MORPH: A Longitudinal Image Database of Normal Adult Age-Progression

Karl Ricanek Jr., IEEE Senior Member
Department of Computer Science
University of North Carolina Wilmington
Wilmington, North Carolina, USA
RICANEKK@UNCW.EDU

Tamirat Tesafaye
Department of Computer Science
Addis Ababa University
Addis Ababa, Ethiopia
TAMIRAT@PROGRAMMER.NET

Abstract

This paper details MORPH a longitudinal face database developed for researchers investigating all facets of adult age-progression, e.g. face modeling, photo-realistic animation, face recognition, etc. This database contributes to several active research areas, most notably face recognition, by providing: the largest set of publicly available longitudinal images; longitudinal spans from a few months to over twenty years; and, the inclusion of key physical parameters that affect aging appearance. The direct contribution of this data corpus for face recognition is highlighted in the evaluation of a standard face recognition algorithm, which illustrates the impact that age-progression, has on recognition rates. Assessment of the efficacy of this algorithm is evaluated against the variables of gender and racial origin. This work further concludes that the problem of age-progression on face recognition (FR) is not unique to the algorithm used in this work.

1. Introduction

It is well known that "data drives research/discovery." For biometrics this statement is evident in the numerous data corpuses available for research. In particular, face based biometrics research has yielded a multitude of face corpuses, so one might ask the following questions: *Why is another face database needed? What does the database offer above and beyond what is offered today? Why should anyone care about this new database?* These questions will be answered in the following sections.

There are dozens of face and gesture data corpuses in existence that focuses on either a single niche area or combination of niches, e.g. pose, gesture, illumination, gender, ethnicity, high resolution, low resolution, three-dimensional, multi-spectral, etc. But, there are only *three* known publicly available databases that include duplicate images of a subject at various ages. Of these three longitudinal databases only the database presented in this work, MORPH, records subject ethnicity, height, weight, and gender, which are critical components for understanding the changes in appearance of the human adult face as it ages. Of the remaining two data corpuses, FERET [6] and FG-NET Aging Database [13], only FG-NET contains longitudinal images in excess of five years. Although FG-NET was designed as a database for age-progression research it has a limited number of subjects, 82 for Part A, and again, it contains no data elements on key parameters that affect the appearance of the face across adulthood.

Section 2 will provide formative knowledge on the changes in the adult human face as it ages, which is quite different from the developmental growth associated with maturational aging of a child's face. In depth details associated with MORPH are found in section 3, while experiments using MORPH to determine effects on recognition performance of a standard face biometric system is examined in section 4. The last section contains discussion points and highlights future work on the database.

2. Morphology of Face Age-Progression

2.1. Types of Morphology

The head and face do continue to change throughout adulthood [1, 12]. Hyper-dynamic facial expressions, influences of gravity, and the changes within the bony (cranial) support structure are causes of adult intrinsic facial morphology. Intrinsic facial morphology is the main causes associated with the formation of physical changes to the skin of the face that appear in the form of wrinkles, lines, and skin sagging. In addition to intrinsic facial morphology, there are extrinsic causes which are behavior driven: photo-aging, which is aging associated with solar radiation damage to the epidermis (skin); heavy cigarette smoking; and alcohol and drug abuse.

Remodeling of the bone is responsible for structural changes in the craniofacial complex of the head which includes craniofacial expansion, cranial size decreases, cranial thickening, jaw and dental arch changes, facial height increases, and contour changes. These cranial structural changes, albeit small, 1.5 mm to 5 mm in deformation, has a significant impact on the overall appearance of the face.

Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)
0-7695-2503-2/06 $20.00 © 2006 **IEEE**

2.2. Variable Rate of Morphology

A notable point is the notion of a variable rate of facial morphology against the adult decade of life, e.g., 20-30 years, 30-40 years, etc. This means that there are periods of significant change in one's appearance, while other periods show no noticeable facial change. This can be a particularly interesting point for face based biometric evaluation, if the evaluating data corpus does not contain enough diversity over age, gender, or ethnicity; thereby, causing artificially high performance. This point is demonstrated in section 4 and in [5].

3. MORPH Face Database

3.1. Source Data

The current source of the images and associated physical properties is from public records. The legacy photographs associated with these records were taken between October 26, 1962 and April 7, 1998. Digital scans of these photographs were collected with legal considerations and IRB approval.

Photographs were digitized with a consumer grade flatbed scanner. Images were scanned in full 48-bit color with a capture resolution of 300 dpi. The digital images were cropped about the face to a size of 400 x 500 pixels. The images were converted to grayscale for database uniformity reasons. The images are stored in the database as a binary PGM (portable gray map) format. Some minimal image enhancements were performed on the images; these include artifact minimization with median filtering and contrast stretching via histogram equalization. Fig. 1 is a sample of three subjects from the database. The sample demonstrates the general appearance changes that are typical with age-progression: filling out of the face due to weight increase, formation of wrinkles and lines, and sagging due to loss of skin elasticity and muscle tone.

A

B

C

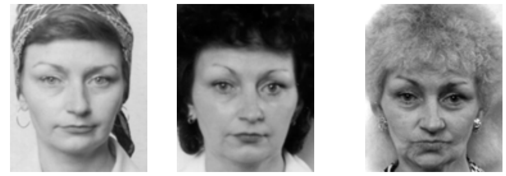

Fig. 1. Sample images from MORPH database. A-Caucasian male age/weight 19 yrs/147 lbs, 20/140 lbs, and 27/188 lbs. B-African-American male age/weight 20 yrs/163 lbs, 30 yrs/200 lbs, and 37 yrs/247 lbs. C-Caucasian female age/weight 35 yrs/140 lbs, 40 yrs/140 lbs, and 44 yrs/118 lbs.

3.2. Statistics

As of this writing, the database contains 1,724 face images of 515 individuals. These images represent a diverse population with respect to age, gender, and ethnicity. There are 1,278 images of individuals of African-American decent, 433 images of individuals of Caucasian decent and 3 images classified as other. There are 294 images of females and 1,430 images of males. For the male images, seventy-six percent have some form of facial hair, usually a mustache.

The average age of the individual at the time of acquisition is 27.3 years, with a standard deviation of 8.6 years, and a maximum age of 68 years. The average maximum age span, based on the age difference of the first enrolled image and subsequent images, of the images is 8.6 years. The minimum are span is 46 days with a maximum of 29 years. Group statistics on decade of life age spans are shown in Table 1. Table 2 provides additional statistics on the datasets with respect to the number of additional images, duplicates, available for each subject. All subjects have at minimum of one duplicate image—two images of subject in total—while sixty three percent have two or more additional images.

Table 1
Statistics by Decade of Age

	Age Span (years)			
	18–29	**30–39**	**40–49**	**50+**
All	*951*	*445*	*126*	*32*
Males	*776*	*366*	*105*	*32*
Females	*175*	*79*	*21*	*0*

Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)
0-7695-2503-2/06 $20.00 © 2006 IEEE

Table 2
Distribution of Duplicate Subject Images

	Number of Image Duplicates			
	1	**2**	**3**	**4 +**
All	*634*	*359*	*85*	*12*
Males	*526*	*297*	*71*	*10*
Females	*108*	*62*	*14*	*2*

3.3. Data Definition

Table 3 highlights the basic information contained for each image. This information is contained within the structure of a database table, and therefore, can be queried. This will allow researchers to build customized datasets for evaluation.

The inclusion of the "Image Quality" field is designed to give researchers the option of selecting images of a particular degree of "quality". The image quality field can be highly subjective in nature; therefore, the authors used an unbiased rating system to determine the image quality. To generate the unbiased rating, all images were ranked by several independent evaluators, the mean score for each image was used to classify the image as poor, fair, or good against a Likert scale of 0 - 10. Fig. 2 is an example ranking for each category of image quality and Table 4 is a break out of quality statistics over the image database.

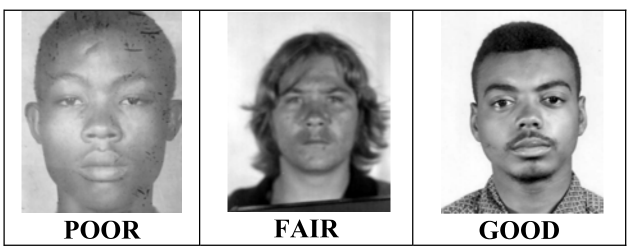

Fig. 2. Example of images ranked as poor, fair, and good.

Table 3
MORPH Primary Table Field Definition

Field	Field Definition
idnum	the identification number for the individual
picture num	the number for the picture for the individual
DOB	date of birth
DOA	date of acquisition (data photo taken)
Weight	weight of the individual
Height	height of the individual
Race	Caucasian, Hispanic, Asian, or African American
Gender	male or female
Facial Hair	yes or no
Glasses	yes or no
Image Quality	Poor, Fair, or Good

Table 4
Distribution of Image Quality

	Image Quality		
	Poor	**Fair**	**Good**
All	*8.4%*	*48.7%*	*42.9%*
Males	*6.7%*	*39.7%*	*35.2%*
Females	*1.8%*	*9.0%*	*7.7%*

4. Face Recognition Age-Progression Impacts

This section attempts to address the following question, "*what quantifiable impact will natural human age-progression have on FR systems?*" This evaluation was done on a specific algorithm, but the findings correlate well to findings of the FRVT 2002 experiment, which was comprised of several commercial and academic FR systems. It is noted that synthesis-based models, e.g. Morphable Models, were not integrated into the FR systems evaluated in [5]. In FRVT2002, a proprietary non-public database was constructed of an ethnically homogenous population of more than 121,000 images. Although, this database is not available to researchers, it was used to formulate conclusions concerning the impact of age-progression against FR systems. The database, although extensive, does not contain longitudinal spans in excess of five years—results indicated that the longest span was approximately 3.2 years. Therefore, the conclusions on impact to recognition rates attributed to age-progression (probe-gallery currency was used in original work) may not be purely linear as highlighted in the work.

4.1. Experiment Methodology

Experiments similar to those conducted in FRVT2002 were performed on the MORPH data corpus using the standard PCA FR algorithm. Experiments were conducted on the entire set of images (experiment T), on males only (experiment M), on females only (experiment F), on African Americans (experiment A), and on Caucasian Americans (experiment W) to determine the impact on recognition rate. The gallery size was 515 for experiment T; this gallery includes the chronological first image for each subject. For subsequent experiments, the galleries are a subset of the T gallery. The M (Male) experiment had a gallery size of 420 individuals, i.e. 420 of the 515 subjects of the MORPH database are males. The remainder was used for F (Female) experiment. The experiments named A (African-American) and W (White) were used to evaluate the impact of age-progression on ethnicity. The size of the gallery for experiment A is 377 where as it is 137 for experiment W. For each experiment the corresponding probe images are divided into groups based on age difference from gallery image of subject. The age spans are organized into three categories: 0-5, 6-10, 11-15 years.

4.2. Experiment Results

In the T experiment (see Table 3.1), for all age ranges, the best performance of the algorithm falls in the age span (0-5), which is of course, the smallest one. This signifies that the performance of the algorithm is optimal when the acquisition time delay is small between the gallery and probe images. Unfortunately, this condition on the gallery and probes can not be enforced in real-world applications of face based biometrics.

With regard to age range, for a specified age span in this experiment, the performance of the algorithm increased from age range (< 18) to age range (40-49), which indicates that younger people are more difficult to recognize than older ones. For the specified age span (0-5), as can be seen from Table 3.1, the performance increased from 0.344 (age range <18), 0.420 (age span 18-29), 0.452 (age range 30-39), and 0.80 (age range 40-49), respectively, which clearly shows that the PCA algorithm performed better in identifying older people with the same age span than younger ones. This is further supported by findings in [5], where this trend was observed over the several FR systems evaluated.

Table 4
Match Results of Experiment T

Experiment (Age range)	Classification of probes based on age span			
	0-5	6-10	11-15	16-20
T1 (<18)	0.344	0.155	0.078	0.125
T2 (18-29)	0.42	0.257	0.134	0.08
T3 (30-39)	0.452	0.3	0.231	*
T4 (40-49)	0.8	*	*	*

Further experiments, which are illustrated in Tables 5, 6, and 7, performed against gender and ethnicity show the same general trend that FR recognition rates decrease as age span increases for all age ranges. Due to the paucity of females in the database across the age spans, the results on female subjects were not included in table form. For the ranges were the data overlaps for both genders—less than 18 and 18-29 ranges against age spans 0-5 and 6-10—males exhibited a higher recognition rate, which is also supported by [5].

Recognition rates across ethnicity and against age-progression have not been reported in the literature. This experiment shows that, at least for the standard PCA algorithm, Caucasian Americans have higher recognition rates than African Americans. Additional research is warranted to validate these findings.

Table 5
Match Results of Experiment M

Experiment (Age range)	Classification of probes based on age span			
	0-5	6-10	11-15	16-20
M1 (<18)	0.354	0.163	0.085	0.083
M2(18-29)	0.431	0.294	0.199	0.073
M3(30-39)	0.513	0.292	0.167	*
M4(40-49)	0.800	*	*	*

Table 6
Match Results of Experiment A

Experiment (Age range)	Classification of probes based on age span			
	0-5	6-10	11-15	16-20
A1 (<18)	0.257	0.051	0.105	0.077
A2(18-29)	0.423	0.252	0.107	0.093
A3(30-39)	0.444	0.304	0.333	*

Table 7
Match Results of Experiment W

Experiment (Age range)	Classification of probes based on age span			
	0-5	6-10	11-15	16-20
W1 (<18)	0.346	0.211	0.077	*
W2(18-29)	0.413	0.25	0.227	*
W3(30-39)	0.571	0.286	*	*

5. Discussion

The MORPH database is a novel longitudinal face corpus that offers images of a large number of diverse subjects over a period of many years. The uniqueness of this database is further elucidated by including linked parameters of exact age, ethnicity, gender, height and weight which have a direct impact on facial appearance [8]. As a result of incorporating these parameters, researchers will now be able to construct better general soft biometric models [10], which can account for the effects of weight on appearance or the role of the ethnicity in age-progression.

As demonstrated in section 4 this database could be used to explore the impacts of age-progression on FR technology. The support of the conclusions from [5] further highlights the need for age-progression evaluation on all commercial FR technology.

The database is expected to triple in number of subjects by the printing of this article. A new source of data has been located that will provide all the elements unique to this database. MORPH is available to researchers by contacting the primary author.

6. References

[1] L.G. Farkas, "Photogrammetry of the face." In Anthropometry of the Head and Face, 2nd edition. LG Farkas (Ed.), Raven Press, Ltd. NY, NY, pp. 79-88, 1994.

[2] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, 2000.

[3] M. Y. Iscan, Global forensic anthropology in the 21st century. Forensic Science International, Vol. 117, Issue 1-2, pp. 1-6, 2001.

[4] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bayer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[5] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT2002: Evaluation Report", http://www.frvt.org/ DLs/FRVT_2002_Evaluation_Report.pdf, March 2003.

[6] P.J. Philips, H. Moon, P. Rauss, and S.A. Rizvi. The FERET evaluation methodology for face-recognition algorithms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997.

[7] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination and Expression (PIE) Database. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition. 2002.

[8] M.S. Zimbler, M.S. Kokosa and J.R. Thomas. Anatomy and pathophysiology of facial aging. Facial Plastic Surgery Clinics of North Anerica, Vol. 9, pp. 179-187, 2002.

[9] T.F. Cootes and C.J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," Proceedings of SPIE Medical Imaging, 2001.

[10] A.K. Jain, S.C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," Proceedings of the International Conference on Biometric Authentication, LNCS 3072, pp. 731-738, Hong Kong, July 2004.

[11] Lanitis, C.J. Taylor, and T.F. Cootes, "Towards automatic simulation of aging effects on face images," Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 4, pp. 442-455, April 2002.

[12] K. Ricanek, E.K. Patterson, and A.M. Albert, "Age-Related Morphological Changes: Effects on Facial Recognition Technologies," UNCW Technical Report, CALL #WRG FSC-$, 2004.

[13] FG-NET Aging Database. www-prima.inrialpes.fr/FGnet/. 2002.

0-7695-2503-2/06 $20.00 © 2006 IEEE