Markovian Mixture Face Recognition with Discriminative Face Alignment

Ming Zhao
Google Inc.
http://www.zhaoming.name

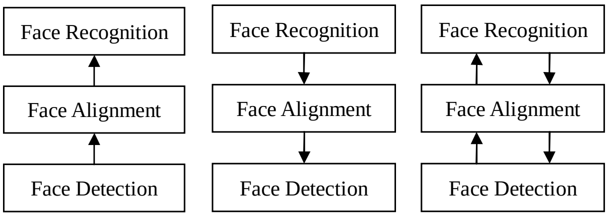
Tat-Seng Chua
National University of Singapore
http://www.comp.nus.edu.sg/∼chuats

Abstract

A typical automatic face recognition system is composed of three parts: face detection, face alignment and face recognition. Conventionally, these three parts are processed in a bottom-up manner: face detection is performed first, then the results are passed to face alignment, and finally to face recognition. The bottom-up approach is one extreme of vision approaches. The other extreme approach is top-down. In this paper, we proposed a Markovian stochastic mixture approach for combining bottom-up and top-down face recognition: face recognition is performed from the results of face alignment in a bottom-up way, and face alignment is performed based on the results of face recognition in a top-down way. By modeling the mixture face recognition as a stochastic process, the recognized person is decided probabilistically according to the probability distribution coming from the stochastic face recognition, and the recognition problem becomes that "who the most probable person is when the stochastic process of face recognition goes on for an infinite long duration". This problem is solved with the theory of Markov chains by properly modeling the stochastic process of face recognition as a Markov chain. As conventional face alignment is not suitable for this mixture approach, discriminative face alignment is proposed. And we also prove that the Markovian mixture face recognition results only depend on discriminative face alignment, not on conventional face alignment. Our approach can surprisingly outperform the face recognition performance with manual face localization, which is demonstrated by extensive experiments.

1. Introduction

A typical automatic face recognition (AFR) system is composed of three parts: face detection, face alignment and face recognition. Given images containing faces, face detection tells where the faces are, face alignment locates the key feature points of faces, and finally face recognition determines who the face is. Many algorithms have been proposed for human face recognition [16]. However, they only focused on each part of the AFR system. Conventionally, these three parts are processed as follows: face detection is performed first, then the detection results are passed to face alignment, and finally results of face alignment are passed to face recognition. This is a bottom-up approach, as shown in Figure 1(a). However, as we know, bottom-up is one extreme of the vision approaches. The other extreme one is the top-down approach [2].



(a) Bottom-Up (b) Top-Down (c) Mixture

Figure 1. Face Recognition Strategies

In the bottom-up approach, each level yields data for the next level. It is a data-driven approach. It uses only class-independent information, and does not rely on class-specific knowledge. For such AFR systems, face detection and face alignment do not use the knowledge about the classes of the persons to be recognized. In order that the bottom-up approach is practical, there are two conditions that it must satisfy [2]: (a) Domain-independent processing is cheap; and (b) For each level, the input data are accurate and it yields reliable results for the next level. However, there are three inherent problems: (1) Class-independent face detection and face alignment may fail for some classes of persons to be recognized. Although face detection is generally good for frontal faces, face alignment is not that good enough. (2) If face detection fails to detect the face or if face alignment can not correctly locate the feature points, the face recognition will usually fail. (3) The recognition process is one-pass and deterministic. Once the recognition fails, it can not be corrected later.

These problems as well as the fact that the vision process does not purely run bottom-up suggest another vision approach: the top-down approach. In the top-down approach,

978-1-4244-2154-1/08/$25.00 2008 IE

the higher level guides the lower level. It makes use of the class-specific knowledge. With class-specific knowledge, the top-down approach could do better for the objects where the knowledge comes from [2, 4]. However, the difficulties with the top-down approach are: (1) There may be large variations within the classes. If the variations can not be properly modelled, they will introduce unexpected errors. (2) In order to model the large variations, various models could be used. The problem is how to choose these models for a particular test example. (3) More efforts are needed to build model for the class-specific knowledge. With the top-down approach for the AFR system, face alignment and face detection can be built based on the classes of persons to be recognized and face recognition guides face alignment and face detection. The top-down face recognition is shown in Figure 1(b).

In order to draw on the relative merits of bottom-up and top-down approaches, a judicious mixture of them will be better [2, 3]. The mixture approach of bottom-up and top-down for face recognition is shown in Figure 1(c). We call it mixture face recognition. In this paper, we propose to incorporate class-specific knowledge to face alignment and combine it with the traditional bottom-up approach. More specifically, this paper will concentrate on combining face recognition and face alignment. Discriminative face alignment (DFA) is proposed to incorporate class-specific knowledge, where a face alignment model is trained for each person. DFA can give good results for itself and bad results for others. So, it can provide discriminative features for face recognition. With the discriminative face alignment, an stochastic mixture face recognition approach is proposed to combine the bottom-up and top-down face recognition, which is properly modeled by a Markov chain.

The rest of the paper is arranged as follows. In Section 2, we present the discriminative face alignment with active shape models. The stochastic mixture face recognition approach with Markov chain is described in Section 3. Experiments are performed in section Section 4 before conclusions are drawn in Section 5.

2. Discriminative Face Alignment

As the top-down approach needs to incorporate class-specific knowledge to face alignment, discriminative face alignment (DFA) is proposed in this paper. It builds a face alignment model for each person to be recognized. This is different from conventional face alignment, which concentrates on general purpose face alignment (GPFA). GPFA builds the model from faces of many persons other than the persons to be recognized in order to cover the variance of all the faces. So it attains the ability of generalization at the cost of specialization. This is to serve for the bottom-up approach. Moreover, GPFA doesn't consider its higher-level tasks. However, the requirements of different tasks may be different, for example, face recognition needs discriminative features whereas face animation requires accurate positions of key points. So it would be better to consider the higher-level task for effective face alignment. As face recognition needs discriminative features, it would be better that face alignment could also give discriminative features. However, the goals of GPFA used in bottom-up approaches is accurate localization. Therefore, the performance of GPFA is not directly related to the performance of the face recognition system.

On the contrary, DFA can provide accurate localization to extract good features to recognize the person on which its model is built. If a being-recognized person is not the person with the discriminative alignment model, the discriminative alignment model will give bad localization so as to extract bad features to prevent the being-recognized person from being recognized as the person with the discriminative alignment model. So, DFA can provide discriminative features for face recognition, which makes it better than GPFA.

2.1. Active Shape Models

Active Shape Models (ASM)[7] and Active Appearance Models (AAM)[5] are most popular face alignment methods. In this paper, ASM is used for DFA. However, similar idea can be also applied for AAM.

ASM is composed of two parts: the shape subspace model and the search procedure. The shape subspace model is a statistical model for the tangent shape space and the search procedure uses the local appearance models to locate the target shapes in the image. Some efforts concentrate on the search procedure [11, 13], while others focus on the subspace model [8, 15]. However, all of these methods only concentrate on GPFA, called GP-ASM in this paper.

To train the ASM shape model, the shapes should first be annotated in the image domain. Then, these shapes are aligned into those in the tangent shape space with the Procrustes Analysis. The ASM shape model is represented by applying principle component analysis (PCA), it can be written as:

$$S = \bar{S} + \Phi_t s \tag{1}$$

where $\bar{S}$ is the mean tangent shape vector, $\Phi_t = \{\phi_1|\phi_2|\cdots|\phi_t\}$, which is a submatrix of Φ (the eigenvector matrix of the covariance matrix), contains the principle eigenvectors corresponding to the largest eigenvalues, and s is a vector of shape parameters. for a given shape , its shape parameter is given by

$$s = \Phi_t^T(S - \bar{S}) \tag{2}$$

2.2. Discriminative Active Shape Model

To build a discriminative active shape model, called D-ASM, for each person, some samples for each person are collected, and they are used to train the ASM model for each person. For an AFR system, if images of each person are labeled during enrollment or registration, the D-ASM model could be built directly from these samples. One problem is that there should be enough variation of each person, otherwise the discriminative alignment model can not generalize well to other faces of the same person. Labelling some images is possible for each person, for example, during enrollment, images can be manually or semi-automatically labeled with the help of constrained search [6] or GP-ASM. And face variation could also be acquired for each person, for example, in the BANCA database [1], each person are recorded 5 images with face variation by speaking some words. Gross *et al.* [10] proposed a person specific face alignment, which is technically similar to D-ASM. However, person specific model is assumed to be built for applications where the identity of the face is known, such as interactive user interface, and it is used to improve the face alignment accuracy. It doesn't provide discriminative features for face recognition, and doesn't consider how to choose between these models.

2.3. Discriminative Features from D-ASM

As D-ASM is able to give good alignment for itself and bad alignment for others, it can provide discriminative features for face recognition, i.e. positions of key feature points. There are small errors of key feature points for good alignment and larger errors for bad alignment. After alignment is performed, key feature points are used to extract the image patch for recognition. As D-ASM can provide accurate alignment of itself and bad alignment for others, the key feature points are discriminative for different persons.

3. Face Recognition with Markov Chain

In this section, the stochastic mixture face recognition is first be introduced. Then, the theory of Markov chains is presented. Finally, the mixture face recognition is properly modeled with a Markov chain and the recognition problem is solved with the basic limit theorem of Markov chains, which also prove that the recognition is only dependent on DFA, not GPFA.

3.1. Stochastic Mixture Face Recognition

The major problem with DFA is how to decide which model to use, which is one of the difficulties of the top-down approach as discussed in Section 1. To deal with this problem, an stochastic mixture approach is proposed to combine DFA and face recognition. The idea is shown in Figure 2. The whole recognition process works in an iterative way: face recognition is performed from the results of DFA in a bottom-up way; then, appropriate DFA models are chosen based on the results of face recognition to further improve face alignment in a top-down way; and face recognition is further improved with the improved face alignment, and the process continues in the same way. Furthermore, the mixture face recognition is performed probabilistically. It can be viewed as a stochastic process, as illustrated in Figure 3:

- For the first-round or initial recognition, GPFA is applied for face alignment. With the initial recognition result, the first-round recognized person P_0 is randomly decided according to an initial recognition probability distribution which comes from the initial recognition. This is different from the traditional deterministic recognition, in which the recognized person is chosen with the highest recognition confidence. The problem with the deterministic recognition is that once the initial recognition is wrong, there is no way to correct it. However, with the probabilistic recognition, the false initial recognition can be corrected later.

- For the second-round recognition, face alignment is performed with the DFA model of person P_0, and the results are used for face recognition. Similar to the first-round recognition, the second-round recognized person P_1 is chosen according to the recognition probability distribution which comes from the second-round face recognition. And the recognition process goes on and on in the same way.

Now, the recognition problem becomes that *"who the most probable person is when the stochastic recognition process goes on for an infinite long duration"*. This problem can be solved with the theory of Markov chains.

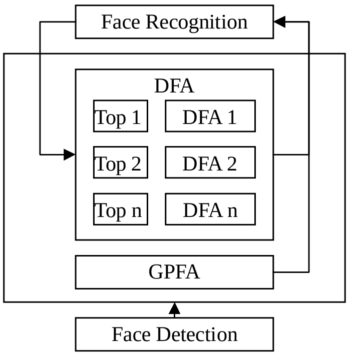


Figure 2. Mixture Face Recognition

3.2. Markov Chains

"A discrete time *Markov chain* is a *Markov process* whose state space is a finite or countable set, and whose (time) index set is $T = (0, 1, 2, \ldots)$.""A *Markov process* $\{X_t\}$ is a stochastic process with the property that, given the value of X_t, the values of X_s for $s > t$ are not influenced by the values of X_u for $u < t$. [12]" In formal terms,

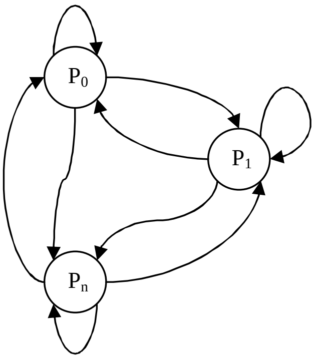


Figure 3. Stochastic Face Recognition with Markov Chain

the Markov property for a discrete time *Markov chain* is that:

$$\begin{aligned} Pr\{X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\ \} \\ = Pr\{X_{n+1} = i_{n+1} | X_n = i_n\} \end{aligned} \tag{3}$$

where $Pr\{\cdot\}$ denotes the probability function.

Given that the chain is in state i at time n, *i.e.* $X_n = i$, the probability of the chain jumping to state j at time $n+1$, *i.e.* $X_{n+1} = j$, is called *one-step transition probability* and denoted by $P_{ij}^{n,n+1}$. That is,

$$P_{ij}^{n,n+1} = Pr\{X_{n+1} = i_j | X_n = i\} \tag{4}$$

If $P_{ij}^{n,n+1}$ is independent of the variable n, we say that the Markov chain has *stationary transition probabilities*, *i.e.* $P_{ij}^{n,n+1} = P_{ij}$. In this case, the matrix $\mathbf{P} = \{P_{ij}\}$ is called the transition probability matrix.

To specify a discrete time Markov chain with stationary transition probabilities, three kinds of parameters are needed:

- State space $\mathbb{S}$. $\mathbb{S}$ is a finite or countable set of states that the random variables X_i may take on. For a finite set of states, the state space can be denoted as $\mathbb{S} = \{1, 2, \dots, N\}$.
- Initial distribution π_0. This is the probability distribution of the Markov chain at time 0. For each state $i \in \mathbb{S}$, we denote by $\pi_0(i)$ the probability $Pr\{X_0 = i\}$ that the Markov chain starts out in state i. $\pi_0(i)$ satisfies the following conditions

$$\pi_0(i) \geq 0 \quad (i \in \mathbb{S}) \tag{5}$$

$$\textstyle\sum_{i \in \mathbb{S}} \pi_0(i) = 1 \tag{6}$$

- Transition probability matrix $\mathbf{P} = \{P_{ij}\}$. P_{ij} is the probability of transition from state i to state j. It satisfies the following conditions

$$P_{ij} \geq 0 \quad (i, j \in \mathbb{S}) \tag{7}$$

$$\textstyle\sum_{j \in \mathbb{S}} P_{ij} = 1 \quad (i \in \mathbb{S}) \tag{8}$$

In the rest of this paper, "Markov chain" represents "discrete time Markov chain with stationary transition probabilities".

3.3. Markov Chain for Mixture Face Recognition

The stochastic mixture face recognition process in Section 3.1 can be properly modeled with a Markov chain in Section 3.2 with the following parameters:

- State space $\mathbb{S}$. The states are the persons to be recognized. Suppose that there are N persons to be recognized. So, the state space can be denoted as $\mathbb{S} = \{1, 2, \dots, N\}$.
- Initial distribution π_0. This is the first-round or initial recognition probability distribution coming from the initial face recognition. GPFA is used for face alignment. Assume that the face recognition algorithm produces recognition distance d_i^r for person i.

  Then, a weight w_i is associated with each person, which is defined as:

$$w_i = 1 - \frac{d_i^r}{\sum_{i=0}^N d_i^r} \tag{9}$$

  Then, the initial probability for person i is

$$\pi_0(i) = \frac{w_i}{\sum_{i=0}^N w_i} \tag{10}$$

- Transition probability matrix $\mathbf{P} = \{P_{ij}\}$. P_{ij} is the probability with which person j will be recognized in the next round when person i is recognized in current round. d_{ij}^r means the recognition distance of person j with the face alignment result from the DFA of person i. A weight w_{ij} is defined between person i and j as follows:

$$w_{ij} = 1 - \frac{d_{ij}^r}{\sum_{i=0}^N d_{ij}^r} \tag{11}$$

  Then, the transition probability from person i to person j is defined as:

$$P_{ij} = \frac{w_{ij}}{\sum_{i=0}^N w_{ij}} \tag{12}$$

With the modeling of Markov chain, the face recognition problem, *i.e. who the most probable person is when the stochastic recognition process goes on for an infinite long duration"*, can be solved by the limiting distribution of Markov chains. The limiting distribution π means that after the process has been in operation for a infinite long duration the probability of finding the process in state i is $\pi(i)$. So, the most probable person is the person with the highest $\pi(i)$.

The following question is whether the limiting distribution exist. Because all the elements are strictly positive, the transition probability matrix for mixture face recognition is regular [12]. According to the basic limit theorem of Markov chains, a Markov chain with a regular transition probability matrix has a limiting distribution π which is the unique nonnegative solution of the following equations:

$$\pi = \pi \mathbf{P} \tag{13}$$

$$\sum_{i \in \mathbb{S}} \pi(i) = 1 \tag{14}$$

Equations (13) and (14) shows that the limiting distribution is only dependent on the transition matrix $\mathbf{P}$, not on the initial distribution π_0. In other words, it proves that the recognition only depends on the DFA (which is used to generating $\mathbf{P}$), not on GPFA (which is used to generating π_0).

Equation (11) and Equation (12) only consider the relative values of recognition distances. On the other hand, the absolute values of recognition distances are also very important because they measure how similar the testing face is to the training faces. So, the absolute values should be combined with the relative values. The final combined distance is

$$d_i^c = d_{ii}^r * (1 - \pi(i)) \tag{15}$$

This equation shows that the more probable this person is recognized and the smaller distance it is to the testing face, the more likely this person is the correct person.

4. Experiments

In this section, we perform experiments on the BANCA face database [1]. The CSU Face Identification Evaluation System [9] is utilized to test the performance of the stochastic mixture face recognition.

Face detection is performed with an AdaBoost face detector. For images with no detected face or more than two detected faces, we manually give the the face detection or manually choose the correctly detected face. Face alignment is performed with unified subspace optimization of ASM[14], which can improve both the accuracy and speed. After face alignment, the images are registered using eye coordinates and cropped with an elliptical mask to exclude non-face area from the image. After this, the grey histogram over the non-masked area is equalized.

4.1. Discriminative Features from DFA

This subsection will validate the statements in Section 2.3 that discriminative face alignment can provide discriminative features. The experiments are performed on the BANCA database. We manually labeled 5 images of each of the 52 persons in session 1 and these images are used to train D-ASM. And GP-ASM is trained on the labeled images in session 1 from the other group, *i.e.* from G1 and G2 alternatively. The testing images are from session 2, each person with 2 images whose faces are manually labeled. So there are totally 104 testing images. The results are evaluated by the average reconstruction error (RecErr) and the average point-to-point errors of all the feature points (AllErr), the key feature points (KeyErr) (including eye centers, nose tip and mouth center) and eye centers (EyeErr). To get the reconstruction error, the texture PCA model is built from another 200 labeled faces. The following experiments are conducted: (1) GP-ASM-A: GP-ASM is used to align all the testing images; (2) D-ASM-O: D-ASM is used to align all the testing images of other persons. (3) D-ASM-S: D-ASM is used to align only the testing images of itself. Results are shown in Table 1. These results show that D-ASM-S gives more accurate results than GP-ASM, and it can give significantly better results than D-ASM-O. This clearly shows that D-ASM can provide discriminative features.

	AllErr	KeyErr	EyeErr	RecErr
GP-ASM-A	4.88	3.33	3.23	15.18
D-ASM-O	9.75	6.93	6.67	36.77
D-ASM-S	3.05	2.20	2.10	13.42

Table 1. Results of GP-ASM and D-ASM

4.2. Mixture Face Recognition on BANCA

The BANCA database contains 52 subjects (26 males and 26 females). Each subject participated in 12 recording sessions in different conditions and with different cameras. Session 1-4 contain data under controlled conditions while sessions 5-8 and 9-12 contain degraded and adverse scenarios respectively. To minimize the impact of illumination and image quality, we choose to use session 1-4. For BANCA, we manually labeled 87 landmarks for session 1, i.e. totally 260 ($= 52 * 5$) faces. Session 2 - 4 client attack faces are used for testing, totally 780 ($= 52 * 15$) faces. So, there are 260 faces as gallery set, and 780 faces as probe set. D-ASM is trained on five images for each person. PCA and LDA are used to extract the feature vector and Euclidean distance is applied. The Makovian mixture recognition results are shown in Figure 4(for PCA) and Figure 5 (for LDA). We selected top 5, top 10, and top all for Markovian mixture recognition. And we compare them with FPFA and manual alignment results. It's clearly shown that Markovian mixture face recognition can consistently and significantly give better results than GPFA. It also shows that Markov mixture face recognition, espcially with PCA, is very close to manual alignment results in the top 1 recognition precision, and it can even give better recognition results for the recognition tail, i.e. it can give better recognition recall than manual alignment.

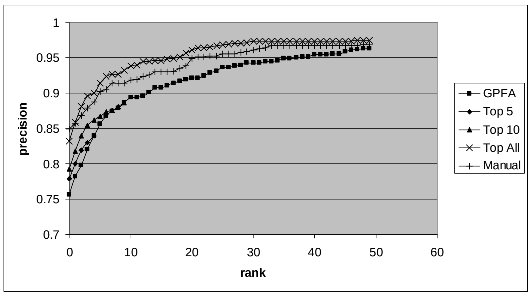


Figure 4. Face Recognition Results with PCA

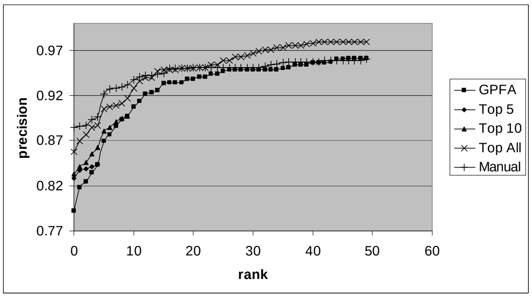


Figure 5. Face Recognition Results with LDA

5. Conclusions

Conventional face recognition is only a bottom-up approach. This paper proposed to use the top-down approach and combine it with the bottom-up approach. In particular, a stochastic mixture approach is proposed for combining face alignment and face recognition. The recognition process works as a stochastic process, and it is properly modeled by a Markov chain, and the recognition problem is solved with the basic limit theorem of Markov chains. Discriminative face alignment is also proposed to incorporate class-specific knowledge, and it can provide discriminative features for better face recognition. Proof is done to show that the face recognition results are dependent only on discriminative face alignment, not on conventional face alignment. Experiments demonstrated that the mixture face recognition algorithms can consistently and significantly improve the face recognition performance, and even surprisingly outperform the performance with manual face localization. Future work includes improving other face recognition algorithms with the stochastic mixture face recognition and incorporating face detection in the whole mixture face recognition framework.

References

[1] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, and J.-P. Thiran. The BANCA database and evaluation protocol. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication, Surrey*, Berlin, 2003. Springer-Verlag. 3, 5

[2] D. H. Ballard and C. M. Brown. *Computer Vision*, chapter 10, pages 340–348. Prentice-Hall, 1982. 1, 2

[3] E. Borenstein, E. Sharon, and S. Ullman. Combining top-down and bottom-up segmentation. volume 04, 2004. 2

[4] E. Borenstein and S. Ullman. Class-specifc, top-down segmentation. In *ECCV*, pages 109–122, Copenhagen, Denmark, May 28-31 2002. 2

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV98*, volume 2, pages 484–498, 1998. 2

[6] T. F. Cootes and C. J. Taylor. Constrained active appearance models. In *Proceedings of IEEE International Conference on Computer Vision*, pages 748–754, Vancouver, Canada, July 2001. 3

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: Their training and application. *CVGIP: Image Understanding*, 61:38–59, 1995. 2

[8] R. H. Davies, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging*, 21:525–537, 2002. 2

[9] B. D.S., B. J.R., T. M., and D. B.A. The csu face identification evaluation system: Its purpose, features and structure. In *Third International Conference on Computer Vision Systems*, pages 304 – 311, 2003. 5

[10] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. In *British Machine Vision Conference*, September 2004. 3

[11] C. Liu, H.-Y. Shum, and C. Zhang. Hierarchical shape modeling for automatic face localization. In *Proceedings of the European Conference on Computer Vision*, number II, pages 687–703, Copenhagen, Denmark, May 2002. 2

[12] H. M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling*. Academic Press, 3 edition, 1998. 3, 5

[13] S. Yan, M. Li, H. Zhang, and Q. Cheng. Ranking prior likelihood distributions for bayesian shape localization framework. In *Proceedings of IEEE International Conference on Computer Vision*, volume 1, pages 51 – 58, Nice, France, October 2003. 2

[14] M. Zhao and T.-S. Chua. Face alignment with unified subspace optimization for active statistical models. In *The 7th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 67–72, Southampton, UK, April 2006. 5

[15] M. Zhao, S. Z.Li, and C. Chen. Subspace analysis and optimization for AAM based face alignment. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, May 2004. 2

[16] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003. 1