SHAPE EVALUATION FOR WEIGHTED ACTIVE SHAPE MDOELS

Ming Zhao[1*] *Stan Z.Li*[2] *Chun Chen*[1] *Jiajun Bu*[1]

[1] School of Computer Science and Technology, Zhejiang University, Hangzhou, China
[2]Microsoft Research China, Beijing Sigma Center, Beijing, China
Contact: zhaoming1999@hotmail.com

ABSTRACT

Active Shape Models (ASM) is a powerful statistical tool for face alignment. However, no evaluation is performed on the final results. Nevertheless, the shape evaluation information is very useful for the search and the final results. In this paper, a shape evaluation method and a new search algorithm, called weighted ASM, are proposed. The shape evaluation is based on the local appearance model of ASM to determine how well the searching shape match models derived from the training set. It is used to guide the search procedure to get more accurate results. The weighted-ASM also uses this evaluation information to project the searching shape into the solution shape space in a weighted way. Compared with ASM's orthogonal projection, the weighted projection can drag the search out of local minima to be more accurate and more robust. Experiments have been done to show the ability of this method to align shapes.

1. INTRODUCTION

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. Extensive research has been conducted in the past 20 years. Kass *et al* [1] introduced Active Contour Models, an energy minimization approach for shape alignment. Kirby and Sirovich [2] described statistical modeling of grey-level appearance but did not address face variability. Wiskott *et al* [3] used Gabor Wavelet to generate a data structure named Elastic Bunch Graph to locate facial features. It is demonstrated very useful. However it is time-consuming and need large computation.

Active Shape Models (ASM) and Active Appearance Models (AAM), proposed by Cootes *et al* [4][5], are two popular shape and appearance models for object localization. They have been developed and improved for years. In ASM [4], 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 nice results in shape localization. AAM [5] combines constraints on both shape and texture in its characterization of face appearance. There are two linear mappings assumed for optimization: from appearance variation to texture variation, and from texture variation to position variation. The shape is extracted by minimizing the texture reconstruction error. According to the different optimization criteria, ASM performs more accurately in shape localization while AAM gives a better match to image texture. On the other hand, ASM tends to be stuck in local minima, dependent on the initialization. AAM is sensitive to the illumination, in particular if the lighting in the test is significantly different from the training. Meanwhile, training an AAM model is time consuming.

In this paper, we present a shape evaluation method and a weighted active shape model (weighted ASM) using such evaluation information. The shape evaluation is based on the local appearance model of ASM to determine how well the searching shape match models derived from the training set. It is used to guide the search procedure to get more accurate results. The weighted-ASM also uses this evaluation information to project the searching shape into the solution shape space in a weighted way. Compared with the original method used in ASM, the weighted ASM can achieve more accurate results. Experimental results demonstrate that weighted ASM achieves better results than ASM does.

The rest of the paper is arranged as follows. The origin ASM algorithm is briefly described in Section 2. In Section 3, we present our method of shape evaluation and weighted ASM. Experimental results are presented in Section 4 before conclusions are drawn in Section 5.

* This paper is supported by National Natural Science Foundation of China (60203013).

2. OVERVIDE OF ASM ALGORITHM

2.1. Statistical Shape Models

Here we describe briefly the statistical shape models used to represent deformable objects.

The ASM technique relies upon each object or image structure being represented by a set of points. The points can represent the boundary, internal features, or even external ones, such as the center of a concave section of boundary. Points are placed in the same way on each example of the training set of examples of the object. This is done manually. One example is shown in figure 1. By examining the statistics of the positions of the labeled points a " Point Distribution Model" is derived. The model gives the average positions of the points, and has a number of parameters that control the main models of variation found in the training set.

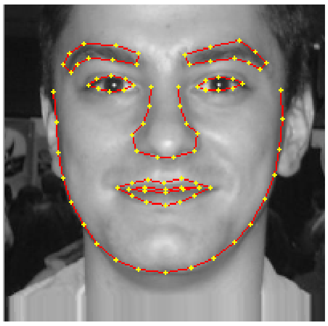

Figure 1. Labeled image with 87 landmarks

The points from each image are represented as a vector x_i and aligned to a common co-ordinate frame. Principle Component Analysis [2] is applied to the aligned shape vectors to generate the ASM model. Three steps are needed for this task.

1) Compute the mean of the aligned shapes

$$\overline{x} = \frac{1}{s}\sum_{i=1}^{s} x_i$$

where s is the number of shapes.

2) Compute the covariance of the data

$$S = \frac{1}{s-1}\sum_{i=1}^{s}(x_i - \overline{x})(x_i - \overline{x})^T$$

3) Compute the eigenvectors, f_i and corresponding eigenvalues l_i of S (sorted so that $l_i \geq l_{i+1}$).

Finally, the ASM model can be written as:

$$x = \overline{x} + \Phi\, b \qquad (1)$$

where $\overline{x}$ is the mean shape vector, $\Phi = \{f_1 \mid f_2 \mid \cdots \mid f_t \mid\}$ contains the t eigenvectors corresponding to the largest eigenvalues, and b is a vector of shape parameters. For a given shape x, its shape parameter b is given by

$$b = \Phi^T (x - \overline{x})$$

The vector b defines a set of parameter of a deformable model. By varying the elements of b we can vary the shape x, using the equation (1). By applying limits of the parameter b we ensure that the shape generated is similar to those in the original training set.

The ASM search procedure is an iteration procedure of two steps: local appearance matching and estimating of shape parameters. Each time it uses local appearance model to find a new shape. Then it updates the model parameter to best fit the new search shape [4].

2.2. Local Appearance Models

The local appearance models, which describe local image features around each landmark, are modeled as the first derivative of the sample's profiles perpendicular to the landmark contour to reduce the effects of global intensity changes [4]. They are normalized by dividing through by the sum of absolute element values.

It is assumed that the local models are distributed as a multivariate gaussian. For the *jth* landmark, we can derive the mean profile g_j^x and the covariance matrix S_{g_j} from the *jth* profile examples. The quality of fitting a new sample g_j^x to the model is given by $f_j(g_j^x) = \left(g_j^x - \overline{g}_j\right)^T S_{g_j}^{-1}\left(g_j^x - \overline{g}_j\right)$, calculating the Mahalanobis distance of the sample from the model mean. It is linearly related to the log of the probability that g_j^x is drawn from the distribution. At the current position, when searching points, the local appearance models find the best candidate in the neighborhood of the search point, by minimizing the $f_j(g_j^x)$, which is equivalent to maximizing the probability that g_j^x comes from the distribution.

Using the local appearance models lead to fast convergence to the local image evidence. However, due to the variation of the illumination and image quality, sometimes the feature point is not accurately located. So some points could be stuck in local minima, despite that they are not the destination points. Fortunately, most of these local minima have bigger distance from the distribution, because it is less probably that they come from this distribution. In this paper, this information is used to drag these points out of the local minima so as to give more accurate results.

3. SHPAE EVALUATION FOR ASM

3.1. Problems from ASM Shape Matching

For the general problems of matching a model instance to an image, there are several approaches which can all be thought of as optimizing a cost function. For a set of model parameters p , we can generate an instance of the model projected into the image. We can compare this hypothesis with the target image, to get a function $F(p)$. The best set of parameters to match the object in the image is then the set which optimizes this measure. For instance, if $F(p)$ is an error measure, which tends to zero for a perfect match, we would like to choose parameters, p , which minimize the error measure. Thus, in theory all we have to do is to choose a suitable fit function, and use a general purpose optimizer to find the minimum. The minimum is defined only by the choice of function, the model and the image, and is independent of which optimization method is used to find it.

As one of such matching approaches, active shape model (ASM) is to find the set of shape parameters which best match the shape model to the shape of the object in the image. However, in the case of ASM, the form of the fit measure for the shape models is harder to determine. If we can assume that the shape model represents boundaries and strong edges of the object, a useful measure is the distance between a given model point and the nearest strong edge in the image strongest nearby edges. This measure relies upon the target points. If some of the shapes are not the strongest edges, it will not be a true measure of the quality of fit. Given no initial knowledge of where the target object lies in an image, we can only use the information in the training images. So in this paper, rather than looking for the best nearby edges, we sample the image around the current model points, and determine how well the image samples match models derived from the training set, which will be described below.

3.2. Shape Evaluation Using Local Appearance Models

As the form of the fit measure for the shape models is hard to determine exactly, we use shape evaluation to approximate it instead. In this paper, shape evaluation is to determine how much the current shape resembles the destination shape. Because shape is represented by a set of points, each of which has a local appearance model, the shape evaluation proposed in this paper is based on this local appearance model. By this method, both the points in the shape and the whole shape can be evaluated.

The local appearance models find the best searching candidates in their neighborhood by minimizing $f_j(g_j^x)$ when ASM searching points, as discussed in section 2.2. For the *jth* point of shape x, the minimum value $f_{\min}^j = Min[f_j(g_j^x)]$ indicates the likelihood that the searching candidate point comes from the local appearance model.

Now, we can evaluation each point in a shape by $f_{\min}^j$. But how can we evaluation the whole shape? As it is composed of all the points, we define $E(x) = \sum_{i=1}^{N} f_{\min}^i$, where N is the number of points in the shape, as the shape's evaluation. It is reasonable because $E(x)$ is linearly related to the log of the product of the probabilities that each g_j^x come from the *jth* point's distribution. This shape evaluation can measure the extent that the current shape gets close to the destination shape. The smaller the evaluation is, the closer the shape gets to the destination.

3.3. Shape Evaluation for Choosing the Best Searching Result within a Multi-resolution Level

The searching procedure is implemented in a multi-resolution way. To get the best searching shape of a level, we must have a measure to decide which one is the best. However, we haven't such a measure. So, we can use the shape evaluation information for this purpose. Though the shape evaluation discussed in section 3.2 can measure the extent that the current shape gets close to the destination shape, it does not always work right. So we have to adjust its results when it is used. Fortunately, we have observed that the best searching shape nearly always lies in the three shapes which have the smallest shape evaluation value. So we can choose the best one from these three shapes. As shown in Figure 2, S_1, S_2, S_3 are the three shapes with the smallest shape evaluation value.

Now we can calculate the distance between these three shapes. We can think that the better two shapes will have smaller distance, because they better shapes tend to converge. Without loss of generality, if we assume $d_1 < d_2 < d_3$, the best shape would be S_1 or S_2. We can choose one from them. The problem is now which is the better of them. If the other shape S_3 is a good shape, the best shape may have smaller distance from it. In this case, the best shape would be S_2. If S_3 is a bad shape, the best shape may have bigger distance from it. IN this condition, S_1 would be the choice. But how can we decide when the other shape S_3 is good or not? We do this by checking the value $\frac{d_2}{d_1}$. If $\frac{d_2}{d_1} < r$, S_3 is thought to be a good shape, otherwise it is a bad shape.

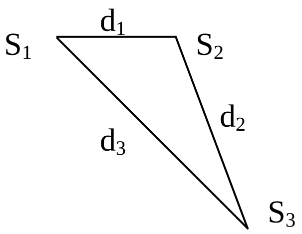

Figure 2. Three shapes with the smallest evaluation value

3.4. Shape Evaluation for Weighted Projection to the Solution Shape Space

The ASM search procedure is an iteration procedure of two steps: local appearance matching and estimating of shape parameters. Local appearance matching is performed using local appearance models derived from the training set and can be evaluated by shape evaluation. Here we will address the problem of the shape parameter estimation.

For the current searching shape x, ASM try to generate a corresponding shape x^p in the solution shape space in order that x^p is similar to those in the original training set. Here we use solution shape space to denote the shape space spanned by equation (1), which represents the global shape model obtained from the training set. To find the solution shape x^p for the searching shape x is a projection from the searching shape space to the solution shape space. So x^p can be denoted as

$$x^p = \bar{x} + \Phi b$$

where $\Phi = \{f_1 | f_2 | \cdots | f_t |\}$ containing the t eigenvectors corresponding to the largest eigenvalues, and b is the t dimensional shape parameter of the solution shape space. The projection shape x^p is determined by Φ and b. Usually Φ is chosen empirically so that the solution shape space represents some proportion (e.g. 98%) of the total variance of the training shapes. So the projection can be finished by only calculating the shape parameter. This is performed by $b = \Phi^T(x - \bar{x})$. In the words of PCA, x^p is the principal component projection of x, which minimizes the squared reconstruction error

$$ERR = \| x - x^p \|^2 = (x - x^p)^T (x - x^p)$$

ASM is an orthogonal projection scheme. We will see that by the means of shape evaluation, a more rational projection scheme can be derived, that is the weighted projection scheme.

As discussed in section 3.2, for a given point *j* in a shape, the searching minimum value $f^j_{\min}$ indicates the likelihood that the searching candidate point comes from the point *j's* local appearance model. The smaller $f^j_{\min}$ is, the closer the searching candidate point might be to the destination position. Based on this observation, the projection shape should be closer to the points with smaller $f^j_{\min}$. Now, in order that the projection shape is closer to the points with smaller $f^j_{\min}$, another shaper x^p_w should be found in the solution shape space to minimize another reconstruction error

$$ERR_w = (x - x^p)^T W (x - x^p)$$

where

$$W = \begin{bmatrix} w_1 & & \\ & w_i & \\ & & w_{2N} \end{bmatrix}$$

is a $2N \times 2N$ diagonal weight matrix with

$$w_{2i-1} = w_{2i} = 1 / f^i_{\min}$$

So the next task is to find a shape x^p_w to minimize ERR_w. As $x^p_w = \bar{x} + \Phi b$, we can have

$$ERR_w = f(b) = (x - \bar{x} - \Phi b)^T W (x - \bar{x} - \Phi b)$$

$$= \sum_{i=1}^{2N} w_i (x_i - \bar{x}_i - \sum_{j=1}^{t} f_{ij} b_j)^2$$

Then we can get

$$\frac{\partial f(b)}{\partial b_k} = \sum_{i=1}^{2N} \frac{\partial \left[w_i (x_i - \bar{x}_i - \sum_{j=1}^{m} f_{ij} b_j)^2 \right]}{\partial b_k}$$

$$= \sum_{i=0}^{n} -2 \times w_i \times (x_i - \bar{x}_i - \sum_{j=1}^{m} f_{ij} b_j) f_{ik}$$

Let $\frac{\partial f(b)}{\partial b_k} = 0 \, (k = 0,..,t)$, we now have,

$$f^T W f \, b = f^T W (x - \bar{x})$$

So

$$b = \left[f^T W f \right]^{-1} f^T W (x - \bar{x}) \quad (2)$$

Now the weighted projection is performed by calculating the shape parameter in the solution shape space with equation (2).

3.5. Adjust the Weight Matrix

In the previous section, we simply choose weight matrix W that $w_{2i-1} = w_{2i} = 1/f^i_{\min}$. There is a problem with this method, i.e. if is wrong in the case that point *j* is a local minima or false alarm, the weight for this for this point will be wrong. So measures must be taken to reduce possibility that such things happen. In this paper, two methods are used. First, the stabilities of the searching points' evaluation and position are used to adjust the weight matrix. Then noises in the weight matrix are filtered.

The stability of the searching point's evaluation $f^j_{\min}$ indicates whether it is correct or not. For example, if point *j* is not close to the destination position, $f^j_{\min}$ will vary a lot. Let $f^j_{\min(t)}$ and $f^j_{\min(t+1)}$ to be the successive evaluation of point *j*, and define $d^j_{ft} = | f^j_{\min(t)} - f^j_{\min(t+1)} |$. If point *j* is close to the destination position, d^j_{ft} will be small, which means that $f^j_{\min}$ has converged. So we can increase its weight when $f^j_{\min}$ is small.

The correctness of $f^j_{\min}$ can also be checked through the point's position. If point *j* has not located near the destination position, its position will vary a lot, and the distance between the same points in successive searching shapes will be large. If it has converged to the destination position, the distance will be small. Let p^j_t and p^j_{t+1} to be the successive positions of point *j*, and define $d^j_{pt} = | p^j_t - p^j_{t+1} |$. If d^j_{pt} is small, it means that point *j* is probably close to the destination position. So we can increase its weight when d^j_{pt} is small.

As $f^j_{\min}$, d^j_{ft} and d^j_{pt} are all negatively related to point *j's* weight, we can combine them linearly and finally define $w_{2j-1} = w_{2j} = 1/(f^i_{\min} + a d^j_{ft} + b d^j_{ft})$, where a, b are related to the local appearance model and the current searching level.

If we look at the diagonal elements of the weight matrix as a sequence of signal, there could be noise in it. As we want to filter the local minima, so we must be careful about the large elements. If w_{2j} is large, then $w_{2(j-1)}$ and $w_{2(j+1)}$ will be large too. We define

$$l = Max(| w_{2j} - w_{2(j-1)} | / w_{2j}, | w_{2j} - w_{2(j+1)} | / w_{2j}$$

If l is small than a threshold, point *j* is probably the local minima. So small weight must be given to it. We must note that as the shape points of different organs in the face (e.g. nose, eye, and mouth) are spatially continuous, their noises must be filtered respectively.

3.6. Weighted-ASM Search Procedure

Our full search procedure is similar to ASM method. It is implemented in a multi-Resolution way. The whole iterative procedure is as follows:

1. Use face detection algorithm to detect face and initialize the shape x
2. Set level L to be the maximum level $L_{\max}$, i.e. $L = L_{\max}$
3. Search each local point and get the new shape x'
4. Use weighted projection(section 3.3) to project x' into the solution shape space, and get the parameter b
5. Apply constraints on b
6. if the shape converged or the maximum iterations have been applied at this resolution
   a) choose the best result of this level using the method described in section 3.2
   b) If $L > 0$, let $L = L - 1$ and go to step 3, otherwise go to step 8
7. Go to step 3.
8. Final result is given by the parameter b at level zero

4. EXPERMENTS

We manually labeled 400 pictures with the size of 200×200. 87 landmarks are labeled on each image. We select 200 images as the training and the other 200 images as the test images. We compare the distance between search shape and the manually labeled shape.

First we calculate the displacement of each single point location result to the corresponding labeled point and get the result as shown in Figure 3. The x-coordinate is the pixel of displacement between search points and the labeled points. The y-coordinate is the percentage of the number of the points whose displacement to the labeled points is x-value. We can see that weighted-ASM gets more accurate results than ASM.

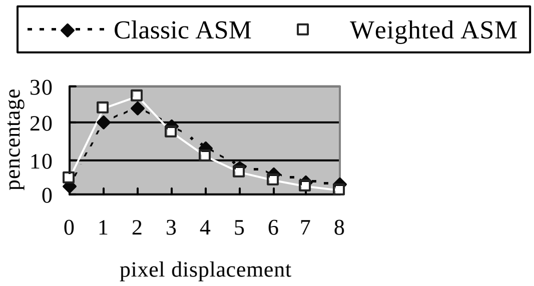

Figure 3. Point displacement of test result.

For each test image, we calculate the overall displacement of the search shape to the labeled shape. The distance of two shapes is defined as follows:

$$Dis = \sum_{j=0}^{P} \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

where P is the total number of landmarks, which is 87 in our system For each test image, we calculate the DisA, which is the distance of ASM search shape to the labeled shape and also the DisW, which is distance of W-ASM search shape to the labeled shape.
We calculate the value

$$m = (DisA - DisW) / DisA \times 100\%$$

Which imply the percentage of improvement of DisW. When m<0, that is *DisW>DisA*, it means that the search result of Weighted ASM is worse than ASM. In the table 1 below, we can see that Weighted ASM works worse in 22 images, and works better than ASM in the remaining 178 images.

m (%)	The number of images
<-5	7
-5<m<0	15
0<m <5	38
5<m <10	50
10<m <15	43
15<m <20	24
20<m	23

Table 1. Overall displacement compare

5. COMCLUSIONS

In this paper, a shape evaluation method and a new search algorithm, called weighted ASM, are proposed. The shape evaluation is based on the local appearance model of ASM to determine how well the searching shape match models derived from the training set. It is used to guide the search procedure to get more accurate results. The weighted-ASM also uses this evaluation information to project the searching shape into the solution shape space in a weighted way. Compared with ASM's orthogonal projection, the weighted projection can drag the search out of local minima to be more accurate and more robust. Experiments have been done to show the ability of this method to align shapes. We compare our weighted ASM with the classic ASM algorithm and better results are achieved.

6. REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos. Snames, "Active contour models." 1st International Conference on Conference on Computer Vision, London, June 1987, pp. 259-268.

[2] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, Jan 1990, pp. 103-108

[3] Laurenz Wiskott, jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg, " Face Recognition by Elastic Graph Matching," Intelligent Biometric Techniques in Fingerprint and Face Recognition, eds. L.C. Jain et al., publ. CRC Press, ISBN 0-8493-2055-0, Chapter 11, 1999, pp. 355-396.

[4] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, " Active Shape Models - their training and application." Computer Vision and Image Understanding, vol.61, no.1, Jan 1995, pp. 38-59.

[5] T.F. Cootes, G.J. Edwards and C.J. Taylor, " Active Appearance Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.6, 2001, pp. 681-685.