Combining Active Shape Models and Active Appearance Models For Accurate Image Interpretation

Wei Wang[1, 2], *Shiguang Shan*[2], *Wen Gao*[2, 3], *Baocai Yin*[1]

[1)] College of computer science, Beijing polytechnic university, Beijing, China, 100022
[2)] ICT-YCNC FRJDL, Institute of Computing Technology, CAS, Beijing, China, 100080
[3)] Vilab, Computer College, Harbin Institute of Technology, Harbin, China, 150001
{ wwang, sgshan, wgao}@jdl.ac.cn, yinbc@bjpu.edu.cn

ABSTRACT

In this paper, we propose to combine Active Shape Models (ASMs) and Active Appearance Models (AAMs), for reliable image interpretation. In the proposed method, ASM local searching strategy is firstly used, then, besides the statistical shape constraints as in the ASMs, global texture constraint based on the subspace reconstruction error is further exploited to evaluate the fitting degree of the currently estimated model to the novel image. And, similar to the strategy used in the AAM approach, global texture is used to predict and update the model parameters. Therefore, by such an interleave iteration of ASM and modified AAM, our method takes the advantages of both ASMs and AAMs, while avoids their shortcomings. Our experiments on the face database significantly show the effectiveness of our method.

Keywords: Image interpretation, Face alignment, Active Shape Model (ASM), Active Appearance Model (AAM),

1. INTRODUCTION

In most pattern recognition and computer vision tasks, it is very important to extract the target from the images. So the localization and alignment of target object from an image is a task of great importance. In recent years many methods have been put forward to deal with the problem including active contour models (snake)[1], elastic bunch graph matching [2], deformable template [3], Gabor wavelet networks [4], Active Shape Models (ASM) [5] and Active Appearance Models (AAM)[6].

Among all these methods, ASM and AAM are both based on statistical models, which are demonstrated to be efficient and effective for image interpretation. In ASMs, local texture on the direction perpendicular to the contour, so called profile, is exploited to model the local texture of each landmark and search for the landmarks locally. The global shape models are then applied to "correct" the local search result according to the statistical shape model. Obviously, in the ASMs, only local texture is used, while the abundant global texture is not utilized to constrain the face shape extraction. Therefore, ASMs are prone to trap into local minimal because of ambiguous local texture or poor starting location.

While in AAMs, global statistical shape and texture constraints are combined to build appearance model. What is more, a linear prediction model is built to predict the appearance parameters for optimization. So not only the shape but also the texture of the target object can be analyzed by the analysis-by-synthesis paradigm. However, since the linear relations existed only in limited range. Therefore, AAMs may not be able to localize individual landmarks accurately due to complex background or target image variations.

Based on the above observations, in this paper we present a new method to combine the ASM's local texture models and AAM's global appearance models. To integrate the local profile and global appearance constraints, the subspace reconstruction residual of the global texture is exploited to evaluate the fitting degree of the current model to the novel image. And, similar to the AAMs, global texture is used to predict and adjust the model parameters. Therefore, by such an interleave iteration of ASM and modified AAM, our method takes the advantages of both ASMs and AAMs, while avoids their shortcomings.

The remaining part of this paper is organized as follows: in section 2, ASM and AAM method are reviewed briefly. In section 3, we describe our method in details. Experimental results are presented in section 4 to verify the effectiveness of our method. A short conclusion is drawn in the last section.

2. BRIEF REVIEW OF THE ASM AND AAM

2.1 Active Shape Model

Shape models are built from a set of annotated images. The annotated points in each image can be represented as a vector. After aligning these vectors into a common coordinate, principal component analysis is applied to get a set of orthogonal basis P. Every aligned shape can be approximately represented as $x \approx \bar{x} + Pb$ where b is shape parameter. Besides shape model, local texture model for each landmark are built also. Cootes use normalized first derivate of profile to build the local texture model. The distance between a new point (its first derivate profile is g_{new}) and the model is $f(g_{new}) = (g_{new} - \bar{g})^T S_g^{-1}(g_{new} - \bar{g})$ where $\bar{g}$ is the mean of the first derivate of the profile, S_g is their covariance.

Based on the two models, search progress can be done. After initialization, each landmark in the model is optimized by selecting the point with minimum distance mentioned above in the direction perpendicular to the contour in a certain range. Since the new shape is possibly implausible, so we need to adjust the shape parameters to as well as affine parameters in 2-D. Such procedure is repeated until no considerable change is observed.

2.2 Active Appearance Model

In this algorithm, shape model as well as texture model is built. By warping the image enclosed by the shape to the mean shape, shape free texture is obtained. Then we apply PCA to the shape free texture, and get the texture model $g = \bar{g} + P_g b_g$ where $\bar{g}$ is the mean value of the trained shape free texture $\{g_1 \cdots g_n\}$, P_g is the matrix containing k principal orthogonal modes of the covariance in $\{g_1 \cdots g_n\}$, and b_g is texture parameters controlling the variance of texture. Appearance vector is built by concatenating the shape parameters and texture parameters $b = \begin{pmatrix} w_s b_s \\ b_g \end{pmatrix}$ where w_s in a diagonal matrix allowing for the difference between the shape and gray models. By applying PCA to the appearance vector $\{b\}$, the correlations between shape and texture can be removed, leading to a further model $b = Qc$ where Q is the eigenvector and *c* is a vector of appearance parameters.

AAM search is based on the linear assumption between the texture difference δg and the error in the model parameters. The linear relations can be denoted as $\delta c = R_c \delta g$ and $\delta t = R_t \delta g$ where δc is the appearance displacement and δt is the position displacement including translation, rotation, and scale displacement. The linear matrix R_c and R_t can be obtained by offline learning. AAM optimizations can then be done with these linear relations.

3. COMBINING ACTIVE SHAPE MODELS AND ACTIVE APPEARANCE MODELS

In this part, a fitting degree evaluation criterion based on global texture is presented and the reasonability of the criterion is also demonstrated by our experiment. Also, a modified AAM search method is presented. With the fitting degree evaluating criterion, we combine ASM and modified AAM search methods together.

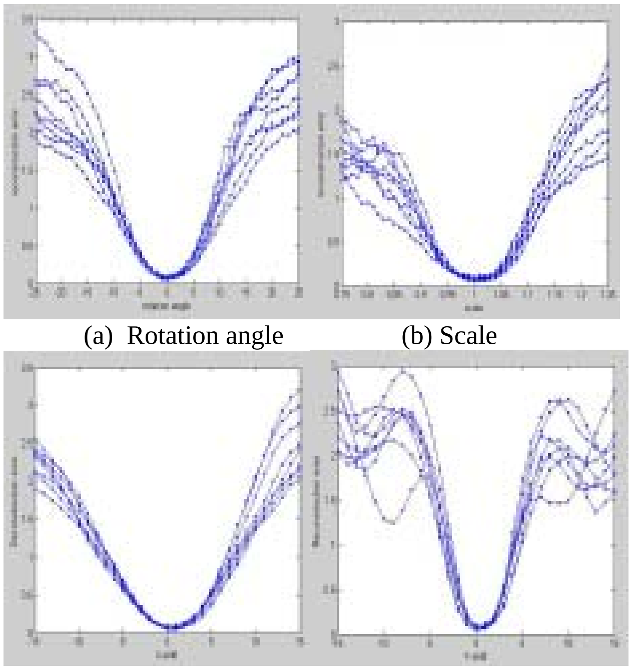
3.1 Fitting Degree Evaluation Criterion

Shape free texture model is built with a number of manually annotated face images by applying PCA to the shape free texture. So the shape free texture of a face image can then be reconstructed as $g = \bar{g} + P_g b_g$ where b_g is the texture parameters calculated as follows: $b_g = P_g^T (g - \bar{g})$. For an input shape in an image, the shape free texture denoted as Φ can be obtained by warping the texture to the mean shape. By making a projection to the trained shape free face texture subspace, we can reconstruct the shape free texture $\Phi' = \bar{g} + P_g (P_g^T (\Phi - \bar{g}))$. The error between the original shape free texture and the reconstructed one can then be calculated by:

$$\varepsilon = \left| \Phi - \Phi' \right|.$$

If shape is well matched to the image, the shape free texture is sure to be a normal face patch without background image and twist in shape. So the reconstruction of such image can achieve good result, and the reconstruction error is relatively small.

In order to find the relations between fitting degree and the reconstruction error, we systematically displace the true shape in scale (from 0.75 to 1.25 to original size), rotation angle (from –25 degree to 25 degree deviation to original) and translation (from –15 pixels to 15 pixels on X and Y axis respectively to original position). 8 face images are used to do the experiments independently. Figure 1 illustrates the corresponding relations between reconstruction error and parameters displacement of our experiment. Each curve represents a series of parameter change for one image. From these figures we can see the approximate monotony relation between reconstruction errors and shape parameters displacement in a certain range. When the displacement is smaller than a certain limit, the monotony relation is quit ideal. So in a certain range, it is reasonable to use the reconstruction error as fitting degree evaluating criterion.



(a) Rotation angle (b) Scale

(c) X-shift (d) Y-shift

Figure 1. Relations between shape parameters variation and reconstruction error

3.2 Updating Shape from Global Texture (modified AAM)

Similar to AAM, We make an assumption that there are linear relations between texture δg (the difference between the trained mean shape free texture and the one enclosed by shape) and the displacement of shape parameter $\delta b = R_b \delta g$ $\delta t = R_t \delta g$, where δb is the shape parameter displacement (in AAM appearance parameter is used), and δt is affine transformation parameters in 2-D. In order to find R_b we generate sets of random displacement $\{\delta b_i\}$ by perturbing the shape parameters of the images annotated manually. So we can get a set of corresponding texture difference $\{\delta g_i\}$. By performing linear regression on these sets of shape parameters displacement and texture difference, matrix R_b can be calculated. As well as perturbations in the shape parameter, small displacements of the position, scale and rotation angle are also modeled to get the matrix R_t . In the optimizing procedure of this mode, shape is optimized using the linear relation mentioned above, which is the similar to the optimization procedure of AAM. But only mean shape free texture is utilized as model texture to calculate the texture difference, for the texture does not change in the procedure as AAM method does.

3.3 Combining ASM and modified AAM based on Fitting Degree Evaluation Criterion

In our search method, to ensure the result of one iteration better fit to the image than the previous one, result of each iteration is evaluated by the fitting degree evaluating criterion, and only the improved one is accepted. First, ASM mode is applied. If the result of one ASM iteration does not satisfy this criterion, the result is abandoned and the optimization mode is changed from ASM to shape prediction method. In turn, the result of shape prediction method is also evaluated by the criterion, and only an improved one may be accepted. Search in this mode will continue until no improvement can be made. The two methods are conducted alternately until no improvement can be made by both methods. Then, convergence can be declared. The whole optimization procedure is outlined as follows:

1. Initializing the mean shape model of ASM.
2. Running one iteration of ASM search.
3. Evaluating the result by our fitting degree criterion. If the fitting result is better than last iteration then go to step 2 for further ASM search, otherwise change the optimizing mode to modified AAM method.
4. Run one step of optimization with modified AAM method.
5. If the result is accepted by the criterion, then go to step 4 and continue to optimize in this method. Otherwise go to step 2 again.
6. If both methods can't improve the fitting degree any longer, convergence is declared.

4. EXPERIMENTS

In this part we introduce our experimental system, and show some experimental results. Performance comparison between standard ASM and our method is made also.

4.1 Face Database

The face database we use includes 500 face images. The ratio between man and woman is about 3:2. Facial expressions include laugh, surprise and angry. The illumination variance is minor, and the pose is near frontal. The size of the images is all 256*256 while face account for about 120*160 pixels. All these images are annotated with a point annotating tool designed by us.

4.2 Experimental Method And Results

In our experiments, the performance is evaluated by calculating the mean point distance between the result shapes and the pre-annotated ones. The distance can be denoted as $d = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n}\sum_{j=1}^{n} dist(P_{ij}, P_{ij}^{'})\right)$ where N is the total number of the test images, n is the number of the landmark points in the shape (for our case, n=103), P_{ij} and $P_{ij}^{'}$ are the j-th landmark point of the i-th test image of pre-annotated shape and the result one correspondingly. The function $dist(p1, p2)$ is the Euclidean distance between the two points.

In our experiment, we randomly select 350 images from the 500 images database to build models of both texture and shape, while use the other 150 for test. 10 times of similar experiments are conducted to evaluate the performance the three methods. So the final evaluation criterion can be denoted as $d = \frac{1}{10}\sum_{i=1}^{10} d_i$. Similarly, the mean variance is calculated. The initialization of mean model is based on the pre-annotated irises location in all these experiments.

Table 1 shows the experimental result. The first row is the mean error and variance after initialization of mean shape model. The following rows present the results of corresponding approaches. From this figure we can see the improvement of our approach with ASM methods. In figure 2, we present some results of our experiment. As is illustrated in this figure, the fitting result of mouth is not ideal with ASM method due to the variation of expression. With our proposed method, the results are much improved.

Table 1. Performance comparison of different methods

Method	Error d (pixel)	Error Variance
Initial shape	3.83	5.06
ASM	3.05	3.23
Our method	2.59	2.71

5.CONCLUSTIONS AND FUTRUE WORK

In this paper we present a new method for image interpretation base on statistical shape and appearance model. The method combine the advantages of ASM and AAM while avoid their shortcomings. Also, a fitting degree evaluation criterion proposed which are proved to be reasonable and effective by our experiment. Experiments show that our proposed method performs better than standard ASM.

Our future work should be focused on more improvements of ASM by global texture constraint and the utilization of our fitting degree evaluation criterion on other tasks such as pose and expression estimation.

Results of ASM
The mouth part falls to local minima

Results of our method

Figure 2. Comparison between our method and ASM

ACKNOWLEDGEMENT

This research is sponsored partly by Natural Science Foundation of China (No.69789301), National Hi-Tech Program of China (No.2001AA114190), SiChuan Chengdu YinChen Net. Co. (YCNC), 100 Talents Foundation of Chinese Academy of Sciences and Natural Science Foundation of Beijing (4011001).

REFERENCE

[1] M.Kass, A.Witkin,and D.Terzopoulos. Snakes: Active contour models. Int. Journal of Computer Vision, pp.321-331, 1988.

[2] L. Wiskott, J.M. Fellous, N. Kruger, C.v.d.Malsburg, Face Recogniton by Elastic Bunch Graph Matching, IEEE Trans. On PAMI, Vol.19, No. 7, pp775-779, 1997

[3] A.L.Yuille, Deformable templates for face detection, J. Cogn. neurosci. 3, 59-70, 1991.

[4] V. Kr"uger and G. Sommer. Gabor wavelet networks for object representation. Technical Report 2002, Institute of Computer Science, University of Kiel, 2000.

[5] T.F.Cootes, C.J.Taylor. "statistical models of appearance for computer vision", http://www.isbe.man.ac.uk/~bin/refs.html, 2001

[6] T.F.Cootes, C.J.Taylor, D.Cooper, and J. Graham. Active shape models--their training and application. Computer vision and image understanding, 61(1): pp38-59, 1995.