A Review of Face Recognition against Longitudinal Child Faces

Karl Ricanek Jr., Ph.D. Senior Member IEEE, Shivani Bhardwaj, & Michael Sodomsky

I3S Institute – Face Aging Group
University of North Carolina Wilmington
601 South College Road
28403 Wilmington
ricanekk@uncw.edu
sb2534@uncw.edu

Abstract: It is an established fact that the face-based biometric system performance is affected by the variation that is caused as a result of aging; however, the question has not been adequately investigated for non-adults, i.e. children from birth to adulthood. The majority of research and development in automated face recognition has been focused on adults. The objective of this paper is to establish an understanding of face recognition against non-adults. This work develops a publicly available longitudinal child face database of child celebrities from images in the wild (ITWCC). This work explores the challenges of biological changes due to maturation, i.e. the face grows longer and wider, the nose expands, the lips widen, etc, i.e. craniofacial morphology, and examines the impact on face recognition. The systems chosen are: Cognitec's FaceVacs 8.3, Open Source Biometric Recognition (SF4), principal component analysis (PCA), linear discriminant analysis (LDA), local region principal component analysis (LRPCA), and cohort linear discriminant analysis. Face matchers recorded low performance: top performance in verification is 37% TAR at 1% FAR and best rank-1 identification reached 25% recognition rate on a gallery of 301 subjects.

1. Introduction

The human face is an important feature of identity recognition. The characteristics of the face that makes it a desirable biometric modality is its uniqueness, universality, acceptability, semi-permanence, and easy collectability [RB11]. Because of its potential and possible variety of application, automated face recognition has received a lot of attention over the last two decades. Face recognition can be accomplished from a distance and via non-contact acquisition, which offer an added advantage over most biometric systems and make it more suitable for security and surveillance systems. Face recognition, may play a vital role in identifying children that go missing and in extensive range of access control and monitoring systems, especially to safeguard children. This technology can provide a whole new approach to protect and support latched-key kid and to provide access control for various internet of things across different age groups. It can be used to protect the non-adult population from predators and illegitimate web contents.

Face recognition is a challenging problem, and a great deal of work has been completed for pose correction, illumination variation, and expression to support face recognition in the wild. However, the majority of the work done has been focused on adults and deals with the dynamics of mature faces. The objective of this paper is to review the current state of facial recognition algorithms with a focus on non-adult stages of growth and development, 2 years to 16 years.

Aging with respect to facial recognition system includes variation in shape, size and texture of the face. These temporal changes will cause performance degradation. Hence, state issued id's, e.g. driving license, has to be renewed every 5-10 years. Mathew Turk stated that "developing a computational model of face recognition is quite difficult, because faces are complex, multidimensional, and meaningful visual stimuli" [TP91]; however, when aging information is added to this problem, it becomes infinitely more difficult. The most challenging problem in developing a solution for childhood face recognition is the formation of a useful dataset. This work addresses this primary concern.

Contributions

This paper provides the following contribution to the research community: 1. provides the baseline for face recognition performance for children against a suite of traditional face recognition techniques and investigate the impact of well-defined structural (skeletal) changes of the face on a suite of FR techniques; 2. establishes the first moderate scale publicly available child face database focused across the growth and development period1, which is one of the key issue in evaluation; and, 3. provides a methodology framework for investigating the problem of face recognition across childhood.

2. Background

Facial recognition is a complex topic that has been researched very heavily and many attempts have been made to understand the effects aging has on facial recognition systems. However, algorithm performance with respect to human aging: as a subject of the growth and development phase of childhood, has just begun to be fully explored by researchers. One of the biggest issues is the vast amount of data that is required to fully understand the human face and its maturation process. As the face changes over time, the ability to recognize the person becomes more challenging. This is further exacerbated if the person under inspection is not known to the observer.

Anthropological and forensic studies have contributed significantly to show that age related changes of non-adults are different from face aging for adults. Human aging can be studied as a two staged process: first involves the growth and development phase and

[1] The FG-Net face database has childhood images of 80 subjects, however many of these images are scanned from photographs. Additionally, CASIA has a twin's dataset that contains a number of child captures but only across a couple of days.

second deals with the effect of maturity as age progress [Rk09]. The performance for non-adult recognition over time spans less than a few weeks, may be on par with adult FR systems; However due to the rapid change that occurs during childhood, temporal displacement has a more profound impact on FR systems.

2.1 Adult aging

Adult aging is dominated by morphological and soft tissue changes i.e. skin texture, wrinkles etc., but some skeletal changes continue to occur [ARP07], [Tk10]. Early adulthood shows the first signs of soft tissue stressing. Hyper-dynamic expressions will start to show wrinkles on the face. Fine facial lines will appear horizontally on the forehead, vertical lines between the forehead and thin lines around the outside corners of the eyes will appear [Ks13]. From ages 40 to 50 there are noticeable changes to skin texture while minimal changes are found in younger years [THB00]. The aging rate of adults differs heavily on the individual, which is not the case for non-adults [ARP07]. These differences can be attributed to genetics and external features. Biological changes in adults alters the shape and texture of a face. As the skull continues to change with age, the eyes appear smaller as they sink in deeper into their orbits. As the skins elasticity begins to degrade wrinkles form, more notably in the eyelids, and the corners of the mouth [Lj74].These feature begin to sag and change in size thus changing the relationships of the features of the face.

2.2 Facial growth and development

Face aging with respect to children majorly involves craniofacial growth and development. This is the phase which is dominated by facial structural development which causes change in shape and size of face. In Karen T. Taylor's book "Forensic Art and Illustration" she describes the changes of the cranium and face year by year from childhood to young adulthood [Tk10].

These underlying skeletal changes will alter the appearance of the face. Cranial growth will not greatly change the features within the face, but it is the cause of change in proportions between them. During developmental changes the features of the face will remain alike to their original. This growth pattern is known as gnomatic growth [Tk10]. Craniofacial growth rate is affected by factors such as puberty and the growth of permanent teeth and there are jumps in growth rate at these periods, which makes growth a non-linear function. The rates of change of face is maximum in non-adult, particularly between birth and 5 years old [Fl92]. For this reason, high rate of change in the 0 to 5 years, face recognition technology may not be an appropriate technology for use, and hence, credentialing systems like national id's and passports should abstain from being used on persons in this age group. Maturation is achieved in males between the ages of 12 and 15 years while the same is true for females between 10 years and 13 years [Fl92]. After maturation the underlying structure of the face will continue to grow, however, not as rapidly.

2.3 Aging effect on performance

Growth & development and aging factors have a great impact on the performance of existing system over time [NG14]. The face develops and ages in numerous ways which pose challenges for face processing techniques. Humans have the ability to recognize a person from years ago; however, the person does look fundamentally different.

Early work in the impacts of face recognition by [LT00] established baselines for performance degradation in the problem of aging for adult faces. The work concluded that the performance does decrease as the time between probe and gallery increases, also it shows that older faces tended to be better recognized than younger faces i.e. individuals in the age range of 40–49 years were better recognized than those in the youngest age range, <18 years. Klare and Jain [RB05] concluded the same results. Later work by NIST concluded that recognition becomes easier with advanced age; however, recognition remains a challenge across large time spans for adults.

A very recent publication by NIST [NG14] evaluated the performance of non-adult face recognition on a suite of commercial FR systems for which the report concluded that significant weakness exists for current commercial systems. Further the report indicates that identification accuracy is strongly dependent on subject age. Where older subjects are easier to identify and easily distinguished from other, the opposite is true with children, being very hard to identify. In case if infants both false negative and false positive rates are much higher [NG14].

3. Dataset

Data is the primary necessity for exploration of face recognition systems whether through algorithm design or algorithm performance. FR technology performs better with the highly constrained images; however most of the time it is not the real scenario where we use this technology. Also to develop a database of this nature is extremely difficult because of human subject requirements and the nature of capturing or finding images across time, of same individual.

Table 1: Outline of Available Aging Datasets

Database	# Subjects	# Images	Images per subjects	Age Range	Image Quality	Label for Age
VADANA [SRK11]	43	2298	3-300	0-78	24-bit colored, 30 scanned	Yes
FGNET [Cf12]	82	1002	6 -18	0-69	Mostly scanned images	Yes
MORPH (Album1) [RT06]	631	1690	1-6	16-69	Digitally scanned at 300dpi, Grey scaled	Yes
MORPH	13673	55608	1-53	16 - 99	8-bit color	Yes

(Album 2) [RT06]					200x240 JPEG or 400x480 JPEG	
Cross-Age Celebrity Dataset [CCH15]	2,000	163,446	--	16 - 62	24-bit colored images	Yes

The recognized public databases that contain child faces is FG-NET; however they did not offer the sufficient number of subjects to evaluate the face recognition systems for children, also most of the images in the dataset are scanned from photographs, which tend to lose anthropometric measures of faces as well as introduces scanning artifacts that are difficult to decouple. That is the reason only few subjects, around 82 are usable from the FG-NET dataset [Cc10]. Adience dataset [EEH] contains non-adult subjects and label its subjects for different age groups; however it is a cross-sectional dataset and does not provide any longitudinal information of subjects.

To support the objective: In-the-Wild Child Celebrity, or ITWCC, dataset was created. It is the largest longitudinal dataset that has been developed to study the present system performance specifically for the non-adults. ITWCC focuses on having large sets of individuals, where the subject growth and development can be observed. As the dataset's name 'In-the-wild' suggests, the images are collected with unrestricted face and the data corpus is designed to emulate a real-life scenario as shown in figure 1. Images were captured by exploiting the fame of the subjects and gathered through open Internet sources, which are free to use. The data was captured until December of 2013. The criteria used to develop this dataset are as follows: 1) The subject must have at least three images to qualify. 2) The subject must have at least two images less than 16 years of age. 3) The date that the photo was taken must be available.

Figure 1: In-The-Wild Child Celebrity Dataset

In addition to the image, other meta-data is also captured. Age, race, gender, data of the photo, subject name, a unique photo identifier, and a conditional makeup and glasses marker, and the URL of the image is recorded for each entry. This information can further illuminate the difference in gender specific aging variations and occlusion's effects on facial recognition systems. In-The-Wild Child Celebrity (ITWCC) dataset is composed of 304 subjects and 1705 images. The subject's age within this dataset range from 5 months to 32 years. The dataset contains 876 female images and 839 male images. The average age of all images is 13.4 years with a standard deviation of 3.4 years. The average age of the first capture for the acquisition into this dataset is 10.2 years with a standard deviation of 3.9 years; furthermore, the average age of final capture is 16.3 years with a standard deviation of 4.467 years.

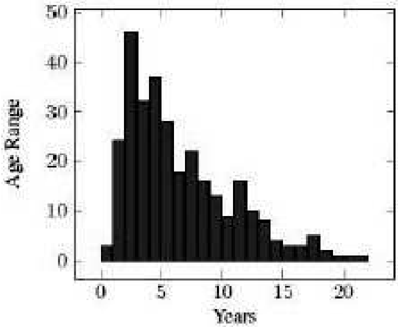

Figure 1: Age Range of Subjects in Year

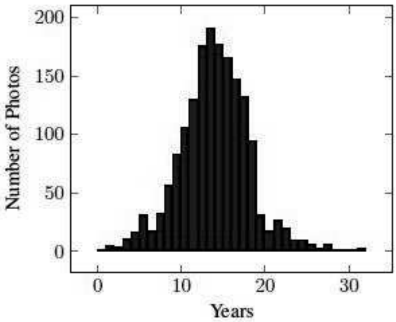

Figure 2: Age of Image

Figure 2 and figure 3, shows the number of subjects with a particular longitudinal age range, i.e. the maximum age less the minimum age and expresses the number of images for each age.

4. Methods

The challenges of using facial recognition techniques on children and adolescent faces were evaluated by running multiple baseline algorithms, an open source matcher and an extensively evaluated commercial system against the ITWCC dataset. Foundational open source algorithms were used because they have been well researched against many different types of dataset, including "in-the-wild" adult dataset, and a commercial system, Cognitec, was chosen due to its strength in nearly all scenarios. Another biometric evaluation toolkit and API is Open Source Biometric Recognition(OpenBR) collaboratory [Kj13]. OpenBR is a collaborative tool that provides a method for researchers to compare algorithms in a controlled environment. Standard face matching systems used were principal component analysis (PCA), linear discriminant analysis (LDA), cohortLDA and local region PCA algorithm (LRPCA). All the foundational algorithms are implemented in open-source environments. Two toolkits implemented by Colorado State University are used in this work: the 2011 Baseline Algorithms, and the CSU Face Identification Evaluation System [Br03], [Ly12], [Pp11].

This work also examines the fully automated face recognition system, Cognitec's, FaceVacs [Gc14]. Each biometric system in this work preprocesses the images to handle illumination, scale, and orientation issues. To achieve this, all the systems use the eye coordinates from image to extract the face and register it into a standardized format and then normalized. Cognitec, OpenBR and CSU's 2011 baselines uses an automated eye detection algorithm to acquire the eye locations. CSU's 2006 biometric toolkit need eye coordinates to be provided and marked by hand for this work. If one of the eye detection algorithms cannot find the eyes in an image the image is considered failure to enroll and is not included into the matching. PCA and LDA do not have any failure to enroll errors.

5. Evaluation technique

Two scenarios are considered to understand the difficulty of temporally displaced data. The scenarios were developed to replicate real world situations, where this type of data would be often used. All six techniques: Cognitec FaceVacs, OpenBR, PCA, LDA, LRPCA and cohort- LDA are used independently, to evaluate the performance of FR systems, under both the scenarios. Each technique detect, preprocess, match and finally evaluate the images. All of the matching information was provided to OpenBR's Face Evaluation toolkit [Kj13]. This toolkit evaluates the matching information and then plots the information in a standardized format.

Fundamentally each system will match at least two biometric templates, one being the stored template and the other being the new users, to produce a score which will decide acceptance or rejection. This match score is a standardized number that shows the likeness between two templates. Both genuine users and impostors are used to evaluate a system. Ideally all genuine users should be accepted while all impostors should be rejected. Important metrics to note are as follows: true accept rate (TAR), true reject rate (TRR), false accept rate (FAR), and finally false reject rate (FRR). True accept rate is the ratio of genuine users whom have been accepted, while the true reject rate is the ratio of impostors who have been correctly rejected. The false accept rate is the ratio of genuine users who are rejected and finally the false reject rate is the ratio of impostors who are mistaken as genuine matchers, i.e. the system grants.. A user is rejected or accepted by comparing the match score to a match threshold. The match threshold is an arbitrary number that each system is tuned to achieve the results it requires. Each system evaluated here used the default threshold values for identification matching.

6. Experiments

6.1 All to all verification

The first experimental scenario designed for this work mimics an access control. The purpose of the All-to-All Verification experiment is to determine how effective face verification performs when matching between temporally displaced non-ideal images. This experiment compares all images within the ITWCC dataset against all other images.

Images of the same individual are matched against the same individual and all others. The Access Control Scenario was conducted to understand how effective, or ineffective, the selected algorithms perform for matching. The entire ITWCC dataset was used in this scenario to generate 2,905,320 matches, with 10,652 genuine matches. Table 2 list the match matrix of all-to-all comparison.

6.2 Young to old identification

Experiment two is an identification task to explore how aging will effect identification performance. This experiment attempts to setup a scenario in which an end-user of a photo tagging tool, such as Facebook, Picasa, etc., would begin adding images over a span of time. The ITWCC dataset is used in this experiment similarly to the first experiment; however, only the first image of each person is used for the gallery and all other images for the individual are used as probes: the youngest image is matched to all of its elder images. The average age of the enrolled faces for all 304 subjects was 10.21 years with a standard deviation of 3.98 years. The minimum age of the gallery was 5 months old. The remaining images were then placed in the probe set; the average age of the probe set was 14.43 years of age , which represents the next chronological age image for every subject. Table 2 list the match matrix of old-to-young comparison.

6.3 Augmented young-to-old identification

The augmented Young-to-Old identification experiment further extends the last experiment by increasing the gallery size. The gallery is augmented with both the CASIA Twins (1,234 images) and the Labeled Faces in the Wild datasets (13,233 images) [Sz10], [Sz10]. By expanding the gallery with these datasets, the scenario will be closer to a real world situation in which a user would upload additional data to match against. This scenario is expected to be much more challenging for identification across non-ideal images. All of the CASIA Twins and LFW are added to the gallery resulting in a gallery size of 14,764 images (14,467 from CASIA twins and LFE, plus 304 first image ITWCC). LFW does not contain images of celebrities younger than 16 years; however, the data is captured in a similar manner to ITWCC, non constrained, public sourced images. The CASIA Twins dataset does contain child and adolescent data, but it is captured in a slightly less varied means. This dataset is not readily available to the general research community.

Table 2: Experiment Match Matrix
- Verification, *- Identity

Experiment	Genuine Matches	Imposter Matches	Ignore Matches	Total Matches
All to All#	10,652	2,894,668	1,705	2,905,320
Young to Old*	1404	421,200	0	422,904
Aug Young to Old*	1404	20,732,868	0	20,734,272

7. Results

To evaluate the performance of the face matchers, True Acceptance Rates (TAR) is compared with the False Acceptance Rates (FAR). Table 3, 4 and 5, shows the performance of FR techiques in All-to-All Verification and the Young-to-Old Identification experiments. It gives an estimation of the accuracy of each algorithm, in particular, how often an impostor gains access to the system vs. how often a true user is accepted.

Table 3: All to All Verification: True Accept Rate at 1% False Accept Rate

	Face Recognition Method					
	Cognitec	SF4	Cohort LDA	LRPCA	LDA	PCA
TAR @ 1%FAR	37%	25%	12.1%	13.5%	12.6%	15%

Table 4: Rank Identification Performance Results Closed-Set with matching on 304/304 Subjects

Algorithm	Young to Old (Exp. #2)		
	Rank-1	Rank-10	Rank-100
Cognitec	**25.0%**	**41.1%**	**73.7%**
S4F	13.7%	32.1%	67.2%
Cohort LDA	6.6%	19.6%	55.8%
LRPCA	6.3%	18.4%	47.9%
LDA	9.9%	23.1%	56.4%
PCA	8.4%	23.8%	64.5%

Table 5: Rank Identification Performance Results Closed-Set with matching on 304/14,764 Subjects

Algorithm	Aug. Young to Old (Exp. #3)		
	Rank-1	Rank-10	Rank-100
Cognitec	0.0%	0.0%	0.6%
S4F	**7.9%**	**19.2%**	**32.8%**
Cohort LDA	3.6%	7.6%	15.9%
LRPCA	4.0%	7.5%	14.8%
LDA	6.9%	13.6%	26%
PCA	5.9%	12.0%	25.6%

8. Conclusion

This work presents the first study of the impacts of craniofacial morphology for infants through to adolocescents on face recognition. This work does not provide a solution; however, it does address the biological phenomonenon responsible for making this area of face recognition extremely difficult. In comparison to adult aging, child aging is far more complex due to the changes in the boney structure as well as in the shape and size of the facial components. This work clearly illustrates the difficult of this problem through the performance metrics against a set of algorithms that have performed reasonably to extraordinarily on other wild datasets.

Three experimental scenarios: Access control, Photo-tagging, augmented photo-tagging were designed to explore the difficulty of non-adult aging. Six algorithms were used to test the hypothesis: Cognitec's FaceVacs, OpenBR's S4F, CohortLDA, LRPCA, LDA, and PCA [Gc14], [Kj13], [Br03], [Ly12], [TP91]. Results on this unique, albeit small dataset, shows that aging on non-adults is a challenging task for facial recognition algorithms.

The most accurate algorithm for verification was Cognitec at a TAR 37.0% at 1.0% FAR and 23.8% at 0.1% FAR. By current standards this level of performance is considered dismal as demonstrated by Klare et al., a true accept rate of 96.3% was achieved on adults with 0-1 years of lapse between images[Cc10]. The identification experiments were far more diasterous with rank-1 identification task rangingfrom 25% to 6.6% on a closed set of 304 subjects (1704 images). The best performer Cognetic became the worst performer when the gallery was increased to 14,767 subjects by augmenting the gallery with images from LFW and CASIA Twins. Cognetic registered a rank-1, rank-10, and rank-100 perfromance of less than 1% true match rate. S4F was the best performing algortihm on this test.

We conclude that non-adult facial recognition is a challenge for conurrent face matchers; with the provided dataset, researchers can start exploring the problem space. The authors of this work will continue to augment the dataset and robust the face recognition system for children's.

9 References

[ARP07] A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. In (Forensic Science International), 2007; p. 1–9

[Bg00] The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[BHK97] Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In (Pattern Analysis and Machine Intelligence): IEEE Transactions , 1997; p. 711–720.

[Br03] The CSU face identification evaluation system user's guide: version 5.0. Computer Science Department. Colorado State University, 2003.

[Cc10] Face age estimation using model selection. In (Computer Vision and Pattern Recognition Workshops (CVPRW)): IEEE San Francisco CA, 2010.

[CCH15] Face Recognition using Cross-Age Reference Coding with Cross-Age Celebrity Dataset. IEEE Transactions on Multimedia. 2015.

[Cf12] The FG-NET aging database. 2007-11-12. http://sting. cycollege. ac. cy/-alanitis/fgnetaging/index. Htm. , 2012.

[EEH] Age and Gender Estimation of Unfiltered Faces. IEEE Transactions on

Information on Forensics and Security.

[Fl92] Growth patterns of the nasolabial region: a morphometric study. The Cleft Palate-Craniofacial Journal, 1992; pp. 18-324.

[Gc14] Facevacs software developer kit. 2014.

[Hg07] Labeled faces in the wild: A database for studying face recognition in unconstrained environments.

[KJ01] Face recognition across time lapse: On learning feature subspaces. In (IEEE Biometrics (IJCB)): International Joint Conference on Biometrics Compendium Washington DC, 2011.

[Kj13] Open source biometric recognition. In (Biometrics: Theory): Applications and Systems (BTAS) 2013 IEEE Sixth International Conference, 2013.

[Ks13] Sketch based face recognition: Forensic vs. composite sketches.

[Lj74] The human face. , Stein and Day, 1974.

[LT00] Robust face recognition using automatic age normalization. In (IEEE Electrotechnical Conference): MELECON 2000. 10th Mediterranean, 2000.

[Ly12] Preliminary studies on the good, the bad and the ugly face recognition challenge problem. In (Computer Vision and Pattern Recognition Workshops (CVPRW)): 2012 IEEE Computer Society Conference, 2012.

[NG14] FRVT: Performance of Face Identification Algorithms. Information Access Division: National Institute of Standards and Technology, 2014.

[Pp11] An introduction to the good, the bad and the ugly face recognition challenge problem. In (Automatic Face & Gesture Recognition and Workshops): FG 2011. 2011 IEEE International Conference

[RB05] The Effect of Normal Adult Aging on Standard PCA Face Recognition Accuracy Rates. In (International Joint Conference on Neural Networks): Montreal Canada, July 2005.

[RB11] What Are Soft Biometrics and How Can They Be Used? In (IEEE Computer); , 2011; pp. 106-108

[Rk09] Craniofacial aging. In (Wiley Handbook of Science and Technology for Homeland Security), 2009.

[RT06] Morph: A longitudinal image database of normal adult age-progression.

[SRK11] VADANA: A dense dataset for facial image analysis. Computer Vision Workshops (ICCV Workshops), IEEE International Conference, 2011; pp. 2175 - 2182.

[Sz10] A study of multibiometric traits of identical twins. SPIE Defense Security and sensing, 2010.

[THB00] Comments on Facial Aging in Law Enforcement Investigation. In (Forensic Science Communications), 2000.

[Tk10] Forensic art and illustration. CRC Press, 2010.

[TP91] Face recognition using eigenfaces. In (Computer Vision and Pattern Recognition), 1991