Biometric Menagerie in Time-Span separated Fingerprint Data

Simon Kirchgasser and Andreas Uhl
Department of Computer Sciences
University of Salzburg
Jakob-Haringer-Str. 2
5020 Salzburg, AUSTRIA
Email: {skirch, uhl}@cosy.sbg.ac.at

***Abstract*—Multiple factors are influencing the performance of fingerprint recognition systems. Some of those depend on the used recognition implementations or datasets, while other factors like fingerprint ageing can be difficult to isolate. The aim of this research is the consideration of user related characteristics which have been introduced as the so called "Doddington's Zoo" to describe possible present template ageing influence. Certain user dependent weaknesses could be influenced by fingerprint ageing such that those system vulnerabilities are amplified or even attenuated. To investigate this aspect, the users in the databases (including a time separation of 4 years) are labelled according to the main model provided by the menagerie concept. The analysis of the single categories revealed that the animal groups are not extended (labelled users in the older datasets are not the same as in the newer ones) regardless which dataset and recognition system is considered.**

I. Introduction

There are different ways to measure the performance of a fingerprint (FP) recognition system [1]. Apart from well known methods including equal error rate (EER), receiver operating curve (ROC) or by describing the tradeoff between the system's false rejection (FRR) and false acceptance rate (FAR), it is also possible to have a look at specific user dependent genuine and impostor score characteristics. The concept of "Doddington's Zoo" [2] introduced the most basic approach to label the users of a given biometric system (originally speaker recognition) as member of different groups (sheep, goats, lambs, wolves - will be explained later into more detail) based on the FRR and FAR of the system. There have been various investigations which tried either to use the basic idea of assigning user properties to improve given concepts of biometric recognition systems or to advance the biometric menagerie in a more theoretical way. The original application domain of "Doddington's Zoo" in speech recognition was extended to many different areas (e.g. FP and face recognition [3], [4]) as discussed before. The possibility to use the menagerie analysis to study the effect of state of art template "self update" procedure on the different animal groups was introduced in [4] on a non-public dataset. [5] is of special interest for the present paper. Investigations wrt. template ageing in hand biometric features (dataset acquisition by a flatbed scanner including a time-span of 5 years) revealed that detected goats in an older dataset seem to be invariant to time progression and can be found in a later acquired dataset as well. The mentioned influenence of ageing in FP recognition is an important aspect in recent studies. Different age groups have been considered in [6]. The investigation revealed that kids' verification performance is lower compared to adults in e.g. higher EER. [7] confirmed a significant recognition performance degradation for FPs in data including a time-span of 10 to 30 years. Another study [8] revealed a decrease of genuine scores for 7 year time separated datasets. The same effect must not be only related to the age of the user. If the imprint quality of the data decreases at the same time an identical effect can be described as well. Due to this a covariate-fit analysis was performed, explaining that image quality describes the genuine score variation better than time-span separation or subject's age can do.

All the studies described so far have not been performed on datasets acquired with commercial off-the-shelf scanners. The aim of this research was to use data which was acquired by off-the-shelf commercial FP sensors and to describe eventually FP ageing related results based on "Doddington's Zoo". It is known that ageing influences skin to become looser and dryer for elderly people due to loss in collagen [9]. This could lead to some degradations which amplifies user specific weaknesses. Those weaknesses can be detected by observing incorrectly low genuine or incorrectly high impostor scores during the matching score calculations. The basic idea will be to find users who are susceptible to such score dependend effects in the first place. Secondly, all those selected persons (we denote them as being "labelled" from now on) from the older dataset are compared to those from the newer datasets. In case that the same user is labelled in both datasets as being of the same biometric menagerie user class, this behaviour indicates time invariance, i.e. ageing effects wrt. biometric menagerie. Following this observation, we define two menagerie class membership phenomena which are considered as indicators for biometric template ageing: 1) Users labelled as belonging to a specific menagerie user class are different in time separated datasets. 2) Users labelled as belonging to a specific menagerie user class are identical in time separated datasets **but** in case matching is conducted involving both time separated datasets, a different set of users is labelled as belonging to this menagerie user class.

The observation of one of these phenomena indicates a high likelihood for biometric template again being present in the used datasets. So the characteristics which are responsible for the labelling of the user in the older dataset are kind of not being preserved across the time-span and influenced by this biological process. We performed experiments which should ascertain the stability of the most important biometric menagerie groups and reflect possible influence of FP ageing. Different analysis methods are applied as discussed in [5]. Especial their robustness in terms of outliers in FP matching (very low or high matching scores) was important. The rest of this paper is organised as follows: In Section II the concept of the so called biometric menagerie is introduced. The used FP recognition systems and datasets will be discussed in Section III. The applied analysis methodologies and the the experimental setup are highlighted in Section IV. The achieved results of the investigations are described in Section IV-A. The final Section V will conclude this paper.

II. Biometric Menagerie

In order to assess a recognition system it is of interest to characterise biometric datasets which display the most important influence on the authentication process. Different techniques for this purpose are investigated in [10]. In order to compare the FRR, confidence interval estimation using parametric and bootstrap-based technique is introduced. This FRR based investigation is displaying the convenience aspect of a given recognition system. An analysis of the FAR behaviour would be the more suited approach in terms of security. Some of those weaknesses in a biometrics authentication system are highlighted in [11]. Among other techniques, also "Doddington's Zoo" was used for analysis and has been extended. A new type of animal was added to the biometric menagerie: chameleons. On the one hand, those users are easy to spoof (passive), and on the other hand they are also good at spoofing others (active) which can cause problems in terms of security. Additionally there are some further classes in the biometric menagerie: worms, phantoms and doves. All of them have been introduced in [12] and display different types of weak user. Because all of the different classes included in the biometric menagerie are characterising specific types of user dependent behaviour it is not ensured that in each dataset and for each recognition system all of those types can be detected. Their existence must be proven for each dataset independently (q.v. Section IV). In addition in [13], [14] a so called "Biometric Menagerie Index" (BMI) describes the characteristics of different animal groups to display an overall dataset behaviour. Quantising the quality or delivering a characterisation of datasets can be done in various manners. A set of 4 well known minutiae based FP recognition systems have been applied on 14 datasets (including FP Verification Contest (FVC) datasets) in order to demonstrate the validity of the so called LOD (Level of Difficulty) models [15]. The presented models have also been compared to the latest menagerie addition revealing that there is a coincidence between the various animal groups and the LOD. Furthermore a classification into the original "Doddington's Zoo" groups was used to improve the results of iris and FP recognition using a user-specific fusion [16]. Exploiting certain persons characteristics results in an investiagtion analysing matching score properties of FP spoof attacks [17]. The usage of a combined matching and fusion schemes like introduced in [16] to increase the robustness of a system against spoof attacks is recommended.

The experiments conducted in this work are based on the definitions of the four classical menagerie user classes (i.e. animals) Doddington originally described in his menagerie concept: sheep, lambs, goats, and wolves.

Sheep: This first class represents the majority of the population and they are generally easy to identify. In the following there will be no consideration of them in particular. The reason for this is that the investigations will concentrate on those animal groups which represent user weaknesses.

Goats: The second user group is generally hard to identify while members are being matched against themselves. They can be identified by considering especially low genuine matching scores. Due to this, goats represent those persons who have a high likelihood of false rejection in a biometric recognition system.

Lambs: If a user is assigned to be a member of this class, than he/she can easily be imitated. So a high likelihood of being falsely accepted if an attacker is represented by a higher impostor score is realistic. Lambs can be called "passive" user. That means, "lambish" users are exceptionally likely to be successfully spoofed (so this is a "passive" property) during biometric identification because of their low impostor scores.

Wolves: The last class represents users who can easily imitate others users (so this is an "active" property). A wolf can also be detected by using the impostor scores of a recognition system similar to the "lambish" persons. While lambs are characterized by low impostor values, wolves on the opposite are described by high impostor match scores. So opposed to lambs are wolves "active" user.

III. Recognition Systems and Datasets

In this study four types of FP recognition systems have been applied to the data. Two of them are minutiae based (NIST Biometric Image Software (NBIS)[1], VeriFinger SDK[2]) and two use non-minutiae approaches (FingerCode (FC), PhaseOnlyCorrelation (POC)). Both non-minutiae applications are based on custom software re-implementing those algorithms [18].

NIST Biometric Image Software (NBIS): In the present work release 5.0.0, implemented by the National Institute of Standards and Technology (NIST) was used.

VeriFinger (NEURO): The *VeriFinger SDK* was developed by Neurotechnology and represents the second minutiae based recogntion system. The current release 8.0 includes algorithmic solutions which enhance the performance of the software

[1]http://www.nist.gov/itl/iad/ig/nbis.cfm
[2]http://www.neurotechnology.com/verifinger.html

focusing on rolled and flat FPs matching, tolerance to FP translation, rotation and deformation as well as adaptive image filtration and scanner-specific algorithm optimisations.

FingerCode (FC): The third method is a ridge feature recognition system. The application of 8 Gabor filters (from 0° to 180°) results in 8 so called "Standard Deviation Maps" which are finally combined to one single map: the Ridge Feature Map. The local orientation and frequency information stored in these ridge feature maps are compared in the matching step by calculating the correlation value in the Fourier space. Subsequently deriving the ITF (Inverse-Fourier-Transformation) the correlation result is weighted due to the overlap of the input imprints. The final score is derived by using the Euclidean distances between the ridge feature values of the considered input FP images. Due to the fact that one of the imprints is rotated during the correlation process finding the best fitting position a list of scores is stored during the matching process. The lower a value in the list, the better is the alignment of the two FPs. Thus the lowest is assigned to be the final match score [19], [20]. In the used implementation a value between 0 (worst) and 125 (best fitting) is representing the final score value.

PhaseOnlyCorrelation (POC): This system uses a holistic correlation based method for the recognition task. According to [21], [22] a modified Phase-Only-Cor- relation function [23], the so called BLPOC (Band-Limited Phase-Only Correlation) function is used to compare FP images and get the matching score, describing the similarity between two imprints. After performing rotation and displacement alignment regions which are detectable in both FPs are selected. So the BLPOC function is used to get similarity values of the selected overlapping regions. The final score of the matching is defined as the maximum value of the output values computed by the BLPOC function. The final result range [0,1] reflects a perfect match if a score value is 1 and a worse one if the value tends to 0.

The biometrics research team at the Center for Biometrics and Security Research (CBSR) at the Chinese Academy of Sciences, Institute of Automation (CASIA) provided two main datasets for the present paper: dataset "CASIA 2009" is a subset of the CASIA-FP V5[3] database. It contains 980 FP images of 49 volunteers acquired using a U.are.U 4000 scanner produced by DigitalPersona. For both hands FP scans of forefinger and second finger, 5 prints per finger, are included. The U.are.U 4000 is an optical scanner with a resolution of 512 dots per inch (dpi). All FP imprints are 8-bit per pixel gray scale images and have a resolution of 328x356 pixels. The so called "CASIA 2013" dataset consists of five different subsets of FP images. Each subset including 980 imprints of the same 49 volunteers as in CASIA 2009. So again for each user 20 imprints acquired of the same fingers are included. The five single datasets have been acquired by a U.are.U 4000 (2 sets), a U.are.U 4500 (2 sets) and a TCRU1C scanner (1 set). The U.are.U 4500 is very similar to the U.are.U 4000.

[3]http://biometrics.idealtest.org/dbDetailForUser.do?id=7

a: CASIA 2013 image captured by TCRU1C sensor. b: CASIA 2013 image captured by U.are.U 4000 sensor. c: CASIA 2013 image captured by U.are.U 4500 sensor.

Fig. 1: Some image impressions from the second dataset.

Due to this fact are the specifications wrt. resolution, image dimensions and bit depth identical. The remaining TCRU1C sensor is a capacitive FP scanner with 508 dots per inch (dpi) resolution. The imprints have a resolution of 256x360 pixels. According the usage of 3 different sensor types an analysis of cross-sensor effects during the experiments is enabled. In Figure 1 example images for each scanner are displayed. Those introduced datasets are restructured for the following ageing analysis into 11 different datasets. Each dataset is labelled with an alphabetic character and in most cases with an index number. The CASIA 2009 dataset is abbreviated as "A" and the different subsets of CAISA 2013 are named as "B" and an index number from 1 to 5. "B1" corresponds to the dataset acquired by the TCRU1C sensor. "B2" and "B3" are the datasets acquired by a U.are.U 4000 scanner. The remaining U.are.U 4500 sensor was used during the acquisition process of "B4" and "B5". Those 6 datasets are so called "single" datasets. To perform a ageing specific analysis 5 so-called "crossed" sets have been constructed as well. They contain the imprints of CASIA 2009 (A) and one of the 5 datasets of CASIA 2013. This leads to a total number of 1960 imprints per crossed dataset. Character "C" and an index value are used as abbreviation: "C1" includes the imprints of CASIA 2009 (A) and the TCRU1C sensor recordings of CASIA 2013 (B1). "C2" and "C3" refer to the combination of CASIA 2009 (A) and the CASIA 2013 U.are.U 4000 sensor FP images (B2 and B3). dataset A and the U.are.U 4500 sensor CASIA 2013 imprints (B4 and B5) are included in "C4" and "C5". In order to receive the matching score information, the procedure used in all four FP Verification Contests (FVC), for example see [1], was applied. Because there are a different number of imprints in the datasets it is obvious that a different amount of genuine and impostor scores is computed for single and crossed databases. 1960 genuine and 95550 impostor matches were computed for datasets A and B1 - B5, respectively. Opposed to this is the number of genuine and impostor scores in C1 - C5 4.5 times and 4 times the size of the single sets, respectively. Because the applied scheme is based on the FVC scheme symmetric calculations have not been executed to avoid correlation. Due to this it is not possible to distinguish between the lambs and wolves classes. For this purpose lambs and wolves will be treated as one combined class in the following because both definitions are equivalent for the used performance experiments.

IV. Experiments and Results

In [2] two different application scenarios for the biometric menagerie analysis are introduced. The first one is trying to get existence information of the various animal classes and the second one uses a set of methods to label specific user as a member of a certain animal class. To perform a statistical existence analysis of goats, lambs and wolves three methods are used. F-Test, Kruskal-Wallis-Test and Durbin-Test have been applied for this purpose on the given data. In the present study two of them have been used as well. On the one hand the F-Test and on the other the Kruskal-Wallis-Test were applied. In both cases the null hypothesis represents that no introduced animal classes can be detected. This indicates t there are no users with special characterizations in the data detectable. A significance level of 0.01 was considered. T null hypothesis must be rejected for all databases and recog tion systems independently for F-Test and Kruskal-Wallis-T respectively. The rejection of the null hypothesis confirms presence of the goats and lambs/wolves classes independen from each other.
Two different methods, which are based on [2], are usec describe the goats and lamb/wolf cases. Both are relativ robust against single outlier matches which are represented much lower/higher impostor/genuine scores compared to majority of all matching scores.
Mean Scores Method (MSM): The first strategy is describ intra user behaviour by the expected value of the user asso ated match score distribution. So the mean of all match sco belonging to a single specific person is calculated. This leads in total to 196 mean values, one for each user included in the used datasets. Those users whose mean score is lower as the 2.5 percentile and a higher as the 97.5 percentile are labelled as e.g. goat. In Figure 2 a graphical representation of the MSM is displayed.
Mean Scores Method 2 (MSM2): Basically the same calculation is performed as mentioned in the mean score method. The only difference is that the user who have a mean score value lower than the 5 percentile are labelled as member of an animal class.

A. Menagerie User Analysis

According to the presence of 196 different users in all datasets (A, B1 - B5, C1 - C5) and the design of the used analysis (MSM, MSM2) always 11 users are classified to have a goat or lamb/wolf characteristic. In the first part of the analysis the number of labelled users has been combined for B1 - B5 and C1 - C5. The same labelled users in A, B1 - B5 and C1 - C5 would cause a rejection of the introduced menagerie class membership phenomena criteria. Otherwise, low correspondence between 2009 and 2013 is describable, confirming the first menagerie class phenomen. In case the labelled users are identical in A and B1 - B5, but different in C1 - C5 compared to A/B1 - B5, the second membership criterion would be verified.
In the following Tables I and II the main results for the menagerie analysis are displayed. The number of users to be in the same animal classes in A, B1 - B5 and in the crossed sets can be found in columns "A", "B" and "C". The total amount of different biometric animal groups in the 2013 and crossed sets excluding those who have been labelled in 2009 before are presented in column "B without A" (BwA) and "C without A" (CwA). The final column "Only Crossed" (OC) is assigned to those who are marked as "goat-like" or "lamb-/wolfish" only in the crossed datasets. The remaining columns representing the relative values where always be named by the short names of those columns which information was used to derive the results.

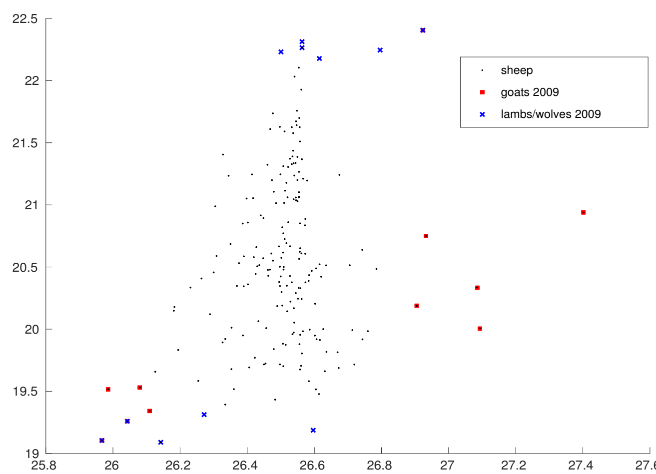

Fig. 2: Average mean genuine scores (x-axis) and average mean impostor scores (y-axis) representing MSM applied on dataset A calculating the matching scores by NBIS (intersection of points represents an intersection of menagerie classes).

Independent of the single results a correspondence between the datasets comparing the values presented in Table I and II can be detected because the total number of assigned biometric animals is clearly lower than 55. Furthermore it is very promising to see *a)* differences in terms of animal classes *b)* various fluctuations between the used recognition systems *c)* obviously some accordances concerning the labelled user comparing the results from dataset A and the crossed datasets. First, the goats and lambs/wolves cases seem to share similarities, e.g. an about equal sized number of animal types, except some bigger differences which depend on the recognition system (cf. NEURO MSM2 result in the lambs/wolves class). This very special case can be explained by a high number of very low impostor scores. Because those scores are 0 in most of the cases it can hardly be differentiated between lamb/wolf characteristics and "normal" user. Thus, the usage of NEURO for the purpose of detecting this class is not very worthwhile for this analysis. Nevertheless there is also

matcher	number of goat-like users								
	A	**B**	**C**	**BwA**	**BwA/B**	**CwA**	**CwA/C**	**OC**	**OC/C**
	mean method								
NBIS	11	33	26	29	87.88%	16	61.54%	8	30.77%
NEURO	11	40	34	38	95.00%	32	94.12%	21	61.76%
FC	11	40	34	39	97.50%	24	70.59%	19	55.88%
POC	11	48	41	46	95.83%	33	80.49%	25	60.98%
	mean2 method								
NBIS	11	31	17	25	80.65%	6	35.29%	6	35.29%
NEURO	11	34	29	33	97.06%	23	79.31%	15	51.72%
FC	11	49	30	45	91.84%	20	66.67%	14	46.67%
POC	11	40	35	39	97.50%	28	80.00%	16	45.71%

TABLE I: Number of users exhibiting a goat-like behaviour in the datasets including all sensors.

matcher	number of lamb/wolf-like users								
	A	**B**	**C**	**BwA**	**BwA/B**	**CwA**	**CwA/C**	**OC**	**OC/C**
	mean method								
NBIS	11	36	25	34	94.44%	16	64.00%	10	40.00%
NEURO	11	29	27	22	75.86%	17	62.96%	16	59.26%
FC	11	36	22	33	91.67%	14	63.64%	10	45.45%
POC	11	32	31	28	87.50%	23	74.19%	16	51.61%
	mean2 method								
NBIS	11	26	22	23	88.46%	17	77.27%	12	54.55%
NEURO	11	12	13	1	0.083%	2	0.153%	1	0.076%
FC	11	32	19	30	93.75%	11	57.89%	4	21.05%
POC	11	26	32	21	80.77%	25	78.13%	22	68.75%

TABLE II: Number of users exhibiting a lamb/wolf-like behaviour in the datasets including all sensors.

an interesting aspect because of this situation. A recognition system delivering very good matching results (sufficiently well separable genuine and very low impostor scores, respectively) are not vulnerable to this kind of weak users.

dataset	User ID
matcher	**NBIS**
	mean method
A	8, 41, 64, 86, 87, 98, 100, 108, 153, 154, 190
B1	18, 19, 32, 50, 76, 98, 101, 137, 155, 178, 192
B2	1, 20, 32, 43, 62, 66, 88, 90, 98, 110, 192
B3	13, 20, 29, 55, 88, 138, 139, 140, 190, 192, 194
B4	3, 4, 49, 20, 55, 86, 88, 98, 100, 139, 192
B5	2, 4, 19, 20, 62, 86, 88, 137, 139, 191, 192
C1	8, 17, 74, 86, 100, 104, 137, 139, 153, 154, 190
C2	3, 4, 8, 41, 42, 56, 64, 86, 87, 100, 190
C3	8, 18, 53, 64, 86, 88, 104, 137, 190, 192, 196
C4	2, 4, 17, 18, 53, 64, 86, 98, 100, 137, 190
C5	2, 8, 17, 18, 64, 86, 88, 100, 137, 189, 190

TABLE III: Goat-like user in A, B1 - B5 and C1 - C5 using mean method and NBIS system.

BwA, CwA and OC are confirming the introduced phenomena for goats and lambs/wolves case. The number of user displayed in the BwA cathegory is representing that there is not much correspondence between those user who are labelled in 2009 and those who are labelled 4 years later. The likelihood of labelling the same user in both years is very low. This result confirms the first phenomenon of this study and disproves the second one simultaneously. Without FP ageing more menagerie dependent information would probably be preserved, which could raise the likelihood of label the same user in 2013 again after labelling the user in 2009 the first time. The high amount of fluctuation in terms of different labelled user, which can also be seen in CwA and OC, is also an indication why the second hypothesis never will be provable in the datasets. The results clearly indicate that the time-span of 4 years was sufficient to distort menagerie characteristics. Of course there are other explanations possible as well. The most reasonable influences are usage of different sensor types or changing acquisition conditions (e.g. misplacement of FPs on the scanner; not cleaning the scanner plate after an imprint was acquired) which leads to fluctuations in quality of the datasets. Due to this there will be a discussion of the quality impact concerning this menagerie analysis.

Before concentrating on the quality aspect of this analysis in Table III an arbitrary example from the performed experiments is displayed. For all datasets, recognition systems and biometric animal methods the same situation can be described - there is not a consistent labelling of the users. To prove those differences two rank order correlation approaches have been used to describe the biometric menagerie classification. The first one is the Spearman rank correlation and the second one the Kendall rank correlation. The user's mean matching scores of dataset A and B1 - B5 are used to calculate the correlation coefficients of both methods. Indepently which approach was considered, the rank correlation coefficient was lower than 0.20 for the most datasets. The only exception is displayed for the lambs/wolves case using NEURO and POC. At least the results for NEURO (around 0.90) were not surprising because of the high number of similar scores. The POC results (around 0.30) did not display a strong correlation as well. Thus, the correlation coefficients confirm the high fluctuation once more. In Figure 3 the described situation is displayed e.g. for dataset A and B4. Apart from the "goat-like" and "lambish/wolfish" users who have been labelled in 2009 (cf. Figure 2) those users labbelled in 2013 are also shown. The goats and lambs/wolves in 2013 are coloured green (goats) and yellow (lambs/wolfes), respectively. As it can be seen they are very different from the 2009 animal types represented by colour red (goats) and blue (lambs/wolves).

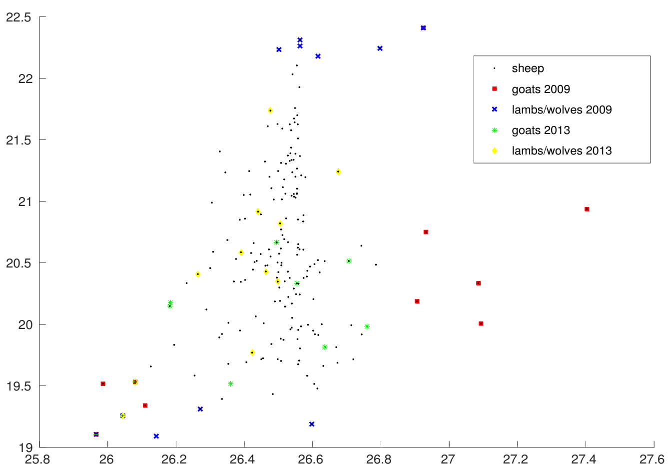

Fig. 3: Average mean genuine scores (x-axis) and average mean impostor scores (y-axis) representing MSM applied on dataset B4 calculating the matching scores by NBIS (intersection of points represents an intersection of menagerie classes).

dataset	**NFIQ**	**IQF**
A	78.40	95.34
B1	85.33	92.85
B2	65.84	95.26
B3	64.09	95.11
B4	69.73	95.25
B5	73.09	95.10
C1	81.87	94.10
C2	72.12	95.30
C3	71.25	95.23
C4	74.06	95.32
C5	75.75	95.22

TABLE IV: NFIQ and IQF values per dataset.

As last analysis step the quality of the used FP images was investigated to exclude or confirm one of the most likely reasons [8] for the described phenomena. For this purpose 2 different well known quality measures have been selected. The first technique is the NIST FP Image Quality (NFIQ)[4] which reflects the quality of an imprint from 1 (best) to 5 (worst) [24]. The second approach, the Image Quality of FP (IQF)[5] [25], is based on the Fast Fourier Transform (FFT) and calculates a score value between 0 (worst) and 100 (best). After measuring the entire dataset by those two methods (for NFIQ a weighted sum approach was used instead of calculating the mean quality value [26]), it is revealed that not quality can be the reason for the interesting observations. The quality values of the datasets are too similar to each other to confirm a quality based influence (cf. Table IV). The quality influence on the ageing phenomena found in the menagerie user class analysis could not be confirmed. Based on this information other aspects of user dependent issues like ageing, sensor type specific varieties and further acquisition conditions (misplacement of the FPs on the sensor plate) are possibly responsible for the results.

V. Conclusion

The influence of user-specific characteristics on FP template ageing data was investigated. The used imprints, acquired by commercial off-the-shelf FP scanner, include a time separation of 4 years. Concentrating on the so called "goat" and "lamb/wolf" classes we found, that there are persons with weak genuine and impostor matching scores across all datasets and that those can be categorised into the two considered biometric animal groups. Furthermore it was possible to confirm that FP ageing is a possibility explanation for a high amount of fluctuation within the performed user specifc analysis. According to the fact that no correlation between user mean genuine/impostor scores from 2009 and 2013 was discerned it can be suggested that ageing is responsible for the detected results. Besides it can be shown that the statement from [5], using the Doddington zoo, cannot be confirmed entirely. Of course there are several other reasons possible: dataset differences (flatbed scanner vs. commercial off-the-shelf scanner, general acquisition conditions, used analysis methods and several other influences could cause probably similar results as described. It could be interesting to focus on that in future work.

References

[1] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer Science & Business Media, 2009.

[2] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, Goats, Lambs and Wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *ICSLP 1998. Proceedings of the 5th International Conference on Spoken Language Processing*, vol. 4. ICSLP, 1998, pp. 1351–1354, [Online; last access 17.July 2015]. [Online]. Available: http://www.itl.nist.gov/iad/mig/publications/storage_paper/icslp_98.pdf

[3] M. Wittman, P. Davis, and P. Flynn, "Empirical studies of the existence of the biometric menagerie in the frgc 2.0 color image corpus," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*. IEEE, 2006, pp. 33–33.

[4] A. Rattani, G. L. Marcialis, and F. Roli, "An experimental analysis of the relationship between biometric template update and the doddington's zoo: a case study in face verification," in *Image Analysis and Processing–ICIAP 2009*. Springer, 2009, pp. 434–442.

[4] http://www.nist.gov/itl/iad/ig/nigos.cfm#Releases
[5] http://www2.mitre.org/tech/mtf/

[5] A. Uhl and P. Wild, "Experimental evidence of ageing in hand biometrics," in *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*. IEEE, 2013, pp. 1–6.

[6] A. Uhl and Wild, Peter, "Comparing verification performance of kids and adults for fingerprint, palmprint, hand-geometry and digitprint biometrics," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–6.

[7] M. Arnold, C. Busch, and H. Ihmor, "Investigating performance and impacts on fingerprint recognition systems," in *Information Assurance Workshop, 2005. IAW '05. Proceedings from the Sixth Annual IEEE SMC*, june 2005, pp. 1 – 7.

[8] S. Yoon and A. K. Jain, "Longitudinal study of fingerprint recognition," *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 112, no. 28, pp. 8555–8560, 2015.

[9] S. K. Modi, S. J. Elliott, J. Whetsone, and H. Kim, "Impact of age groups on fingerprint recognition performance," in *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*. IEEE, 2007, pp. 19–23.

[10] R. M. Bolle, S. Pankanti, and N. K. Ratha, "Evaluation techniques for biometrics-based authentication systems (frr)," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2. IEEE, 2000, pp. 831–837.

[11] R. M. Bolle, J. H. Connell, and N. K. Ratha, "Biometric perils and patches," *Pattern Recognition*, vol. 35, no. 12, pp. 2727–2738, 2002.

[12] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*. IEEE, 2007, pp. 1–6.

[13] N. Poh and J. Kittler, "A methodology for separating sheep from goats for controlled enrollment and multimodal fusion," in *Biometrics Symposium, 2008. BSYM'08*. IEEE, 2008, pp. 17–22.

[14] N. Poh and Kittler, Josef, "A biometric menagerie index for characterising template/model-specific variation," in *Advances in Biometrics*. Springer, 2009, pp. 816–827.

[15] S. Li, H. Kim, C. Jin, S. Elliott, and M. Ma, "Assessing the level of difficulty of fingerprint datasets based on relative quality measures," *Information Sciences*, vol. 268, pp. 122–132, 2014.

[16] A. Ross, A. Rattani, and M. Tistarelli, "Exploiting the 'doddington zoo' effect in biometric fusion," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–7.

[17] A. Rattani, N. Poh, and A. Ross, "Analysis of user-specific score characteristics for spoof biometric attacks," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 124–129.

[18] J. Hämmerle-Uhl, M. Pober, and A. Uhl, "Towards standardised fingerprint matching robustness assessment: the stirmark toolkit– cross-database comparisons with minutiae-based matching," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 111–116.

[19] A. Ross, A. Jain, and J. Reisman, "A hybrid fingerprint matcher," *Pattern Recognition*, vol. 36, no. 7, pp. 1661–1673, 2003.

[20] A. Ross, J. Reisman, and A. Jain, "Fingerprint matching using feature space correlation," in *Biometric Authentication*. Springer, 2002, pp. 48–57.

[21] K. Ito, A. Morita, T. Aoki, H. Nakajima, K. Kobayashi, and T. Higuchi, "A fingerprint recognition algorithm combining phase-based image matching and feature-based matching," in *Advances in Biometrics*. Springer, 2005, pp. 316–325.

[22] H. Nakajima, K. Kobayashi, and T. Higuchi, "A fingerprint matching algorithm using phase-only correlation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 87, no. 3, pp. 682–691, 2004.

[23] C. Kuglin, "The phase correlation image alignment method," in *Proc. Int. Conf. on Cybernetics and Society, 1975*, 1975, pp. 163–165.

[24] E. Tabassi, C. Wilson, and C. Watson, "Nist fingerprint image quality," Tech. Rep., nISTIR7151.

[25] N. B. Nill, "IQF (Image Quality of Fingerprint) software application," Tech. Rep., 2007, mTR 070053.

[26] E. Tabassi and P. Grother, "Quality summarization," Tech. Rep., 2007, nISTIR7422.