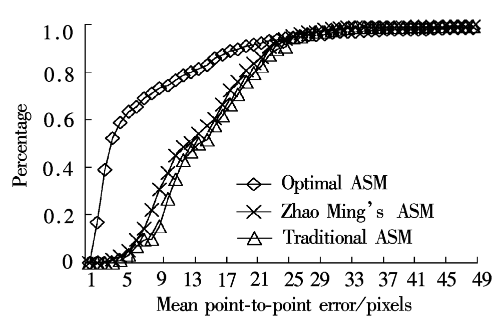

Fig. 5 Comparison of different kinds of ASM

4 Conclusion

In this paper, we have discussed optimal shape subspace in ASM training and optimal searching in ASM searching. With the best experimental results, it can be concluded that containing the basement shapes and the basement shapes' variance at the same time in the shape subspace is very important. Because the shape to be searched varies very much, the subspace to describe them should give enough shape information. We call this kind of shape subspace the optimal shape subspace. Because of using evaluation and constraints in the search process in the proposed method, which has never been done in all the work before, not only a more reliable and accurate result but also a controllable search process are acquired, which is why we call this optimal searching. All kinds of experimental results show that a search with OSS and/or optimal searching has much more efficiency and accuracy.

References

[1] Cootes T F, Taylor C J, Cooper D H, et al. Active shape models: their training and application [J]. *Computer Vision and Image Understanding*, 1995, **61**(1): 38 − 59.

[2] Rogers M, Graham J. Robust active shape model search [A]. In: *Proceedings of the European Conference on Computer Vision* [C]. Copenhagen, Denmark, 2002. 517 − 530.

[3] Huang Xiangsheng, Li S Z, Wang Yangsheng. Evaluation of face alignment solutions using statistical learning [A]. In: *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition* [C]. Seoul, South Korea, 2004. 213 − 218.

[4] Zhao Ming, Chen Chun, Li S Z, et al. Subspace analysis and optimization for AAM based face alignment [A]. In: *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition* [C]. Seoul, South Korea, 2004. 290 − 295.

[5] Martinez A, Benavente R. The AR face database [R]. Barcelona, Spain: Computer Vision Center (CVC), 1998.

[6] Phillips P J, Moon H, Rizvi S A, et al. The FERET evaluation methodology for face-recognition algorithms [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(10): 1090 − 1104.

[7] Li S Z, Zhang Zhenqiu. FloatBoost learning and statistical face detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1112 − 1123.

动态形状模型中最优搜索空间和最优搜索过程

何良华[1,2]　邹采荣[2]　赵　力[2]　胡　蝶[2]

([1] 东南大学学习科学研究中心,南京 210096)
([2] 东南大学无线电工程系,南京 210096)

摘要:首先提出了最优形状子空间概念,它由主要形状子空间和主要形状变化子空间联合构成,最大程度上包含了搜索形状上的变化,更贴近现实中要搜索的目标. 然后,在仔细研究了经典算法中的搜索过程后,通过引入代价函数和反馈机制,提出了一种最优搜索的概念,使在搜索过程中搜索、评价、反馈不断地进行,最后得到最佳的搜索结果. 实验表明:提出的最优形状子空间在保证主要形状的基础上给出了最大限度的形状变化,最优搜索过程可保证搜索到局部的惟一最优形状. 它们的综合大大改善了动态形状模型的性能,并提高了搜索的精确性.

关键词:动态形状模型;形状子空间;搜索子空间;主分量分析

中图分类号:TP391