An Experimental Study on the Accuracy Issue of Automatic User Verification Based on a Single Fingerprint Image

Qinghai Gao

Abstract— In crime scene investigation, it is often the case that a single fingerprint image is found. In fingerprint authentication system it is common that a single fingerprint image is captured and applied during the verification stage. However, two questions can be asked: (1) How reliable is it to identify an individual based on a single image? (2) Does short-time interval have any effect on matching? In this paper experiments with individual fingerprint images from six databases are carried out and the findings are reported.

Index Terms— Single image, Fingerprint minutiae template, Automatic user verification, Genuine match rate, Accuracy, Database, False non-match rate, Short-time interval.

1 Introduction

BIOMETRICS is defined as the recognition of an individual based on his/her peculiar physiological and/or behavioral characteristics. In general, physiological features, such as iris, fingerprint, hand geometry, face, are more stable and thus can be utilized to identify a person more accurately than behavioral features, such keystroke, signature, gait, etc.

Among various commonly used physiological biometrics iris is still regarded as one of the most stable and accurate biometrics [6], [7], in spite of the fact that ageing does degrading the performance of iris recognition, as indicated by a few researchers [13], [14], [15], [16]. Even though face biometrics is widely used for recognition, many researchers [17], [18], [19], [20], [21] have studied the ageing effects on face recognition due to its long-term instability.

Fingerprint is a broadly employed biometric in criminal investigation and physical access control. Like other biometric technologies, fingerprint verification system works with two stages: registration and verification. During registration a user's live fingerprint will be measured to obtain fingerprint images, from which a template is extracted and then stored in a database. During verification, the person must provide the same live finger for new image capturing. A new template will be generated from the new image and then matched against the registered template. The matching score will be compared with a predefined threshold to determine whether the new template matches the stored template or not.

As Jain [1] indicated that the fundamental premise of biometric recognition is that biometric traits are unique and permanent with very small intra-class variability and very large inter-class variability.

The large intra-class variability could be caused by short-term effects such as changes in moisture, temperature, sensor type, illumination, finger pressure and position, etc, or by long-term effects such as ageing. A few papers have been concentrated on accounting for the intra-class variability of fingerprints due to ageing. Kang et al. [3] found that four different types of sensors perform differently and optical sensor showed the best results. Modi et al. [4] studied the impacts of ageing on fingerprint matching and found that error rates increase for fingerprints of older individuals from which it is more difficult obtaining quality images. Uhl and Wild [5] found that the verification performance of different fingers are different from each other, such as middle finger often has lower equal error rates than little finger. Overall, there is limited literature concentrated on the intra-class variability of fingerprint.

In recent years the fingerprint based misidentification cases, such as Brandon Mayfield [8], Shirley McKie [9], Rick Jackson [10], and Lana Canen [11], have raised the doubts on the effectiveness of a latent or partial fingerprint for identification. However, it is a common belief of the biometric community that a single fingerprint image with reasonably good quality can always verify a user automatically. In this paper, to check the validity of such a belief we investigate the accuracy of fingerprint based verification by matching fingerprint minutiae template constructed from a single fingerprint image of good quality. Also we make an attempt to check if short-time interval has any effect on fingerprint matching.

The rest of the paper is organized as the following. In section II, fingerprint minutia and matching algorithm are briefly described. Section III presents the experimental results with six publicly available fingerprint databases. In Section IV we summarize the paper and propose future research direction.

2 METHODS

Currently, majority of automatic fingerprint recognition systems are based on minutia points, which are the ridge endings and ridge bifurcations (Refer to Fig. 1). Each point is represented with one triple (x, y, θ), where (x, y) is a minutia's Cartesian coordinates, and θ is the orientation of ridge flow at the

• *Qinghai Gao, is currently an Associate Professor in the Department of Criminal Justice & Security Systems at Farmingdale State College, A Campus of the State University of New York, Farmingdale, New York. Email: GaoQJ@farmingdale.edu*

point. A fingerprint template consists of a number of such points extracted from one or more fingerprint images.

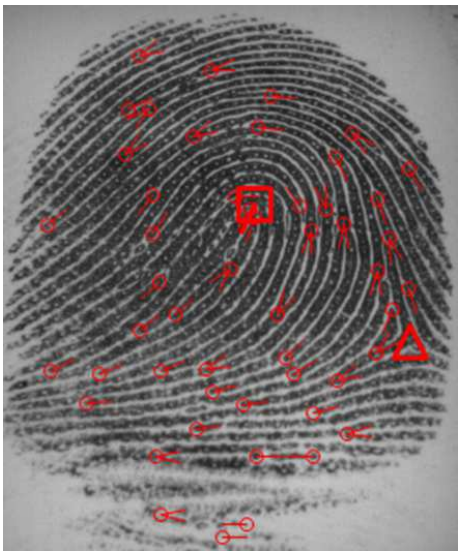

Fig. 1. Fingerprint minutiae

The matching algorithm consists of three major steps [27], [28]:

•Construct intra-fingerprint minutia comparison tables (CT)

One CT is for the registered fingerprint and one CT is for the query fingerprint. For each pair-wise minutia comparison, an entry including the length (d) of the line segment between the two minutiae and the angles (β1, β2) between each minutia and the line connecting the two minutiae, is made into a CT.

•Construct inter-fingerprint minutia compatibility tables

Compare each entry in the registered fingerprint's CT to the entries in the query fingerprint's CT. If the differences between the lengths of the two segments and those of the corresponding angles are smaller than or equal to the predefined thresholds, an entry representing the two pairs of minutiae will be made into the inter-fingerprint minutia compatibility table.

•Traverse the inter-fingerprint compatibility tables to obtain a matching score between the two fingerprints.

More details can be obtained from NIST Biometric Image Software [27], [28]. All the matching results in this paper are obtained with this algorithm, based on the threshold value 40.

In this study six fingerprint databases are selected, as given in Table 1. According to [24], the original datasets were constructed with three sessions separated by intervals of more than two weeks. During each session four images were captured from each finger. In all there were 12 images per finger from three sessions. However, the downloaded databases contains only 8 images from each finger in both DB1_A and DB1_B of FVC2004. Therefore, it is unclear about the sessions from which the images were captured.

TABLE 1
DATABASES UTILIZED FOR EXPERIMENTS

Fingerprint database	DB1_B FVC2000	DB1_B FVC2002	DB1_B FVC2004	DB2_B FVC2006	DB1_A FVC2004	DBII PolyU
Source	[22]	[23]	[24]	[25]	[24]	[26]
Sensor type	Optical	Optical	Optical	Optical	Optical	Optical
Image size	300x300	388x374	640x480	400x560	640x480	640x480
Resolution	500dpi	500dpi	500dpi	569 dpi	500dpi	1200dpi
No. Fingers	10	10	10	10	100	148
Images/per finger	8	8	8	12	8	10
Imaging sessions	1	1	N/A	1	N/A	2

For the DBII from PolyU [26], there were two image-capturing sessions separated by two weeks. In the first session a set of five images were captured from each finger. In the second session another set of five more images were captured from the same finger. The ten images were separated into two groups according to their capturing sessions.

All the experimental results in this study are obtained from fingerprint images in these databases.

3 EXPERIMENTAL RESULTS

First we conducted testing with the databases from Fingerprint Verification Competition (FVC). Then, experiments were carried out with the high-resolution fingerprint database DBII from PolyU [26]. The results are given below.

3.1 FVC2000~2006 [22], [23], [24], [25]

As given in Table 1, for the following three databases, DB1_B of FVC2000, DB1_B of FVC2002, and DB1_B of FVC2004, each contains 80 images captured from 10 fingers, while DB2_B of FVC2006 contains 120 images captured from 10 fingers.

For all the images from each finger, we do the following steps:

•Process image for minutiae extraction: each image is enhanced, binarized, and then compressed with wavelet scala quantization (wsq) at a ratio of 15:1.

•Construct a minutia template for every image: minutiae extraction procedure is carried out on the wsq-compressed fingerprint image.

•Match each template against other templates belonging to

the same finger to obtain matching scores with the matching algorithm from [28].

•Categorize each matching score as a match or a non-match based on the threshold.

•Calculate the matching rate (or non-matching rate) based on the number of matching scores (or non-matching scores).

In this paper, if the matching score is greater than or equal to the chosen threshold, 40, it is considered a match. Otherwise, it is a non-match.

For the DB2_B of FVC2006, there are 12 images for each finger, which give 66 matching scores (12×11/2 =66). One calculation example is given here. For Finger#2, 41 out of 66 matching scores are greater than or equals to the threshold 40. Therefore, the genuine match score for Finger#2 is calculated as the following:

$$(41\div66)\times100\% = 62.1\%$$

For the other three databases, there are 8 images from each finger, which give 28 matching scores (8×7/2 =28). The genuine match rates are calculated accordingly. The results are summarized in Table 2.

TABLE 2
AVERAGE GENUINE MATCH RATES OF DIFFERENT IMAGES OF SAME FINGER

FVC Database	2000	2002	2004	2006
Finger#	DB1_B	DB1_B	DB1_B	DB2_B
1	***3.6%***	32.1%	46.4%	12.1%
2	57.1%	50.0%	50.0%	62.1%
3	39.3%	***14.3%***	57.1%	93.9%
4	85.7%	89.3%	85.7%	43.9%
5	57.1%	71.4%	42.9%	37.9%
6	78.6%	100.0%	21.4%	77.3%
7	78.6%	100.0%	71.4%	***7.6%***
8	28.6%	96.4%	53.6%	80.3%
9	60.7%	64.3%	32.14%	98.5%
10	7.1%	82.1%	***7.1%***	92.4%
Average	49.6%	70.0%	46.8%	60.6%
Total Average	56.8%			

From Table 2 we can identify the following fingers that give the lowest genuine match rate for each database: Finger#1 of FVC2000 DB1_B (3.6%), Finger#3 of FVC 2002 DB1_B (14.3%), Finger#10 of FVC2004 DB1_B (7.1%), and Finger#7 of FVC2006 DB2_B (7.6%), as highlighted in the table. These results tell us that the genuine match rates for some fingers can be very low. Therefore, false non-match becomes an issue for these fingers.

Looking at the average genuine match rates for all the fingers in a database (columnwise) in Table 2, we can see that DB1_B of FVC2002 has the highest average genuine match rate 70.0%, while DB1_B of FVC2004 has the lowest average genuine match rate 46.8%. Since the overall average genuine match rate for all the databases in Table 2 is 56.8% (in the last row), which gives a false non-match rate 43.2%, the accuracy of such a fingerprint verification system is generally considered as unacceptable.

By dissociating the original finger number with its corresponging score, we sorted the matching scores for each database and plotted the results into Fig. 2.

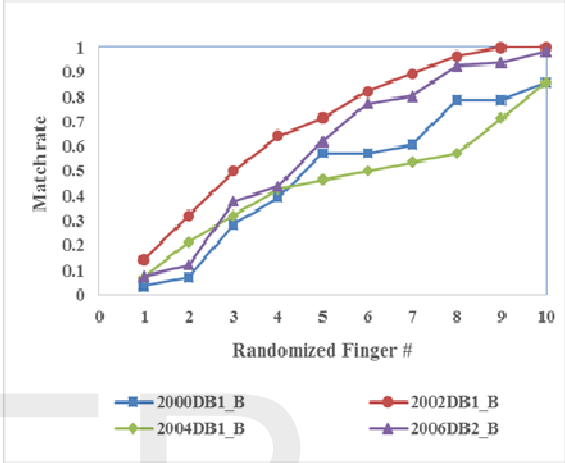

Fig. 2. Reordered genuine match rates for the individual fingers in FVC databases

From Fig. 2 we can see that DB1_B of FVC2002 has the best genuine match rates, followed by DB2_B of FVC2006, DB1_B of FVC 2000, and DB1_B of FVC2004 in that order. These findings are consistent with the results reported by Li et al. [12].

Since the number of fingers covered by the four databases listed in Table 2 is small, the validity of these results can be questioned. Therefore, we further carried out experiments with DB1_A of FVC2004, which contains 800 images from 100 fingers (8 images from each finger). The results are plotted in Fig. 3.

From Fig.3 we can estimate that the overall average of matching scores for the the entire database should locate somewhere between 0.5 and 0.6, as indicated by the horizontal line in the figure. In fact, we found by calculation that the lowest average genuine match rate is 14.3% (Finger#96) and the highest average genuine match rate is 96.4% (Finger#31 and Finger#100). The overall average genuine match rate for the 100 fingers in the database is 54.4%, which is very close to the overall average given in Table 2.

With these results we conclude that it is inaccurate to verify users based on a single fingerprint image. To check the validity of this conclusion, we carried out experiments with another database containing high-resolution fingerprints. The results are given in 3.2.

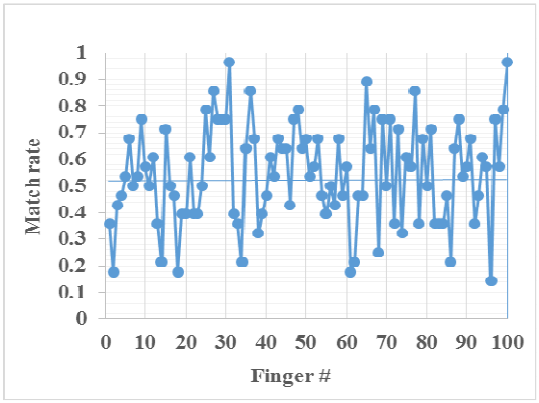

Fig. 3. Genuine match rates of the individual fingers in DB1_A of FVC2004

3.2 PolyU HRF Database DBII [26]

In this section the fingerprints in the high-resolution fingerprint database DBII prepared by the Hong Kong Polytechnic University [26] were utilized. Details about images in the database are given in Table 1.

Here we carried out matching experiments with the following three categories (after constructing the minutiae templates):

•Category 1 (cat-1): self matching - a template matches against itself. In reality it is impossible to regenerate a biometric template with 100% accuracy. However, this type of matching represents an ideal circumstance that can be used for the purpose of comparison a contrast with non-ideal or practical situation.

•Category 2 (cat-2): same day matching - a template matches against a different template extracted from an image captured in the same session. This is the scenario when a user starts using the finger for verification right after registration.

•Category 3 (cat-3): 2-week interval matching - a template matches against a different template extracted from an image captured two weeks after registration. This represents a more common scenario on when a user would use his register fingerprint for verification that category 2. It is a testing on short-term stability of fingerprint.

Since there are two image-capturing sessions (separated by two weeks), each of which produces five fingerprint images, there are 10 cat-1 matching scores (10 different images), 20 cat-2 matching scores (5x4/2=10, 10x2=20), and 25 cat-3 matching scores (5x5=25).

Experiments were conducted with 148 fingers in the database. The matching results of four fingers, Finger#32, #33, #95, and #112, are given in Fig. 4. The finger-wised average matching scores for all 148 fingers in DBII are plotted in Fig. 5 and the finger-wised non-self matching scores are replotted in Fig. 6.

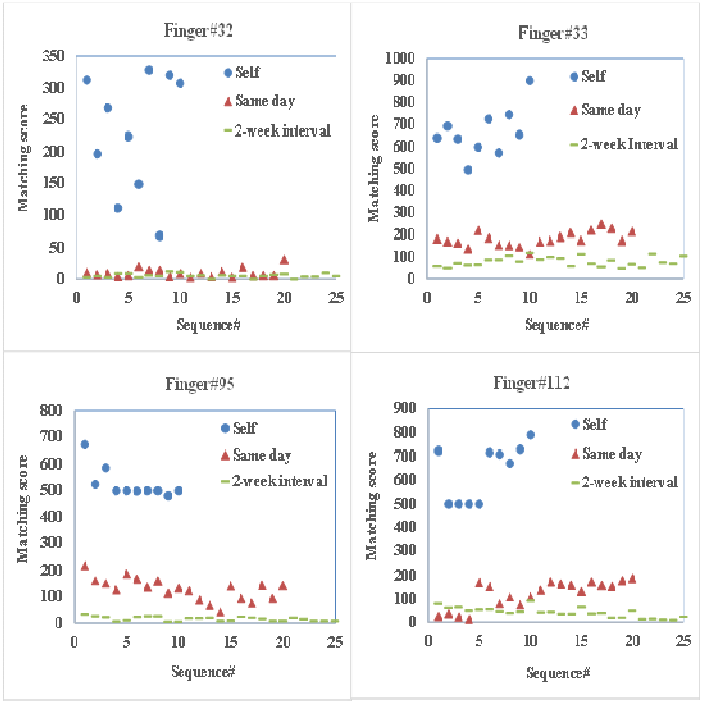

Fig. 4. Matching results of four fingers in HRF DBII PolyU

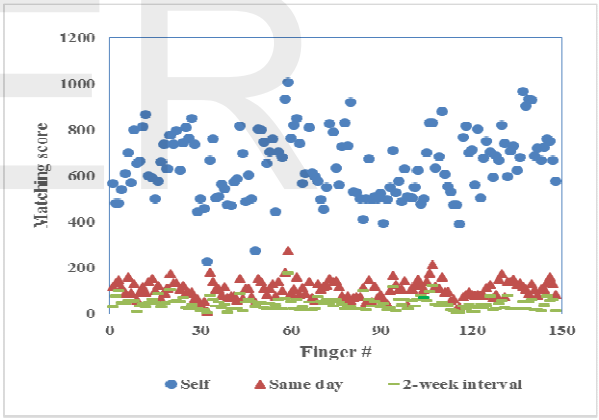

Fig. 5. Individual finger based average matching scores of the fingerprints in DBII

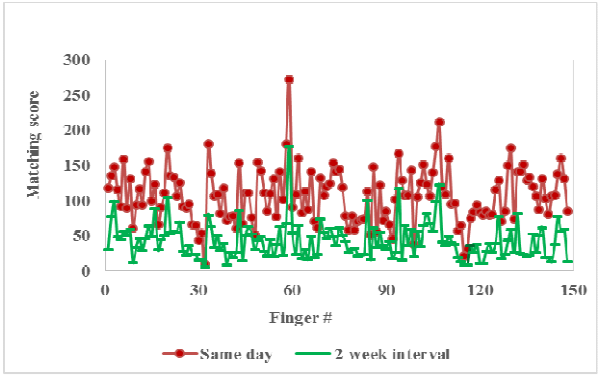

Fig. 6. Individaul finger based non-self average matching scores of the fingerprints in DBII

From Fig. 4, Fig. 5, and Fig. 6, we can draw the following two conclusions:

•The self (cat-1) matching scores are much higher than the non-self matching scores (cat-2 and cat-3).

•The same day matching scores are generally higher than those of the 2-week interval.

The matching results of the four fingers in Fig. 4 and the overall average for all the fingers in DBII are summarized in Table 3, based on the selected threshold.

TABLE 3
MATCHING RESULTS SUMMARY FOR FINGERS IN DBII

#	Average & Standard deviation (Std)						Matching rate	
	cat-1	Std	cat-2	Std	cat-3	Std	cat-2	cat-3
32	228.4	94.2	9.9	7.0	5.4	3.2	0	0
33	664.7	109.6	180.5	33.0	79.3	21.5	100	100
95	525.4	58.9	128.8	42.4	15.2	8.4	100	0
112	633.4	119.2	119.8	57.9	41.3	21.6	80	52
All	645.3	95.4	106.9	42.5	42.8	16.0	88.6	45.2

From Table 3, it can be seen that:

•For Finger#32, the matching rates for both cat-2 and cat-3 are 0%. Therefore, a single image of this finger cannot be used for user verification.

•For Finger#33, even though the 2-week interval lowers the matching scores significantly, the matching rates for both cat-2 and cat-3 are 100%. Therefore, a single image of this finger is adequate for user verification.

•For Finger#95, the cat-2 matching rate is 100% while the cat-3 matching rate is 0%. That is to say, the fingerprint images changed dramatically during the 2-week interval. Therefore, this type of images cannot be used for user verification due to their lackness of stability.

•For Finger#112, the cat-2 matching rate is 80% while the cat-3 matching rate is 52%, which means the fingerprint images changed significantly during the 2-week interval. Applying this type of images for user verification can result in very high false rejection rate.

•For the entire database, The average cat-2 (Same day) matching rate is 88.6%, which mean that if users register their fingerprints and then use them on the same day (or immediately), 11.4% of the users will not be able to authenticate themselves successfully with a single image. Perhaps it is okay for controlling access to high-security facility to have such a high false rejection rate. However, it is generally unacceptable for commercial applications with a large customer population. Since the average cat-3 (2-week interval) matching rate is 45.2%, such a system will fail to verify its registered users with a single image in most circumstances.

These results are comparable to those given in 3.1.

In sum we reiterate our conclusion that even with high resolution fingprint, it is inaccurate to verify a user based on a single fingerprint image.

4 CONCLUSIONS

In this paper we utilized the most widely adopted open source fingerprint software to investigate the accuracy of user verification with a single fingerprint image. Six online accessible databases were selected for testing, including DB1_B of FVC2000, DB1_B of FVC 2002, DB1_B of FVC2004, DB2_B of FVC2006, DB1_A of FVC2004, and HRF DBII of PolyU.

In the experiments, one minutiae template is extracted from one fingerprint image. The template is matched against another template originated from the same finger to generate a matching score, which is then compared with a predefined threshold to determine the matching results: a match or a non-match.

Our experimental results show that for the fingerprints in the three DB1_B and one DB2_B databases of FVC, the average genuine match rate based on a single image is about 56.8%, while the rate becomes 54.2% for the fingerprints from DB1_A of FVC2004 (The corresponding false non-match rate is 45.8%)

With the fingerprints from the DBII of PolyU, we carried out experiments with the following three categories: cat-1 (Self matching, one image), cat-2 (different images captured on the same day/session, non-self), and cat-3 (different images captured with a 2-week interval, non-self). The results show that the self (cat-1) matching scores are much higher than the non-self (cat-2 and cat-3) matching scores. And the same day (cat-2) matching scores are generally higher than the 2-week interval (cat-3) matching scores. The overall average genuine match rate is about 45.2% (The corresponding false non-match rate is 54.8%). Therefore, the short-time interval does have a significant effects on the matching for the fingerprints in the database.

Due to the low genuine match rates, we conclude that automatic user verification with minutiae template generated from a single image may have unacceptable high false non-match rate. Future research will be directed towards extracting other distinct information from an image, in addition to its minutiae, to reduce false non-match rates and improve matching performance.

ACKNOWLEDGMENT

The author wishes to thank Dr. Javier Ortega-Garcia from Universidad Autonoma de Madrid for allowing us to use the DB2_B database from FVC2006, and to thank Dr. Lei Zhang from Hong Kong Polytechnic University for allowing us to use the PolyU HRF Database.

REFERENCES

[1] Anil K. Jain, "Biometric Recognition: A New Paradigm for Security," Information Trust Institute Distinguished Lecture Series, Univ. of Illinois, Urbana-Champaign, October 18, 2007.

[2] U. Uludag, A. Ross and A. K. Jain, "Biometric Template Selection and Update: A Case Study in Fingerprints," Pattern Recognition, vol. 37, no. 7, pp. 1533-1542, July 2004.

[3] H. Kang, B. Lee, H. Kim, D. Shin, and J. Kim, "A Study on Performance Evaluation of Fingerprint Sensors," Audio- and Video-Based Biometric Person Authentication. Springer Berlin Heidelberg, pp. 574-583, January 2003.

[4] S.K. Modi, S.J. Elliott, J. Whetsone, and K. Hakil, "Impact of Age Groups on

Fingerprint Recognition Performance," IEEE Workshop on Automatic Identification Advanced Technologies, vol., no., pp. 19-23, 7-8 June 2007, doi: 10.1109/AUTOID.2007.380586.

[5] A. Uhl, and P. Wild, "Comparing Verification Performance of Kids and Adults for Fingerprint, Palmprint, Hand-geometry and Digitprint biometrics," IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, vol., no., pp.1-6, 28-30 Sept. 2009, doi: 10.1109/BTAS.2009.5339069.

[6] J. Daugman, "Recognizing persons by their iris patterns," Biometrics: Personal Identication in Networked Society, A. K. Jain, R. Bolle, and S. Pankanti, Eds. Norwell, MA, USA: Kluwer, pp. 103-121, 1998.

[7] P.J. Grother, J.R. Matey, E. Tabassi, G.W. Quinn, and M. Chumakov, "IREX VI. Temporal Stability of Iris Recognition Accuracy," NIST Interagency Report 7948, July 24, 2013.

[8] Office of the Inspector General, Oversight and Review Division, "A Review of the FBI's Handling of the Brandon Mayfield Case," January 2006.

[9] J. Fisher, "Shirley McKie Case: Fingerprint Misidentification," January 9, 2012. Access online on August 15, 2013 from: http://jimfishertruecrime.blogspot.com/2012/01/shirley-mckie-case-fingerprint.html

[10] J. Fisher, "Rick Jackson Fingerprint Misidentification Case," November 25, 2012. Access online on August 15, 2013 from: http://jimfishertruecrime.blogspot.com/2012/11/rick-jackson-fingerprint.html

[11] M. Godsey, "Fingerprint Misidentification Leads to Wrongful Conviction in Indiana (State of Indiana v. Lana Canen)," October 17, 2012. Accessed online on August 15, 2013 from: http://wrongfulconvictionsblog.org/

[12] S. Li, H. Kim, C. Jin, S. Elliott, and M. Ma, "Assessing the Level of Difficulty of Fingerprint Datasets Based on Relative Quality Measures," International Conference on Hand-Based Biometrics, pp. 1-5, 17-18 November 2011, doi:10.1109/ICHB.2011.6094295.

[13] S.P. Fenker, E. Ortiz, and K.W. Bowyer, "Template Aging Phenomenon in Iris Recognition," Access, IEEE, vol.1, no., pp.266-274, 2013.

[14] C. Rathgeb, A. Uhl, and P. Wild, "State-of-the-Art in Iris Biometrics," Iris Biometrics, Springer New York, pp. 21-36, 2013.

[15] S.E. Baker, K.W. Bowyer, P.J. Flynn, and P. J. Phillips, "Template aging in iris biometrics: evidence of increased false reject rate in ICE 2006," Handbook of Iris Recognition, M. Burge and K. Bowyer, eds., Springer, pp. 205-218, 2013.

[16] N. Sazonova, F. Hua, X. Liu, J. Remus, A. Ross, L. Hornak, and S. Schuckers, "A study on quality-adjusted impact of time lapse on iris recognition," Proc. SPIE 8371, Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX, 83711W, May 1, 2012, doi:10.1117/12.919642.

[17] S. Gurumurthy, "Age Estimation and Gender Classification based on Face detection and feature extraction," International Journal of Management & Information Technology, vol. 4, no. 1, pp. 134-140, 2013.

[18] P. Phillips, J. Beveridge, B. Draper, G. Givens, A. O'Toole, D. Bolme, J. Dunlop, Y. Lui, H. Sahibzada, and S. Weimer, "The Good, the Bad, and the Ugly Face Challenge Problem," Image and Vision Computing, vol. 30, no. 3, pp. 177-185, 2012.

[19] A. Lanitis and N. Tsapatsoulis, "Quantitative Evaluation of the Effects of Aging on Biometric Templates," IET Computer Vision, Special Issue: Future Trends in Biometric Processing, vol. 5, no. 6, pp. 338–347, 2011.

[20] A.M. Albert, K. Ricanek, and E. Patterson, "A Review of the Literature on the Aging Adult Skull and Face: Implications for Forensic Science Research and Applications," Forensic Science International, vol. 172, no. 1, pp. 1–9, 2007.

[21] J.S. Nayak and M. Indiramma, "Efficient face recognition with compensation for aging variations," Fourth International Conference on Advanced Computing, vol., no., pp. 1-5, 13-15 December 2012, doi: 10.1109/ICoAC.2012.6416839.

[22] FVC2000, available at: http://bias.csr.unibo.it/fvc2000/

[23] FVC2002, available at: http://bias.csr.unibo.it/fvc2002/

[24] FVC2004, available at: http://bias.csr.unibo.it/fvc2004/

[25] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano and J. Gonzalez-Rodriguez, "BioSec baseline corpus: A multimodal biometric database", Pattern Recognition, Vol. 40, n. 4, pp. 1389-1392, April 2007.

[26] PolyU HRF Database II, available at: http://www.comp.polyu.edu.hk/~biometrics/HRF/HRF.htm

[27] C. L. Wilson, C. I. Watson, M. Garris, and A. Hicklin, "Studies of fingerprint matching using the NIST verification test bed (VTB)", 2003. Available at: ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7020.pdf

[28] NIST fingerprint software. Available at: http://fingerprint.nist.gov/nfis/