Biometric Key Computation using Handwriting Features

*MORADEYO Oluwatomilola Motunrayo and OLANIYAN Abolade Shekinah

Abstract-Biometrics is the measurement of a biological characteristic such as fingerprint, iris pattern, retina image, face or hand geometry; or a behavioural characteristic such as voice, gait or signature. It can also be said to be the science of using matchless human characteristics for personal authentication based on a person's biological and behavioural characteristics. Therefore, the process involved in transforming a piece of live biometric data into a biometric key is biometric-key computation. In this paper, biometric keys are to be generated from a behavioural biometric variety - handwriting biometric. Although, behavioural biometrics are not unique enough to deliver steadfast human identification; they have been shown to provide suitably high accuracy identity verification. They also exhibit several qualities that make them attractive for key generation. For example, whereas an adversary can passively extract physiological biometrics, behavioural biometrics do not provide themselves as easily to deceitful capture as they require a user to consciously perform an action. The signals enrolled from this biometric feature are concatenated to form one single signal and each signal is then compressed with the Discrete Wavelength Transform – Discrete Fourier Transform (DWT-DFT). Intra and inter class analysis are going to be carried out on the keys generated from handwriting captured from users.

Keywords-Biometrics, Handwriting, Key Computation, Key generation, Security, Random number generation and Verification

1 Introduction

Biometrics is the use of distinctive biological or behavioural characteristics to identify people [1]. It is a measurement of a biological characteristic such as fingerprint, iris pattern, retina image, face or hand geometry; or a behavioural characteristic such as voice, gait or signature. It is also the science of using matchless human characteristics for personal authentication based on a person's biological and behavioural characteristics [2]. [1] listed thirteen characteristics by which good biometric can be judged. Amongst these properties are primary factors such as universality, uniqueness, permanence, and measurability, i.e. the biometric should be possessed by all members of the relevant population. It should also be distinctive to each person and should remain distinctive with the ability for the data to be collected in a appropriate form. Further considerations include social, practical, and systems design issues.

Biometrics offer automated method of identity verification or identification on the principle of this measurable physiological or behavioural characteristic [3]. Collection of behavioural data often does not require any special hardware and is so very cost effective. While behavioural biometrics are not unique enough to provide reliable human identification they have been shown to provide sufficiently high accuracy identity verification.

- *Moradeyo, Oluwatomilola M.: Computer Science Department, The Ibarapa Polytechnic, Eruwa, Oyo State, Nigeria. PH-08055909700. E-mail: tokenny2003@yahoo.com*
- *Olaniyan, Abolade S. Computer Science Department, The Ibarapa Polytechnic, Eruwa, Oyo State, Nigeria. PH-08057745242. E-mail: 308mobeni@gmail.com*

Behavioural biometrics exhibit several qualities that make them attractive for key generation. For instance, whereas an adversary can passively extract physiological biometrics, behavioural biometrics do not lend themselves as easily to deceitful capture as they require a user to consciously perform an action. Additionally, while physiological biometrics cannot change, behavioural biometrics naturally changes with the action that is performed. This property is useful for security applications such as key generation, where key compromise necessitates the creation of a new key [4].

When handwriting is enrolled online, handwriting data are collected and encoded as time-varying parameters such as x and y components concerning the pen-position at time, t, the status of pen-down or pen-up, s, the pen-pressure, pr, the pen-altitude, ϕ and pen-azimuth, ϕ. So the raw handwriting data are represented as a seven dimensional feature vector $\{x, y, t, s, pr, \Theta, and\ \phi\}$ at each sampling point [2]. Therefore, data captured by sampling the position of a stylus tip over time on a digitizing tablet or pen computer are referred to as *online* handwriting, whereas inputs that are presented in the form of scanned image is referred to as *offline* handwriting [5].

Biometric key generation is the direct generation of bits out of information contained in the biometric data. The basic purpose of generating biometric-based keys in security is for the user authentication. On the other hand, biometric key release requires access to biometric template for biometric matching, this happens where the biometric secrete key and biometric template are stored in the system. The key is only released after a valid biometric match [6].

2 State of the art

Several people have worked building some models that can generate biometric keys from biometric features. [7] proposed the first biometric hash on dynamic hand

signature which made use of a 50-feature-parameter set from dynamic hand signature and an interval matrix to store the upper and lower thresholds acceptable for correct identification. Since these methods are parameter-based, the feature extraction is limited and short, and are small in key space, the keys are not cancelable and more importantly, they are generally low in entropy. They are also not secure due to storage of user-specific statistical boundaries that could be used to recover the biometric features.

[8] combined the methods of [9] and [10] to enable longer and cancelable or replaceable keys but however, the user-specific key statistics required to correct the feature vector allows an adversary to easily guess the most probable combination from the compromised user boundaries information and reduced number of segments.

Likewise, [11] proposed a user-specific, likelihood ratio based quantizer (LQ) that allows extraction of multiple bits from a single feature. The bits generated from every feature are concatenated to form a fixed length binary string that can be hashed to protect its privacy. The keys were derived directly from biometrics data as keys to be used in various cryptosystems. However, in the event of compromised keys, the user has to change his biometrics, which is not feasible for biometrics like face, iris, fingerprint and even hand signatures.

Later on, [8] came back in their research using BioPhasor mixing and 2^N discretization on dynamic hand signatures compute biometric hash. This process offers a one-way transformation that prevents exact recovery of the biometric vector from compromised biometric hashes and stolen tokens. The 2^N discretization also performs both as an error correction step as well as a real-to-binary space converter.

[12] explores the extraction of a reproducible bit string referred to as biometric key from biometric data using signature. Claiming that biometric key generation would be easier to use without auxiliary data they took up the challenge of generating a biometric key without the use of any auxiliary data. It is therefore deduced that storing biometric template locally increases the risk of stealing of the biometric data. Instead of storing the original biometric signal in the system database, only its transformed version is stored. Likewise, if feature extraction is parameter-based as we have in [7], the feature extraction is limited and short, and could not support use in cryptographic systems as they are small in key space. Also, the keys are not cancelable and more importantly, they are generally low in entropy. They are also not secure due to storage of user-specific statistical boundaries that could be used to recover the biometric features.

Furthermore, [13]'s method only considers static signature features, and the robustness against change of pixel numbers in particular allows the reconstruction of the biometric key from a printed signature image. Lastly, instead of trying to find a single unique feature, biometric key needs to find only a collection of rather unique features or parameters that when assembled collectively create a unique profile for an individual. This calls for injection of random numbers into the key because biohashes exposes the statistic information about biometric feature, which can be used to estimate the original feature. These motivated the research.

3 Proposed Scheme

This paper proposes a method for biometric key computation and generation from handwriting biometric. A schematic representation or proposed biometric key generation method from handwriting biometrics is shown in Fig. 1.

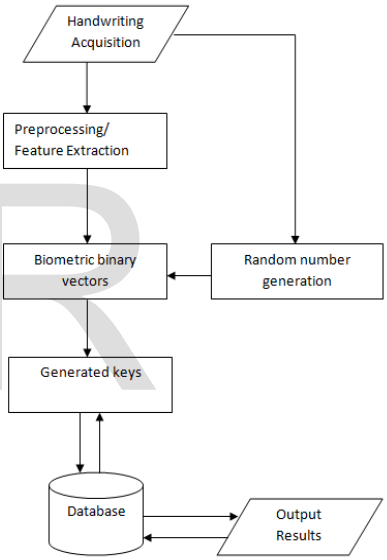

Figure 1 : *The Design Architecture for key generation from handwriting biometrics*

Step I : A pressure-sensitive pen and tablet personal computer was used for capturing the online signature signals in terms of pressure information (p), pen altitude (a), and azimuth (az) for each point.

Step II

The pen-down segments (detected as points between two pen downs) are concatenated to form one single signal and each signal is then compressed with the Discrete Wavelength Transform – Discrete Fourier Transform (DWT-DFT) method as described method as described in equation (1). Each dynamic handwriting signal can be modeled as function $f(t) \in L^2(R)$

IJSER

$$f(t) = \sum_{j=1}^{L} \sum_{k=-\infty}^{\infty} d(j,k)\psi\,(2^{-j}t - k) + \sum_{k=-\infty}^{\infty} a(L,k)\phi(2^{-L}t - k) \quad (1)$$

where j and k are integers and t =time functions

$\psi(t)$ = the mother wavelet at level L

$\phi(t)$ = scaling function

From *(1)*, the wavelet decomposition at any level L, $f_L(t)$ can be obtained from approximation coefficient $a(L, k)$ and layers of detail coefficients $\{d(j, k) \mid j \le L\}$

Each compressed wavelet $F(t)$ = compress ($f_2(t)$) can then be represented by a Fourier integral of form as

$$g(w) = \int_{-\infty}^{\infty} F(t)\, e^{-jwt}\, dt \quad (2)$$

The DFT is performed using FFT and the resulting $g(w)$ is then normalized via division by $\sqrt{\sum g_i^2}$ so that $|g| = 1$.

Step III

The objective of mixing random numbers as described by [12] is to inject randomness into this biometric features by using tokenised Pseudo Random Numbers (PRN) with the handwriting feature vector that was derived. In addition to this it enables the changeability of this biometric features by altering the external secrete r_{ij}. jth user formulation is given as

$$\alpha_{ij} = \frac{1}{n}\sum_{k-1}^{n} \tan^{-1}\left(\frac{s_j^k}{r_{ik}^j}\right), \; i = 1, \quad (3)$$

r_{ik} = *PRN* independently drawn from *N(0,1)*

The output is a set of m phasor values $\{\alpha_i \mid i=1,\ldots,m\}$ with range $[-\pi,\pi]$ where m can be set either equal to or smaller that the original biometric feature lenght, n.

The outline of the mixing is as follows :

(i) Random Numbers T are generated.

(ii) To compute the random basis, generate $m < n$ number of random vectors $t_i \in RnR$ with subscript R denoting that the number is generated randomly using T as the seed, n as the length of the biometric feature, and an integer m. Then, orthonormalize $\forall\, ti$ using the Gram-Schmidt method.

(iii) Compute $h_i = [\sum_{j=1}^{n} \arctan((b_j)^q / t_{i,j})] / n$ (4)

where $q \in Z$ for $i = 1, \ldots, m$. The parameter q tunes the magnitude of the biometric feature element .

Since $\arctan(x)+\arctan(x^{-1}) = \pi/2$, the projected vector can be rewritten as

$$h_i = [\sum_{j=1}^{n} \frac{\pi}{2} - \arctan((ti, j)/((b_j)^q))] / n)]/n \quad (5)$$

with $q = 2$, which has a more complicated transformation than random projection using iterative inner product used in earlier work [13]. In particular, the effect is a one-to- one arctan transformation of the random projection of the inverse of biometric vector b onto bounded range of $(-\pi/2, \pi/2)$, followed by reflection of the arctan projected space along the x-axis and displacement of $\pi/2$.

4 Experiment Process

The experimental data was captured from 10 users in three sessions by a HP Tablet PC. In other to refrain from capturing signatures which is the traditional way of evaluating handwriting since it has been shown by [15] that the usage of such alternative contents may lead to similar results as the usage of the signature in context of online handwriting based authentication performance. Therefore, we asked each user to provide 3 samples of 3 different semantics: Telephone number, pseudonym and an answer to the question "Your favourite colour" A pseudonym is a name that a person or group assumes for a particular purpose, which differs from his or her original or true name (orthonym).Pseudonyms include stage names, screen names, pet names, nicknames, aliases, gamer identifications, and reign names of emperors, popes and other monarchs.

The samples collected from these sessions were enrolled into the database and they form the original user samples. Impostor's handwriting were also captured from and stored in the database for the purpose of this experiment. An attempt of one user to be verified as another user is considered as an impostor trial. For the experiment proper, the performance measure was done in two ways namely : the percentage comparison and matching distance analysis of the biometric keys generated which compare the matching rate of the biometric keys generated. Authentication performance of a biometric system cannot be measured directly, it has to be determined empirically. Although the final accept/reject decision of the system is based on the comparison between the biometric keys which highly simplifies the experiments. In essence, hamming distance between the semantics of the same writer is calculated. Likewise, hamming distance between the impostor's semantics and the original writer's semantics.

In equation (6), x and y represents the biometric hash vectors of dimension k to be compared, and x_i and y_i are the corresponding elements of x and y at index i. The direct comparison of x_i and y_i is *0* if the two elements are equal and *1* else. The Hamming Distance between the hashes x and y is the sum of the results of all single comparisons [7]. The percentage comparison between the keys generated by the same writer was also calculated and also the keys between the impostor's and the original writer.

In this context, the Hamming Distance measure determines the number of positions, where two biometric keys are different and returns a value between *0* and the number of elements.

$$hd(x,y) = \sum_{i=0}^{k-i} dist(xi, yi) \quad dist(x_i, y_i) \qquad (6)$$

$$\text{with } dist(x_i, y_i) = \begin{cases} 0 \; if \; xi{=}yi \\ 1 \; else \end{cases}$$

This is done with the form as shown in figure 2

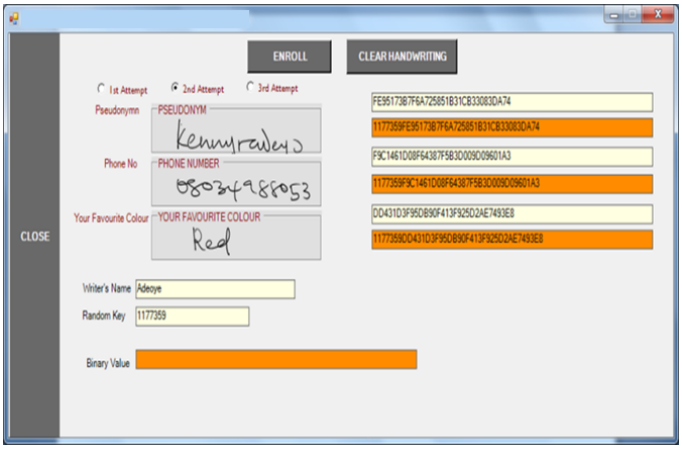

Figure 2: Input panel for the handwriting

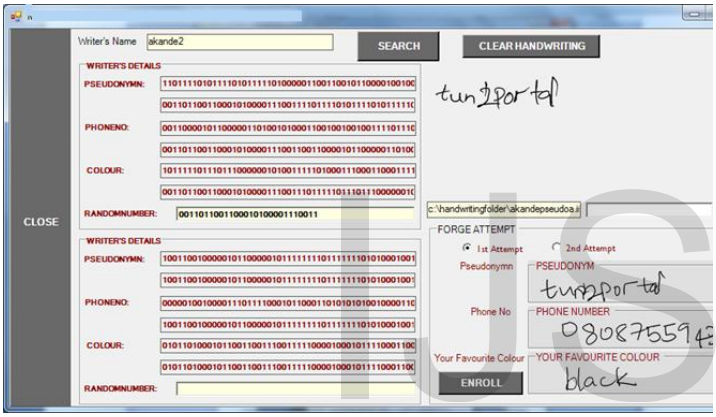

Figure 3: Input panel for the forger's handwriting

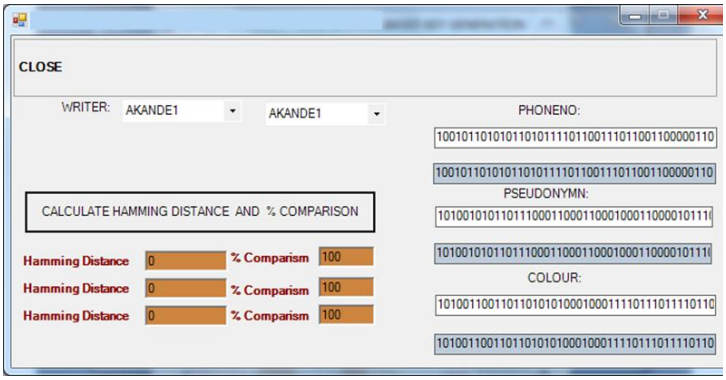

Figure 4: Window form that implement hamming distance and percentage comparison of the handwriting semantics .

5 Experimental Results

This subsection presents the results for the hamming distance and percentage comparison calculations. The corresponding tests are carried out on every semantics of each writer covering both the interclass and intraclass. The keys generated are in binary form of 128 bits.

Table 1: Intra Class Percentage Comparison in % per semanti c class

Writer	Semantics		
	Pseudonym (%)	Phone Number (%)	Colour (%)
U1	67	61	62
U2	55	59	75
U3	66	54	64
U4	65	56	70
U5	66	69	55
U6	58	58	64
U7	68	63	68
U8	75	70	72
U9	62	60	56
U10	62	60	56

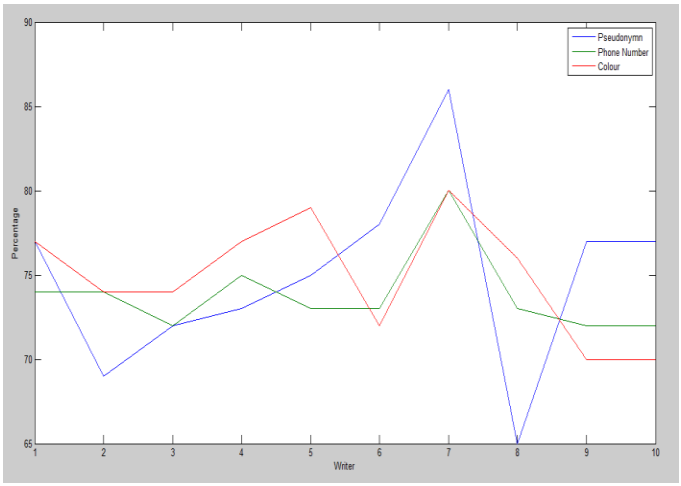

Figure 5: Percentage comparisons of writer's two successive handwriting semantics

Writer	Percentage Comparison			Hamming Distance		
	Pseudonym (%)	Phone Number (%)	Colour (%)	Pseudonym	Phone Number	Colour
U1	73	80	68	54	77	62
U2	72	75	73	70	72	67
U3	74	70	69	76	67	76
U4	71	74	74	71	66	57
U5	80	77	71	69	59	61
U6	74	72	77	62	69	60
U7	73	80	78	63	65	61
U8	68	78	72	62	64	64
U9	68	68	77	64	50	60
U10	73	80	68	54	77	62

Table 2: Percentage comparison and hamming distance of forgers forging the handwriting semantics

In the experimental evaluation, we have practically shown the possibility of generating biometric keys from online handwriting biometrics. The semantics used revealed that for phone numbers forgers made a huge success of average of 79.023%. While pseudonym give in to like 76.578% average. The colour semantic is the lowest at 55.876%. On one side, the hamming distance shows a better method for this evaluation of the key generated from handwriting semantics. Pseudonym has the average of 64.5 while the Phone number semantic recorded 66.6 average and colour 63. This affirms that the strength of these keys can be maximized by the multi-semantic key generation of the semantics which is the simple concatenation of two biometric keys based on different semantics.

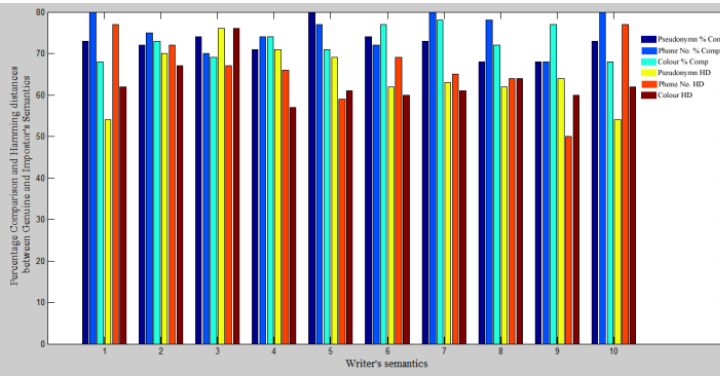

Figure 5: Percentage comparison and Hamming distances of genuine and forged semantics

6 Conclusion

A handwriting based key generation system has been designed. The system is designed with user friendly interface that enrolls the handwriting semantics of three types. The supplied biometric feature was enrolled and were not stored directly so as to avoid being stolen. The performance has been evaluated using the system designed with various handwriting styles and semantics supplied to generate keys. Random numbers were also generated and injected into the generated keys to further strengthen the generated keys. The generated keys are tested to confirm percentage comparison for semantics from the same user. Also, forgers were made to forge the handwriting and the generated keys were compared for both hamming distance and percentage comparison. The security strength of this scheme lies in the fact that the transformation from real-valued biometric feature to index space and finally to binary bit strings, which can be seen as a form of error correction to compensate for noisy biometric data as well as lossy compression. Also, the irreversible extraction of the biometric information from the supplied semantics.

IJSER

References

[1] T. Dunstone and N. Yager. Biometric System and Data Analysis Design, Evaluation, and Data Mining. Springer, 2009.

[2] A.B.J. Teoh and W.K. Yip. Secure Dynamic Signature-Crypto Key Generation. In L. Wang and X. Geng, editors, Behavioural Biometrics For Human Identification: Intelligent Applications. IGI Global, 2010.

[3] Dipti Verma and Ankit Arora, Preprocessing and Feature Extraction Method for Static Signature Recognition Using Hough Transform and Other Important Parameters International Journal of Computer, Electronics & Electrical Engineering (ISSN: 2249 - 9997) Volume 1 – No.1. 2011

[4] Beng, A., Yonsei, Jin Teoh,Kar-Ann Toh (2008) Secure biometric-key generation with biometric helper, Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on June 2008

[5] Lucas B., Fabian M. and Daniel L. Biometric authentication revisited: understanding the impact of wolves in sheep's clothing,in USENIX-SS'06 Proceedings of the 15th conference on USENIX Security Symposium - Volume 15 Article No. 3. 2006

[6] Freire, M. R., Fierrez, J., Galbally J. and Javier Ortega-Garcia J. "Biometric Hashing Based on Genetic Selection and Its Application to On-Line Signatures". 2007

[7] Vielhauer, C., Biometric User Authentication for IT Security: From Fundamentals to Handwriting, Springer, New York. 2006

[8] Yip W. K., Teoh B.J, and Ngo C. L. Secure Hashing of Dynamic Hand Signatures Using Wavelet-Fourier Compression with BioPhasor Mixing and 2^N Discretization,Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume 2007, Article ID 59125. 2007

[9] Goh, A and Ngo, C.L, Computation of Cryptographic Keys from Face Biometrics, 7th IFIP CMS 2003, Torino, Springer-Verlag LNCS2828. 2003

[10] Chan, C.W. and Hao, F., 'Private Key Generation from On-Line Handwritten Signatures' in Information Management & Computer Security, Vol. 10, No. 2, pp. 159-164, 2002.

[11] Schomaker. L Writer identification and verification In N. Ratha and V. Govindaraju, editors, Advances in Biometrics: Sensors, Systems and Algorithms, pages 247 {264. Springer-Verlag}. 2007

[12] Dirk Scheuermann, Bastian Wolfgruber, and Olaf Henniger, On biometric key generation from handwritten signatures. BIOSIG, Vol. 191GI, p. 103-114. 2011

[13] Yip W. K., Goh, A., Ngo, D. and Teoh, A. Cryptographic keys from dynamic hand-signatures with biometric security preservation and replaceability. In Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pages 27–32, Los Alamitos, CA, 2005. IEEE Computer Society. 2005

[14] C.F. Colmas: The Writing Systems of the World. Blackwell, 1980.

[15] Kenny, C. (2005). Random number generators: An evaluation and comparison of random.org and some commonly used generators. Retrieved 10/16/2011 from http://www.random.org/analysis/Analysis2005.pdf

[16] Marsaglia,G. (2005). Random number generation. Retrieved 10/16/2011 fromdl.acm.org/ft_gateway.cfm?id=1074752&ftid=634693&dwn=1&CFID=6016096 5&CFTOKEN=82346114

IJSER