# PARAMETER OPTIMIZATION FOR ACTIVE SHAPE MDOELS

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

[1] School of Computer Science and Technology, Zhejiang University, Hangzhou, China
[2]Microsoft Research China, Beijing Sigma Center, Beijing, China
Contact: chenc@zju.edu.cn

**ABSTRACT**

Active Shape Models (ASM) is a powerful statistical tool for extracting objects, e.g. face, from images. It is composed of two parts: ASM model and ASM search. In ASM, these two parts are treated separately. First, ASM model is trained. Then, ASM search is performed using this model. However, we find that these two parts are closely interrelated. The performance of ASM depends on both of them. Improvement on one of them does not consequentially improve the overall performance, for it may worsen the other. In this paper, we find the key parameter that relates these two parts: subspace explanation proportion. By optimizing subspace explanation proportion, the overall performance of ASM can improve by a percentage of about 20 in our experiments. Furthermore, this paper proposes to decompose the ASM overall error into ASM model subspace reconstruction error and ASM search error, proving that the square of the subspace reconstruction error is linearly related with the subspace explanation proportion and finding that the square of the search error is a piecewise function of the explanation proportion. This decomposition is a new method for further analysis and possible improvement. Based on this decomposition, we propose a method to estimate the optimal explanation proportion. Experiments show that the estimation is satisfactory.

## 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. 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. 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. In this paper, we will concentrate on ASM.

ASM is composed of two parts: ASM model and ASM search, which are treated separately. First, ASM model is trained. Then, ASM search is performed using this model. However, we find that these two parts are closely interrelated. The performance of ASM depends on both of them. Improvement on one of them does not consequentially improve the overall performance, for it may worsen the other. Unfortunately, this relationship is often neglected by previous work. Some work attempted to improve the ASM model [6]; others attempted to improve the ASM search procedure [7][8][9][10]. In this paper, ASM model and search are considered together. We first find the key parameter that relates these two parts: subspace explanation proportion, a proportion which the subspace can explain of the variance exhibited in the training data. The performance of both ASM model and ASM search is affected by the subspace explanation proportion. Then, we decompose the ASM overall error

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

into ASM model subspace reconstruction error and ASM search error, proving that the reconstruction error is linearly related with the subspace explanation proportion and finding that the search error is a piecewise function of the explanation proportion. With this decomposition, the ASM overall error becomes a function of the explanation proportion. So, to minimize the ASM overall error is to optimize a parameter, the explanation proportion. Finally, we propose a parameter optimization method to find the optimal explanation proportion.

The rest of the paper is arranged as follows. The analysis of ASM algorithm is described in Section 2. In Section 3, we present the decomposition of ASM overall error. And in section 4, estimation for optimal explanation proportion is discussed. Experimental results are presented in Section 5 before conclusions are drawn in Section 6.

## 2. ANALYSIS OF ASM ALGORITHM

ASM is composed of two parts: ASM model and ASM search. ASM model is a statistical shape model. It is also called point distribution model (PDM). ASM model is to build a PCA subspace to approximate the object's shape space. ASM search is to use ASM model to locate the target object.

### 2.1. ASM Model - Statistical Shape Model

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 for face is shown in figure 1. By examining the statistics of the positions of the labeled points, a " Point Distribution Model" is derived [4]. 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.

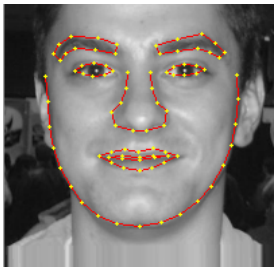

Figure 1. Labeled image with 87 landmarks for face.

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

$\bar{x} = \frac{1}{s}\sum_{i=1}^{s} x_i$ , where $s$ is the number of shapes.

2) Compute the covariance of the data

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

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:

$$x = \bar{x} + \Phi b \quad (1)$$

where $\bar{x}$ 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

$$b = \Phi^T (x - \bar{x}) \quad (2)$$

The vector $b$ defines a set of parameters 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 data. The statistical shape model is a PCA subspace of the object's shape space. In this paper, we find that the size of this PCA subspace model is critical. It can be represented by the number of eigenvectors or the subspace explanation proportion.

The number of modes (eigenvectors), $t$, to retain can be chosen in several ways:

- The usual way is to choose $t$ so as to explain a given proportion (e.g. 98%) of the variance exhibited in the training data. We call this proportion as (subspace) explanation proportion. Let $l_i$ be the eigenvalues of the covariance matrix of the training data. Each eigenvalue gives the variance of the data about the mean in the direction of the corresponding eigenvector. The total variance in the training data is the sum of all the eigenvalues $V_T = \sum l_i$ . We can then choose the $t$ largest eigenvalues such that $\sum_{i=1}^{t} l_i \geq a V_T$ , where $a$ defines the explanation

proportion of the total variation (for instance, 0.98 for 98%).

- Another way is to choose $t$ so that the residual terms can be considered as noise. And an alternative approach is to choose enough modes that the model can approximate any training example to within a given accuracy.

No matter which approach is used, we can use the explanation proportion $a = \sum_{i=1}^{t} l_i / V_T$ as the characteristic of the ASM PCA subspace model. The higher is the explanation proportion, the smaller is the subspace reconstruction error. So, in the usual way, $a$ is chosen as high as 95%~98%. Thus the ASM model can approximate the object's shape accurately. The underlying assumption is that if the reconstruction error is smaller, the ASM overall error will be smaller too. But ASM results are not solely decided by ASM model. They are decided by ASM search too. What's more, ASM search is also affected by the explanation proportion. In section 2.2, we will see that ASM search embeds ASM shape model in its searching procedure. So the explanation proportion $a$ influences both ASM model and ASM search.

Now it is clear that ASM model and ASM search interact with each other by the explanation proportion. Things become more complex. The underlying assumption for choosing $a$ is not right. The ASM overall error depends on ASM model and ASM search. In section 3, we will decompose ASM error into ASM reconstruction error and ASM search error. Unfortunately, this is often neglected by previous work. Some work attempted to improve the ASM model; others attempted to improve the ASM search. In this paper, ASM model and ASM search are considered together. We find that for the best performance of ASM, the explanation proportion $a$ can not be as high as 95%~98%. It is much lower. In our experiments, the optimal explanation proportion is 72%~75%.

### 2.2. ASM Search

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 shape parameter to best fit the new search shape [4].

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. It is assumed that the local models are distributed as a multivariate Gaussian distribution. When searching new shape, every shape point uses its local appearance model to find a best matching point in its neighborhood. All the best matching points form a new shape.

When the new shape is found, it is probably not a plausible object shape. So the ASM model is used to transform the new shape to a plausible object shape. To do this, equation (2) is used to project the new shape to the ASM subspace model to get the shape parameter.

Though ASM search is a very complex procedure, we can hardly find the mathematic description for it. In this paper, we use the search error function to describe its performance.

## 3. DECOMPOSITION OF ASM ERROR

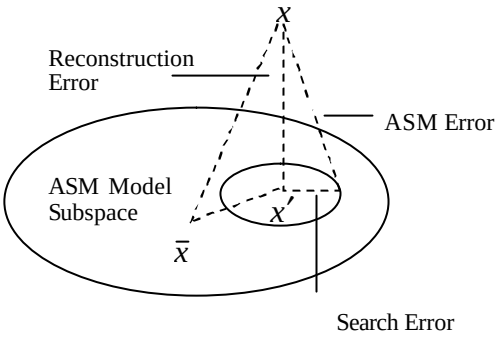


Figure 2. The relationship between ASM error, reconstruction error and search error

As stated in section 2.1, ASM overall error is affected by ASM model and ASM search. In this section, we attempt to decomposition ASM overall error into reconstruction error and search error. The reconstruction error is introduced by the ASM model, and the search error is introduced by ASM search. We use $E_{ASM}$ for ASM overall error, $E_{rec}(a)$ for reconstruction error, and $E_s(a,g)$ for search error, where $a$ is the explanation proportion, and $g$ represents the ASM search procedure. As shown in figure 2, the relationship between them is:

$$E_{ASM}^2 = E_{rec}^2(a) + E_s^2(a,g)$$

So, to minimize $E_{ASM}$, we only need to minimize $E_{rec}^2(a) + E_s^2(a,g)$. If we know the formulation of $E_{rec}(a)$ and $E_s(a,g)$, the work is done.

In this paper, we prove that the square of reconstruction error is linearly related with the explanation proportion, i.e. $E_{\text{rec}}^2(a) = (1-a)V_T$. As shown in figure

2, we can have

$$\| x-\bar{x} \|^2 = \sum_{i=1}^{n} b_i^2 \quad \text{and} \quad \| x'-\bar{x} \|^2 = \sum_{i=1}^{t} b_i^2$$

where $b_i$ is the $x$ 's projection coordinate on the $ith$ eigenvector in ASM model subspace, $t$ is the number of eigenvectors retained, and $n$ is the total eigenvectors of the object's shape space. Then we can get square of the reconstruction error for point $x$

$$E^2_{rec(x)}(a) = \| x-x' \|^2 = \| x-\bar{x} \|^2 - \| x'-\bar{x} \|^2 = \sum_{i=t+1}^{n} b_i^2$$

So the square of the total reconstruction error is

$$E^2_{rec}(a) = \frac{1}{M}\sum_x E^2_{rec(x)}(a)$$

$$= \frac{1}{M}\sum_x \sum_{i=t+1}^{n} b_i^2 = \sum_{i=t+1}^{n}\left(\frac{1}{M}\sum_x b_i^2\right)$$

$$= \sum_{i=t+1}^{n}\left(\frac{1}{M}\sum_x (b_i-0)'(b_i-0)\right) = \sum_{i=t+1}^{n}(l_i)$$

$$= \left(1-\sum_{i=1}^{t}(l_i)\Big/\sum_{i=1}^{n} l_i\right)\cdot\sum_{i=1}^{n} l_i = (1-a)V_T$$

where $M$ is the number of samples.

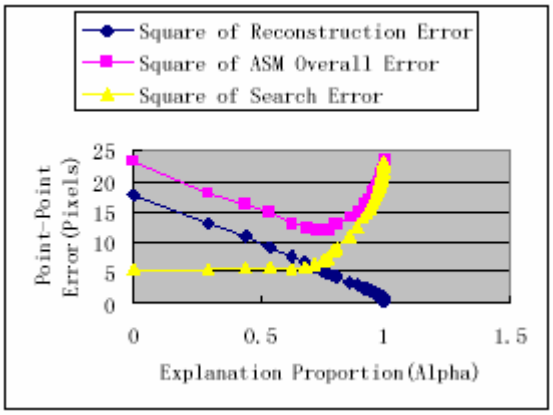


Figure 3. the relationship between the squares of ASM error, reconstruction error, search error and the explanation proportion.

Unfortunately, we can not find the mathematic formulation between the search error and explanation proportion, for the ASM search is a dynamic procedure and can hardly be mathematically described. So we can find their relationship by experiments. We train the ASM model on 200 images and test it on other 200 images. Then we can get the square of ASM overall error and reconstruction error. The square of search error is the difference between them. Figure 3 shows the relationship between the three errors and the explanation proportion. We can find that the square of the search error is a piecewise function of the explanation proportion with two pieces. The first piece is a linear function and the second piece is a quadratic function.

## 4. ESTIMATION FOR OPTIMAL EXPLANATION PROPORTION

As found in section 3, the square of the search error is a piecewise function of the explanation proportion with two pieces. The first piece is a linear function and the second piece is a quadratic function. To estimate the optimal explanation, we need to estimate this piecewise function. But estimation of the quadratic function is neither robust nor efficient. So we try to decrease the order of this function. We first consider the three errors, not the square of them. But we can not find linear relationship either. So we need to find other relationship.

We define a new search error

$$E'_s(b,g) = E_{ASM} - E_{rec}(b)$$

where $b = \sum_{i=1}^{t}\sqrt{l_i} \Big/ \sum_{i=1}^{n}\sqrt{l_i}$ , $t$ is the number of eigenvectors retained, and $n$ is the total eigenvectors of the object's shape space. The relationship between these three errors and $b$ is shown in figure 4. We can see that although none of them is an exact linear function, they are nearly linear or piecewise linear. This property is very useful for the estimation of the optimal explanation proportion.

Now the estimation method is fairly simple. First, we calculate two or three values for the three errors near both $b=0$ and $b=1$ . Second, we linearly fit the reconstruction error and piecewise linearly fit the search error to get the reconstruction error function and search error function. Finally, we minimize the sum of these two functions to get the optimal $b$ . With this optimal $b$ , the number of eigenvectors in the ASM model can be calculated and thus the explanation proportion is there.

## 5. EXPERIMENTS

Our database contains 400 face images in the FERET database, the AR database and other collections. 87 landmarks are labeled on each face. We randomly select 200 images as the training and the other 200 images as the testing images. Multi-resolution search is used, using 4 levels with resolution of 1/8, 1/4, 1/2, 1 of the original image in each dimension. At most 5 iterations are run at each level. The ASM uses profile models of 9 pixels long (4 points on either side) and searches 2 pixels either side.

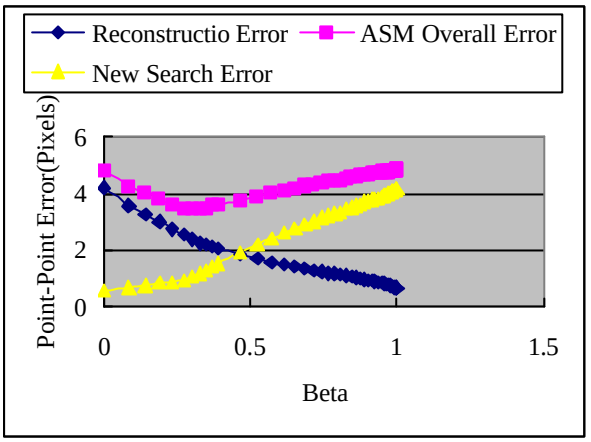


Figure 4. the relationship between ASM error, reconstruction error, new search error and beta.

**5.1. Point Location Accuracy Comparison**

On each test image, we initialize the starting mean shape with displacements from the true position by $\pm 10$ pixels in both $x$ and $y$, 9 initializations in total. Within these displacements, most of the search will converge to the target shape. The searching results are compared with the labeled shapes. We use point-to-point error and point-to-boundary error (the distance from the found points to the associated boundary on the labeled shape) as the comparison measure. The comparison results are shown in Figure 5. We can see that the optimal explanation proportion is 0.72~0.75 for both measure.

The point-to-point error is 4.56 for explanation proportion 98%, 4.27 for explanation proportion 95%, 3.462 for explanation proportion 72.6%, 3.468 for explanation proportion 75.8%. The point-to-point accuracy improvement is 0.8~1.1 pixel, with improvement rate 18~24%.

The point-to-boundary error is 2.62 for explanation proportion 98%, 2.47 for explanation proportion 95%, 2.138 for explanation proportion 72.6%, 2.135 for explanation proportion 75.8%. The point-to-boundary accuracy improvement is 0.3~0.5 pixel, with improvement rate 14~19%.

Figure 6 shows the percentage of the located shapes whose point-to-point errors are less than a given threshold with different eigenvectors (explanation proportion). Five kinds of eigenvectors/explanation proportion are plotted.

**5.2. Capture Range Comparison**

On each test image, we initialize the starting mean shape with displacements from the true position by up to $\pm 30$ pixels in $x$. Then a search is performed to attempt to locate the target shapes. Figure 7 shows the point-to-point errors with different initial displacements and different eigenvectors (explanation proportion). We can see that eigenvector 6 (explanation proportion 72%) is the best within displacements of $\pm 20$ pixels in $x$.

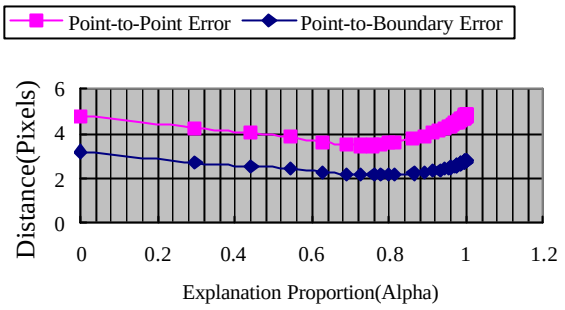


Figure 5. ASM error with different explanation proportion

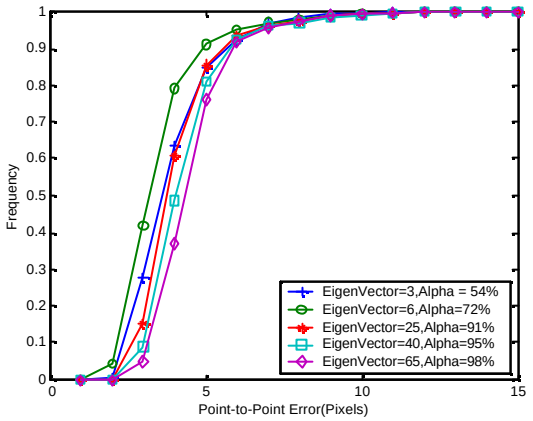


Figure 6. the percentage of located shapes whose point-to-point errors are less than a threshold with different eigenvectors (explanation proportion).

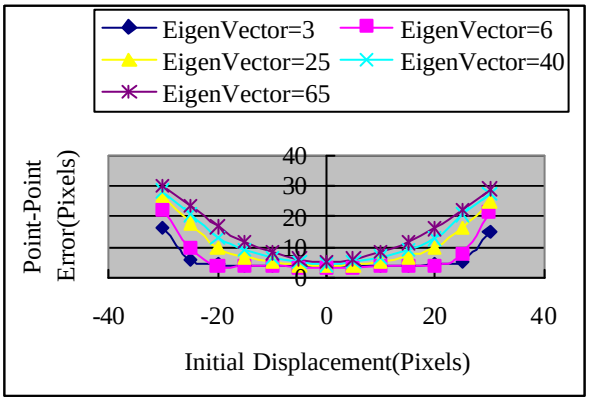


Figure 7. the point-to-point errors with different initial displacements and different eigenvectors.

**5.3. Estimation for Optimal Explanation Proportion**

We use the proposed method in section 4 to estimate the optimal explanation proportion. Three values are calculated for the three errors near both $b = 0$ and $b = 1$. The values are listed in Table 1 and also refer to figure 4.

If we only use $b = 0, 0.082, 0.98, 1$, the piecewise linearity function of the new search error is:

$$y = \begin{cases} 1.11b + 0.59 & (b \geq 0) \\ 5.17b - 0.98 & (b \leq 1) \end{cases}$$

The estimated optimal $b$ value is 0.387 corresponding to the number of eigenvector 10, the explanation proportion 0.81.

If we use all these values, the piecewise function linearity of the new search error is:

$$y = \begin{cases} 1.25b + 0.59 & (b \geq 0) \\ 4.90b - 0.71 & (b \leq 1) \end{cases}$$

The estimated optimal $b$ value is 0.28 corresponding to the number of eigenvector 5, the explanation proportion 0.69. The true optimal number of eigenvector is 6, corresponding to the explanation 0.72~0.75. So the estimated results are satisfactory.

| Beta | Eigenvectors | Reconstruction Error | New Search Error |
|---|---|---|---|
| 0.000 | 0 | 4.212 | 0.590 |
| 0.082 | 1 | 3.593 | 0.681 |
| 0.140 | 2 | 3.276 | 0.763 |
| 0.902 | 85 | 0.954 | 3.740 |
| 0.953 | 120 | 0.828 | 3.954 |
| 1.000 | 170 | 0.650 | 4.196 |

Table 1. the reconstruction error and new search error with different eigenvectors.

## 6. CONCLUSIONS

ASM is composed of two parts: ASM model and ASM search, which are treated separately. Some of the improvements on ASM are for ASM model, others are for ASM search. However, the relationship between them is neglected. This paper finds the key parameter that relates these two parts: subspace explanation proportion. By optimizing subspace explanation proportion, the overall performance of ASM can improve by a percentage of about 20 in our experiments. Furthermore, this paper proposes to decompose the ASM overall error into subspace reconstruction error and search error, proving that the square of the subspace reconstruction error is linearly related with the subspace explanation proportion and finding that the square of the search error is a piecewise function of the explanation proportion. This decomposition is a new method for further analysis and possible improvement. Based on this decomposition, we propose a method to estimate the optimal explanation proportion. Experiments show that the estimation is satisfactory.

## 7. 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.

[6] Shuicheng Yan, Ce Liu, Stan Z. Li, Hongjiang Zhang, Heung-Yeung Shum, Qiansheng Cheng, "Face alignment using texture-constrained active shape models", Image and Vision Computing 21(2003): 69–75.

[7] Ce Liu, Heung-Yeung Shum, Changshui Zhang, "Hierarchical Shape Modeling for Automatic Face Localization." ECCV (2) 2002: 687-703.

[8] M. Rogers and J. Graham "Robust Active Shape Model Search," ECCV (4) 2002: 517-530.

[9] Z. Xue, S.Z. Li, E.K. Teoh. "AI-EigenSname: An Affine-invariant Deformable Contour Model for Object Matching," Image and Vision Computing. Vol. 20, No. 2, pp 77-84, Feb 2002.

[10] Yi Zhou, Lie Gu, Hong-Jiang Zhang, "Bayesian Tangent Shape Model: Estimating Shape and Pose Parameters via Bayesian Inference," CVPR2003.