

# Improvements and Performance Evaluation Concerning Synthetic Age Progression and Face Recognition Affected by Adult Aging

Amrutha Sethuram, Eric Patterson, Karl Ricanek, and Allen Rawls

Face Aging Group, Computer Science Department, UNCW, USA  
{sethurama,patterson,ricanek,rawlsa}@uncw.edu  
<http://www.faceaginggroup.com>

**Abstract.** Aging of the face degrades the performance of face recognition algorithms. This paper presents recent work in synthetic age progression as well as performance comparisons for modern face recognition systems. Two top-performing, commercial systems along with a traditional PCA-based face recognizer are compared. It is shown that the commercial systems perform better than the baseline PCA algorithm, but their performance still deteriorates on an aged data-set. It is also shown that the use of our aging model improves the rank-one accuracy in these systems.

**Keywords:** Facial aging, synthetic age progression, performance evaluation.

## 1 Introduction

Of the many challenges to the performance of modern face-recognition (FR) algorithms, the robustness to age variation has perhaps been studied the least. Algorithms have made significant strides in the past decades toward dealing with real-world conditions such as pose and lighting. However, temporal variance particularly due to longer spans where aging changes are evident remains a difficulty. Overcoming this difficulty would bring useful improvements in robustness to biometrics and would also benefit related fields of forensics and medicine.

In this work, we demonstrate the severity of performance loss due to adult aging and discuss steps towards improving the related robustness of face-based biometrics. The organization of the paper is as follows. In section 2, a brief discussion of biometric-related work on aging of the human face is provided. In section 3, recent improvements concerning synthetic age progression are presented. Comparison and performance evaluation of two commercial face-recognition systems on longitudinal images are presented in section 4. The final section gives conclusions and a brief discussion on continued research.

## 2 Related Studies in Face Aging

Some attempts have been made at unifying anthropological and forensic studies that consider aspects of human aging and resulting changes in the face [3] [4], but

this knowledge has not been used in all cases to inform the relatively few studies conducted regarding the effects of growth and development and aging on synthetic progression and face recognition technologies [5] [6] [7] [8] [9]. Concerning simulation of aging in facial images or models, a few approaches have been considered. They can broadly be divided into geometric and/or physical-based [6] and analysis-synthesis model-based methods [5] [6] [7]. Some of the most promising analysis-synthesis methods have involved a variation of active appearance models (AAM) [10], [11] for growth and development [12], [13] or adult aging [8]. Several of these did not distinguish between childhood growth-and-development and adult effects of aging. It should be stressed, however, that these are distinct processes; pre-adult processes cause largely shape-based changes in the mid and lower face. Aging, however, occurring after roughly twenty years of age, consists of shape deformation from weight change, minor bone remodeling, and tissue degeneration combined with gravity effects along with significant textural change due to a variety of complex phenomenon [3]. It is this textural change that is critical to photo-realistic synthesis of age-progression and its use as a synthetic template to improve face recognition.

### 3 Improvements in Synthetic Age Progression

In this section, we discuss recent improvements made in synthetic age progression of images of adults. Our goal is to produce very accurate estimates, across human variation, of the changes in appearance of an individual due to the effects of aging over a number of years and to create synthetic images that realistically portray in a data-informed manner, the actual effects that occur in humans as documented by medical and anthropological literature. For the most part, all current approaches are still “general” models of aging because truly accurate models of individualized, idiosyncratic aging requires a large set of representative data for the same individual across adult lifespan. Such data is not currently available. Our ultimate aim, though, is to create models that may be altered specifically and automatically to improve accuracy of predictive images by accounting for personal individuality and lifestyle as possibly indicated by accompanying meta-data.

The improvements that we have achieved in our recent work are due in part to three areas of focus – improving quality and quantity of data, improving labeling of landmarks used for models, and improving methods for training and simulation in our synthetic age-progression technique. We include several images age-progressed from the most notable longitudinal databases FG-NET [16] and MORPH [1] for human comparison of recognizable images.

Our work uses an analysis-synthesis face-model approach, using Active Appearance Models (AAM) as the method of representation for both shape and texture information. Our recent methods have used support vector regression (SVR) to attempt to learn age-based properties of AAM parameters across the entire training set. Also, gradient-regression-based AAMs are used due to earlier work that suggests these may best represent texture information in faces.

Once the AAM is created, a Monte-Carlo simulation is performed (with 500,000 iterations) to garner the feature information learned by the support vectors, generating random face images within the range of acceptable face images, and classifying them based on their age estimated by the SVR solution. Bins are created and averaged for each age, ultimately creating a table of AAM parameters that may be indexed and differenced to synthetically age progress or regress a facial image. Particular care was taken to very accurately label (manually) the hundreds of images used to build these models. As our method for age-progression is data-driven, large numbers of clear images of sufficient resolution are necessary to build accurate image-based face models of what occurs over years of aging. We began by augmenting some of our past training sets and building a European-descent male data-set of approximately 200 images, demonstrating visually comparable performance improvements in the modes of the AAM, the simulated table of representative age parameters, and synthetic aging of images of similar individuals. Pictured first in Figure 1 are the representative faces for each decade of adult life as learned by the aging model. For comparison of synthesis improvements on an individual, Figure 2 shows an age-progression using the newer male model versus an earlier version shown in previously published work [15]. Better representative shape changes and finer texture changes are evident upon comparison of stills and movies of progression. Our most recent work has expanded the training set of images to that of 541 distinct individuals, currently of neutral expression and for the most part lacking facial hair, glasses, or other obstructions. Both the male-only model and the larger, varied model show an obvious representation of age changes in the first few modes of the AAM parameters. The larger, varied model also demonstrates observable gender and ethnicity representation in the modes. For example, one of the first few modes of the most recent, varied model is shown in Figure 3.

For additional comparison, images from FG-NET and MORPH are shown in Figures 4 and 5. These images are actually produced completely without previously incorporated information in either the AAM or the aging model, as they were not used in training at either level. It is possible that including them for the AAM training could improve the texture quality of the reconstructed and synthetically aged images, and future work may include comparisons of models aged that are not present in the AAM versus those that are (as this would certainly be possible in real-world scenarios to use all available information). In general, the images produced by the most recent methods appear to better illustrate texture and shape differences created by skin wrinkling and tissue sag over time, and these correspond well to those documented in the anthropological and medical literature [3] [4]. Minor shape changes, the appearance of rhytids, deepening of lines such as the nasio-labial, and ptosis of the orbital tissue are all well demonstrated. Some areas such as the forehead, though, do not demonstrate as noticeable and possibly not sufficiently accurate texture detail, but it is not clear at this point where the weakness may lie. In general, the progression of images demonstrate well the expected changes in shape and texture due to tissue and skin changes from adult aging and are better representative than past models.



**Fig. 1.** The representative, average male of European-descent for each decade ( $2^{nd}$  -  $8^{th}$ ) as learned by the SVR training on approximately 200 images and generated by the Monte-Carlo simulation

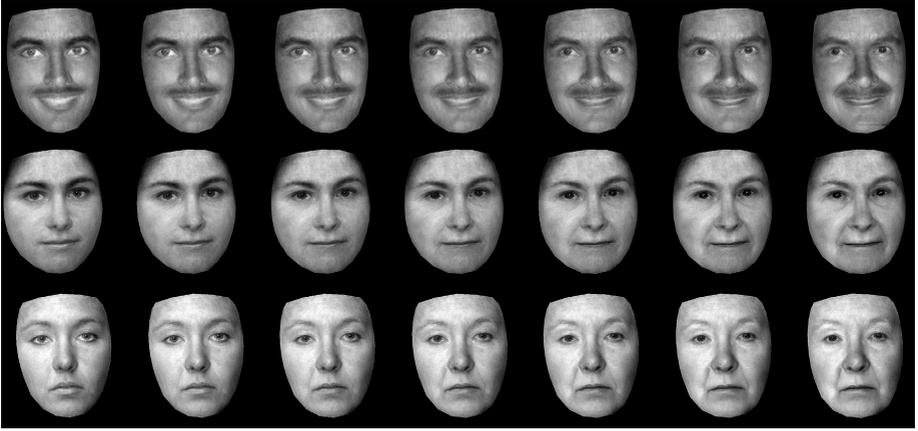


**Fig. 2.** Images from earlier work included for comparison, one of the authors age-progressed through the major decades. The top two rows are earlier versions (at original age 30 then 40,50,60,70), the bottom row aged using the aforementioned newer model (synthetic from 20 to 80).

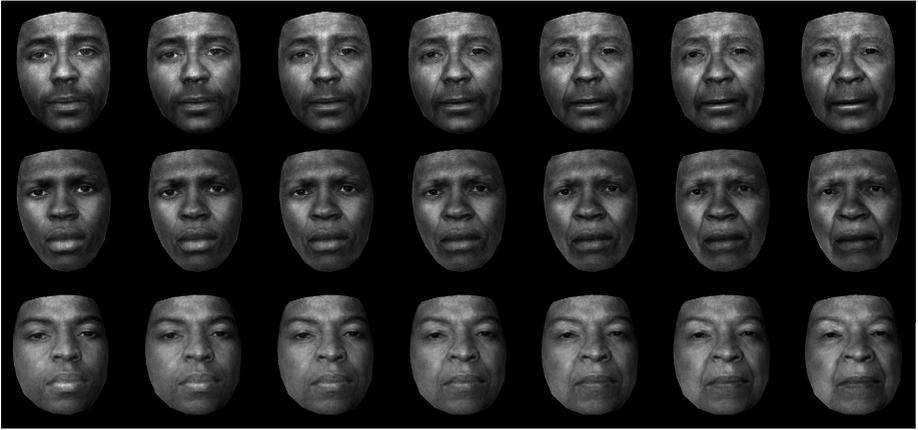


**Fig. 3.** One of the first principal modes of the AAM trained on 541 images representative of adult aging, demonstrating representation of age, gender, and ethnic variation

The current model also seems capable of reasonably reconstructing a wide variety of faces of different gender and ethnicity and representing characteristics of general aging within each of those. In the following sections, face-recognition performance and the application of these synthetic techniques will be discussed.



**Fig. 4.** Samples from the FG-NET database aged through the decades 20 to 80 years



**Fig. 5.** Samples from the MORPH database aged through the decades 20 to 80 years

## 4 Performance Evaluation and Comparison

Two modern, high-performance, and commercially available face recognition systems were evaluated for performance on face images of individuals separated by several years of aging. For purposes of anonymity, the systems will be referred to as SDK *A* and SDK *B*. Methods and related results achieved using the SDKs provided by the two vendors are discussed in the following subsections.

### 4.1 Database

Currently, FG-NET [16] and MORPH [1] are the only two publicly available face-image databases that attempt to represent longitudinal changes. The FG-NET

aging database consists of 1002 images of 82 subjects with a span of ages ranging from 0-69. Morph [17] is an ongoing data corpus in two albums that includes facial images of numerous individuals and with essential metadata, such as age, gender, ancestry, height, and weight. MORPH Album 1 contains digital scans (much like FG-NET) of photographs of 625 individuals with a total of 1690 images. MORPH Album 2 at present, contains 55,608 images of 13,673 individuals. In this work, a dataset comprising of both the FG-NET and MORPH Album 1 is used for training and testing the algorithms. 1299 images of individuals older than 18 and younger than 50 years of age were used to construct the training gallery and testing probes.

## 4.2 Experimental Setup

The goal of the experiments was to evaluate and compare the performance of advanced commercial systems relative to aging. A baseline PCA face recognition algorithm from the Colorado State University’s Face Identification Evaluation System [18] was used as well. Two different experiments were conducted to evaluate the performance.

**Experiment 1.** This experiment was set-up to evaluate the recognition rate of the SDKs and the baseline PCA algorithm on individuals as they aged. A combined dataset of Morph 1 and FG-NET databases was used for this experiment. The gallery and probes were constructed as outlined in Table. 1 w.r.t the age groups.

**Table 1.** Age range for the *gallery* and *probes* for Experiment 1

G/P	Age-Range(years)
Gallery (G)	18-30
Probe A ( $P_a$ )	18-30
Probe B ( $P_b$ )	31-40
Probe C ( $P_c$ )	41-50

The gallery data,  $G$

$$G = [g_1, g_2, \dots, g_n]; \quad (1)$$

was constructed such that youngest image of each person in the age range of 18-30 years available in the dataset was selected. Let  $P$  be the probe data defined by

$$P = [p_1, p_2, \dots, p_m]; \quad (2)$$

which was constructed with aged images of the subjects from the Gallery. Three different probes  $P_a$ ,  $P_b$  and  $P_c$  were generated to include aged images in different age-groups as shown in Table 1.

A quality rating of 80% or higher (based on the more rigorous SDK quality rating method) was used as a threshold in choosing images for both the gallery

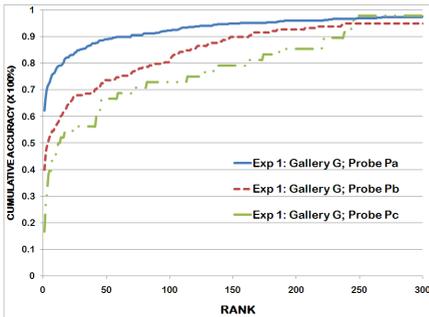
and probes. Input images were converted to 8-bit grayscale for all the experiments. Having set-up the data, SDK *A* and SDK *B* were trained on the images in the gallery *G*. For every experiment, scores were obtained for each image in the probe compared with every image in the gallery. The match list obtained was then ranked with decreasing order of the scores generated.

In keeping with a standard face evaluation system, Colorado State University’s Face Identification Evaluation System (Version 5) [2] was used. This system provides a standard set of well known algorithms for face identification that assist in providing a baseline metric and are used to compare and contrast relative methods for identification. For this baseline comparison, Principle Component Analysis (PCA) was used. Images were prepared using CSU’s preprocessing normalization techniques. The Euclidean distance measure was used to rank the probes against the gallery. To compare the performance directly with the scores generated with SDK *A* and SDK *B*, these distance measures were converted to similarity-like measures and were normalised using the following equation:

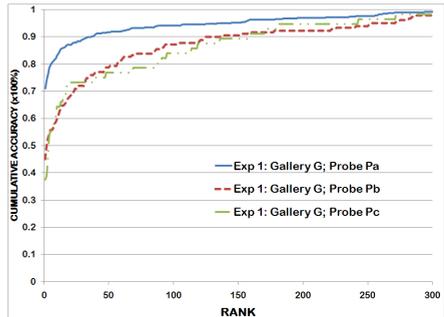
$$CSU\_Score = 1 - \left[ \frac{Distance_i - \min(Distance)}{\max(Distance) - \min(Distance)} \right] \quad (3)$$

### 4.3 Results and Discussion

Cumulative Match Characteristic (CMC) curves were obtained for the different probe sets. Performance of SDK *A* for different probes is shown in Figure 6, SDK *B* in Figure 7. The baseline PCA performance is shown in Figure. 8. For the same probe, (*P<sub>a</sub>*), CMC curves were obtained for the two SDKs and baseline PCA algorithm. Results are shown in Figure 9. The experiment was repeated on probe, (*P<sub>b</sub>*), the results of which are shown in Figure 10.



**Fig. 6.** CMC curves for SDK-A for different Probes



**Fig. 7.** CMC curves for SDK-B for different Probes

It is clear from figures 6- 8 that the performance of the algorithms decreases with increasing ages of the probes. In other words, the algorithms fail to identify subjects properly as they age versus their younger gallery images, indicating a

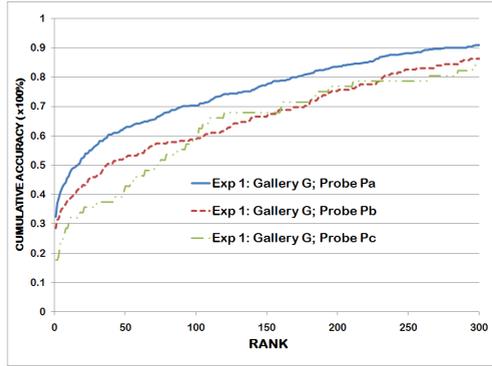


Fig. 8. CMC curves for PCA for different Probes

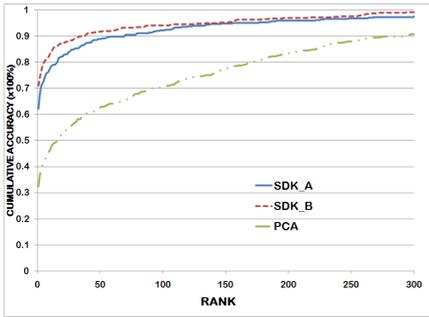


Fig. 9. Comparison of CMC curves of the three algorithms for Probe A

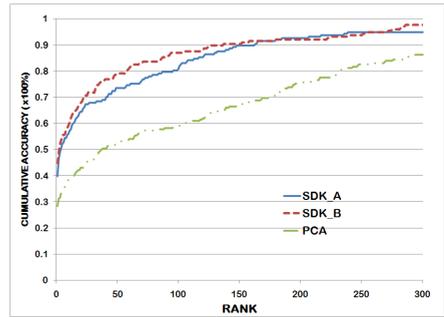


Fig. 10. Comparison of CMC curves of the three algorithms for Probe B

general lack of robustness to face aging even in these advanced systems. From Figure 9 and Figure 10, it can be concluded that performances of both SDK A and SDK B were expectedly better than the baseline PCA algorithm. In general, also, SDK B performed slightly better than SDK A.

**Experiment 2.** The goal of this experiment was to quantify possible performance improvements with the introduction of synthetic aged images using our most recent techniques discussed earlier. This experiment was conducted on SDK A. One image of each of these subjects with an age-range between 18-24 years was used to build the gallery. The probe was created using an older image of each of these subjects whose actual ages were in the range of 35-50 years. Next, the images in the Gallery  $G_0$  were age-progressed to match the approximate ages in the Probe  $P$ . These synthetic aged images were used to build another Gallery  $G_A$ . Rank-1 accuracy was computed for the probe matched with each of the galleries separately as shown in Table.2 This experiment, although still relatively small, demonstrates a notable gain of recognition performance by

**Table 2.** Rank-1 Accuracy for Probe  $P$  matched with the original younger gallery  $G_0$  and synthesized older gallery  $G_A$  for Experiment 2

Exp	Rank-1 Accuracy
Original Images Only	18.75%
Including Age-Progressed Images	31.25%

injecting synthetically age-progressed images into a modern commercial system. With additional data and improved methods, future work should be able to further quantify this improvement and suggest specific techniques for the best performance gains.

## 5 Conclusions and Suggested Direction

In this paper, we have presented improved results in synthetic age-progression and also evaluated, relative to longitudinal data, the performance of two modern face recognizers. We have shown that even the performance of these advanced, commercially available systems degenerate with increasing age spans of test probes. We have also demonstrated the possibility of improvement, even in these systems, with the addition of synthetically age-progressed gallery images. We plan to extend and evaluate this experiment on a larger data-set when possible. Our future research will also focus on developing quantitative measures to evaluate performance in a more detailed manner. To conclude, this work has emphasized the need to develop robust algorithms invariant to facial aging as well as the need to continue improvements in synthetic age progression. Also, research efforts should be focused on collecting and processing necessary datasets to further improve and evaluate both age-progression and face-recognition technologies that seek to be age invariant.

## References

1. Ricanek, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: 7th International Conference on Automatic Face and Gesture Recognition, April 2006, pp. 341–345 (2006)
2. Beveridge, et al: The CSU Face Identification Evaluation System User’s Guide: Version 5.0. Colorado State University (May 1, 2003)
3. Albert, A.M., Ricanek, K., Patterson, E.: A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. In: Forensic Science International (April 2007) (in Press)
4. Behrents, R.G.: Growth in the Aging Craniofacial Skeleton. University of Michigan, Ann Arbor (1985)
5. Lanitis, A., Taylor, C.J.: Towards automatic face identification robust to ageing variation. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 391–396 (2000)
6. Ramanathan, N., Chellapa, R.: Face verification across age progression. In: IEEE Conference Computer Vision and Pattern Recognition (2005)

7. Ricanek, K., Boone, E., Patterson, E.: Craniofacial aging impacts on the eigenface face biometric. In: Proceedings of the Sixth IASTED International Conference on Visualization, Imaging, and Image Processing, Palma de Mallorca, Spain, August 2006, pp. 249–253 (2006)
8. Patterson, E., Ricanek, K., Albert, A.M., Boone, E.: Automatic representation of adult aging in facial images. In: Proceedings of the Sixth IASTED International Conference on Visualization, Imaging, and Image Processing, Palma de Mallorca, Spain (August 2006)
9. Wang, J., Shang, Y., Su, G., Lin, X.: Age simulation for face recognition. In: 18th International Conference on Pattern Recognition (2006)
10. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic identification and coding of human faces using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7) (July 1997)
11. Edwards, G.J., Taylor, C.J., Cootes, T.F.: Interpreting face images using active appearance models. In: Proceedings of the Fifth European Conference on Computer Vision, pp. 581–595 (1998)
12. Lanitis, A., Taylor, C.J., Cootes, T.F.: Simulating aging on face images. In: Proceedings of the Second International Conference on Audio and Video-based Biometric Person Authentication (1999)
13. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4) (April 2002)
14. Cootes, T.F., Kittipanya-ngam, P.: Comparing Variations on the Active Appearance Model Algorithm. In: Proc. BMVC 2002, vol. 2, pp. 837–846 (2002)
15. Patterson, E., Sethuram, A., Albert, M., Ricanek, K., King, M.: Aspects of Age Variation in Facial Morphology Affecting Biometrics. In: Proceedings of the IEEE Conference on Biometrics: Theory, Applications, and Systems, Washington, DC (September 2007)
16. FG-NET Aging Database, <http://www.fgnet.rsunit.com>
17. Face Aging Group, <http://www.faceaginggroup.com>
18. CSU FFace Identification Evaluation System, <http://www.cs.colostate.edu/evalfacerec/>