MULTI-COMPONENT/MULTI-MODEL AAM FRAMEWORK FOR FACE IMAGE MODELING

Muhammad Aurangzeb Khan, Costas Xydeas, Hassan Ahmed

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

ABSTRACT

An image face modeling framework is proposed that aims to enhance the face modeling capability of the well known Active Appearance Model (AAM). AAM has been used successfully in person-specific related applications but it poses significant limitations when employed in generic face modeling. Thus this work is focused on the development of new face models which are generic in nature and which accurately fit unseen image faces, both in terms of shape and texture. For this purpose, images are decomposed into face related components which are subsequently clustered on the basis of shape similarities. Experimental results show that models generated through this novel framework can be significantly more effective than conventional AAM, in terms of both shape and texture.

Index Terms— Active Appearance Models, Image Face Analysis and Synthesis.

1. INTRODUCTION

The Active Appearance Model (AAM) introduced by Cootes et al. [1] is an algorithm for constructing a synthetic image that is a close match to an input face image, in terms of both shape and appearance. Furthermore AAM can be considered as an optimization problem that minimizes the difference between the synthesized image and the real texture of the input image. AAM's ability of differentiating and modeling shape and texture helps in the synthesis of more photorealistic images. AAM applications can be found in a variety of areas such as object tracking [2], medical image analysis [3], age estimation and synthesis [4, 5], facial recognition and modeling systems [6], and gait analysis [7]. A comprehensive survey of AAM based modeling techniques can be found in [8].

Note that there are two types of application scenarios for modeling face images [9]. One relates to applications such as Gaze Estimation, Head Pose Estimation or Expression Recognition and involves person-specific models. The second type deals with the construction of unseen faces and involves generic face models. The processes used in the construction of Active Appearance Models for these two types are completely different. Authors in [9] have shown that person-specific AAMs are easier to build, whereas generic AAMs appear to be problematic in texture modeling. Modeling a face in age estimation and synthesis comes under the later of the two scenarios discussed above and is the main driver in this work.

Thus the proposed face modeling framework, named Multi-Component/Multi-Model AAM (MC/MM-AAM), aims at the creation of generic AMM based face models, which are robust to unconstrained input conditions and can preserve discriminative information when generating "unseen" face images.

The modeling phase of MC/MM-AAM involves three major steps. In the first step face images, taken from the training dataset, are decomposed into face related components e.g. eyes, mouth, nose, etc. to form facial component specific datasets. This decomposition aims to exploit the local characteristics of each component and can result in better model fitting as suggested as [10, 11]. In the second step, clustering is applied on each individual facial component dataset. This is achieved on the basis of face similarities and as a result, each original facial component set can be represented by several subsets (or clusters). Note that the idea of grouping face images into a number of clusters is also presented in [12, 13], but clustering is done on the basis of shape orientation (pose) only, whereas clustering here caters for both face orientation and expression. Finally, in the last step a conventional AAM is applied to each cluster of a facial component that results in more than one model for each component.

The synthesis phase of MC/MM-AAM, allows for more than one shape for each component of the input face image to be synthesized. The best component shape offering minimum average mean square error, which is formed between original and synthesized textures, is then chosen for each component. Finally, the selected shapes of all components are combined to form a whole face shape, which when presented into a whole face conventional AAM [1] delivers the synthesized texture of the input face image.

Experimental results show that the proposed framework produces more accurate models of unseen face images, in

terms of both shape and texture, as compared to the conventional AAM model .

The rest of this paper is organized as follows. Section 2 explains the design and structure of the Modeling and Synthesis parts of MC/MM-AAM. A discussion on experimental results is given in Section 3, whereas concluding remarks are provided in Section 4.

2. MULTI-COMPONENT / MULTI-MODEL AAM (MC/MM-AAM)

The proposed MC/MM-AAM framework comprises of two parts i.e. a Modeling and a Synthesis phase.

2.1. Modeling Phase

Modeling involves three major steps, as it is explained below and shown in figure 1.

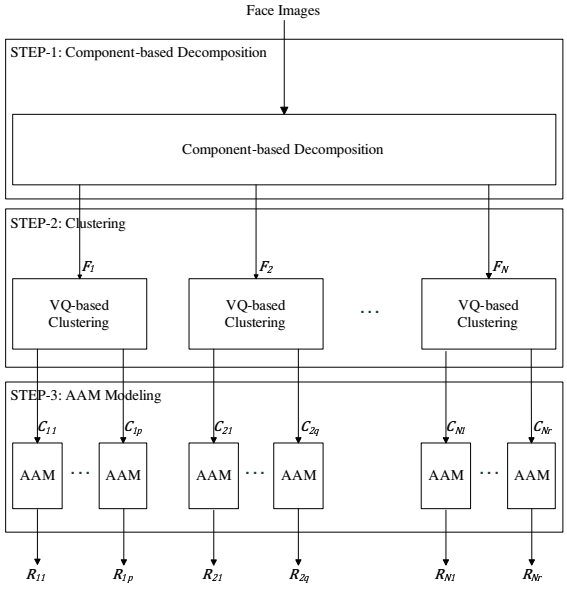


Fig. 1. System Diagram of proposed MC/MM-AAM. Here $\mathbf{F_1}, \mathbf{F_2}, \ldots, \mathbf{F_N}$ are $(N = 4)$ components-based datasets and $\mathbf{C_{nk}}$'s are corresponding component clusters. These are used to produce component based model matrices $\mathbf{R_{nk}}$'s. The number of clusters employed for each component can be the same i.e. $p = q = r = 8$ or it can differ.

STEP-1: Involves the component-based decomposition of images into facial components. Face images taken from a training dataset are decomposed on the basis of N facial components ($N = 4$ in our case i.e. cheeks + eyebrows, eyes, mouth and nose). This yields N component datasets $\mathbf{F_n} = \{\mathbf{S_n}|\mathbf{G_n}\}$ for $n = 1, 2, \ldots, N$, each corresponding to a specific face component and containing shape $\mathbf{S_n}$ information, in form of landmark points and texture $\mathbf{G_n}$ in form of intensity values. This component-based decomposition is being used to account for the local shape and texture variability that characterizes different facial components. Shape in a face image $i = 1, 2, \cdots, L$ is represented by a vector f containing the M landmark points outlining the different facial components

$$f^i = [x_1, x_2, \ldots, x_M, y_1, y_2, \ldots, y_M]^T, \tag{1}$$

here $\{(x_m, y_m)\}$ are the coordinates of the mth point. In this step, the shape vector f^i of the ith face image is decomposed into N sub vectors of different lengths, such that

$$\begin{aligned} f_1^i &= [x_{11}, x_{12}, \ldots, x_{1a}, y_{11}, y_{12}, \ldots, y_{1a}]^T, \\ f_2^i &= [x_{21}, x_{22}, \ldots, x_{2b}, y_{21}, y_{22}, \ldots, y_{2b}]^T, \\ &\vdots \\ f_N^i &= [x_{N1}, x_{N2}, \ldots, x_{Nc}, y_{N1}, y_{N2}, \ldots, y_{Nc}]^T, \end{aligned} \tag{2}$$

where f_n^i is the shape vector of nth facial components of the ith image. Figure 2 shows sample shapes of all four components. After decomposition of all training face images, shape and texture vectors belonging to the same component, are grouped into separate sets to form N component-based datasets as given by:

$$\begin{aligned} \mathbf{S_n} &= [f_n^1, f_n^2, f_n^3, \cdots, f_n^L], \\ \mathbf{G_n} &= [g_n^1, g_n^2, g_n^3, \cdots, g_n^L], \\ \mathbf{F_n} &= \{\mathbf{S_n}|\mathbf{G_n}\}, \end{aligned} \tag{3}$$

where $\mathbf{F_n}$ is the nth component-based dataset containing shape vectors $\mathbf{S_n}$ and texture vectors $\mathbf{G_n}$ from all L training images.

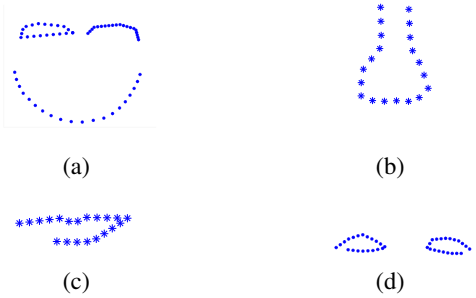


Fig. 2. Example shapes of four components i.e. cheeks + eyebrows, nose, mouth and eyes.

STEP-2: Each component-based dataset $\mathbf{F_n}$ is divided into a number of clusters by using LBG-Vector Quantization (VQ) [14].

This is an iterative algorithm that starts with taking the average of whole training set to be the initial code vector. This

is subsequently split into two code vectors which are subsequently optimized and divide the initial set into two clusters. These two clusters are split into four and the LBG-VQ process continues until the desired number of clusters is obtained. Thus each component-based dataset $\mathbf{F_n}$ is split into a number of clusters, using shape information. In general, the number of clusters for each component i.e. p, q, etc. can be different as shown in figure 1. Note that in this work eight clusters are employed per facial component, i.e. $p = q = \cdots = r = 8$, so

$$\mathbf{C_{nk}} = VQ\{\mathbf{S_n}\}, \tag{4}$$

where $\mathbf{C_{nk}}$ for $k = 1, 2, 3, \cdots, p$ is the kth cluster corresponding to the nth facial component and is obtained by employing VQ on shape vectors $\mathbf{S_n}$. Note that every cluster contains both shape and texture information so $\mathbf{C_{nk}} = \{\mathbf{S_{nk}}|\mathbf{G_{nk}}\}$ and the union of p $\mathbf{C_{nk}}$ clusters gives $\mathbf{F_n}$. Figure 3 shows sample shapes of cheeks + eyebrows taken from two different clusters. Obviously samples from same cluster will have similar shapes.

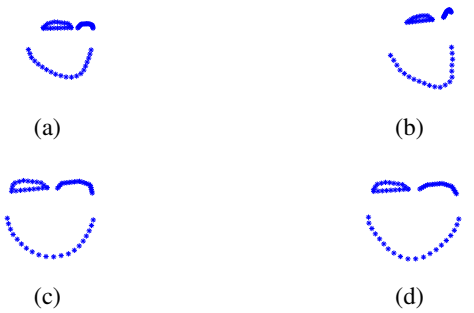


Fig. 3. Example shapes of cheeks + eyebrows component taken from two different clusters: (a) and (b) belong to one cluster, whereas (c) and (d) are from another cluster. Intra cluster similarities and inter cluster variations can be observed.

STEP-3: Here a corresponding component model and a model matrix $\mathbf{R_{nk}}$ is constructed for each cluster using conventional AAM optimization [1], see figure 1. Thus a training process produces $(p+q+\cdots+r)$ model matrices $\mathbf{R_{nk}}$ which are stored and can be subsequently used in the synthesis of an unseen input face image.

Note that the purpose of employing the second and third steps is to increase modeling accuracy by exploiting similarities in the shape characteristics of different facial components. This in turn can be viewed as an attempt to bridge the existing gap between the observed relatively low modeling accuracy of generic AAMs and the much higher accuracy required in person specific AAMs.

2.2. Synthesis Phase

Following the previously generated component models, the best model for each component is selected and these are subsequently fused to form a single set of parameters that represents the whole face. The proposed model fitting process can be explained as follows:

1. Input face image t is decomposed into components to form $f_1^t, f_2^t, \cdots, f_N^t$ shape vectors and their corresponding texture vectors $g_1^t, g_2^t, \cdots, g_N^t$.

2. Apply p conventional iterative AAM fitting algorithms based on model matrices $\mathbf{R_{nk}}$ $(k = 1, 2, \cdots, p)$ for each $n = 1, 2, \cdots, N$ component. Obtain same number of model parameterśvectors $c_{n1}^t, c_{n2}^t, \cdots, c_{np}^t$ for each component.

3. For each component, select the best model parameters' vector on the basis of a minimum average Mean Square Error (MSE). MSEs are formed between the original texture and the textures associated with the p models of each component and calculated across all iterations of the fitting algorithm [1].

4. Synthesize only the shape vectors $\hat{f}_1^t, \hat{f}_2^t, \cdots, \hat{f}_N^t$ of all N components with the best model parameters selected above. Combine all component-based shape vectors to form one single vector that represents the whole face shape i.e.

$$\hat{f}^t = [\hat{f}_1^t, \hat{f}_2^t, \hat{f}_3^t, \cdots, \hat{f}_N^t]. \tag{5}$$

5. Finally in the last step, the whole face texture is synthesized for the shape vector obtained in previous step. For this purpose, the corresponding texture of the best shape vector is projected in the Eigen space obtained from a whole face conventional AAM. The resulting model parameters are then used to synthesize the whole face texture.

3. EXPERIMENTAL RESULTS

Performance assessment and thus comparisons between MC/MM-AAM and conventional AAM has been performed using the publically available facial dataset IIM [15]. IIM consists of 240 annotated images (6 images per person). Each image is 640×480 pixels and comes with 58 hand labeled shape points which outline the face contours. From these images system training has been performed using 180 images (30 persons with 6 images per person). The remaining 60 images of 10 persons have been used for synthesis purposes.

Thus MC/MM-AAM and conventional AAM have been compared with respect to both shape and texture. In case of

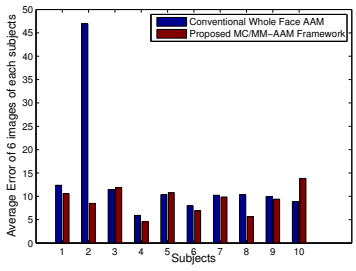
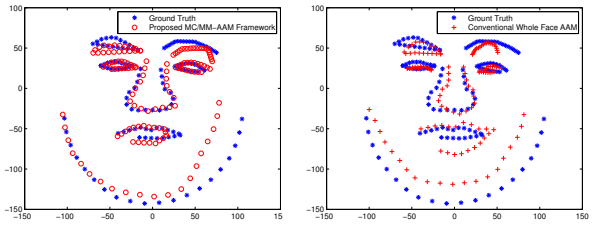


Fig. 4. Each bar in the graph represents **Average Shape Error** over six test images per person. In 7 out of 10 persons MC/MM-AAM outperforms AAM.



(a) (b)

Fig. 5. (a) modeled shape (Red-circles) of MC/MM-AAM are superimposed on Ground Truth Shape, (b) modeled shape of Conventional AAM (Red-plus signs) are superimposed on Ground Truth Shape.

shape, model fitting is evaluated using the point-to-point errors between modeled shape and ground truth point coordinates as suggested in [8] and given below:

$$E(\hat{f}^t, f^{gt}) = \frac{1}{M}\sum_{m=1}^{M}\sqrt{((\hat{x}_m^t - x_m^{gt})^2 + (\hat{y}_m^t - y_m^{gt})^2)} \quad (6)$$

where $E(f^t, f^{gt})$ is the point-to-point error between modeled shape $\hat{f}^t$ and ground truth shape f^{gt}. $E(\hat{f}^t, f^{gt})$ is then averaged over the 6 test images available per person, see figure 4. Note that MC/MM-AAM outperforms the conventional AAM system in 7 out of 10 cases. A further illustration of the potential performance of MC/MM-AAM is shown in figure 5. Figure 6 shows errors between modeled and original textures for both systems. Again MC/MM-AAM outperforms AAM in 7 out of 10 persons.

Finally, figure 7 illustrates visually and possibly more effectively, the MC/MM-AAM advantage over AAM, by offering a comparison between "Target", MC/MM-AAM modeled textures and AAM modeled textures.

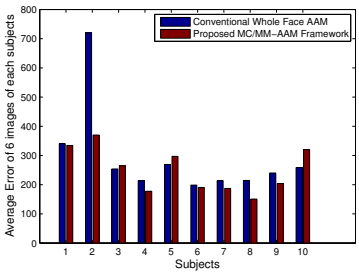


Fig. 6. Each bar in this graph represents Average Texture Error over six test images per person.

(a) (b) (c)

Fig. 7. (a) "Target" i.e. original image textures corresponding to ground truth shapes, (b) modeled textures corresponding to MC/MM-AAM modeled shapes, and (c) modeled textures corresponding to AAM modeled shapes.

4. CONCLUSION

In this paper, a novel face modeling framework is proposed which can successfully synthesize unseen face images. This has been effectively achieved by introducing two processing steps prior to the use of conventional AAM. One is component-based decomposition through which local face characteristics can be better accounted for and preserved. The second step involves shape based clustering of facial components into groups. The resulting Multi-Component/Multi-Model system offers in most cases significant modeling gains both in terms of shape and texture.

5. REFERENCES

[1] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *Computer VisionECCV98*, pp. 484–498, 1998.

[2] M.B. Stegmann, "Object tracking using active appearance models," in *Proc. Danish Conf. Pattern Recog. Image Anal*, 2001, vol. 1, pp. 54–60.

[3] 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," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 5, pp. 415–423, 2001.

[4] E. Patterson, A. Sethuram, M. Albert, and K. Ricanek, "Comparison of synthetic face aging to age progression by forensic sketch artist," in *The Seventh IASTED International Conference on Visualization, Imaging and Image Processing*. ACTA Press, 2007, pp. 247–252.

[5] X. Geng, Z.H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 12, pp. 2234–2240, 2007.

[6] 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.

[7] X. Li, S.J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 2, pp. 145–155, 2008.

[8] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 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] C. Zhang and F. Cohen, "Component-based active appearance models for face modeling," *Advances in Biometrics, Lecture Notes in Computer Science*, pp. 206–212, 2005.

[11] R.C. Luo, C.Y. Huang, and P.H. Lin, "Alignment and tracking of facial features with component-based active appearance models and optical flow," in *Advanced Intelligent Mechatronics (AIM), 2011 IEEE/ASME International Conference on*. IEEE, 2011, pp. 1058–1063.

[12] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol. 20, no. 9, pp. 657–664, 2002.

[13] 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.

[14] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *Communications, IEEE Transactions on*, vol. 28, no. 1, pp. 84–95, 1980.

[15] 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.