Boosted Dynamic Mixture Active Shape Model For Multi-View Face Alignment

Yu Chen
School of Computer Science and Engineering
The University of New South Wales
Sydney, Australia
Email: yuc@cse.unsw.edu.au

Arcot Sowmya
School of Computer Science and Engineering
The University of New South Wales
Sydney, Australia
Email: sowmya@cse.unsw.edu.au

Xiongcai Cai
School of Computer Science and Engineering
The University of New South Wales
Sydney, Australia
Email: xcai@cse.unsw.edu.au

***Abstract*—A novel optimization scheme in the popular Active Shape Model (ASM) framework is presented, which increases the accuracy and robustness of segmentation on multi-view face images. The shape model is trained on a Gaussian Mixture Model and then combined with a visibility model of each point for searching faces. The determininistic fitting scheme in traditional ASM is substituted by a probabilistic estimation approach. A set of weighted particles is used to represent each salient feature point. Integration of the dynamic model with a mixture model not only improves performance toward the local minima problem, but also addresses the improper initialisation problem. Furthermore, combination of the boosted observation model and visibility model potentially solve the problem of occlusion. The proposed algorithm is much more robust in multi-view face alignment and not affected by initialization conditions, as are current ASM optimization algorithms. Experiments show that the developed approach provides higher accuracy and increases the convergence rate in face segmentation on the high resolution PUT [1] public test data set.**

I. Introduction

There has been significant research into model-based approaches for the interpretation of face images due to their capability to represent large variations and different expressions. In particular, the Active Shape Model(ASM) [2] has been successfully applied to model human faces in many computer vision tasks such as face tracking and facial motion recognition. A number of works [3]–[5] based on the Gaussian Linear Model with ASM demonstrate the effectiveness and robustness of using ASM on face images in the frontal view. However, when the view changes dramatically, the Gaussian Linear Model fails to model shape variations properly. View based, non-linear and 3-D approaches have been proposed to solve the multi-view face registration problem. View based methods represent large shape variation by sets of small variations and model each set separately [6], however, such methods do not regularize shape in a multi-modal framework and the training data is highly restricted for each distinct view. Non-linear methods use a nonlinear function to model the variations of shape and assume that all the feature points are visible during view changes, while some are occluded practically [7]. An alternative method is to create a 3-D shape model instead of a 2-D one to estimate the view and indicate the occluded points [8], but this suffers from high computation cost and limited improvement compared to 2-D methods.

The state-of-art non-linear method Bayesian Mixture Active Shape Model [9](BMASM) involves a unified framework for combining multiple shape models into a mixture shape model. Self-occlusion can be handled by learning visibility from training data. However, since the texture model used in the local search of each label point depends on the view, it is sensitive to the estimation of the hidden view parameter. With an incorrect initial view, the results become unreliable.

In order to improve the accuracy and solve the initialization problem, the BMASM method is extended and a novel Boosted Dynamic Mixture Active Shape Model(BDMASM) approach is proposed for achieving a multi-view Active Shape Model. First, a set of weighted particles are used to represent the location of each feature point, which are varied according to a boosted regression function on different views. In the optimization phase, the sequential importance re-sampling technique is combined with the EM algorithm to estimate the model parameters, and the particles are propagated randomly and moving towards higher probabilities, which enlarges the search range and avoids the improper initialization problem. A view-based dynamic model is used to predict new locations for each view based particle set. The predicted sets of particles are updated based on the observation model and the suggested shape is output for each view. The visibility model for each view is further optimized the observation model to solve the occlusion problem.

The rest of the paper is organized as follow: related work and preliminary knowledge are reviewed in Section II and Section III. The proposed novel BDMASM algorithm is introduced in Section IV. Experiments and comparisons are presented to support the novel idea in Section V with conclusion following in Section VI.

II. Background

The ASM proposed by Cootes et al. [2] is one of the most popular methods for wrapping an initialized shape around

image features. It models the statistical shape and its variation over a labeled training set containing a set of salient feature points such as eyes, nose and mouth in the case of faces. There are other deformable models such as snakes [10], but unlike snakes, which have little prior knowledge incorporated, ASM uses an explicit shape model to place global constraints on the generated shape. The Active Appearance Model (AAM) also proposed later by Cootes et al. [11] integrates a global appearance model into the shape model. Traditional ASM assumes that the shape variations have a Gaussian distribution, however, when the pose changes dramatically, the assumption becomes unreliable due to non-linear variations.

In recent years, many modifications of the general ASM scheme have been proposed to solve the non-linear problem. Romdhani et al. [7] use Kernel Principal Components Analysis and a Support Vector Machine to model nonlinear changes in 3D pose variations. Zhou et al. [12] estimate shape and pose parameters using Bayesian inference after projecting shapes into a tangent space. Wang et al. [13] introduce a resampling scheme to incorporate a pose-dependent appearance model for multi-view face alignment. A mixture of gaussian models is employed by Cootes et al. [14] to represent the shape variation in contrast to assuming a single gaussian distribution in the classical ASM. Zhou et al. [9] also propose a unified multi-modal distribution based on a Bayesian mixture model. However, the estimation of pose parameters depends only on the observation model due to its local Markov property and also has high requirements on the initialization. To improve the existing methods, a boosted dynamic ASM [4] was proposed that combined with the Bayesian mixture model to optimize the local search and parameter estimation. The proposed method, namely Boosted Dynamic Mixture Active Shape Model (BDMASM), avoids some of the bottlenecks in the previous methods, as well as achieving a comparable results.

III. Preliminary

A. Original Active Shape Model

The Active Shape Model (ASM) consists of statistical shape modeling and an optimization process which deforms the shape model to best fit the image data. For general facial analysis, the shape model is trained with a set of labeled face images and aligned from one shape to another with a similarity transform containing three pose parameters - translation, scaling and rotation. The average of all the aligned images is the mean shape of the training set. Principal Component Analysis (PCA) is used to model shape variation from the covariance matrix of those aligned images. Any shape can be approximated using:

$$x = \bar{x} + Pb \tag{1}$$

where x is a vector of model points in shape space, $\bar{x}$ is the mean shape of the training samples, P is a matrix of orthogonal eigenvectors and b is the shape parameter containing a vector of weights corresponding to the eigenvectors in P. A shape vector X in image space is a euclidean transformation of the shape model x.

$$X = M(s, \theta)[x] + t \tag{2}$$

where s, θ and t are the pose parameters of scaling, in-plane rotation and translation respectively. ASM iteratively updates the pose and shape parameters to find a shape vector X that best matches the image information to the local appearance model. Classical ASM matches the local appearance model by minimising the Mahalanobis distance of the local eigen patches.

ASM begins from initialising a start shape X_t at iteration t either manually or by using some global detector to determine the face location. It then updates the shape and pose parameters to best match the image until convergence conditions are met.

B. Bayesian Mixture Active Shape Model (BMASM)

To solve the problem of multi-view face alignment, a Bayesian mixture model [9] has been proposed to model the multi-modality and variable feature points in a unified framework, which involves two novel ideas:

1) The original ASM model is extended to a mixture model to describe the shape distribution:

$$p(x|b) = \sum_{i=1}^{m} \pi_i (2\pi\sigma_i^2)^{-N} exp\{-\frac{1}{2\sigma_i^2}\|x - \mu^{(i)} - \phi^{(i)} b^{(i)}\|\} \tag{3}$$

where m is the number of mixture models, $\sigma^{(i)}$ is the covariance matrix, $\mu^{(i)}$ is the mean shape of each cluster and $\phi^{(i)}$ is the principal matrix whose columns are the eigenvectors.

2) A visibility vector is added to all n training data $\{(\mathbf{x}_1, \mathbf{v}_1), (\mathbf{x}_2, \mathbf{v}_2) \cdots (\mathbf{x}_n, \mathbf{v}_n)\}$, where $\mathbf{v}$ is a vector of 0 or 1 indicating each point's visibility. Thus the prior visibility of the kth point is formulated as:

$$q_k^{(i)} = \sum_{t=1}^{n} \mathbf{v}_k^t f_i(x^(t)) / \sum_{t=1}^{n} f_i(x^{(t)})$$

and the visibility model is defined as a mixture Bernoulli model:

$$p(\mathbf{v}) = \sum_{i=1}^{m} \pi_i \prod_{k=1}^{N} [(q_k^{(i)})^{v_k} (1 - q_k^{(i)})^{1-v_k}],$$

3) With the learned mixture model and visibility model, an EM framework is used to estimate the shape and pose parameters (b, θ):
   a) Given an hypothesis shape y, the cluster weight w_i is estimated.
   b) With w_i and y, point visibility in current iteration is estimated.
   c) estimate the hypothesis shape $\hat{x}$.
   d) update the shape parameters.
   e) update the pose parameters.

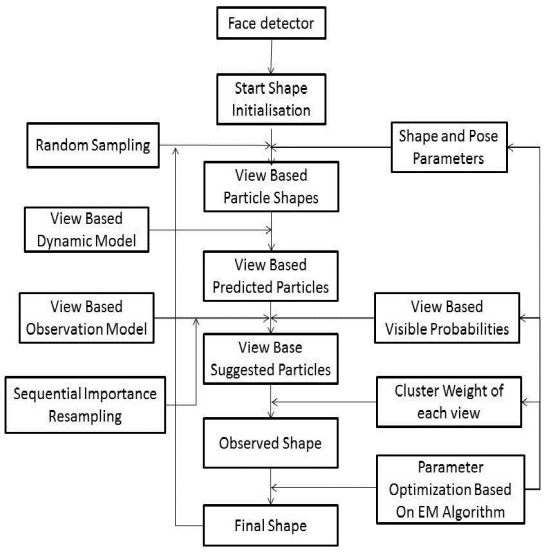

Fig. 1. Illustration of BDMASM algorithm

IV. BDMASM

Generally, BMASM has outstanding improvements for multi-view face alignment. However, it may simply fail due to improper initialization or be affected by non-linear noisy images. Thus, probabilistic based ASM becomes an option as it can estimate better possibilities rather than find an optimal solution, especially particle filter that is able to approach an Bayesian optimal estimate on complex distributions. The combination of BDASM with a mixture model forms the novel idea of the boosted dynamic mixture active shape model (BDMASM) framework. The integration of dynamic information achieves better results for large shape variations in multi-view facial segmentation. The boosted observation model is enhanced by a visibility information, which leads to high performance on occlusion data. More specifically, a particle based mixture shape prior is introduced instead of a single shape prior, a boosted regression function is learned on the haar-wavelet feature for the observation model to find the hypothesis shape, and a dynamic displacement is added to each prior shape based on the previous shape variation. Sequential importance sampling method is used to optimize the shape and pose parameter. The interaction between the particle filter and mixture shape model is shown in Fig. 1. Given a start shape x, a set of particle shapes was sampled around x, each particle shape is displaced based on the view based dynamic model and the weights of particles are calculated according to the observation model and visible states for each cluster. The posterior distribution is reconstructed by sequential importance sampling to focus on particles with larger weights. The observed shape is formed with weighted mean shape of the particles. Finally, EM algorithm is applied to updated the shape and pose parameter.

A. View Based Prior Shape Model

The process of initializing particle shapes from a start shape is regarded as the fundamental change from deterministic to probabilistic methods. Unlike BDASM, the mixture model is composed of a number of independent clusters, where each cluster has its corresponding shape and appearance model. Therefore, a set of particle shapes are generated randomly for each cluster. The view based particle shape model is interpreted as:

$$x_m^k = x + rand * R; \tag{4}$$

where x_m^k is the kth particle in each cluster m, rand is a random number distributed as Gaussian with zero mean and variance of between -1 and 1 and R controls the range of the noise distribution. Thus the location of the particles of mth view set are in the range of a circle of radius R and centred at that feature point.

The particle sets for each view are independent of each other. Each set has a weight ϕ indicating the importance of this set. However, all the particles will be combined together into one observed shape, which is used to update the pose and shape parameters to contribute the final shape in the current iteration. For the next iteration, the final shape becomes the prior shape and repeated from the random sampling scheme.

B. View Based Dynamic Model

As stated in the previous subsection, each particle set is independent of the others, thus a dynamic model is used for each view based particle set. The dynamic model follows the second order autoregressive model. For each view, the dynamic information is based on the weighted displacement of the previous two iterations as follows:

$$\begin{aligned} Pred(x_{m|t+1}^k) = x_{m|t}^k + A(x_{m|t}^k - x_{m|t-1}^k) + \\ B(x_{m|t-1}^k - x_{m|t-2}^k) + \sigma \end{aligned} \tag{5}$$

where t is the number of the iteration, $Pred(x_{m|t+1}^k)$ is a predicted location of the kth particle in the mth cluster at iteration t+1, $x_{m|t}^k$ is the current location, $x_{m|t-1}^k$ the location of the previous iteration and $x_{m|t-2}^k$ that of the iteration two before, and σ is a white noise process with zero mean and variance. A and B are two dynamic parameters (coefficients), which can be determined by a default setting or learned using the most common least squares method and used to control the displacement of the dynamic. With the dynamic prior information, each view based particle set moves according to the nature of its view, and is used for local texture matching independently.

C. View Based Observation Model

In the learning phase, the database has been categorized into several view sets, and the observation model is learned for each view set based on the boosted haar wavelet function. The processes of learning each observation model are the same as in previous work [4].

During the search phase, the initialized particle sets in all the mixtures are random by sampled from the same

shape, but evolve based on different dynamic models. Particle sets are derived from their corresponding observation model, i.e. the confidence values of particles are modeled by their own view appearance model. However, the confidence values cannot represent the visible state of the points. Therefore, the confidence value for each particle is multiplied by its visible probability for the current iteration. Then the updated particle sets are re-sampled according to the posterior confidence value based on the SIR method.

D. Particle Based Observed Shape and Parameter Optimization

After re-sampling, the observed shape is derived from the updated view based particle sets, where the mean shape of each particle set is calculated for each cluster:

$$x_m = \frac{\sum_{k=1}^{q} x_m^k}{q} \tag{6}$$

and the cluster shapes summed with respect to their weights

$$y = \sum_{m=1}^{M} \omega_m \times x_m \tag{7}$$

The primary goal of shape optimization is to update the underlying shape in shape space according to the observed shape, then the shape and pose parameters are updated based on the changes of the underlying shape. During search, the observed particles are used to update the pose and shape parameters by the Expectation Maximization algorithm. Basically, the EM algorithm iterates between an E(Expectation) Step and an M(Maximization) step. In the E step, the cluster weight and visible state are re-estimated given the observed shape. In the M phase, the underlying shape is updated and parameters determined by variation of the underlying shape. These steps can be summarized as follows:

1) Given a hypothesis shape y, the cluster weight w_i is estimated:
$$p(\omega_m|y) \propto p(\omega_m)p(y|\omega_m) \tag{8}$$
2) With w_i and y, point visibility is estimated:
$$p(v_m|\omega_m, y) \propto p(v_m|\omega_m) \int p(y|x, v, \theta) f_i(x|b) dx \tag{9}$$
3) the underlying shape $\hat{x}$ is estimated:
$$\hat{x} \propto f_i(x|b)p(y|x, v, \theta)p(x|y, v, \omega) \tag{10}$$
4) the shape parameters are updated by projecting the shape to the PCA subspace, which is regularized by shrinking its weight along each principal direction:
$$b = \frac{\omega_m \Lambda}{\omega_m \Lambda + \sigma^2}(\phi x)) \tag{11}$$
5) the pose parameters are updated by weighted least square method:
$$\theta = min\{\sum_{m=1}^{M} \omega_m (y - T_\theta(x))\} \tag{12}$$

TABLE I
DATABASE SPLIT

Database \ View clusters	Training Dataset	Test Dataset	Total
Frontal view	308	292	600
Right side view	199	101	300
Total	507	393	

E. BDMASM

The whole process of BDMASM is summarised here including the learning and search phases. In the learning phase, the database is split into training and test datasets, with each dataset containing images that are categorised into different views. Learning of the shape model are described in Algorithm 1

Algorithm 1 LEARNING SHAPE AND APPEARANCE MODEL

1) For a set of training data, learn a specified number of view clusters based on Gaussian Mixture Model and EM algorithm, and combine all the clusters into a unified mixture model framework as in E.q. 3.
2) For each view based training dataset, extract the haar-like features for each training example and learn the boosted regression function by Gentleboost.

After the learning phase, an M cluster mixture shape model is obtained where μ_i, ϕ_i, w_i, P_i are the centre, covariance, weight and eigen-matrix for the ith cluster respectively. Given a start shape, the complete process of BDMASM search is shown in Algorithm 2.

V. EXPERIMENTS

For evaluation of the proposed method, the BDMASM framework has been implemented as a set of scripts in Matlab platform. The database used to evaluate BDMASM contains both frontal and right side view face images, however, the situation could be complex as illumination, expression, wearing glasses, beard or moustache are introduced in the images. BDMASM was trained and tested on the public PUT database [1] which contains about 10000 face images of 100 different persons with high resolution of 2048×1536 pixels. In order to train the two cluster mixture model, a subset database of 900 images with 600 near front view images and 300 right side view images is chosen, and the sub database was randomly split into 507 training data and 393 test data. The number of images in each view contained in each set is shown in Table I. The test criterion that used to evaluate the results is:

$$m_e = \frac{1}{ns}\Sigma d \tag{13}$$

where d is the point to point error for each individual point feature location, s the groundtruth inter-ocular distance, and n is the number of total feature points. There are also some parameters to be set before the experiment. In the training phase, a 21×21 patch was chosen to extract the haar-like features and 50 features were selected through boosting

Algorithm 2 BDMASM ALGORITHM

1) Initialise start shape X_0 by a global face detector with the average point projected into the bounding box:

$$X_0 \leftarrow M[\bar{x} + Pb_0](s_0, \Theta_0) + t_0$$

2) Initialise cluster weights, visible probabilities, and dynamic information.

repeat
randomly sample m particle sets $\{q^i_{1,2,\cdots N}, \omega^i_{1,2,\cdots N}\}$ around each salient feature point i for each view.
for $t = 1$ to k **do**
for $j = 1$ to N **do**
predict a new location for every particle set with its corresponding dynamic model.
for $i = 1$ to m **do**
update weight ω^j_i for each particle according to the view based boosted observations and visible probabilities for each view.
end for
end for
re-sample the particles according to the weight for each cluster to a new posterior distribution.
save the displacement between the two iterations $X_{t+1} - X_t$ for evolving the particles in the next iteration.
calculate the observed shape by the summation of the mean of each view particle set.
estimate cluster weights, visible probability and a posterior observed shape based on EM, update pose and shape parameters based on EM algorithm.
align the model from the updated pose and shape parameters to get the final shape X_{t+1} and repeat from the random sample step.
end for
until Converged
output the final shape as the result.

iterations; in the search phase, The initial shape was calculated by wrap the mean shape to the bounding box that was detected by SNoW face detector [15], 50 particles were randomly generated from the start shape and α and β in the dynamic model are set to 1 and 0.5 respectively.

The detection and displacement test methods in [4] were employed to evaluate and compare the results between BDASM and BDMASM. The experiments were performed in the same system environment based on the default parameters. Besides the whole test data, the split test data of frontal and right side images were also used for comparison. The results demonstrate that BDMASM outperforms BDASM in multi-view face segmentation.

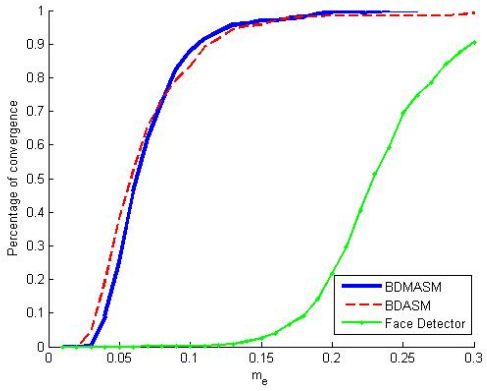

Fig. 2. BDMASM and BDASM detection search results for front and right side images

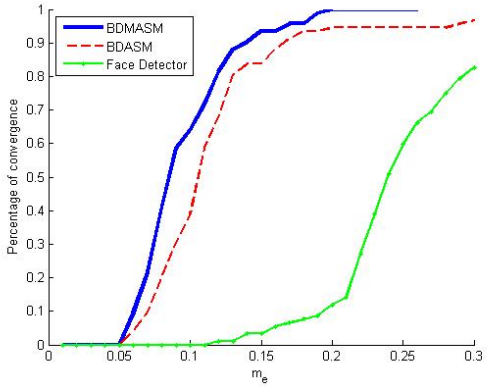

Fig. 3. BDMASM and BDASM detection search results over right side images

A. Detection Based Search

The BDASM and BDMASM detection search results on the PUT dataset are shown in Fig. 2, where the complete test data set with front and right side images was evaluated. For the evaluation of the whole data set, the accumulative error graph is calculated, where the x coordinate represents the m_e error rate, and the y coordinate shows the percentage of images within that error rate. The results are too close to tell the difference as both BDASM and BDMASM achieve high accuracy on the whole dataset, however, BDASM is slightly better than BDMASM for $m_e < 0.05$, but slightly worse for $0.05 < m_e < 0.1$. The reason for such performance is that BDASM is better than BDMASM on frontal view data, but worse on right side view data. To demonstrate performance on a single view, the results considering only the right side face images are shown in Fig 3. The graph strongly demonstrates that BDMASM has higher convergence rate than BDASM on right view facial segmentation.

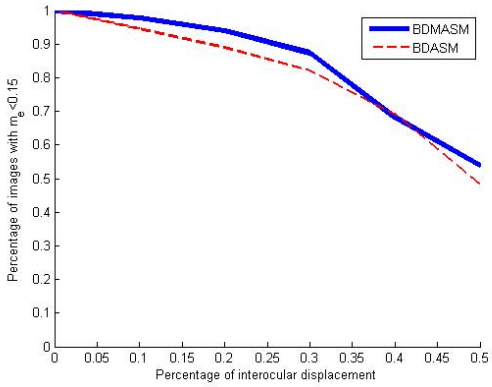

Fig. 4. BDMASM and BDASM displacement search results for front and right side images

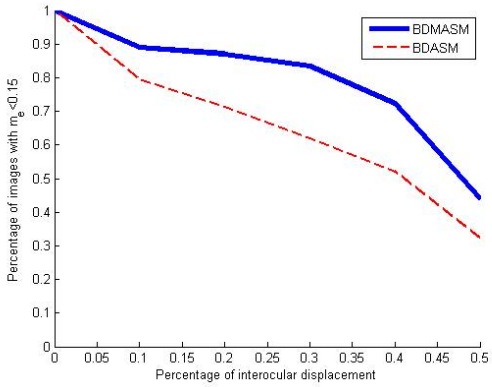

Fig. 5. BDMASM and BDASM displacement search results over right side images

B. Displacement Based Search

To demonstrate that the BDMASM algorithm is not affected by close initialisation and outperforms BDASM in multi-view cases, displacement based search experiments were then conducted. The results of displacement search on the entire test data are shown in Fig. 4. The curve represents the percentage of images with $m_e < 0.15$, where the x coordinate is the displacement represented by 10%, 20%, 30%, 40% and 50% of inter-ocular distance, and the y coordinate is the percentage of images within the specific error rate. The results also indicate that BDMASM is better than BDASM. For 10% of inter-ocular displacement, 98% of the images converge with $m_e < 0.15$ compared to 95% for BDASM, then there is a slight decrease for 20% displacement to 95% and 89% respectively. However, the results of BDMASM drop fast at 40% displacement, and both the curves intersect at 65%, but BDMASM still remains better at 50% displacement.

The curves in Fig 5 also show how displacement affects the search performance for right side images only. It is clear that BDMASM completely outperforms BDASM on right side view data.

VI. Conclusion

This paper combines a mixture shape model with the BDASM framework to achieve multi-view face segmentation. The nature of solving multi-view face segmentation problems has been analysed and the novel framework of BDMASM has been proposed to integrate those solutions. The contributions of BDMASM to multi-view facial segmentation include a solution to the problems of large shape variation and initilisation based on dynamic model, and an accurate and robust observation model based on view based boosted particle filter to solve non-linear and occlusion problems. Experiments are conducted to demonstrate the accuracy and robustness of BDMASM on PUT public multi-view face databases against BDASM. The evaluation of the results shows that BDMASM outperforms BDASM in multi-view facial segmentation.

References

[1] A. S. A. Kasinski, A. Florek, "The put face database," in *Image Processing and Communications*, vol. 13, no. 3-4, 2008, pp. 59–64.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, January 1995.

[3] D. Cristinacce and T. Cootes, "Boosted regression active shape models," in 18^{th} *British Machine Vision Conference, Warwick, UK*, 2007, pp. 880–889.

[4] Y. Chen, X. Cai, and A. Sowmya, "Boosted dynamic active shape model," in *Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference*, nov. 2009, pp. 215 –220.

[5] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 504–513.

[6] L. Zhang and H. Ai, "Multi-view active shape model with robust parameter estimation," in *Proceedings of the 18th International Conference on Pattern Recognition - Volume 04*, ser. ICPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 469–468.

[7] S. Romdhani, S. Gong, A. Psarrou, and R. Psarrou, "A multi-view nonlinear active shape model using kernel pca," in *In British Machine Vision Conference*. BMVA Press, 1999, pp. 483–492.

[8] Y. Su, H. Ai, and S. Lao, "Multi-view face alignment using 3d shape model for view estimation," in *Proceedings of the Third International Conference on Advances in Biometrics*, ser. ICB '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 179–188.

[9] Y. Zhou, W. Zhang, X. Tang, and H. Shum, "A bayesian mixture model for multi-view face alignment," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 741–746.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. V1, no. 4, pp. 321–331, January 1988.

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *Proceedings of the European Conference on Computer Vision*, vol. 2, pp. 484–498, 1998.

[12] Y. Zhou, L. Gu, and H.-J. Zhang, "Bayesian tangent shape model: estimating shape and pose parameters via bayesian inference," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1, June 2003, pp. I–109–I–116 vol.1.

[13] Z. Wang, X. Xu, and B. Li, "Bayesian tactile face," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[14] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," in *Image and Vision Computing*. BMVA Press, 1997, pp. 110–119.

[15] M. H. Yang, D. Roth, and N. Ahuja, "A snow-based face detector," in *Advances in Neural Information Processing Systems 12*. MIT Press, 2000, pp. 855–861.