

A Comparative Study of Active Appearance Model Annotation Schemes for the Face

Amrutha Sethuram^{*}
Face Aging Group
UNCW, USA
sethurama@uncw.edu

Karl Ricanek
Face Aging Group
UNCW, USA
ricanekk@uncw.edu

Eric Patterson
Face Aging Group
UNCW, USA
patterson@uncw.edu

ABSTRACT

Active appearance models (AAM), a group of flexible deformable models, have been widely used in various applications such as, object tracking, medical image segmentation and synthesis. AAMs are statistical models which model the shape and texture of an object. There has been much published work in this field to improve the speed and fitting accuracy. However, there has not been any significant study related to the quantity and selection of annotation points (landmarks) used to define the object and its texture. This paper proposes four different annotation schemes used for modeling the human face and evaluates each scheme in regard to reconstructing face images. In pursuit of choosing a particular annotation scheme for age progression and synthesis, this paper presents qualitative and quantitative methods for evaluation.

Keywords

face aging, active appearance models, annotation schemes

1. INTRODUCTION

Active Appearance Models (AAM) have been successfully used in many applications such as tracking, medical image segmentation, recognition and synthesis. AAM is a flexible and powerful learning-based deformable model proposed by Cootes et al. [24]. A primary advantage of AAM is that both the shape and texture of the deformable object is modeled through a set of training examples and a range of valid instances of the object can be synthesized. There has been a vigor in the research community involving AAMs owing to the model's flexibility and the simple framework.

The literature found on AAMs can be broadly categorized into theory, extensions and improvements of AAM algorithms and finally applications of AAM. The introduction, theory and implementation of the AAM algorithm is described in [24]. Many improvements on the basic AAM

have been proposed to strengthen the ability of the AAM. Cootes et al.[3] introduced prior information to guide the model fitting and obtain higher accuracy. Combining AAM with ASM, Yan et al. [28] obtain the model robustness to illumination. Stegmann et al. [22] discuss the details of extension and applications of the AAM. Variations to the AAM and a discussion on the different methods for estimating the update matrix is provided by Cootes et al. [2]. Applications of the AAM are plentiful. Due to the discrimination capabilities of AAM, it is used for feature extraction for recognition tasks. Xiao et al. and Tao et al. [26] [23] use AAM for face and expression recognition. AAMs have been used for tracking in real time situation [20]. In addition, AAMs have been used in important computer aided diagnosis applications such as segmentation of MRI of the knee [24] and segmentation of cardiac images for diagnosing heart diseases [12]. Finally, one of the main applications is in synthesis [7] [16]. Recently, a comprehensive review of the active appearance models based on efficiency, discrimination and robustness has been provided by Gao et al. [5].

Modern day face modeling and facial extraction methods are extensively based on AAMs. Lee et al. [9] attempt to find a suitable facial appearance modelling method for AAMs by a comparative study. Although AAMs are 2D, they can still be used to model 3D phenomena such as faces moving across pose. Xiao et al. [27] study the representational power of AAMs and show that they can model anything a 3D Morphable Model can, albeit with additional shape parameters. They propose a real-time algorithm, Combined 2D+3D AAM, for fitting the AAM while enforcing 3D shape constraints.

In the recent past, AAMs have been widely used in the field of face biometrics. AAMs are used to model the shape and texture of face images of the training set. A new set of face images are then synthesized based on the AAM parameters. AAMs are used to model the pattern of facial aging by Lanitis et al. [8]. Patterson et al. [15] [16] [19] [14] extend the concept of facial aging to adults by utilizing AAMs for face modeling. Luu et al. [10] and Ricanek et al. [18] use AAM models for age estimation based on the face model.

It should be noted that AAM is a statistical technique which models shape and texture of an object. The first step in modeling the shape is to obtain a set of annotation points or labeled landmarks. These annotation points are then concatenated into a shape vector while the texture, represented by pixel intensities is captured by sampling a suitable image warping function e.g. a piece-wise affine warp based on Delaunay triangulation is used for texture model formu-

^{*}Corresponding author

lation. Thus, a training set of images with corresponding annotation points is required for AAM model training. The selection and the optimum number of such annotation points has not been studied thus far.

1.1 Our Contribution

In this paper, we address the selection of landmark points by evaluating four schemes for annotating the face. The schemes are based upon anthropometric landmarks of the soft tissue and the skull. Additional points were selected based upon aging trends of the face. These schemes are evaluated in regard to proper reconstruction of a face and image quality of synthesized reconstructed faces. Although this study is geared towards texture reconstruction for face images and particularly aging, similar ideas may be useful in recognition tasks using other related biometric modalities e.g. periocular recognition. Periocular recognition is based upon the analysis of the eye region for authentication [25] [13]. It is universally accepted that the region around the eye changes as adults age. Further, this approach can be generalized for any object recognition problem that exhibits temporal variability. A comparative evaluation of the schemes is performed by visual analysis of how best the source images can be reconstructed using the different annotation schemes. Quantitative evaluation metrics viz. the root mean square error (RMSE) and best match scores on experiments conducted utilizing a commercial face recognition SDK are used to analyze the results.

The remainder of the paper is structured as follows. In section 2, a brief review of the AAM model is presented. The problem statement and our contribution is presented in section 3. In section 4, we present the experimental results and evaluation of the annotation schemes. Finally, we summarize our work, draw conclusions and outline future work in section 5.

2. ACTIVE APPEARANCE MODELS

Active appearance model was first proposed by Cootes et. al [24]. AAM decouples and models shape and pixel intensities of an object. The latter is usually referred to as texture. The basic steps involved in AAM is as shown in Figure 1. A very important step in building an AAM model is identifying a set of landmarks and obtaining a training set of images with the corresponding annotation points either by hand, or by semi- to completely automated methods. As described in [24], the AAM model can be generated in three main steps: (1) A statistical shape model is constructed to model the shape variations of an object using a set of annotated training images. (2) A texture model is then built to model the texture variations, which is represented by intensities of the pixels. (3) A final appearance model is then built by combining the shape and the texture models.

2.1 Statistical shape model

A statistical shape model is built from a set of annotated training images. In a 2-D case, a shape is represented by concatenating n point vectors $\{(x_i, y_i)\}$

$$x = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)^T \quad (1)$$

The shapes are then normalized by Procrustes analysis [6] and projected onto the shape subspace created by PCA

$$x = \bar{x} + P_s \cdot b_s \quad (2)$$

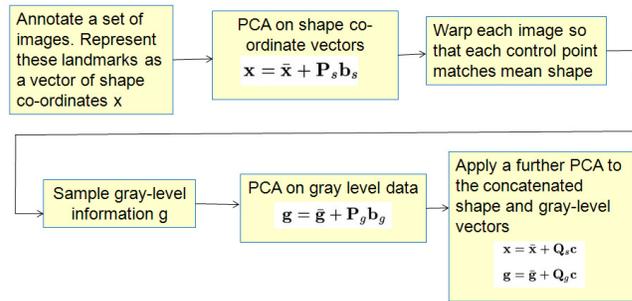


Figure 1: Active Appearance Model - Methodology

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 describing the modes of variations derived from training set, and b_s includes the shape parameters in the shape subspace. Subsequently, based on the corresponding points, images in the training set are warped to the mean shape to produce *shape-free patches*.

2.2 Statistical texture model

The texture model is generated very similar to the shape model. Based on the shape free patch, the texture can be raster scanned into a vector g . Then the texture is linearly normalized by the parameters $u = (\alpha, \beta)^T$ and g is given by

$$g = \frac{(g_i - \beta \cdot 1)}{\alpha} \quad (3)$$

where α and β are, respectively, the mean and the variance of the texture g , and $1 = [1, 1, \dots, 1]^T$ is the vector with the same length of g_i . The texture is ultimately projected onto the texture subspace based on PCA

$$g = \bar{g} + P_g b_g \quad (4)$$

where \bar{g} denotes the mean texture, $P_g = \{g_j\}$ is 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.

2.3 Combined appearance model

Finally, the coupled relationship between the shape and the texture is analyzed by PCA and the appearance subspace is created. At the end, the shape and the appearance can be described as follows:

$$x = \bar{x} + Q_s \cdot c \quad (5)$$

$$g = \bar{g} + Q_g \cdot c \quad (6)$$

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

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s (P_s)^T (x - \bar{x}) \\ (P_g)^T (g - \bar{g}) \end{pmatrix} \quad (7)$$

and Q is the matrix of eigen vectors of b .

3. METHODOLOGY

One of the many applications of AAM is facial age progression and age estimation. As already discussed in the above section, the first step in the modeling process is selecting an *annotation scheme*. Figure 2 shows an application of AAM where a source image of an individual is age progressed to several decades. For such an application, an annotation scheme should be chosen such that the AAM model can generate accurate reconstructions of original images. Hence, for a specific task of synthetic age progression, the problem statement can then be described as “*What is a good annotation scheme to be adopted for synthetic age progression?*”. Unfortunately, there is a not a *gold standard* to address such a problem. In this work, we analyze and compare four annotation schemes to address such a problem.

3.1 Shapes and Annotations

A shape can be defined as a *collection of corresponding border points*. Alternately, as defined by D.G.Kendall, **shape** is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object. One way to describe a shape is by locating a finite number of points on the outline. A **landmark** or an **annotation** can then be defined as a point of correspondence on each object that matches between and within populations. Dryden et. al. [4] further discriminate landmarks into three subgroups:

1. Anatomical landmarks: Points assigned by an expert that correspond between organisms in some biologically meaningful way
2. Mathematical landmarks: Points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extremum point
3. Pseudo landmarks: Constructed points on an object either around the outline or between landmarks

Although the concept of landmarks is conceptually useful, the process of acquisition can be cumbersome. In addition to placing annotations manually, the process usually involves comparing annotations to ensure correspondence across the training set.

The choice of annotation scheme is important when AAMs are used to develop models for synthetic face aging i.e. the capability of the AAMs to reconstruct facial images and synthesize older facial images, with minimum texture loss, is also dependent on the choice of annotation scheme. From the literature found on dynamics of craniofacial aging [1] and consultation with an anthropologist, in addition to the above guidelines on landmarks, four annotation schemes were devised. As already mentioned, the schemes are based upon anthropometric landmarks of the soft tissue and the skull with additional points selected based upon aging trends of the face.

Figure 3 shows the different annotation schemes that are reviewed in this paper. The number of annotation points for each of these schemes are as shown in Table 1. It should be noted that Scheme_A and Scheme_B are detailed mark-up schemes. Scheme_C is derived from Scheme_A and Scheme_D is derived from Scheme_B by keeping the anatomical and mathematical landmarks. The goal was to choose an optimum annotation scheme for the face. Experimental set-up, evaluations and results are presented in the next section.

Table 1: Number of landmark points

Annotation Scheme	Number Landmark Points
A	161
B	252
C	114
D	138

Table 2: Dataset details

Group	Number of images
Caucasian American Female	95
Caucasian American Male	101
African American Female	95
African American Male	95

4. EXPERIMENTAL RESULTS

Stegmann’s AAM-API [21] was used for implementation and evaluation of the annotation schemes. The model was set to capture 95% of the variance in the training data. Four groups of datasets - Caucasian American Female (CAF), Caucasian American Male (CAM), African American Female(AAF) and African American Male(AAM) - were created based on ethnicity and gender. The images were chosen to be clean and without any occlusions from the Morph [17] and the PAL database [11]. In addition, the ethno-gender groups were formulated with equal distribution of images between 18 and 65 years. The number of images in each group is as shown in Table 2.

Evaluation

Individual ethno-gender AAM models were built using the AAM-API for the training data set for the four different schemes. The number of AAM parameters obtained for the models for the various schemes are as shown in Table 3

4.1 Reconstructed Images

For each of the ethno-gender models and the annotation scheme, active appearance parameters were generated for each training image. The image was then reconstructed back using the appearance parameters and the model. Some example images and their reconstructions for each of the ethno-gender groups and the annotation schemes are as shown in Figures 4 - 11

On careful visual inspection of the reconstructed images, it can be observed that Scheme_B and Scheme_D can better preserve the shape of the physical features such as the eyebrows, eyes, nose and the mouth. Also, in some of the images, the texture reconstruction seems to be better in Scheme_B and Scheme_D. In addition to this qualitative measure, a couple of quantitative measures were evaluated as explained in the next section.

4.2 Quantitative analysis - RMSE

Table 3: Number of AAM parameters

	Scheme_A	Scheme_B	Scheme_C	Scheme_D
CAF	51	49	48	46
CAM	50	51	42	48
AAF	52	48	50	45
AAM	52	49	52	46



Figure 2: Age progression: an application of AAM

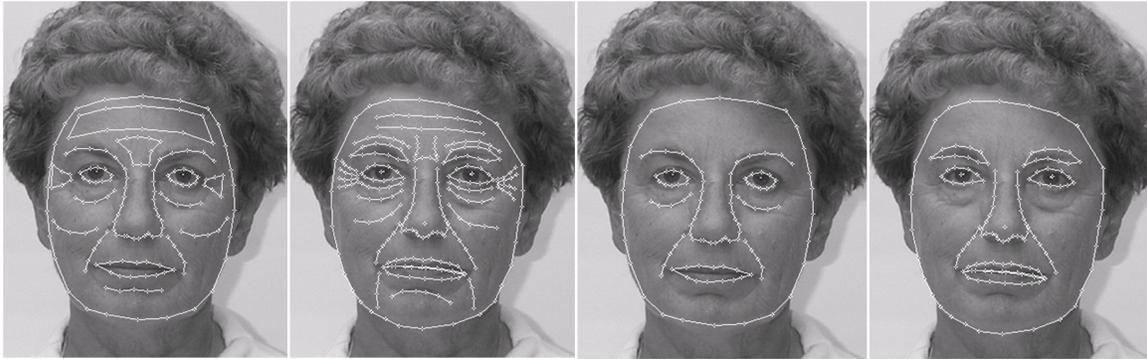


Figure 3: Annotation schemes: (L-R) - Scheme_A, Scheme_B, Scheme_C and Scheme_D



Figure 4: CAF image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D



Figure 5: CAF image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D



Figure 6: CAM image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D

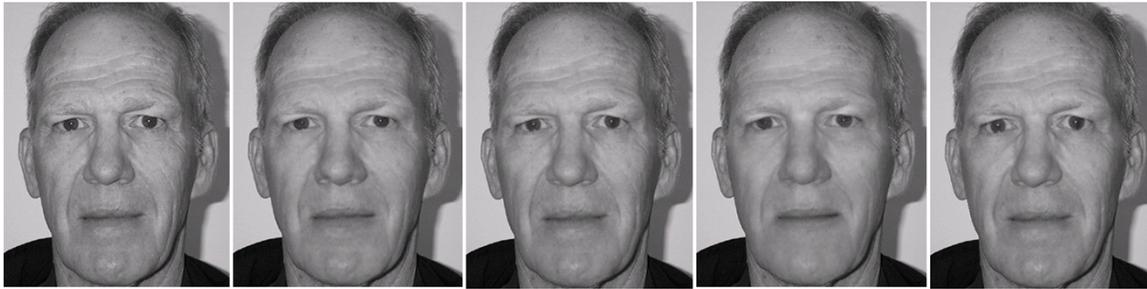


Figure 7: CAM image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D



Figure 8: AAF image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D



Figure 9: AAF image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D



Figure 10: AAM Image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D

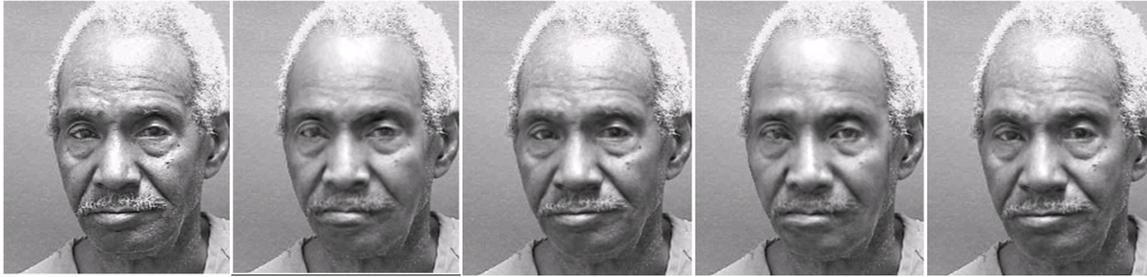


Figure 11: AAM Image reconstructions: (L-R) Original Image, Scheme_A, Scheme_B, Scheme_C, Scheme_D

Although in practical scenarios, qualitative judgement is better for such experiments, some quantitative evaluation of the results is possible. In particular, we calculated:

1. Root mean squared error (RMSE) between the original image and the reconstructed image for the four annotation schemes of each ethno-gender group. Let I_S and I_R be the original source image and the respective reconstructed image. The RMSE for each image can then be calculated as

$$RMSE = \sqrt{\text{mean}(|I_S - I_R|^2)} \quad (8)$$

2. Frequency of minimum RMSE occurrences: For each training image, RMSE values were compared across all annotation schemes. A voting scheme was then adopted to flag the scheme with minimum RMSE value for each example image. The number of such votes was then counted for each annotation scheme which is termed as the *frequency of minimum RMSE* occurrences. A plot of such frequency of minimum RMSE occurrence across all the ethno-gender models is as shown in Figure 12.

From Figure 12, it is clear that Scheme_B and Scheme_D have the highest number of minimum RMSE values across all the ethno-gender models. This further suggests that Scheme_B and Scheme_D are better choices of annotation schemes for this application.

4.3 Quantitative analysis - Best match score

To further evaluate the efficiency of the annotation schemes, FaceVACS ©SDK from Cognitec, a commercial face recognition toolkit was used to generate the match scores. The experiment was set-up such that

1. For each of the training images, a gallery G was formed

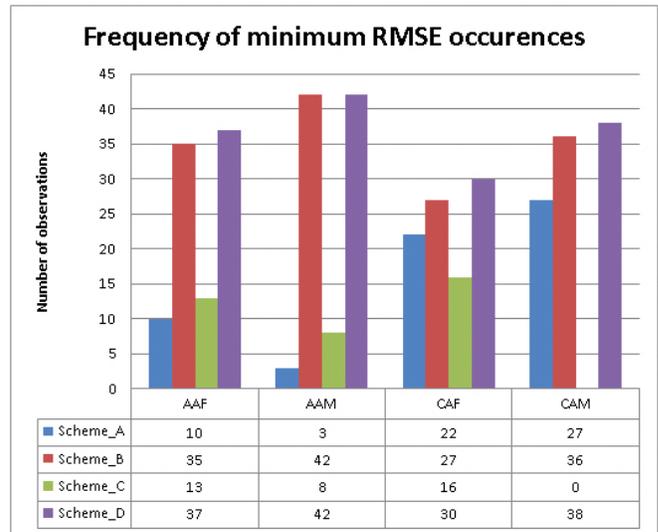


Figure 12: Frequency of minimum RMSE occurrences

from the reconstructed images of each of the schemes.

$$G = \{R_A, R_B, R_C, R_D\} \quad (9)$$

where R_A, R_B, R_C and R_D are reconstructed images under the different annotation schemes. The original image I was set as the probe image.

2. The probe image was then matched to each image in the gallery to get the corresponding match scores.
3. A voting scheme to flag the annotation scheme with the highest match score was adopted.
4. This was then repeated for all the examples across the ethno-gender models. Figure 13 shows the frequency of best match score using the FaceVACS SDK. Only examples for which a valid score was generated by the SDK for each of the schemes were included in the analysis.

From Figure 13, it is observed that Scheme_B and Scheme_D have higher frequency of best match scores than the other two schemes. Owing to all the qualitative and quantitative evaluation measures as discussed above, it can be inferred that Scheme_B and Scheme_D, which are comparable in performance, are better annotation schemes when compared to the other two schemes. It should be noted that one of the desired features of annotation schemes is the ability to automate the process. Since Scheme_D is a subset of Scheme_B and for the following reasons: 1) the number of annotation points is much lesser in Scheme_D than Scheme_B, but with comparable performance; 2) Scheme_D is a much better candidate for automation; 3) Due to fewer annotation points, Scheme_D is less prone to human/automated annotation errors; Scheme_D can be chosen as the best annotation scheme among the schemes chosen for modeling the human face.

5. CONCLUSIONS

In this paper, we have presented four annotation schemes for AAM modeling of the human face. Two of the annotation schemes are derivatives of the other two schemes. AAM models based on these four annotation schemes are evaluated on a set of four ethno-gender models - Caucasian American female, Caucasian American male, African American female and African American male models. We have presented a qualitative (visual inspection) and quantitative (RMSE and best match scores) analysis to evaluate the four annotation schemes. Based on this evaluation, the annotation scheme with 232 landmark points and its derivative performs better both qualitatively and quantitatively. However due to desired features such as the ability to automate and the number of annotation points which are lesser when compared to the original scheme, the derived scheme from the original 232 annotation scheme was chosen as the optimum scheme for building synthetic models for the face using AAM. Future work will involve comparing the performance of the 232 landmark annotation scheme and its derivative for age progression.

6. ACKNOWLEDGMENTS

This work is supported by the Intelligence Advanced Research Projects Activity, Federal Bureau of Investigation and the Biometrics Task Force. The opinions, findings and

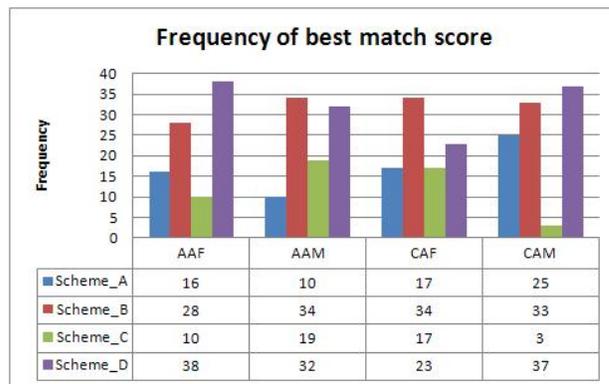


Figure 13: Frequency of best match scores using FaceVACS SDK

conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

7. REFERENCES

- [1] A. M. Albert, K. R. Jr, and E. Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, 172(1):1–9, October 2007.
- [2] T. F. Cootes and P. Kittipanyangam. Comparing variations on the active appearance model algorithm. *Proc. Brit. Mach. Vis. Conf.*, 2:837–846, 2002.
- [3] T. F. Cootes and C. J. Taylor. Constrained active appearance models. *Proc. IEEE Int. Conf. Comput. Vis.*, 1:748–754, 2001.
- [4] I. Dryden and K. Mardia. *Statistical shape analysis*. John Wiley and Sons, 1986.
- [5] X. Gao, Y. Su, X. Li, and D. Tao. A review of active appearance models. *IEEE Transactions On Systems, Man, And Cybernetics-UPART C: Applications and Review*, 40(2), March 2010.
- [6] C. R. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. B*, 53(2):285–339, 1991.
- [7] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):743–756, July 1997.
- [8] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4), April 2002.
- [9] S. J. Lee, K. R. Park, and J. Kima. A comparative study of facial appearance modeling methods for active appearance models. *Pattern Recognition Letters*, 30(14):1335–1346, October 2009.
- [10] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen. Age estimation using active appearance models and support vector machine regression. *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, pages 314–318, September 2009.

- [11] M. Minear and D. Park. A lifespan database of adult facial stimuli. *Behavior research methods, instruments and computers : a journal of the Psychonomic Society, Inc.*, 36:630–633, 2004.
- [12] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C. Reiber, and M. Sonka. Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac mr mages. *IEEE Trans. Med. Imag.*, 20(5):415–423, May 2001.
- [13] U. Park, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2009.
- [14] E. Patterson, A. Sethuram, M. Albert, and K. Ricanek. Comparison of synthetic face aging to age progression by forensic sketch artist. *Proceedings of the seventh IASTED International Conference on Visualization, Imaging, and Image Processing*, pages 247–252, August 2007.
- [15] E. Patterson, A. Sethuram, A. M, K. Ricanek, and M. King. Aspects of age variation in facial morphology affecting biometrics. *1st IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1–6, 2007.
- [16] E. Patterson, A. Sethuram, K. Ricanek, and F. Bingham. Improvements in active appearance model based synthetic age progression for adult aging. *3rd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 104–108, November 2009.
- [17] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. *7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 341–345, April 2006.
- [18] K. Ricanek, Y. Wang, C. Chen, and S. J. Simmons. Generalized multi-ethnic face age-estimation. *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, pages 127–132, September 2009.
- [19] A. Sethuram, E. Patterson, K. Ricanek, and A. Rawls. Improvements and performance evaluation concerning synthetic age progression and face recognition affected by adult aging. *Proceedings of the 3rd IEEE International Conference on Biometrics*, 5558(5):62–71, June 2009.
- [20] M. B. Stegmann. Object tracking using active appearance models. *Proc. Danish Conf. Pattern Recog. Image Anal.*, 1:54–60, 2001.
- [21] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. Fame: A flexible appearance modeling environment. *IEEE Trans. Med. Imag.*, 22(10):1319–1331, October 2003.
- [22] M. B. Stegmann, R. Fisker, and B. K. Ersbøll. Extending and applying active appearance models for automated, high precision segmentation in different image modalities. *Proc. Scand. Conf. Image Anal.*, pages 90–97, 2001.
- [23] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. Maybank. Bayesian tensor approach for 3-d face modeling. *IEEE Trans. Circuits Syst. Video Technol.*, 18(10):1397–1410, October 2008.
- [24] T.F.Cootes, G.J.Edwards, and C.J.Taylor. Active appearance models. *Proc. Eur. Conf. Comput. Vis.*, 2:484–498, 1998.
- [25] D. L. Woodard, S. Pundlik, J. Lyle, and P. Miller. Periocular region appearance cues for biometric identification. *IEEE Conf. Computer Vision and Pattern Recognition 2010 /IEEE Biometrics Council Workshop on Biometrics*, pages 162–169, 2010.
- [26] B. Xiao, X. Gao, D. Tao, and X. Li. A new approach for face recognition by sketches in photos. *Signal Process.*, 89(8):1576–1588, 2009.
- [27] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 535–542, June 2004.
- [28] S. Yan, C. Liu, S. Z. Li, H. Zhang, H. Shum, and Q. Cheng. Face alignment using texture-constrained active shape models. *Image Vis. Comput.*, 21(1):69–75, 2003.