

# ON THE APPLICATION OF AAM-BASED SYSTEMS IN FACE RECOGNITION

*Muhammad Aurangzeb Khan, Costas Xydeas, Hassan Ahmed*

School of Computing and Communication  
Infolab21, Lancaster University, Lancaster, LA1 4WA  
United Kingdom

## ABSTRACT

The presence of significant levels of signal variability in face-portrait type of images, due to differences in illumination, pose and expression, is generally been accepted as having an adverse effect on the overall performance of i) face modeling and synthesis (FM/S) and also on ii) face recognition (FR) systems. Furthermore, the dependency on such input data variability and thus the sensitivity, with respect to face synthesis performance, of Active Appearance Modeling (AAM), is also well understood. As a result, the Multi-Model Active Appearance Model (MM-AAM) technique [1] has been developed and shown to possess a superior face synthesis performance than AAM. This paper considers the applicability in FR applications of both AAM and MM-AAM face modeling and synthesis approaches. Thus, a MM-AAM methodology has been devised that is tailored to operate successfully within the context of face recognition. Experimental results show FR-MM-AAM to be significantly superior to conventional FR-AAM.

**Index Terms**— Face Recognition, Multi-Model Active Appearance Model.

## 1. INTRODUCTION

The Multi-Model Active Appearance Model (MM-AAM) technique proposed in [1], produces generic face models that allow the successful synthesis of unseen face images, both in terms of shape and texture. MM-AAM is an extension of the well-established Active Appearance Model (AAM) face modeling and synthesis (FM/S) algorithm [2]. Furthermore, AAM has been studied extensively, see [3-8], and within the context of different applications including face recognition (FR). However, as suggested in [1] and [9] AAM lacks accuracy in creating generic face models, a weakness that has been addressed successfully by MM-AAM, particularly when synthesizing “unseen” face images. Note that grouping of training images into clusters is also employed in [10, 11], but their clustering is based only on shape orientation (i.e. pose) whereas the VQ-based clustering proposed in [1] caters for both face pose and expression. This means that in MM-AAM, images with similar pose are

grouped into several different clusters, each characterized by relatively low variability.

This paper examines the applicability of AAM and MM-AAM in face recognition and proposes an FR-MM-AAM system that is far less dependent on input data variability and thus outperforms FR-AAM. Here, input data variability relates to differences in face illumination, pose and expression [12]. Furthermore, input data variability is generally been accepted as having an adverse effect on the overall performance of both face modeling/synthesis (FM/S) and face recognition (FR) systems.

This proposed FR-MM-AAM framework has been developed and evaluated in two phases, that is: i) a design phase based on system training and using a wide range of input face images and ii) a system performance testing phase, using both “seen” and “unseen”, by the face modeling and synthesis processes, input image data.

System training involves three major steps: Firstly, face image models, created according to MM-AAM procedures given in [1], are used to model the shape information of all faces included in the training input dataset. Shape information is then employed to extract, from corresponding training images, the actual facial texture, i.e. those face pixels contained within corresponding shapes. It is this actual face texture information that forms the basis for information discrimination and face recognition.

Secondly, Principal Component Analysis (PCA) [13] is employed on this texture information in order to obtain a lower dimensionality feature/texture space than that created in the previous step.

Finally, a given classification system is designed via training to operate on PCA derived information. Note that two classification methods have been employed in this work, i.e. Linear Discriminant Analysis (LDA) [14] and Multi-Class, linear kernel Support Vector Machines (SVM) [15].

Testing system performance involves the synthesis of face shape information for a given input image and the subsequent extraction of corresponding texture information. This face texture is projected on the PCA Eigen-Space that is created during the training phase, to yield a lower dimensional feature vector, which then is classified to one of a number of possible subjects.

Face recognition system performance has been evaluated using i) “seen” and ii) “unseen” input face images. In the first case, face shape models were defined during training using a set of 655 input images (taken from 95 subjects), whereas system recognition performance was evaluated based on a subset of 175 out of the 655 images (35 subsets). In the second case the previous subset of 175 images were used only during recognition and they were therefore “unseen” by the shape modeling process.

The remainder of this paper is organized as follows: section 2 explains in more detail the design and structure of the FR-MM-AAM system training and testing phases. Experimental results and discussion are presented in Section 3, whereas concluding remarks are provided in Section 4.

## 2. MM-AAM AS APPLIED TO FACE RECOGNITION

MM-AAM face modeling and synthesis has been proposed by the authors in [1, 16] and shown to outperform conventional AAM. However, both schemes combine shape and texture parameters within the same set of information. This is then transformed to another set of “hybrid” face model coefficients, with each coefficient representing both shape and texture. Note however that, although such hybrid model coefficients can successfully be employed in face modeling and synthesis, they exhibit a reduced discrimination between different input faces capability. This characteristic of model parameters is detrimental to FR and as a result we have restricted the application in FR for both AAM and MM-AAM only to shape information. This gave rise to the proposed FR-MM-AAM system which is described below in terms of i) System Training and ii) System Operation. Note that FR-AAM can be easily derived from FR-MM-AAM.

### 2.1. FR-MM-AAM System Training

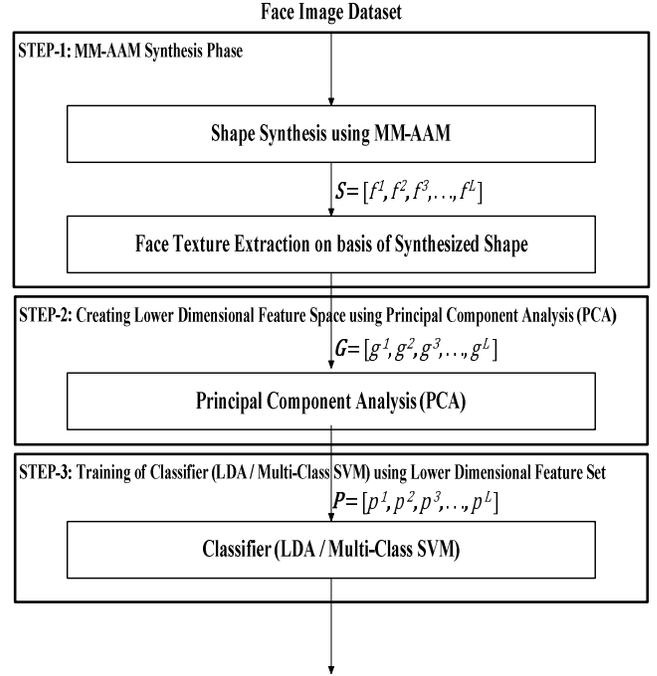
Training encompasses three major steps. These are shown in Figure 1 and discussed below. Note that MM-AAM operation is a precursor to FR-MM-AAM training.

1. Consider that the MM-AAM synthesized shape information of the  $i^{\text{th}}$  input face image is represented as

$$f^i = [x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_M]^T, \quad (1)$$

where  $\{(x_m, y_m)\}$  are the coordinates of the  $m = 1, 2, \dots, M$  landmark points outlining different facial components. Furthermore, the set  $\mathbf{S}$  of shape information obtained from all the  $L$  training images of dataset  $\mathbf{F}$  is given by

$$\mathbf{S} = [f^1, f^2, f^3, \dots, f^L]. \quad (2)$$



**Fig. 1.** FR-MM-AAM Training. This involves the synthesis of face image shapes for the complete training image dataset  $\mathbf{F}$ , using the MM-AAM procedures described in detail in reference [1]. MM-AAM training is a precursor to FR-MM-AAM training.

Now, for each face image  $i$ , face texture information  $g^i$  is extracted (i.e. pixels contained within shape defined face outlines) using corresponding shape coordinates  $f^i$ . Thus, the set  $\mathbf{G}$  of face texture information obtained from all  $L$  training images is represented by

$$\mathbf{G} = [g^1, g^2, g^3, \dots, g^L]. \quad (3)$$

2. Principal Component Analysis (PCA) is applied on dataset  $\mathbf{G}$ , to obtain a lower N-dimensionality representation, texture only related,  $p^i$  for each texture vector  $g^i$ . This is achieved by PCA selecting a relatively small number of N Eigen-Vectors on the basis of percentage energy captured by corresponding Eigen-Values [13].
3. Finally, a classifier (LDA or Multi-Class SVM) is trained using the set  $\mathbf{P}$  of feature vectors  $p^i$   $i = 1, 2, \dots, L$ .

### 2.2. FR-MM-AAM System Operation

System operation involves the following three procedures:

1. Given an input face image  $t$ , MM-AAM synthesis is applied that derives shape coordinate information  $f^t$ . This information allows the corresponding face texture  $g^t$  to be obtained from the input image.
2. Texture  $g^t$  is projected onto the N-dimensional Eigen Space produced during training, in order to obtain the N-dimensional feature set  $p^t$ .

- Finally,  $p^t$  is given as input to the classifier, that has been designed during training, the output of which returns a label class i.e. person assigned to image.

### 3. EXPERIMENTAL RESULTS

Experimentation was performed using computer simulation of FR systems and involved two different publicly available facial datasets i.e. i) FG-NET, an age related dataset [17] which contains face images taken at different subject ages, and ii) the IMM face dataset [18]. In particular, FG-NET contains 1002 colour or grey-scale face images of 82 persons (12 images per person on average) with an age range from infant to 69 years. Furthermore each image comes with 68 hand labeled shape points.

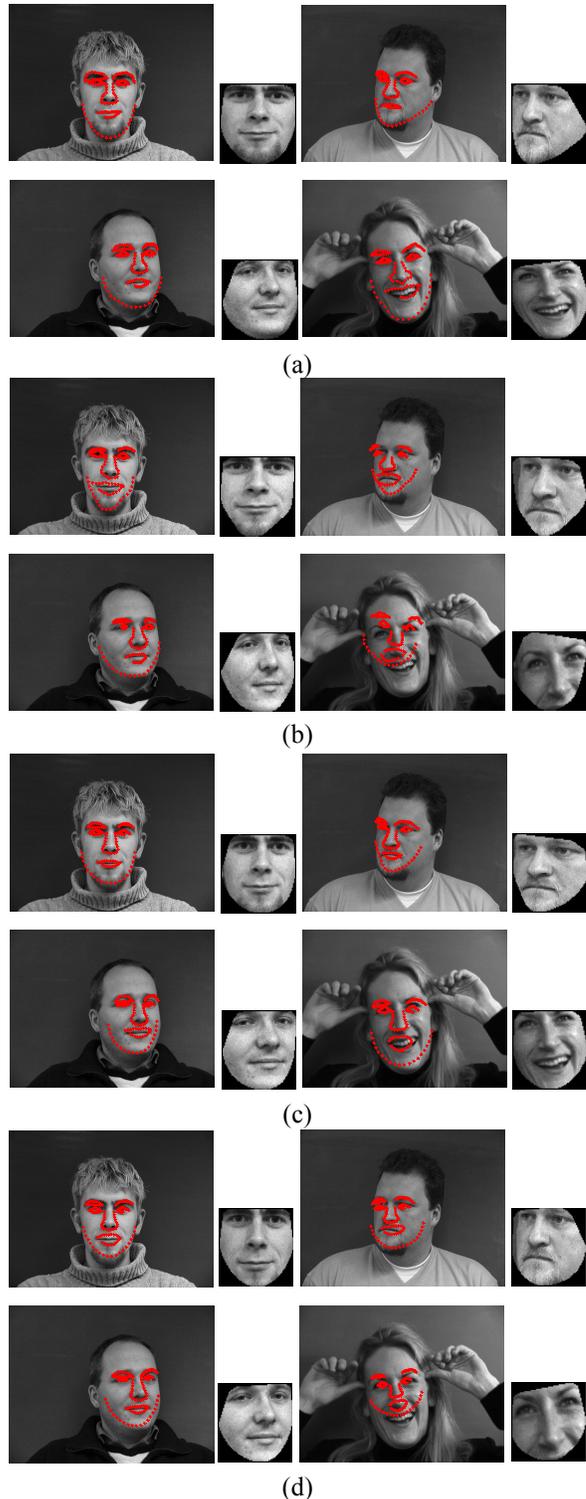
The IMM dataset consists of 240 images, that is, 6 images for each of 40 persons. Each image is  $640 \times 480$  pixels in size and comes with 58 hand labeled shape points that outline face components.

The FR performance of systems under consideration was assessed and compared in two experiments.

In the first experiment, MM-AAM and AAM training, involved i) 480 images of 60 persons (8 images per person) from FG-NET dataset and ii) 175 images of 35 persons (5 images per person) from the IMM dataset. The number  $M$  of hand labeled face shape points was kept to  $M = 58$  in both datasets. Recall that MM-AAM and AAM training, as specified in [1], is a precursor to the FR-MM-AAM three steps training process described in the previous section.

In this experiment, face recognition system performance was evaluated using a “seen” dataset composed of the above 175 IMM images (used in shape modeling) plus another 35 image version of the same IMM subjects.

In the second experiment, MM-AAM and AAM training involved only the 480 images taken from FG-NET, whereas during face recognition the “unseen” dataset of 210 IMM images was employed as input data. Examples of MM-AAM and AAM synthesized shapes with their corresponding extracted textures, for both “seen” and “unseen” input datasets are shown in Figure 2. These images are indicative of the improved shape modeling accuracy of MM-AAM, as compared to AAM, particularly in the case of “unseen” images. Furthermore, and in both experiments, FR system training and testing has been performed using the Leave One Image Out (LOIO) fold approach, i.e. in each fold one image taken from each person is left out for testing whereas all remaining images are used for system training. Notice that in addition to the above techniques, a conventional FR-PCA system has been simulated and tested. Here cropped image face information i.e. pixels, see Figure 3, are used directly in dimensionality reduction (PCA) and resulting coefficients are classified. The purpose for using manually cropped image data is to compare the FR-PCA



**Fig. 2.** Examples of synthesized shapes and their corresponding extracted textures: a) and b) are examples obtained from previously “seen” data and correspond to MM-AAM and AAM respectively, whereas c) and d) are results with MM-AAM and AAM operating on “unseen” data.

system performance with that obtained from the FR-AAM and FR-MM-AAM systems, where texture is obtained via shape modeling.

FR system performance bars are plotted in Figures 4, 5 and show recognition rates for systems using two different classification techniques i.e. LDA and Multi-Class SVM.

Note that throughout this work, retained Eigen-Vectors correspond to the largest Eigen-Values while 90% of cumulative signal energy is captured.

In both figures, results indicate clearly that FR performance based on texture extracted using MM-AAM modeled shapes and with the system operating on “seen” input data is significantly better than that obtained from both FR-AAM and FR-PCA.

In the case of LDA classification, see Figure 4, FR-MM-AAM offers a recognition rate of 94.29% and 69.05% for “seen” and “unseen” datasets, respectively. Furthermore FR-MM-AAM outperforms FR-AAM by 42.39% and 25.50%, for “seen” and “unseen” datasets, respectively. This is indicative of AAM’s inability to cope with input data variability, even in the case of previously “seen” inputs. Furthermore, and in the case of “seen” input images, FR-MM-AAM outperforms FR-PCA by 21.91%. The above noted general trends in FR performance are valid in Figure 5, with the additional observation that systems employing LDA classification demonstrate FR gains of the order of 20% to 25% over the corresponding Multi-Class SVM systems. Note that a simple base line, linear kernel, SVM has been used in these experiments, which in turn explains SVM’s inferior performance as compared to that of LDA.

#### 4. CONCLUSION

In this paper, the applicability of Active Appearance Model (AAM) and Multi-Model-AAM is considered within the context of face recognition (FR) and for input datasets exhibiting significant levels of variability in face illumination, pose and expression. Experimental results clearly demonstrate that under such input conditions, conventional AAM face modeling is unable to support FR, whereas the recently proposed MM-AAM approach can be tailored to yield a highly successful FR operation.

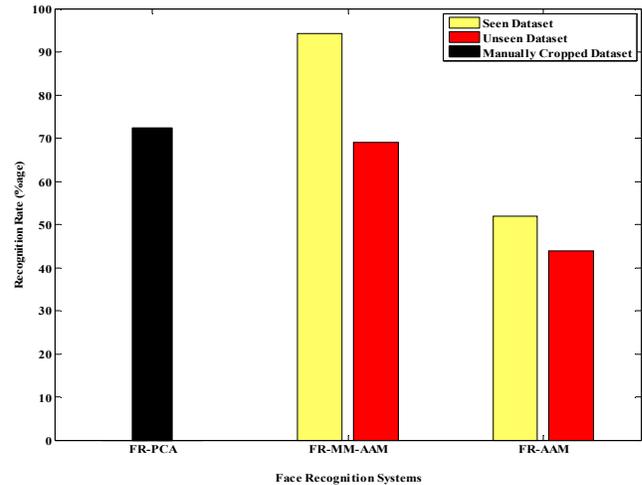
Furthermore, MM-AAM modeling is amenable to decomposition into face components (i.e. eyes, mouth, nose, etc.) [17], which in turn allows FR to operate with partial input image face information. This will be the topic of a subsequent paper.

#### REFERENCES

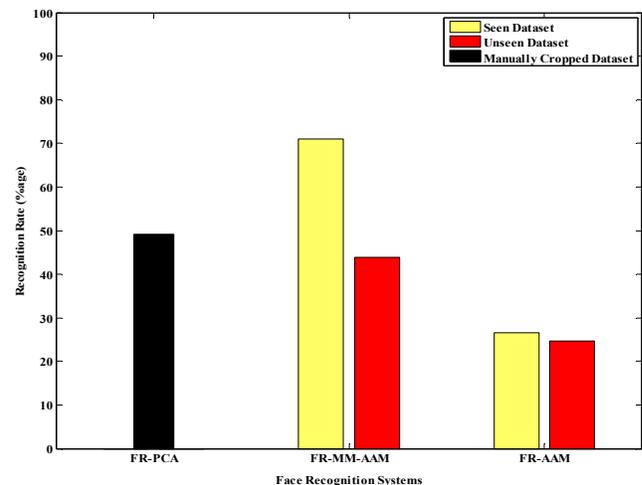
[1] M.A. Khan, C. Xydeas, and H. Ahmed. "Multi-model AAM framework for face image modeling," in *Proc. IEEE. International Conference on Digital Signal Processing (DSP)*, July 2013, pp. 1-5.



**Fig. 3.** Examples of manually cropped images used in the performance evaluation of conventional FR-PCA. The purpose of using cropped images is to generate input images which are therefore comparable to those used in FR-MM-AAM and FR-AAM system performance experiments, see Figure 2.



**Fig. 4.** LDA Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Yellow Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Black Bar).



**Fig. 5.** Multi-Class SVM Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Yellow Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Black Bar).

[2] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” in *Proc. Computer Vision ECCV98*, 1998, pp. 484–498.

- [3] M.B. Stegmann, "Object tracking using active appearance models," in *Proc. Danish Conf. Pattern Recog. Image Anal*, 2001, vol. 1, pp. 54–60.
- [4] S.C. Mitchell, B.P.F. Lelieveldt, R.J. Van Der Geest, H.G. Bosch, JHC Reiver, and M. Sonka, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *IEEE Transaction on Medical Imaging*, vol. 20, no. 5, pp. 415–423, 2001.
- [5] E. Patterson, A. Sethuram, M. Albert, and K. Ricanek, "Comparison of synthetic face aging to age progression by forensic sketch artist," in *Proc. The Seventh IASTED International Conference on Visualization, Imaging and Image Processing*. ACTA Press, 2007, pp. 247–252.
- [6] X. Geng, Z.H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.
- [7] B. Xiao, X. Gao, D. Tao, and X. Li, "A new approach for face recognition by sketches in photos," *Signal Processing*, vol. 89, no. 8, pp. 1576–1588, 2009.
- [8] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 2, pp. 145–158, 2010.
- [9] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image and Vision Computing*, vol. 23, no. 12, pp. 1080–1093, 2005.
- [10] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor, "Viewbased active appearance models," *Image and Vision Computing*, vol. 20, no. 9, pp. 657–664, 2002.
- [11] L. Teijeiro-Mosquera and J.L. Alba-Castro, "Performance of active appearance model-based pose-robust face recognition," *IET Computer Vision*, vol. 5, no. 6, pp. 348–357, 2011.
- [12] Abate, Andrea F., et al. "2D and 3D face recognition: A survey." *Pattern Recognition Letters* 28.14 (2007): 1885-1906.
- [13] M.A. Turk, and A.P. Pentland. "Face recognition using eigenfaces," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586-591.
- [14] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, 1992.
- [15] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, V. Vapnik, "Comparison of classifier methods: a case study in handwritten digit recognition," in *IEEE. International Conference on Pattern Recognition*, 1994, pp. 77–87.
- [16] M.A. Khan, C. Xydeas, and H. Ahmed. "Multi-Component/Multi-Model AAM framework for face image modelling," in *Proc. IEEE. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 2124-2128.
- [17] "The fg-net aging database," <http://www.fgnet.rsunit.com/>, [Online: accessed 27th Nov. 2011].
- [18] M.M. Nordstrøm, M. Larsen, J. Sierakowski, and M.B. Stegmann, "The imm face database-an annotated dataset of 240 face images," *Informatics and Mathematical Modeling*, 2004.