A Biometric Database with Rotating Head Videos and Hand-drawn Face Sketches

Hanan A. Al Nizami, Jeremy P. Adkins-Hill, Yong Zhang, John R. Sullins,
Christine McCullough, Shaun Canavan, and Lijun Yin

Abstract

The past decade has witnessed a significant progress in biometric technologies, to a large degree, due to the availability of a wide variety of public databases that enable benchmark performance evaluations. In this paper, we describe a new database that includes: (i) Rotating head videos of 259 subjects; (ii) 250 hand-drawn face sketches of 50 subjects. Rotating head videos were acquired under both normal indoor lighting and shadow conditions. Each video captured four expressions: neutral, smile, surprise, and anger. For each subject, video frames of ten pose angles were manually labeled using reference images and empirical rules, to facilitate the investigation of multi-frame fusion. The database can also be used to study 3D face recognition by reconstructing a 3D face model from videos. In addition, this is the only currently available database that has a large number of face sketches drawn by multiple artists. The face sketches are valuable resource for many researches, such as forensic analysis of eyewitness recollection, impact assessment of face degradation on recognition rate, as well as comparative evaluation of sketch recognitions by humans and algorithms.

I. Introduction

The importance of having public datasets to allow for performance evaluations of existing biometric systems and to stimulate the development of new algorithms has long been recognized in biometrics community, as exemplified by the FERET tests [1, 2]. Since then, a considerable amount of efforts has been devoted to building more databases. Some of the noticeable projects are: Face Recognition Grand Challenge [3], Iris Challenge Evaluation [4], Notre Dame Biometric Database [5], Fingerprint Verification Competition Database [6], USF Gait Dataset [7], CUHK Sketch Database [8], CMU Multi-PIE Database [9], NIST Mugshot Identification Database [10], and many others (see [11] for descriptions). Some of the new databases can be found in [12]. One of the trends is that the size of a database is becoming increasingly larger. For example, a database with an enrollment of over 100 subjects is common. Another trend is that databases have become more diversified, often with a focus on a special type of imaging sensor (optical devices, infrared cameras, or 3D laser scanners), a particular biometric feature (fingerprint, face, ear, iris, or gait), or an application domain (data security, access control, health care registration, or forensic investigation).

The database presented here includes two unique types of data: (i) Rotating head videos with strong shadows; (ii) Hand-drawn face sketches. In addition, video frames of ten face pose angles were determined manually using the Reference Images and empirical rules. Although the initial motivation of building such a database is to study multi-frame fusion for video-based face recognition and quantitative assessment of sketch effectiveness in criminal investigations, the database is useful for a wide range of research topics.

In video-based face recognition, experiments have shown that multi-frame fusion is an effective method to improve the recognition rate [13, 14, 15]. The performance gain is probably related to the use of 3D face geometry embedded in video sequences. However, it is not clear how the inter-frame variation has contributed to the observed performance increase. Will the multi-frame fusion work for videos of strong shadows? Can we find an upper-bound of multi-frame fusion by exploring the connection between the shadow areas on a face and its 3D pose? How many frames are necessary for maximizing the recognition increase without incurring a heavy computational cost? To address these issues, a set of rotating head videos is needed.

Face sketches are used by law enforcement agencies to identify and capture fugitives and suspects. Face sketches can be drawn either by a trained police artist or using a composite software kit [16, 17, 18]. Both types of sketches have been studied in the context of searching or matching a sketch to a subject's face in a database of photos or mug-shots [8, 19, 20, 21, 22]. However, we are not aware of any public database that consists of a large number of sketches that were drawn by multiple artists. Such a sketch database is essential for advancing the state-of-the-art in the fields of forensic arts, human visual perception, cognitive psychology, as well as 3D sketch reconstruction and recognition.

II. Database Overview

To meet the needs of various research projects, videos and images in the database are presented in three formats: (i) Video clips; (ii) Video frames of ten pose angles; (iii) Scanned sketch images. The database structure for a subject in a single collection session is illustrated in Fig. 1.

This work was supported in part by Youngstown State University Research Council Grant No. 08-8 and New York State Foundation for Science, Technology, and Innovation (NYSTAR).

H. A. Al Nizami, J. P. Adkins-Hill, Y. Zhang, and J. R. Sullins are with Department of Computer Science and Information Systems, Youngstown State University, Youngstown, OH 44555.

C. McCullough is with Department of Arts, Youngstown State University, Youngstown, OH 44555.

S. Canavan and L. Yin are with Department of Computer Science, State University of New York at Binghamton, Binghamton, NY 13902.

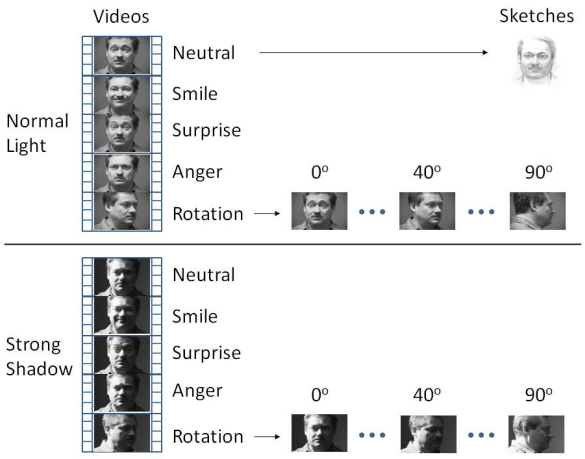


Fig. 1. Data formats for a subject enrolled in a collection: two video clips, twenty labeled frames (ten for each clip), and a sketch of a frontal view.

Each subject has at least two video clips, one being captured under a normal indoor lighting condition and the other being captured with strong shadows (if a subject enrolled in two collection sessions, he or she would have four video clips). Each video clip contains five segments that correspond to four facial expressions and a 90-degree head rotation sequence. All video clips were broken into individual frames. For a rotation sequence, frames of ten pose angles were determined manually (0°, 10°, 20°, …, 90°). A face sketch was drawn using a frame that has a frontal face with a neutral expression.

III. Data Acquisition

A. Rotating Head Videos

Videos were obtained in an indoor environment with two illumination conditions: normal indoor lights and strong shadows (see Fig. 2). Shadows were created using a headlight that was projected toward one side of a subject's face. A Canon XL1s digital camcorder was used, with a default capture rate of 30 frames per second. The resulting video frame resolution is 720 x 480 pixels. The acquisition protocol is as follows: (1) Prior to the first collection, all participants signed a consent form; (2) During an acquisition session, a subject was seated on a rotating chair facing the camcorder, with a blue curtain background. The distance between the camcorder and the subject was about 4 feet; (3) Under a normal lighting condition, the subject was instructed to open his or her mouth slowly twice; (4) The subject then showed four expressions: neutral, smile, surprise, and anger; (5) The subject slowly rotated his or her body by 90 degrees, from the fontal view to the profile view; (6) After the lighting condition being switched to strong shadows, steps 3-5 were repeated. It took about 5 minutes for a subject to complete a session.

Fig. 2. Acquisition set-up for rotating head videos with two illumination conditions: (a) Normal indoor lights; (b) Strong shadows.

Fig. 3 shows a few sample frames illustrating the shadow effect. The appearance of a face was altered in terms of the highly polarized intensity values, with almost half of the face being severely "darkened". This type of faces poses a great challenge to recognition methods. As demonstrated in Section V-*A*, a multi-frame fusion can effectively reduce the adverse impact of shadows on recognition rate by exploring the coherent intensity variation in a rotating head video sequence.

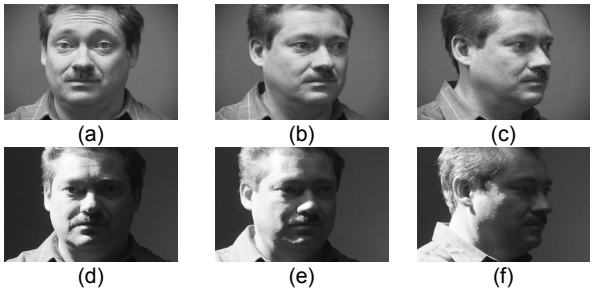


Fig. 3. Sample frames of rotating head videos that were acquired under a normal lighting condition (a, b, c) and a strong shadow condition (d, e, f).

B. Frame Selection

To quantify how the inter-frame variation of a head rotation video affects the performance of a multi-frame fusion, it is necessary to know face pose angles in frames. Determining a face's pose angle can be done automatically or manually. An automatic method utilizes the geometric relationship between a head rotation degree and the projected nose position in a frame to calculate the face pose angle [13]. But an automatic method requires that a subject kept his/her head and upper-body strictly along a vertical axis during a rotation, a condition that was not always satisfied in video collections. A manual method relies on human volunteers to select a frame in which a face has a desired pose angle. Although a manual method is tedious and time consuming, it has the advantage of being able to handle non-ideal videos. To minimize the errors and uncertainties caused by human bias, we used the Reference Images and a set of empirical rules to guide the selection process, as being shown in Fig. 4 and Table I.

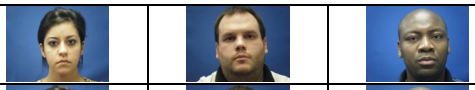
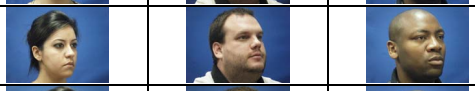
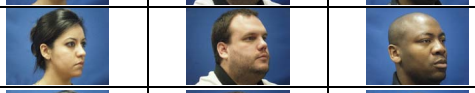
Angle	Reference Images
0°	
10°	
20°	
30°	
40°	
50°	
60°	
70°	
80°	
90°	

Fig. 4. The Reference Images of three subjects for ten pose angles

TABLE I
EMPIRICAL RULES FOR DETERMINING FACE POSE ANGLES

Angle	Ears	Eyes and eyebrows	Nose, cheek and lips
0°	Two ears are visible.	Two eyes are of equal length.	Nasal tip is in the center of face.
10°	Right ear disappears. Left ear is visible.	Left eye is slightly longer than right eye.	Nasal tip is slightly away from the center.
20°	More details of left ear are visible.	Left eye is longer the right eye.	Nasal tip is aligned vertically with the corner of right eye.
30°	Parts of cymba and fossa areas are visible.	Right eye is on the edge of the face line.	Nasal tip is below the center of right eye.
40°	More of cymba and fossa areas are visible.	Half of the right eye disappears.	The nasal tip is close or on the cheek line.
50°	Most of Cymba and fossa areas are visible	More than half of the right eye disappears, but with its eyelash and eyelid visible.	Nasal tip is across the cheek line that starts to disappear.
60°	Auditory canal is not visible	Right eye almost disappears, with its eyebrows visible.	Cheek line disappears. Philtrum is visible.
70°	Auditory canal is barely visible	Right eye fully disappears, with its eyebrows barely visible.	Philtrum is barely visible.
80°	Auditory canal is partially covered by tragus.	The upper eyelid of left eye is close to the top of the nose.	Philtrum is invisible.
90°	Auditory canal is fully visible.	The upper eyelid of left eye is on the top of the nose.	Philtrum is invisible.

The Reference Images were taken when a subject sit in a chair that had a *preset* rotation degree (a head rotation degree can be regarded the same as a face pose angle, if a subject did not tilt or lean his/her head during a rotation). We acquired the Reference Images of three subjects with ten pose angles, from the frontal view (0 degree) to the profile view (90 degree), with a 10 degree increment. Using these images, we summarized a set of empirical rules that characterize ten pose angles in terms of the visibility and relative positions of several facial features, including ears, eyes, nose, and cheek line (Table I).

For example, to select a frame that has a 20° face pose angle, a volunteer would compare a few candidate frames to the corresponding Reference Images of 20° in Fig. 4, and then pick the best one. In case that some frames have non-ideal properties (faces were covered by hairs, subjects leaned forward or titled backward in a rotation, faces had shadows or features of unusual sizes and shapes, etc.), a volunteer can resort to the empirical rules to make a decision.

Two student volunteers performed the frame selection, with each volunteer processing all video clips independently. Their results are quite consistent in terms of the frames being chosen for a particular pose angle, which is also confirmed by two statistical similarity measures (see Section IV-*B*).

C. Drawing Face Sketches

50 subjects were randomly selected from the database. For each subject, a frame that shows his or her frontal face of a neutral expression under a normal lighting condition was used. The color frame was then converted into a grayscale photo to be used for sketch drawing.

Five trained artists took part in sketch drawing sessions. All of them have at least 3 years of college-level education and multiple years of experiences in figure drawing, painting, sketching and sculpting. Prior to the first drawing session, all artists attended a forensic art workshop given by a police sketch artist from the Ohio Bureau of Criminal Identification and Investigation.

The sketches were drawn on Bristol papers. They were then scanned to digital images with a resolution of 100 dpi. For each subject, both the sketch of original size and the sketch of normalized dimension using eye coordinates are provided. A few normalized sketches are shown in Fig. 5.

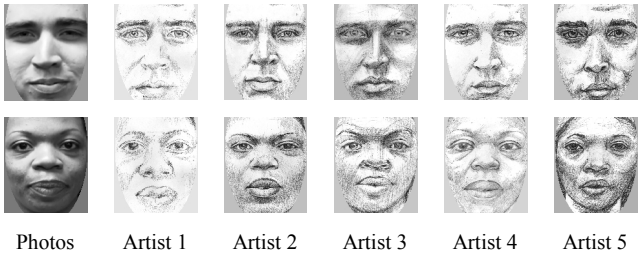


Fig. 5. Face photos and sketch images (normalized using eye coordinates).

IV. DATABASE STATISTICS

A. Subjects in Rotating Head Videos

Two video collection sessions were arranged, with a break of 40-60 days between the first and second sessions. Each session lasted for 6 to 8 weeks, depending on the number of subjects involved. A total of 259 subjects enrolled in the first collection session, of which 169 returned for the second session. Subjects came from Youngstown State University, primarily undergraduate students and a small number of faculty and staff members. The demographic information of participating subjects in terms of gender, ethnicity, and age are summarized in Table II, Table III, and Fig. 6.

In both collection sessions, male subjects accounted for about two-thirds of the total number of subjects, partially because that the majority of subjects were students from engineering and science departments. The distribution of ethnic groups is similar to that of the overall YSU enrollment. The age demographics are dominated by the younger groups, which is typical for a college population.

B. Selected Frames

To quantify the consistency between the frames selected by two volunteers, we computed two similarity metrics. The first metric uses an intensity-based correlation coefficient:

$$\rho(A,B) = \frac{Cov(A,B)}{\sqrt{Var(A)\,Var(B)}},$$

where ρ denotes the correlation coefficient, A is the frame of a particular pose angle selected by the first volunteer, and B is the frame selected by the second volunteer. The second metric computes the mutual information of two frames:

$$I(A,B) = \sum_{a,b} p(a,b) \log \frac{p(a,b)}{p(a)\,p(b)},$$

where $I(A,B)$ is the mutual information of frame A and frame B, $p(a)$ and $p(b)$ are marginal probability distributions, $p(a, b)$ is the joint probability distribution, a and b are intensity values in A and B, respectively. Table IV gives the percentages of frames that fall into four groups of metric values. It is clear that that the majority of frames selected by the two volunteers are very similar, as indicated by the high percentages of frames that have larger similarity values, though frames of shadows show slightly more disparities between the two volunteers.

C. Face Sketches

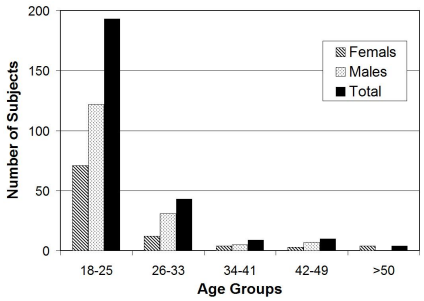
The demographics of subjects in sketches are summarized in Table V. Since those subjects were randomly selected (all from the first collection session), their demographic distributions show resemblance to that of subjects in videos.

TABLE II
GENDERS OF SUBJECTS IN ROTATING HEAD VIDEOS

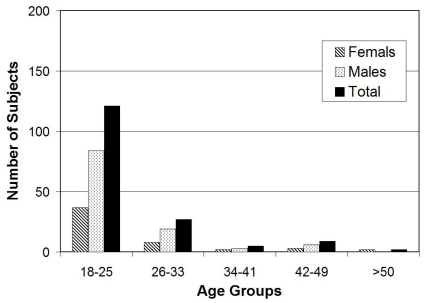
	Male	Female	Total
First Session	165	94	259
Second Session	114	55	169

TABLE III
ETHNICITY OF SUBJECTS IN ROTATING HEAD VIDEOS

	Caucasian	African American	Asian	Others
First Session	203	35	3	18
Second Session	132	19	2	16



(a) Subjects in the first session.



(b) Subjects in both the first and second sessions.

Fig. 6. Age distributions of subjects in videos: (a) Subjects in the first session; (b) Subjects in both the first and second sessions.

TABLE IV
PERCENTAGES OF FRAMES IN FOUR GROUPS OF SIMILARITY VALUES

p(a, b)	**0.5 – 0.7**	**0.7 – 0.8**	**0.8 – 0.9**	**0.9 – 1.0**
Normal light	4%	6%	26%	64%
Strong shadow	6%	10%	22%	62%
I (A, B)	**1.0 – 1.5**	**1.5 – 2.0**	**2.0 – 2.5**	**2.5 – 8.0**
Normal light	0%	18%	30%	52%
Strong shadow	8%	11%	38%	43%

TABLE V
DEMOGRAPHICS OF SUBJECTS IN FACE SKETCHES

Gender	Male	Female	Total	
	37	13	50	
Ethnicity	Caucasian	African American	Asian	Others
	39	8	1	2
Age	18-25	26-33	34-41	>42
	39	6	3	2

V. EXPERIMENTS USING THE DATABASE

A. Multi-Frame Fusion

Several research projects using the database have been reported and more are currently undertaken. For example, in a recent study of using multi-frame fusion to improve face recognition in videos [13], it was found that, given the videos taken under a normal lighting condition, the rank one recognition rate increased from 91% with a conventional single frame method to 100% with a 7-frame fusion method (Fig. 7). More importantly, using the videos taken under a strong shadow condition, the rank one recognition rate increased from 63% with a single frame to 85% with a 7-frame fusion (Fig. 8). Although the fusion tests were done on image level to minimize the information loss, it can be expected that a score level fusion would achieve the same degree of improvement.

One possible explanation of the observed performance gain is that the 3D face geometry was encoded in the continuous intensity variation of a rotating head video sequence, and was picked up implicitly by the multi-frame fusion. More thorough investigations will be carried out to establish a quantitative connection between the inter-frame variation and the fusion efficiency. One related issue to be addressed is whether and when the performance of a multi-frame fusion method will reach a "plateau".

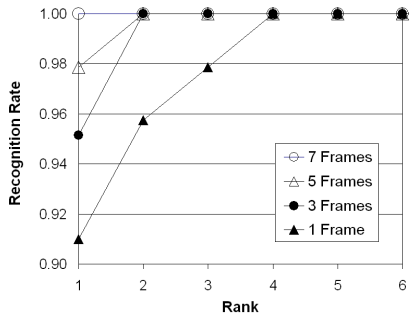


Fig. 7. Performance improvement of multi-frame fusion as measured by the CMC curves. Both gallery and probe frames were from videos taken under a normal indoor lighting condition.

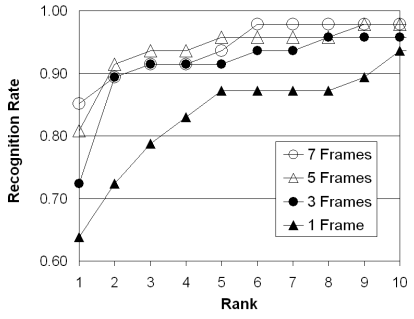


Fig. 8. Performance improvement of multi-frame fusion as measured by the CMC curves. Gallery frame were from videos taken under a regular indoor lighting condition, while probe frames were from videos taken under a strong shadow condition.

B. Face Sketch Recognition

Face sketching has been used in criminal investigations for a long time, but it lacks the kind of scientific rigor that has been developed in other forensic techniques such as DNA testing or fingerprint matching. Although efforts have been made to design more accurate computerized face composite systems [23], relatively little is known about how to improve the validity and effectiveness of sketches drawn by forensic artists. This is partially due to the lack of a database of an adequate number of sketches.

To address those issues, we performed a comparative evaluation using 250 hand-drawn face sketches [24] (Fig. 9). The primary findings of the study are: (i) There exists a large inter-artist variation with respect to sketch recognition rate; (ii) Multi-sketch fusion can greatly improve the recognition performance; (iii) Humans show a better performance in recognizing face sketch of distortions (non-perfect sketches) than a PCA-based algorithm.

We plan to extend the face sketch project in two directions: (i) More forensically realistic sketches will be added to the database, i.e., the sketches that are drawn based on verbal descriptions of eyewitnesses. This type of sketches is more challenging for both artists and recognition algorithms; (ii) In addition to untrained volunteers, trained professionals will participate in sketch evaluations, such as psychologists who specialize in visual perception and cognition, experts in forensic investigations, as well experienced police officers.

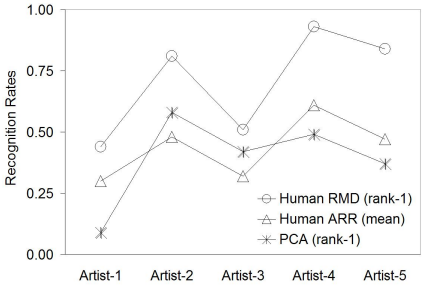


Fig. 9. Comparison of humans and PCA in sketch recognition. RMD is Relative Matching Distance. ARR is Absolute Recognition Ratio.

C. 3D Sketch Modeling

3D face recognition has attracted much attention recently (see [25] for a comprehensive review and in-depth discussions). 3D faces have the advantage that they are less affected by the pose and illumination changes. Along the same vein, 3D sketch models reconstructed from 2D sketches may also improve sketch recognition rate.

To build 3D sketch models, we developed a new scale-space topographic feature representation approach to model the facial sketch appearance explicitly. We initially selected 92 key facial landmarks, and then interpolated them to 459 using a Catmull-Rom spline method [26]. From the interpolated landmarks, we used a 3D geometric reference model to create individual faces. The reference model consists of 3,000 vertices. Based on the topographic

labels [27] and curvatures obtained from the sketch images, we then applied a spring-mass motion equation [27] to converge the reference model to the sketch topographic surfaces in both horizontal and depth directions [26]. This procedure was performed based on a series of numerical iterations until the node velocity and acceleration were close to zero. Such a mesh adaptation method was applied to sketch regions to instantiate the model. Fig. 10 shows an example of 3D sketch model constructed from a 2D sketch. The 3D model can be rotated to different poses that otherwise would be impossible with a 2D sketch.

.

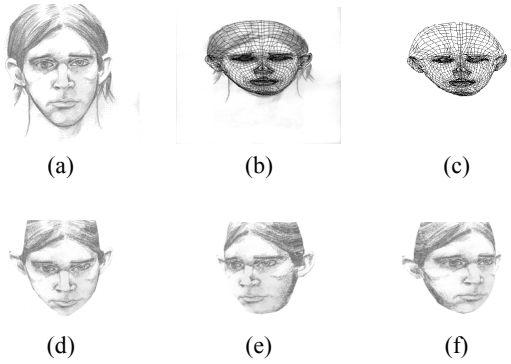


Fig. 10. A 3D face sketch model that was reconstructed from a 2D sketch. (a) The original 2D sketch. (b) The 3D model adapted to the sketch. (c) The generated 3D wireframe model; (d-f) The 3D sketch model in different views (leaned forward, left, and right) with sketch textures.

VI. SUMMARY

In this paper, we describe a biometric database that consists of rotating head videos and face sketches. In addition to some commonly seen data types such as static frontal-view images and facial expressions, this database has three unique features: (i) Rotating head videos under a strong shadow condition were acquired. This type of dataset is more challenging and has implications for certain realistic tasks such as video surveillance and criminal investigations; (ii) Frames of ten pose angles per subject were manually determined. These labeled frames will facilitate the investigations of multi-frame fusion and model-based 3D face recognition; (iii) Sketches of 50 subjects by five artists were included. To our best knowledge, this is the only database that has a large number of hand-drawn face sketches by multiple artists. Because of these features, the database can be used for a wide variety of biometric researches.

REFERENCES

[1] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", *Img. & Vis. Comp.*, 16(5), pp. 295-306, 1998.

[2] P. J. Philips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms", *IEEE Trans. PAMI*, 22(10):1090-1100, 2000.

[3] P. J Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and *et al.*, "Overview of the face recognition grand challenge," *IEEE CVPR05*, pp. 947-954, Washington DC, 2005.

[4] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, and W. T. Scruggs, "The iris challenge evaluation 2005", *IEEE Conf. on BTAS08*, Washington, DC, 2008.

[5] K. W. Bowyer, P. J. Flynn, Notre Dame Biometrics Database, http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html.

[6] http://bias.csr.unibo.it/fvc2006/.

[7] S. Sarkar, P. J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. W. Bowyer, "The human ID gait challenge problem: Data sets, performance, and analysis", *PAMI*, 27(2): 162-177, 2005.

[8] X. Tang, and X. Wang, "Face sketch recognition", *IEEE Trans. on Circuits and Sys. for Video Tech.*, 14(1), pp. 50-57, 2004.

[9] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker, "Multi-PIE", *Auto. Face and Gesture Recog.*, 2008.

[10] http://www.nist.gov/srd/nistsd18.htm

[11] R. Gross, "Face Databases", *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, ed., Springer-Verlag, 2005.

[12] http://www.face-rec.org/databases/

[13] S. Canavan, M. Kozak, Y. Zhang, J. R. Sullins, M. Shreve, and D. Goldgof, "Face recognition by multi-frame fusion of rotating heads in videos", *IEEE Conf. on BTAS*, 2007.

[14] D. Thomas, K. W. Bowyer, and P. J. Flynn, "Multi-frame approaches to improve face recognition," *IEEE Workshop on Motion and Video Computing*, Austin, TX, 2007.

[15] D. Thomas, K. W. Bowyer, and P. J. Flynn, "Strategies for improving face recognition from video", in *Advances in Biometrics: Sensors, Systems and Algorithms*, N. Ratha and V. Govindaraju, editors, Springer, 2007.

[16] K. T. Taylor, *Forensic Art and Illustration*, CRC Press, 2000.

[17] L. Gibson, *Forensic Art Essentials: A Manual for Law Enforcement Artists*, Academic Press, 2007.

[18] C. D. Frowd, V. Bruce, A. McIntyre, D. Ross, and *el al.*, "Implementing holistic dimensions for a facial composite system", *Journal of Multimedia*, 1(3), pp. 42-51, 2006.

[19] R. G. Uhl, and N. V. Lobo, "A framework for recognizing a facial image from a police sketch", *IEEE CVPR*, 1996.

[20] P. Yuen, and C. Man, "Human face image searching system using sketches", *IEEE Trans. on SMC-A,* 37(4):493-504, 2007.

[21] V. Bruce, H. Ness, *et al.*, "Four heads are better than one: combining face composites yields improvements in face likeness", *J. of Applied Psychology*, 87(5):894-902, 2002.

[22] C. D. Frowd, D. McQuiston-Surrett, S. Anandaciva, and *et al.* "An evaluation of US systems for facial composite production", *Ergonomics*, 50:562-585, 2007.

[23] C. Frowd, P. Hancock, and D. Carson, "EvoFIT: A holistic evolutionary facial imaging technique for creating composites", *ACM Trans. on Applied Perception*, 1: 1-21, 2004.

[24] Y. Zhang, C. McCullough, J. R. Sullins, and C. R. Ross, "Human and computer evaluations of face sketches with implications for forensic investigations", *IEEE Conf. on BTAS*, Washington DC, 2008.

[25] K. W. Bowyer, K. Chang, and P. J. Flynn. "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *CVIU*, 101(1):1-15, 2006.

[26] S. Canavan and L. Yin. "Dynamic face appearance modeling and sight direction estimation based on local region tracking and scale-space topo-representation," *IEEE Conf. on Multimedia and Expo.,* New York, NY, June, 2009.

[27] L. Yin and K. Weiss. "Generating 3d views of facial expressions from frontal face video based on topographic analysis," *ACM Multimedia,* pp. 360-363, 2004.