Research Article
A Review Article for Active Appearance Model Fitting Based on Optimization Methods

[1]Mohammed Hasan Abdulameer, [1]Dhurgham A. Mohammed, [2]Hassan Farhan Rashag and
[3]Hasanein D. Rjeib
[1]Department of Computer Science, Faculty of Education forWomen, University of Kufa,
[2]Foundation of Technical Education,
[3]I.T Reseaqrch and Development Center, University of Kufa, Iraq

Abstract: Active Appearance Model (AAM) is one of the most popular techniques that extract features by precise modelling of human faces under various physical and environmental circumstances. In such active appearance model, fitting the model with original image is a challenging task. State of the art shows some of the optimization methods that are applicable to resolve the optimization problem. This encourages us to review the recent improvements of AAM fitting based on optimization methods. This study provides an up-to-date review of researches as regards to AAM optimization; studies are reviewed and discussed in four different sections. This review evidently explores the limitations of the previous works to facilitate better understanding different kind of approaches. The aim of this study is to help as a guide for further studies.

Keywords: Active appearance model, fitting, face recognition

INTRODUCTION

AAM is a computer vision algorithm, which matches a statistical model of object shape and appearance to a new image. The Active Appearance Model (AAM) has been proposed by Cootes *et al.* (1998) and Abdulameer *et al.* (2013); of late, a number of applications employ AAMs for feature extraction, which include, face modelling for understanding human behaviour and medical imaging projects, such as segmentation of cardiac MRIs or the diaphragm in CT data and registration in functional heart imaging (Bosch *et al.*, 2001; Wang and Chen, 2010).

Basically, AAM model could be developed in four primary phases (Gao *et al.*, 2010; Sethuram *et al.*, 2010):

- Initially, a statistical shape model is built to model the differences in the shape of an object, by utilizing a number of illustrated training images
- Next, a texture model is constructed to model the variations of texture, which is depicted by pixel intensity.
- Ultimately, an appearance model is constructed by merging the shape and the texture models (Cootes *et al.*, 1998).

In this study, we provide a critical review of active appearance model fitting based on optimization methods and we investigate their limitations to help as a guide for further studies.

ACTIVE APPEARANCE MODEL THEORETICAL BACKGROUND

Essentially, AAM model can be composed from three primary phases:

- Statistical shape model
- Texture model
- Combined appearance model (Cootes *et al.*, 1998).

Statistical shape model: A statistical shape model is constructed from a set of annotated training images. In a 2-D case, a shape is represented by concatenating *n* point vectors {(xi; yi)}:

$$x = (x1, x2, \ldots, X_n, y1, y2, \ldots, y_n)T \qquad (1)$$

The shapes are then normalized by Procrustes analysis (Goodall, 1991) and projected onto the shape subspace created by PCA:

$$x = \bar{x} + P_s . b_s \qquad (2)$$

where, $\bar{x}$ denotes the mean shape, $P_s = \{s_i\}$ is the matrix consisting of a set of orthonormal base vectors s_i and outlining the modalities of variants extracted from training set and b_s consists of the shape variables in the shape subspace. Consequently, depending on the corresponding points, images in the training set are warped to the mean shape to produce shape-free patches.

Texture model: The texture model is developed more identically as the shape model. Depending on the shape free patch, the texture could be raster scanned into a vector g, followed by linear normalization of the texture, by the parameters $u = (\alpha, \beta)T$ and g is given by:

$$g = \frac{(g_i - \beta.1)}{\alpha} \tag{3}$$

where, α and β are, respectively, the mean and the variance of the texture g and $1 = [1; 1, \ldots.1]$ T is the vector with the same length of g_i. Eventually, based on PCA, the texture is projected upon the texture subspace

$$g = \bar{g} + P_g.b_g \tag{4}$$

where,
$\bar{g}$: The mean texture
$P_g = \{g_j\}$: The matrix consisting of a set of orthonormal base vectors g_j and describing the modes of variation derived from training set and b_g includes the texture parameters in the texture subspace.

Combined appearance model: Finally, the coupled relationship between the shape and the texture is analysed by PCA and the appearance sub space is created. At the end, the shape and the appearance can be described as follows:

$$\begin{aligned} x &= \bar{x} + Q_s.c \\ g &= \bar{g} + Q_g.c \end{aligned} \tag{5}$$

where, c is a vector of appearance parameters controlling both the shape and the texture and Q_s and Q_g are matrices describing the modes of variation derived from the training set. Thus the final appearance model can be represented as $b = Qc$ where:

$$b = \begin{pmatrix} w_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} w_s(p_s)T(x - \bar{x}) \\ (p_g)T(g - \bar{g}) \end{pmatrix} \tag{6}$$

And Q is the matrix of eigen vectors of b.

After the model is created, it is important to fit the model to new images, which is crucial to identify the most appropriate parameters of the model for an object. Nevertheless, this is an unconstrained optimization problem, which is challenging to solve. Generally, it could be addressed by the gradient descent algorithm. Let p signify the AAM's parameter vector($p^T = c^T/t^T/u^T$), which is the combination of the appearance parameters c, pose parameters t and the texture transformation parameters u. g_c is the sampled texture vector of the current image, which is projected to the texture model frame and g_m is the texture vector, generated by the model. It can be supposed that there is a linear relationship between the two amounts, the
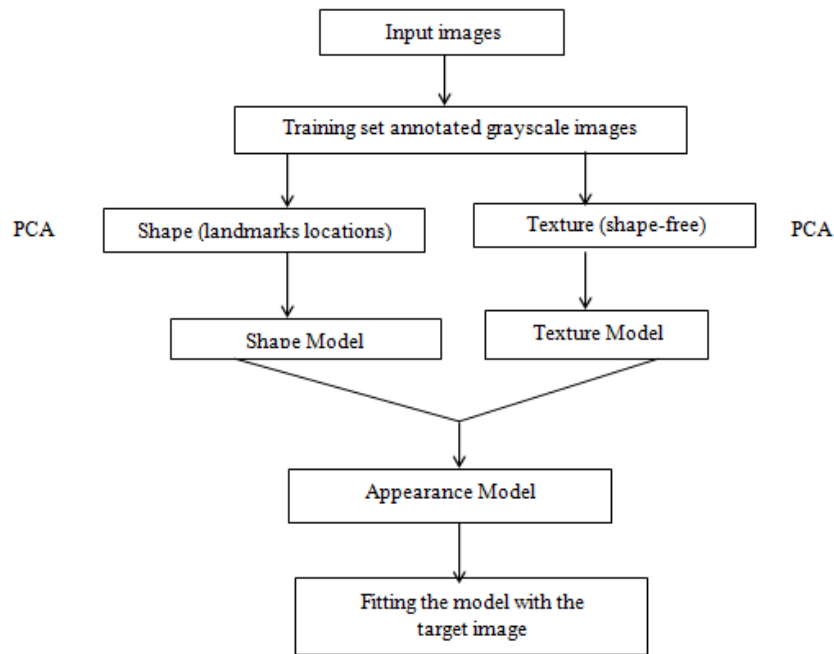
Fig. 1: The basic structure of active appearance model

variance of p and the texture difference between an image and the model, which is given by:

$$\delta_p = R.r(p) \qquad (7)$$

$$r(p) = g_c - g_m \qquad (8)$$

where, δp is a small variance of p and R is the linear relationship (or gradient matrix) between δp and r. AAM assumes R to be fixed and pre-computes it by multivariate linear regression techniques. Since the relationship R is pre-computed, fitting is carried out as an iterative procedure as follows (Gao *et al.*, 2010) (Fig. 1):

- Sample the texture of an image and project it to the texture model space.
- Calculate the residual texture vector, $r = g_i - g_m$ and evaluate the fitting accuracy using $E = |r|2$, where |.| means the norm (2-norm generally).
- Predict the variance of the parameters of the model by
  $\delta_p = -R.r(p)$.
- Update the model parameters $p \rightarrow p + k\delta p$, where $k = 1$ initially.
- With the new model parameters, calculate the new model texture g_m^- and resample the texture,g_i^-.
- Calculate the new residual vector, $r^- = g_i^- - g_m^-$ and $E^- = |r^-|2$.
- If $E^- < E$, then accept the update; otherwise, try at k = 0.5, 0.25, etc., then go to the first step.

These procedures iterate until there is no more improvement.

AAM Fitting based on optimization methods: According to Peyras *et al.* (2007) choosing fitting algorithm is an important concern of AAM. The existence of fitting problem in the face recognition approaches affects their efficiency, as well as generates a poor recognition precision (Baek *et al.*, 2009). Of late, several research approaches have been formulated for resolving the fitting problem in AAM. The techniques for enhancing the fitting efficiency of AAM could be categorized into four sections. Figure 2 shows the four categories.

Manifold deformable model: The first approach uses manifold deformable model. Furthermore, Christoudias *et al.* (2004, 2006) have defined the manifold appearance of human face due to external lighting, where each model has been acquired based on particular view of 2D shape of each face, in a specific light field. In order to fit the light-field model to the input image, they have initially selected a particular view of the light-field model, which is most adjacent to the view of the input image and later employed the direct search to coordinate the input image to a level in the manifold light-field. Based on employing manifold estimation in the consolidated view-based feature spaces, (Huang *et al.*, 2010) have proposed and technique for recognizing face features across poses, which is capable of incorporating undetectable views, even there is a big change in pose. Nonetheless, they have evaluated their work on only 25 models of 3D and this might be not adequate to confirm their work.

Combines the current models: The second approach combines the current models. Wang *et al.* (2003) have hybridized Active Shape Models (ASMs) and Active Appearance Models (AAMs), for efficiently interpreting image. In the proposed approach, firstly, ASM local searching approach has been employed, later, apart from the statistical shape limitations similar to the ASMs, the global texture constraint depending on the sub-space reconstruction error is further utilized for assessing the fitting level of the presently projected model to the new image. Furthermore, identical to the technique used in the AAM approach, the global texture is employed to estimate and update the parameters of the model. Yan *et al.* (2002) have proposed a Texture-Constrained Active Shape Model (TC-ASM), which acquired the local appearance model of ASM, as a result of its strength in diverse light conditions. They have also obtained the global texture model of AAM to restrict the shape and to present an optimization measure for identifying the shape parameters. Making use of the texture-constrained shape has enabled the
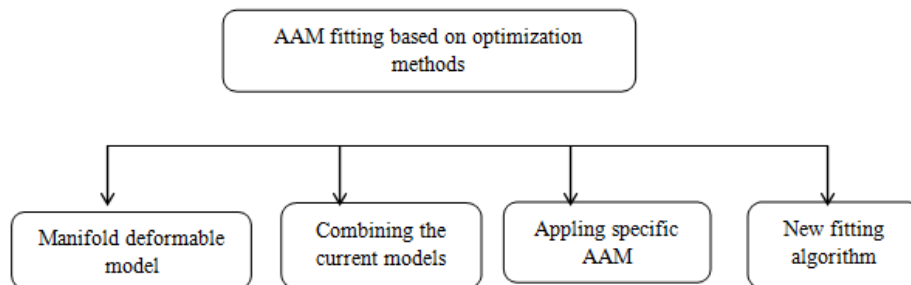
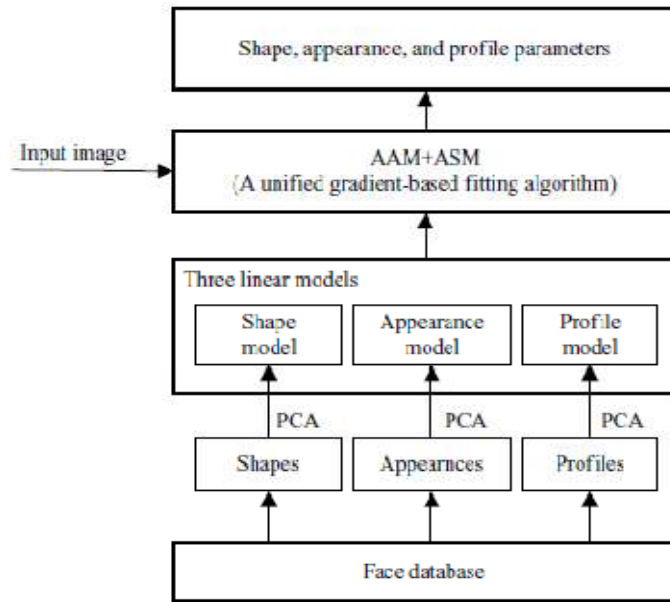

Fig. 2: AAM optimization approaches

Fig. 3: The proposed fitting method Sung *et al.* (2007)

search approach to escape from getting trapped in local minima of the ASM search, leading to the enhanced fitting outcomes.

For addressing the disadvantages in AAM fitting algorithm, Sung *et al.* (2007) have hybridized Active Shape Models (ASM) with AAMs, where the later attempts to identify appropriate landmark points making use of the local profile model. Considering that the authentic objective function of the ASM search is not suitable for merging these methods, a gradient based iterative method has been derived by enhancing the objective function of the ASM search. Later, they have proposed a novel fitting approach, which hybridizes the objective functions of both, ASM and AAM into a single objective function in a gradient based optimization framework as explained in Fig. 3. The AAM utilizes gradient-based optimization approaches for model fitting and consequently, it is very sensitive to initial model parameters. For addressing this issue Sung *et al.* (2008) have combined the active appearance models and the Cylinder Head Models (CHMs), where the global head motion parameters obtained from the CHMs have been employed as the indicators of the AAM parameters for an excellent fitting or re-initialization:

Employ specific AAM: The third approach employs a specific AAM, which is suitable for various individuals and external variations. Lucey *et al.* (2006) have employed the person-specific AAM for tracking the object and have employed the generic AAM to determine features independent to subjects. As the traditional AAM is always viewpoint-generic, which minimizes the fitting precision; Kawarazaki *et al.*

(2011) have proposed a viewpoint-specific AMM for extracting robust facial point under multi-viewpoint. The proposed technique performs the following functions:

- Gathers the training samples and separates them into various groups based on the viewpoint
- Builds the AMMs for each training group
- If a new test image is given, it is fitted based on the trained AAMs;
- Compares the fitting error and chooses the minimum one as the final fitting results.

Berry *et al.* (2011) have proposed a new image pyramid fitting structure exclusively developed to be applied in an Audio-Visual Speech Recognition (AVSR) system, where the area described by the AAM will be minimized in the course of the iterations. This enables the fitting technique to become more precise as the fitting advances. The new fitting structure is applied with the Fixed Jacobian algorithm and then it compared to a conventional strategy, where the shape of mouth is extracted from a full face AAM. In the above technique the appearance model able to symbolizea shape within the limit of 4 pixels. Nevertheless, in case of specific individuals, the model is not being capable of to represent the shape within this boundary.

Generally, the traditional AAM often deviates, when the input image have differences in pose, appearance and lighting effects, which were not included in the training set. Lee and Kim (2009) are proposed tensor-based AAM for addressing the above problem, which employs multi-linear algorithm to the shape and appearance models of the traditional AAM,

for enhancing the fitting efficiency. Tensor-based AAM comprises of an image tensor and a model tensor. Image tensor determines the pose, appearance and illumination of the input image; on the other hand, the model tensor creates variation specific AAM basis vectors, making use of a single trained model tensor and the projected variations, on line. To estimate the variants from the image tensor, they have proposed two distinct techniques:

- Discrete variation estimation
- Continuous variation estimation

Where, the former estimates the pose, expression and lighting of the input image, based on one of the poses, expressions and lighting effects of the training images. On the other hand, the continuous variation approach approximates the pose, expression and lighting effects of the input image, by a combining its corresponding basis vectors. Nevertheless, it consumes more time in case of tensor-based AAM with continuous variation estimation, which demands more pre-processing time for determining the image differences.

Introducing new fitting algorithm: The forth technique modifies the AAM fitting algorithm itself, by introducing a novel fitting algorithm or enhancing the current fitting algorithm. Donner *et al.* (2006) have proposed a rapid AAM utilizing the Canonical Correlation Analysis (CCA), which models the relation between the differences of image and parameter, for enhancing the convergence speed of fitting algorithm. Moreover, Matthews and Baker (2004) have proposed arapid AAM fitting algorithm depending on the inverse compositional image alignment algorithm, which does not need a linear relationship among the differences of image and model parameter. They have accomplished better fitting precision as against the conventional AAM; furthermore, the model also has faster convergence. Andreopoulos and Tsotsos (2005) have proposed an extended fitting 2-D AAM approach, based on the inverse compositional image alignment algorithm. The expansion of the algorithm involves the fitting of 3-D AAMs on short axis cardiac MRI. Moreover, Xiao *et al.* (2004) have proposed a 2D+3D AAM, which has an additional 3D shape model. They have also offered the effective fitting algorithm of2D+3D AAM, by including 3D constraints to the cost function of 2D AAM. Furthermore, (Hu *et al.*, 2004) have proposed an additional extension of 2D+3D AAM fitting algorithm, which is called as multi-view AAM fitting (MVAAM) algorithm, which fits a single 2D+ 3D AAM to images with different views, which were simultaneously acquired from multiple cameras. Nevertheless, the assessment process was not adequate to demonstrate the efficiency of the proposed technique.

Table 1: AAM approaches based onan optimization methods

Authors, year	Optimization concept	Approaches	Disadvantages and limitations
Huang *et al.* (2010)	Manifold deformable model	View-based feature space	25 models of 3D and this not adequate to confirm their work.
Wang *et al.* (2003)	Combining the current models	ASM+AAM	lack of the ability to entirely exploit the detailed information present in the given image and sensitive to initial model parameters
Yan *et al.* (2002)	Combining the current models	TC-ASM	The dataset was not enough to assess the generalization of the shape prediction from the texture
Sung *et al.* (2007)	Combining the current models	ASM-AAMs	Model Fitting utilizes gradient-based optimization approaches and consequently, it is very sensitive to initial model parameters
Lucey *et al.* (2006)	Specific AAM	Person-specific AAM	As the traditional AAM is always viewpoint-generic, which minimizes the fitting precision
Kawarazaki *et al.* (2011)	Specific AAM	Viewpoint-specific AMM	The improvement was limited in the proposed method because it is based on only thee viewpoints
Berry *et al.* (2011)	Specific AAM	Pyramid fitting structure based on Fixed Jacobian algorithm	In case of specific individuals, The model is not able to represent the shape within this boundary.
Baek *et al.* (2009)	Specific AAM	Tensor-based AAM	It consumes more time in case of tensor-based AAM with continuous variation estimation, which demands more pre-processing time for determining the image differences
Donner *et al.* (2006)	New Fitting Algorithm	CCA-AAM	CCA suffers from small size problem (SSS), linearity and it is not completely involved the class information of the samples
Matthews and Baker (2004)		Inverse compositional image alignment algorithm	The inverse compositional AAM fitting algorithm can only be applied to *independent* AAMs.
Andreopoulos and Tsotsos (2005)	New Fitting Algorithm		It cannot be applied to *combined* AAMs
Xiao *et al.* (2004)	New Fitting Algorithm	2D+3D AAM	Computational time per iteration for the 2D algorithm is lesser than the time for the proposed 2D+3D algorithm.
Hu *et al.* (2004)	New Fitting Algorithm	Multi-view AAM fitting (MVAAM) algorithm	The assessment process was not adequate to demonstrate the efficiency of the proposed technique since in some cases the algorithm shows instability.
Abdulameer *et al.* (2014)	New Fitting Algorithm	AAM+Adaptive ABC	Time consuming because ABC algorithm spends long time to fit the generated model with the target image.

Moreover, Sung and Kim (2006) have proposed a new fitting algorithm of 2D+3D AAMs for a multi-focal camera system, referred to as stereo AAM (STAAM), for improving the stability of the fitting of 2D + 3D AAMs. Furthermore, STAAM minimizes the amount of model parameters and demonstrates a superior fitting steadiness, as against the present multi-view AAM. Finally, the AAM fitting problem had been solved by introducing a new adaptive ABC algorithm (Abdulameer *et al.*, 2014). The introduction of adaptive ABC has fastened the AAM fitting and hence the effectiveness of recognition has been enhanced without compromising the recognition performance. Table 1 show and summaries the relevant studies as regards to the four sections of fitting optimization methods with their limitations.

CONCLUSION

This study has presented a critical review about fitting optimization in active appearance model technique which is widely used with face recognition applications. In this review, several research approaches have been formulated for resolving the fitting problem in AAM. The techniques for enhancing the fitting efficiency of AAM have been categorized into four sections which were: manifold deformable models, combining the current models, using specific AAM and introducing new fitting algorithm. In addition, this review has been explored the limitations and the drawbacks of the previous studies clearly. From what has been reviewed so far, we can conclude that introducing a new fitting algorithm with AAM could have much influence on the fitting process.

REFERENCES

Abdulameer, M.H., S.N.S. Abdullah and Z.A. Othman, 2013. Face recognition technique based on active appearance model. Int. Rev. Comput. Software (IRECOS), 8(11): 2733-2739.

Abdulameer, M.H., S.N. Sheikh Abdullah and Z.A. Othman, 2014. A modified active appearance model based on an adaptive artificial bee colony. Sci. World J., 2014: 879031.

Andreopoulos, A. and J.K. Tsotsos, 2005. A novel algorithm for fitting 3-D active appearance models: Applications to cardiac MRI segmentation. Proceedings of the 14th Scandinavian conference on Image Analysis (SCIA'05). Joensuu, Finland, pp: 729-739.

Baek, S.Y., B.Y. Kim and K. Lee, 2009. 3D face model reconstruction from single 2D frontal image. Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry. New York, pp: 95-101.

Berry, C., A. Kokaram and N. Harte, 2011. An extended multiresolution approach to mouth specific AAM fitting for speech recognition. Proceeding of the 19th European Signal Processing Conference. Barcelona, Spain, pp: 1959-1963.

Bosch, H.G., S.C. Mitchell, B.P.F. Lelieveldt, F. Nijland, O. Kamp, M. Sonka and J.H. Reiber, 2001. Active appearance motion models for endocardial contour detection in time sequences of echocardiograms. Proceeding of SPIE 4322, Medical Imaging 2001: Image Processing, pp: 257-268.

Christoudias, C.M., L.P. Morency and T. Darrell, 2004. Light Field Appearance Manifolds. In: Pajdla, T. and J. Matas (Eds.), Computer Vision - ECCV 2004, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 3024: 481-493.

Christoudias, C.M., L.P. Morency and T. Darrell, 2006. Non-parametric and light-field deformable models. Comput. Vis. Image Und., 104(1): 16-35.

Cootes, T.F., G.J. Edwards and C.J. Taylor, 1998. Active appearance models. Proceeding of European Conference on Computer Vision, 2: 484-498.

Donner, R., M. Reiter, G. Langs, P. Peloschek and H. Bischof, 2006. Fast active appearance model search using canonical correlation analysis. IEEE T. Pattern Anal., 28(10): 1690-1694.

Gao, X., Y. Su, X. Li and D. Tao, 2010. A review of active appearance models. IEEE T. Syst. Man Cy. C, 40(2): 145-158.

Goodall, C., 1991. Procrustes methods in the statistical analysis of shape. J. Roy. Stat. Soc. B Met., pp: 285-339.

Huang, X., J. Gao, S.C.S. Cheung and R. Yang, 2010. Manifold estimation in view-based feature space for face synthesis across poses. Proceeding of the 9th Asian Conference on Computer Vision (ACCV'09), pp: 37-47.

Hu, C., J. Xiao, I. Matthews, S. Baker, J. Cohn and T. Kanade, 2004. Fitting a single active appearance model simultaneously to multiple images. Proceeding of the British Machine Vision Conference, September, pp: 1-10.

Kawarazaki, Y., G. Duan, T. Shinkawa and Y.W. Chen, 2011. Viewpoint-specific active appearance model for robust feature point extraction. Proceeding of the 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT, 2011), pp: 883-886.

Lee, H.S. and D. Kim, 2009. Tensor-based AAM with continuous variation estimation: Application to variation-robust face recognition, IEEE T. Pattern Anal., 31(6): 1102-1116.

Lucey, S., I. Matthews, C. Hu, Z. Ambadar, F. de la Torre and J. Cohn, 2006. AAM derived face representations for robust facial action recognition. Proceeding of the 7th IEEE International Conference on Automatic Face and Gesture Recognition (FGR, 2006), pp: 155-160.

Matthews, I. and S. Baker, 2004. Active appearance models revisited. Int. J. Comput. Vision, 60(2): 135-164.

Peyras, J., A. Bartoli, H. Mercier and P. Dalle, 2007. Segmented AAMs improve person-independent face fitting. Proceeding of the 18th British Machine Vision Conference (BMVC'07), pp: 1-10.

Sethuram, A., K. Ricanek and E. Patterson, 2010. A comparative study of active appearance model annotation schemes for the face. Proceeding of the 7th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '10), pp: 367-374.

Sung, J. and D. Kim, 2006. STAAM: Fitting a 2d+3d AAM to stereo images. Proceeding of the International Conference on Image Processing. Atlanta, GA, pp: 2781-2784.

Sung, J., T. Kanade and D. Kim, 2007. A unified gradient-based approach for combining ASM into AAM. Int. J. Comput. Vision, 75(2): 297-309.

Sung, J., T. Kanade and D. Kim, 2008. Pose robust face tracking by combining active appearance models and cylinder head models. Int. J. Comput. Vision, 80(2): 260-274.

Wang, W., S. Shan, W. Gao and B. Yin, 2003. Combining active shape models and active appearance models for accurate image interpretation. ICASSP, pp: 1-4.

Wang, Y.K. and W.R. Chen, 2010. Facial feature alignment by manifold learning of active appearance model. Proceeding of the 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP, 2010), pp: 224-227.

Xiao, J., S. Baker, I. Matthews and T. Kanade, 2004. Real-time combined 2d+3d active appearance models. Proceeding of the 2004 IEEE Computer Society Conference on Computer vision and Pattern Recognition (CVPR'04), pp: 535-542.

Yan, S., C. Liu, S.Z. Li, H. Zhang, H.Y. Shum and Q. Cheng, 2002. Texture-constrained active shape models. Proceeding of the 1st International Workshop Generative-Model-Based Vision, pp: 107-113.