Biometric Security with Hashing

Aron Chan
University of Calgary
Calgary, AB T2N 1N4
archan@ucalgary.ca

Charlie Cheung
University of Calgary
Calgary, AB T2N 1N4
ccheu@ucalgary.ca

Jiashan Li
University of Calgary
Calgary, AB T2N 1N4
jiashan.li@ucalgary.ca

Jie Hu
University of Calgary
Calgary, AB T2N 1N4
jiehu@ucalgary.ca

ABSTRACT

This literature study encompasses the conclusions drawn by our research of biometric security and the hashing of biometric functions. The use of biometrics in information security has allowed humans to protect and authenticate the current technology applications we have in use today whether in daily life or in high security databases. As the technology advances, so have the technologies created to capitalize on their vulnerabilities. In order to protect biometric identification, hashing functions were created and implemented. This report will introduce the main types of biometric security currently used to authenticate and protect users. Then, we will address their current vulnerabilities and how hashing can effectively improve their security. For this report we will focus on the current method of hashing fingerprints as a basis of the usage of hashing to improve security. We will continue by reviewing the current limitations of biometric security in its respective disciplines and argue the beneficial uses of hashing for authentication and access control with regards of iris recognition and facial recognition.

CCS Concepts

E.m [**Data**]: Miscellaneous – *biometrics, security, hashing.*

Keywords

Authentication; Biometrics; Hashing; Security; Privacy; Fingerprints; Voice recognition; Iris recognition

1. INTRODUCTION

When dealing with user authentication, while traditional password based authentication systems and the use of physical tokens can be considered secure enough, they run into their own set of issues that may impact its overall security. People can forget their passwords, or an attacker could guess a user's credentials. Tokens are susceptible to loss or theft, and with these potential problems in mind the use of biometrics is considered as an alternative method for identification and authorization.

Within the research community, the application of biometrics to authentication systems has become widespread. With the proposal to use biometric data for identification and authentication systems, there are privacy concerns because compromised biometric templates cannot be revoked and reissued. A person cannot be given a new set of fingerprints the same way they could with a new password, and research papers investigate into how to acquire these biometrics safely while preserving performance.

To deal with this issue, researchers have looked for ways to prevent a template database of biometric data from being compromised as well as developing cancellable biometrics, which allow biometric templates to be cancelled or replaced. Our paper will cover hashing methods of fingerprints, voice recognition, and iris recognition to implement a one-way transformation of cancellable biometrics. After biometric data becomes hashed via a mathematical function, it becomes computationally infeasible to recover, making it difficult for an attacker to obtain the original biometric data from its template. Through the efforts of these studies, biometric based authentication can perform effectively while preserving the privacy of its sensitive data.

In regards to fingerprinting, a symmetric hash function can be implemented which hashes fingerprint minutiae without requiring pre-alignment of the fingerprint. Another hashing algorithm, BioHashing, has been successful in further reducing the error rates of authentication while preserving the features of symmetric hashing.

Other popular uses of biometric security involve the application of iris recognition and facial recognition. Recent development in biometric security has increased the efficiency in which facial recognition can be used as a means of security. This has been accomplished through the use of pattern recognition and computer vision communities. As machine learning develops and technology in the computer graphics vision communities continue to grow, so has the use of facial recognition in biometric security.

The use of iris recognition involves an automated method of biometric identification. The iris is so distinct that the use of photographs as identification has been proposed over the use of fingerprints. The complex patterns of an individual's eyes are complex and unique as well as stable, much unlike a person's facial features that can change dramatically with age. Patterns found in a person's iris are encoded into a template to allow for a person's identification and authentication. Due to its uniqueness, iris recognition has risen as a prominently used method of biometric security.

This report will also explore the applications of facial recognition for biometric security. We will review the current implementation methods of facial recognition and discuss their advantages and disadvantages. This report will introduce the benefits of hashing facial recognition to further enhance security as compared to the use of hashing in relation to other methods of biometric security.

The report will continue with this approach for iris recognition. After a thorough review of the advantages and disadvantages of iris recognition, a review of biometric hashing used over iris images will be done. This report will also propose how biohashing can benefit iris recognition so that it is less susceptible to attacks that target its exploitable vulnerabilities.

Finally, this report will conclude by exploring the possibilities of future works in biometric security. The future works component of this report will heavily involve the improvements that hashing algorithms propose to further improve security and perhaps generalize them to work with any biometric. As we conclude this report, we hope that the use of biometric security will continue to

grow and develop as technology advances to improve the trust users can place in biometrics.

2. FACIAL RECOGNITION

In recent years, facial recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities. The machine learning and computer graphics communities are also increasingly involved in facial recognition. This common interest among researchers working in diverse fields is motivated by our remarkable ability to recognize people and the fact that human activity is a primary concern both in everyday life and in cyberspace.

2.1 Predominant Approaches

There are two predominant approaches to the facial recognition problem: geometric (feature based) and photometric (view based) [5]. As researcher interest in face recognition continues, a variety of different algorithms were developed, three of which have been well studied in facial recognition literature will be introduced in the following subsections.

2.1.1 Principal Components Analysis (PCA)

Principal Components Analysis or 'PCA' was invented in 1901 by Karl Pearson and is commonly referred to as the use of eigenfaces. The eigenface algorithm uses the Principal Component Analysis (PCA) for dimensionality reduction to find the vectors which best account for the distribution of facial images within the entire image space. PCA computes the basis of a space which is represented by its training vectors. These basis vectors, otherwise known as eigenvectors, computed by PCA, are in the direction of the largest variance of the training vectors. When a particular face is projected onto the face space, its vector into the face space describe the importance of each of those features on the face. The face is expressed in the face space by its eigenface coefficients (or weights). We can handle a large input vector, facial image, only by taking its small weight vector in the face space. To identify a test image, it requires the projection of the test image onto the face space to obtain the corresponding set of weights. By comparing the weights of the test image with the set of weights of the faces in the training set, the face in the test image can be identified.

To create a set of eigenfaces, one must prepare a training set of face images [14]. These images must be taken under the same lighting conditions and be normalized to have the eyes and mouths aligned across all images. Each image should be treated as a vector by a simple concatenation of the rows of pixels in the original image. Next, the mean is to be subtracted and the average image has to be calculated and then subtracted from each original image. The eigenvectors and eigenvalues can then be calculated into a matrix called the eigenfaces. The principals components are chosen from these images and sorted so that in the future, these eigenfaces can be used to represent both existing and new faces.

In most applications of eigenfaces, faces can be identified using a projection between 100 and 150 eigenfaces.

2.1.2 Linear Discriminant Analysis (LDA)

LDA is a statistical approach for classifying samples of unknown classes based on training samples with known classes [15]. This technique aims to maximize between-class (i.e., across users) variance and minimize within-class (i.e., within user) variance. The fisherfaces are faster than eigenfaces, in some cases and have lower error rates [13]. It works well even if there are different illumination and facial expressions. When dealing with high dimensional face data, this technique faces the small sample size problem that arises where there are a small number of available training samples compared to the dimensionality of the sample space. This algorithm works best when the goal of the system is classification rather than representation.

2.1.3 Elastic Bunch Graph Matching (EBGM)

EBGM is an algorithm in computer vision for recognizing objects or object classes in an image based on a graph representation extracted from other images [4]. EBGM relies on the concept that real face images have many nonlinear characteristics that are not addressed by the linear analysis methods discussed earlier, such as variations in illumination (outdoor lighting vs. indoor fluorescents), pose (standing straight vs. leaning over), and expression (smile vs. frown).

A Gabor wavelet transform creates a dynamic link architecture that projects the face onto an elastic grid. The Gabor jet is a node on the elastic grid, notated by circles on the image below, which describes the image behavior around a given pixel. It is the result of a convolution of the image with a Gabor filter, which is used to detect shapes and to extract features using image processing. Recognition is based on the similarity of the Gabor filter response at each Gabor node. This biologically-based method using Gabor filters is a process executed in the visual cortex of higher functioning mammals. The difficulty with this method is the requirement of accurate landmark localization, which can sometimes be achieved by combining PCA and LDA.

2.2 Current Applications

Current uses of facial recognition technology includes law enforcement agencies using this software as a crime fighting tool. The use of facial recognition has been issued to scan driver's licenses to prevent identity fraud. The same technology has also be integrated to scan faces in CCTV footage to identify persons of interest. Another popular application involves the usage of facial recognition in border control deployments to verify the identities of travellers.

The software for facial recognition has been used to authenticate users to their bank accounts so they are able to complete actions such as online banking and payment processing.

Facial recognition has also proven useful to prevent voter fraud. In cases where some individuals have been registering to vote under several different name in an attempt to place multiple votes, facial recognition software was used to compare new facial images to those already in the voter database. Authorities were able to reduce duplicate registrations using this authentication method.

Systems created using facial recognition have been developed to unlock software on mobile devices. Applications take advantage of the phone's built-in camera to take a picture of the user. Facial recognition is used to ensure only this person can use certain applications which they choose to secure.

2.3 Strengths and Weaknesses

The following section will review the strengths and weakness of face recognition biometric security.

2.3.1 Strengths

One of the strengths of facial recognition is having no more time fraud. It will be impossible for buddy punching to occur, since

everyone has to have go through face scanning biometrics devices to clock in.

Another advantage is an increase to security. A user will enjoy better security with a facial recognition biometrics system. Not only can a system track employees through biometric time attendance tracking, but any visitors can be added to the system and tracked throughout the area as well. Anyone that is not in the system will not be given access.

Having an automated facial system is another strength of this type of biometric security. Many companies can benefit from the fact that biometric imaging systems are automated. Therefore, managers won't have to worry about having someone there to monitor the system 24 hours a day manually.

Facial recognition software is also easily integratable. Integrated biometric facial systems are easy to program into a company's current computer system. Usually they will work with existing software that is already in place resulting in easy setup and quicker up time.

Finally, facial recognition biometric security benefits from having a high success rate, making it a reliable method for securing information. It is extremely difficult to fool the system, so you can feel reasonably secure knowing that your biometrics computer security system will be successful at tracking time and attendance while providing better security when identifying and authenticating users.

2.3.1 Weaknesses

Current facial recognition still often misidentifies people which can sometimes led to controversy. One of the most prominent and controversial examples of misidentification occurred when Google was criticized for racism in it's system when a black couple were misidentified as gorillas. Facial recognition software generally doesn't do as well in identifying minorities. Accurate facial recognition can also be compromised by poor lighting, sunglasses, hats, scarves, beards, long hair, makeup or other objects partially covering the subject's face, and low resolution images. Another serious disadvantage is that many systems are less effective if facial expressions vary. Canada now allows only neutral facial expressions in passport photos. Finally, though facial recognition has many of the benefits of other biometric security methods, the human face is frequently exposed to the environment and changes much more dramatically than a person's iris or fingerprint. These changes can contribute to errors in the system.

2.4 Hashing Algorithm

A face recognition system namely focuses on facial identification and authentication. The system must be able to determine whether an input image matches a person's data in the database. The idea of face hashing has been extracted from early graphical feature mapping techniques. A proposed method of doing so is by calculating numerical values of a face image for fast recognition [11]. These values should be unique for each face in a database. The calculated hash value of the face is then saved into a database with a corresponding face ID. This will aid in modifying a binary search to recognize the faces from the database. To further decrease the searching time, the database should contain the hash values separated into different files corresponding to the first character of the hash value. Therefore, the entire process will consist of two phases: database generation phase and binary search for face recognition phase.

The person's face will be segmented into four partitions and their centroid values will be calculated. The sum of these four centroid values is calculated and considered as the face hash value. This value will be saved into the database corresponding to the first digit of hash value. Each entry will consist of six fields: four for the centroid of clusters (the hash value), one field for the sum of centroids, and one for the picture ID to keep track of pictures in the database.

As for the face recognition process, the proposed hash and database will help speed up the recognition rate. The system will take a segmented face region as an input image then convert the image to grayscale and segment it to extract the face region. The image is then normalized and scanned to find the centroid values. The sum of the centroid values will result in the face hash value and find the first digit of this hash value. Once a match is found, return that image and grant them access to the system. Otherwise, if the input image does not match with any image in the database then deny access.

Not only does this hashing technique promote security, it also promotes fast face recognition that can be helpful if the database is large. The original data is reduced to a few values that can be easily compared. This in turn will make a hash value that reduces the searching time efficiently and maximizes the recognition rate.

3. IRIS RECOGNITION

Iris recognition is an automated method of biometric identification. It uses mathematical pattern-recognition techniques on video images of an individual's eyes [1]. Those patterns are complex and unique, as well as stable and can be seen from some distance. Iris recognition uses video technology with a subtle underlying of near infrared illumination to acquire images of the intricate structures of the iris which can be seen externally. These patterns are encoded into a template through the use of algorithms and allow for a person's identification. These templates have a remarkably low false match rate. However, these templates need to be stored in the hundreds of millions scale when used in places such as passport-free automated border-crossings at airports. The iris is so distinct that the use of photographs as identification has been proposed over the use of fingerprints. Unlike other biometrics, iris recognition comes from randomly distributed features. This leads to its high reliability for personal identification. The iris images are typically stored in a database where they can be matched with a user in need of identification or verification.

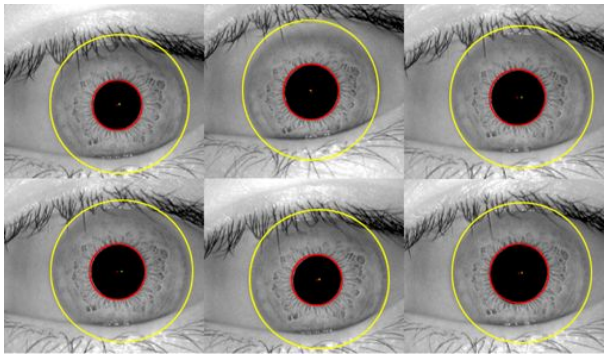

Figure 1: Sample of iris images taken from an iris database.

3.1 Current Applications

Iris recognition systems are typically deployed by acquiring images of an iris illuminated with the light by the near infrared wavelength and of the electromagnetic spectrum. The majority of people have dark brown eyes which appear richly structured in the NIR band. For identification or verification, a template created by imaging an iris is compared to stored templates in a database. This presents an interesting approach where iris patterns can be made reliable visual recognition utilities of persons when imaging can be done at distances of less than a meter. This is especially true when there is a need to search very large databases without having any false matches despite a large number of possibilities. There are many advantages of iris recognition that all illustrate the usefulness it has in regards to biometric security.
The following figure is the current approach in iris image processing.

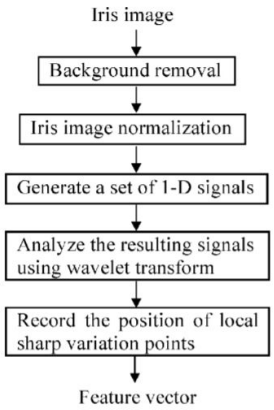

Figure 2: Diagram to iris image processing approach.

The following subsections will introduce some of the new advancements made to improve the usability of iris recognition.

3.1.1 Active Contours

Iris recognition begins with finding an iris in an image and omitted its inner and outer boundaries as well as excluding any eyelashes or reflections from the cornea or eyeglasses. Collectively, these processes can be called segmentation [2]. The active contours enhance iris segmentation because they allow for non-circular boundaries and enables flexible coordinate systems. The active contour models for the inner and outer iris boundaries support an isometric mapping of the iris tissue between them, regardless of the actual shapes of the contours. Since iris boundaries might have poor or unclear edges, active contours can increase both the segmentation and recognition accuracy.

3.1.2 Off-axis Gaze

One of the current limitations of iris recognition cameras is that they require an on-axis image of an eye. This is usually achieved through something known as a "stop and state" interface [2]. This is when a user must align their optical axis with the camera's optical axis. Sometimes, while using this method, the standard cameras acquire images for which the on-axis assumption is inaccurate. To combat this limitation, gaze estimation has been designed so that an iris can still be recognized off gaze. This is an important step to correct an off-angle iris image. The eye gaze direction must be estimated in absence of other information such as a well-controlled light source and multiple cameras.

3.2 Advantages

Advantages of the iris include how it is an internal organ, well protected against damage and wear. This is an advantage over fingerprints which could be difficult to recognize after years of certain manual labourers. The iris is also mostly flat and controlled by two muscles making it more predictable than the face with its many angles and planes. The chance of issuing a false positive is also extremely low even though there is virtually no way to prove iris's to be unique. It is also not very intrusive as the scan can be taken from 10cm to a few meters away. They can also be done through eyeglasses, non-mirrored sunglasses, and clear contact lenses. Finally, the ease of localizing eyes in faces and the distinctive shape of the iris, allows for a reliable and precise isolation of this feature and the creation of a size-invariant representation. This also presents a great mathematical advantage since the iris' pattern variability among individuals is enormous. The complex pattern contains distinctive features such as arching ligaments, furrows, ridges, crypts, rings, corona, freckles, and a zigzag collarette.

3.3 Disadvantages

Commercial iris scanners can be easily fooled by a high quality image of the iris in place of the real thing. It is also still a relatively new technology and incompatible or simply too expensive in places it can see purpose. The iris is a small target for a system to use as an identification method and it also requires the user to remain perfectly still. This could prove difficult if used with young children. The scan can also be made difficult but eyelashes and all things that could cause a reflection.

3.4 Iris Biometric Hash Generation

Biometric data is hashed so that the calculated hash points to a location in the database where a match can be found. To apply a hash to iris recognition, the approach involves an application of the biometric hashing for the database indexing to locate biometric templates. This is done so that no complex sorting of templates is required due to the use of low-dimensional hashes. These are generally generated out of biometric data. To create a biometric hash of the iris, parts of the ring are discarded during preprocessing since these parts are often affected by the eyelids or eyelashes. The hash will be generated from the remaining parts and into a texture. To do so, the remaining parts of the preprocessed texture image are employed to extract a hash for the database indexing.

Several requirements regarding the calculated hashes need to be fulfilled such as collision freeness and efficient computation [10]. A image is processed of the iris and $x * y$ pixel blocks located in the upper half of the iris texture are analyzed. This is because the

upper half of the iris proves to be more consistent than the outer band. When handling the problem of variance, a number of pixel blocks need to be processed. A number of *n* pixel blocks in a set *M* of mean values, defined by $M = \{m_1, m_2,...,m_i\}$ is extracted. Therefore, each mean value m_i is compared against a predefined threshold *k* to generate a binary hash $B = \{b_1, b_2, ..., b_i\}$ such that,

$$b_i = \begin{cases} 1, & \text{if } m_i \geq k, \\ 0, & \text{if } m_i < k. \end{cases} \forall i = 1...n$$

The process of hash generation is computationally efficient and offers several advantages. To begin with, the proposed hashing generation extracts similar hashes for similar biometric data. This will fulfill the requirement of a hash used for coarse level database indexing. These hashes are scalable with respect to length by varying block dimensions and the number of applied blocks.

Hashing generation is applied during template generation representing a single-sensor scenario. Each user has a unique biometric template as well as a hash generated for them. An example of this hash generation for irises was carried out comprising of the iris images of 250 persons [7]. The following chart shows the average matching for different block dimensions and hash lengths.

Block Dim.	Hash Length	Avg. Matchings
8 × 8	32 bit	13.211
12 × 8	32 bit	10.555
16 × 8	32 bit	7.695
16 × 12	32 bit	9.671
32 × 4	32 bit	9.344

Figure 3: Average matchings for different block dimensions and hash lengths.

The evaluation of this method assumed that the underlying iris recognition algorithm worked perfectly. This is to reveal the actual performance gain. The above table summarizes the average number of matchings for several parameters which had to be performed until successful authentication was achieved. Based on the values above, the best result occurred with block dimensions 16x8 with a hash length of 32 bits because it gives the lowest average of matches.

The use of hashing proves to greatly improve the efficiency of authentication. Based on the applied block dimensions, an appropriate number of pixel block rows are chosen to achieve the according hash length. This allows for a system to have more accurate matches, increased security, and operate at a greater efficiency.

4. FINGERPRINT

Biometrics are automated methods of recognizing a person based on physiological or behavioral characteristic. Securing a biometric template is vital part in the successful implementation of biometrics based authentication systems. In this paragraph, the new method of how to hash and secure the fingerprint template would be introduced.

As we know, a hash function H is a transformation that takes an input m and returns a value h (called the hash value); h = H(m). it is also called one-way function if it is hard to invert. For example, it is hard to find a message M'≠ M such that H(M') = H(M). Also, biometric matching is performed using hashed features instead of the original template, which can prevent the potential attacks; Similar to the attack on the password authentication, such as eavesdropping attacks or attacks on the transmission over a network.

4.1 Challenges

In case the biometric data is hashed, even a slight change in the acquisition of the biometric can lead to a totally different hash value. Thus, it may not match the original within the same matching threshold as that for the straight unhashed scenario. When securing a fingerprint, the following four requirements have to be solved.

· Similar fingerprints should have similar hash values
· Different fingerprints should not have similar hashes
. Rotation and translation of the original template should not have a big impact on hash values
· Partial fingerprints should be matched if sufficient minutiae are present

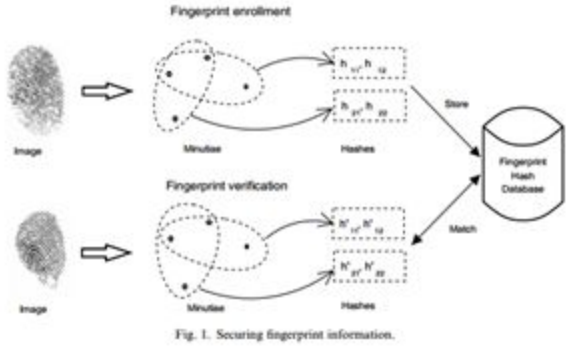

Figure 4: Securing fingerprint information

4.1 Minutiae based matching

In fingerprint based biometric authentication systems, minutiae based matching has become a De facto standard. A fingerprint consists of a series of ridges and furrows on the surface of the finger. The earlier securing method is to check whether two prints are aligned, and the number of matching minutiae points are dominated as an important role to verify the subject. The new method focuses on 'localized matchings' into the fingerprint recognition algorithm to avoid global alignment. Localized matchings consists of matching minutia triplets, minutia angles at a ridge and bifurcation (x, y, Θ). A secondary feature vector $S_1 = \{r_1 0, r_1 1, \Theta_1 0, \Theta_1 1, \varphi\}$ is generated based on the Euclidean distances and orientation difference between the central minutia and its nearest neighbors.

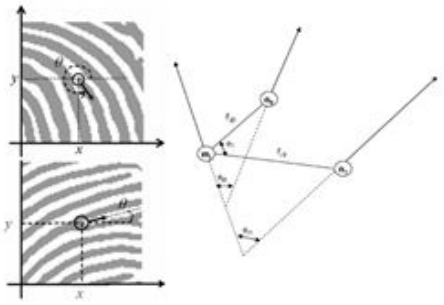

Figure 5: Left: Minutia angles at a ridge and bifurcation.

Right: Secondary features based on the minutia m_i and its nearest neighbors n_0 and n_1.

4.2 Symmetric hash functions

We use symmetric hash functions because of the case if the order of inputs changes and then it will influence the result. For example, $H(X) = k_1x_1 + k_2x_2 + \ldots\ldots + k_nx_n$, $k_1 \neq k_2 \neq \ldots. \neq k_n$.
In this hash function, the sequence of k is fixed and it will change the hash value if the order of a series of inputs x is changed [12].

$H^m_{sym}(x) = x^m_1 + x^m_2 + \ldots\ldots + x^m_n$
In this hash function, changing the order of input won't bring the change to the result, and this type of symmetric hash functions can be used for hashing the fingerprint minutiae.

4.3 Hashing functions of minutiae points

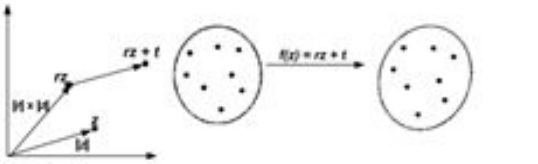

Fig. 3. Minutiae as represented in the complex plane. *f* represents the accidental shifting of minutia points.

Figure 6: Minutiae as represented in the complex plane. f represents the accidental shifting of minutia points.

We want to compare the fingerprint with the original one and there are three factors, different position, rotation and scale, coming from different scanners and different positioning of the finger on the scanner. We use $\{c_i\}$ to represent minutia points, the transformation of one fingerprint to another can be presented by the complex function $f(z) = rz + t$ where r and t represent the scalar rotation and translation parameters of the accidental shift of points under the registration and authentication scans.

Given n minutia points $\{c_1, c_2, \ldots, c_n\}$ and take them into the following m symmetric hash functions:

$H_1(c_1, c_2, \ldots, c_n) = c_1 + c_2 + \ldots. + c_n$
$H_2(c_1, c_2, \ldots c_n) = c_1^2 + c_2^2 + \ldots. + c_n^2$
…
$H_m(c_1, c_2, \ldots c_n) = c_1^m + c_2^m + \ldots. + c_n^m$

Suppose another image of the fingerprint is obtained using the above described transformation $f(z) = rz + t$. locations of the corresponding minutia points are $c_i' = f(c_i) = rc_i + t$. Hash functions of the transformed minutia can be presented as:

$H_1(c_1', c_2', \ldots, c_n')$
$= c_1' + c_2' + \ldots + c_n'$
$= (rc_1 + t) + (rc_2 + t) + \ldots + (rc_n + t)$
$= rh_1(c_1, c_2, \ldots, c_n) + nt$

$H_2(c_1', c_2', \ldots, c_n')$
$= c_1'^2, c_2'^2, \ldots, c_n'^2$
$= (rc_1 + t)^2 + (rc_2 + t)^2 + \ldots + (rc_n + t)^2$
$= r^2h_2(c_1, c_2, \ldots, c_n) + 2rh_1(c_1, c_2, \ldots, c_n) + nt^2$
…

Let $h_i = h_i(c_1, c_2, \ldots, c_n)$ be the hash value of the one fingerprint and let $h_i' = h_i(c_1', c_2', \ldots, c_n')$ be the another fingerprint. The new equation can be written as:

$H_1' = rh_1 + nt$

$H_2' = r^2h_2 + 2rth_1 + nt^2$

$H_3' = r^3h_3 + 3r^2th_2 + 3rt^2h_1 + nt^3$

…

if we add the errors into consideration, the relation between the hash values of the enrolled fingerprint $\{h_1, \ldots, h_m\}$ and the hash values of the test fingerprint $\{h_1', \ldots, h_m'\}$ can be presented as:

$h_i' = f_i(r, t, h_1, \ldots, h_n) + \varepsilon_i$

During implementation we have considered minimization of error functions $\varepsilon = \sum \alpha_i|\varepsilon_i|$, where weights α_i are chosen empirically.

In terms of security of the algorithm, since the number of hash values for each local minutia set is less than the number of these minutia, it is not possible to get the locations using only the information of any one local set. Also, it is not known which minutia participated in the creation of a particular hash value. Thus the non-inevitability is still maintained.

5. BIOHASHING

In another study, cancellable biometrics is applied using the BioHashing approach. The proposed cancellable biometric approach, BioHashing, involves creating a non-invertible hashing algorithm for fingerprints which integrates random numbers with the features of a fingerprint [6]. Like with the minutiae based symmetric hash function, BioHashing does not require pre-alignment on the registration point of the fingerprint image to work properly because all the minutiae are translated into a predefined two dimensional space based on a reference minutia. One defining feature of cancellable biometrics is that it most effectively allows the reuse of a biometric template. So, even if the biometric template is lost it can still be cancelled and replaced, dealing with a common issue that biometric authentication systems face.

In the proposed approach the image of the fingerprint pattern is preprocessed so it is not exposed outside of the client computer, and is distorted using a non-invertible transformation. Given various parameters, a single person's fingerprint could generate thousands of different virtual fingerprints, verifying its ability to be replaced. The randomly generated key needs to be stored in a secured manner for it to be effective.

To avoid high rejection rates, this two-factor authentication involving tokenized pseudo-random number and fingerprint features is implemented, resulting in a set of user specific codes known as the biocode. One important thing to note with this method is that both the random data and user fingerprint feature are required to recover the biocode as the code is a combination of both data sets. As a result, biometric fabrication is protected against and unauthorized users can be detected.

5.1 High Level Overview

The progression of BioHashing can be seen in Figure 1 of the paper [9], where the fingerprint first is subjected to some transformation, extracting the features of the fingerprint. These features are separated into vectors, and are combined with the random numbers to generate the hash code. Simply changing the

transformation will create a new variant for re-enrollment of users, making the entire system replaceable.

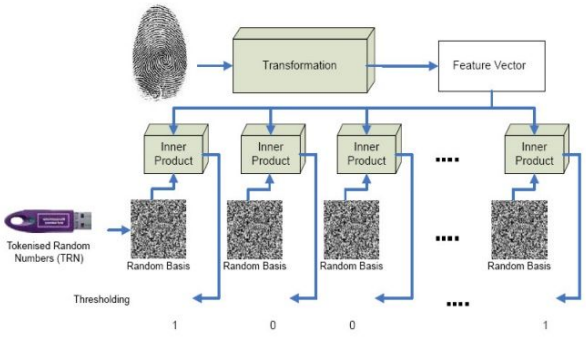

Figure 7: Progression of Biohashing

5.2 Algorithm of the Transformation

Let f be a many-to-one function which takes a raw biometric image as input and its output is its distorted virtual image. Its features are then laid out in many vectors, calling them feature vectors F. An input token is then used to generate a set of pseudo-random vectors V, where V is $\{P_i \in \mathbb{R}^M \mid i = 1, \ldots, M\}$ based on the seed. The Gram-Schmidt process is applied to V to obtain a set of orthonormal vectors O, where O is $\{r_I \in \mathbb{R}^M \mid i = 1, \ldots, M\}$. Then, the dot product of F and O is calculated, P_i such that $<v, r_i>$ and finally a threshold τ to obtain biocode $B = \{b_1, \ldots, b_i, \ldots, b_m\}$ where its elements are defined as:

$b_i = \{0$ if $<v,r_i> \leq \tau$, 1 if $<v,r_i> > \tau\}$, where i is between 0 and m, the dimensionality of B. The two biocodes are compared by hamming distance.

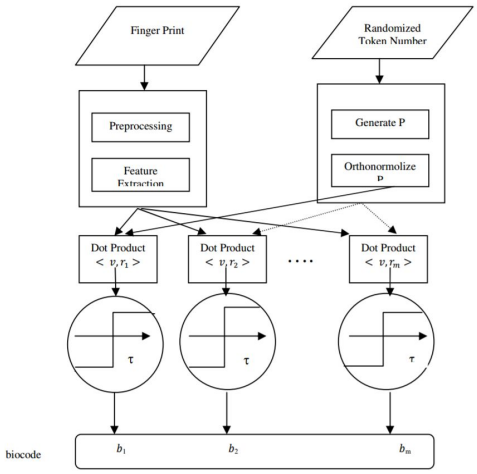

Figure 8: The architecture of a biohash.

5.3 Experimental Results

The experiment done by the paper is set up using Eigen feature to act as the fingerprint feature extractor. Imposters are done by matching each subject's Biohash with every other subject generated, with the assumption of worse case scenario where the imposter always manage to steal the random token.

From this, the equal error rate (EER), the average of false acceptance rate (FAR) and false rejection rate (FRR), is calculated. This procedure was done ten times and the following results were shown, where pca denotes Eigenvalues and pcab-m denotes Biohash with m bit length:

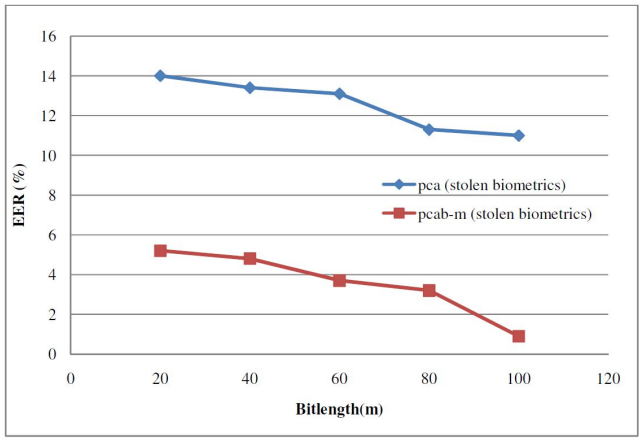

Figure 9: The performance comparisons for pcab-m, pca for Stolen Biometrics.

The results observe that FAR is almost zero and FRR is very minimum, making EER less than 1%. Also, since the EER of pcab-m is less than pca, the Biohashing approach is preferable compared to traditional biometric approaches. It was shown that this is true regardless of the size of bit length. Overall, the genuine distribution mean and standard deviation between pca and pcab-m do not differ much.

Biohash as a non-invertible cancellable biometric approach provides a solution for protecting the privacy of users while maintaining exceptional performance in a low EER rate. Biohashing also fulfills the requirement of the need for a revocable system, allowing it to be removed and replaced if the data is compromised. Since it is non-invertible, it is computationally hard to obtain the original data without the unique key.

5.4 Applications

Other researchers apply the use of BioHashing to sustain its properties of security, functionality, and non-invertibility. A method to properly secure the randomly generated key was done using Reed-Solomon error code and threshold secret sharing scheme [3].

Another approach, the fuzzy vault scheme [8], has been presented and implemented in the past involving the use of biometric cryptosystems to protect both the secret key and the biometric template, similarly to BioHashing. The use of a cryptographic framework intends to preserve the secrecy of the secret keys. However, this method has difficulties with the alignment of the fingerprint, possibly requiring multiple fingerprint impressions during enrollment and verification to achieve intended results. Fingerprint minutiae is compared by obtaining high curvature points to help align the template minutiae with the query minutiae. BioHashing handles this problem by focusing more on minutiae descriptors and neighboring minutiae.

Future work on BioHashing can be done with the focus of improving the EER even when the key is exposed. The BioHashing approach can also be extended by using different transformations such as the Baker Non-Invertible Transform, as suggested. This makes the feature points strongly mixed, providing even better security.

6. FUTURE WORKS

Future work in facial and iris recognition could aid in the technology being more widely used to protect and identify users. Enterprise and government both widely acknowledge the use of both these biometric systems in the near future. In facial recognition, future technology could eliminate the need of geofencing as a means of marketing to shoppers. Software will help serve shoppers with more personalized deals, promotions, product suggestions. In terms of security, the future of facial recognition could see to highlighting a person's facial features and predicting changes that could occur when the face is modified with facial hair, glasses, or age.

In iris recognition, the future work for this technology should focus on combating its current weaknesses. The technology will need to become more flexible in order to succeed as a method of biometric security by improving recognition given off-gaze images and handle variations that can occur from obstructions such as eyelashes and eyeglass reflections.

Once these challenges are combatted, facial and iris recognition must develop more efficient ways of applying symmetric hash functions. The future hashing functions will need to account for an imperfect image of the face or iris and recognize the similarities in an image to correctly match them.

Performance and accuracy regarding measures of false acceptance rate and false rejection rate can be optimized by proposing extended work to current methods. This can be done in implementation of the system by using multiple biometric sources. In fingerprinting, an assortment of other minutiae data not limited to the location or orientation of minutiae points can be used such as ridge curvature or ridge density.

The methods mainly involving fingerprint techniques can be generalized to other biometric modalities that use locations of specific image features. The challenge then is to apply symmetric hash functions for biometrics that examine a pattern of natural sequencing of features, such as voice recognition or signatures. The behavioral attributes associated with them work differently than non-behavioral biometrics we investigate with fingerprints, iris recognition, and facial recognition.

7. CONCLUSION

With a growing concern of information security, it is important to have reliable methods of identification and authentication in our systems. The use of biometric systems requires a high degree of trust and reliability that can be compromised with the threat of outsider attacks. While traditional methods of cryptography are sufficient with other authentication techniques, the importance of non-invertibility is highlighted because of the intention to mask user biometric data. In this final report, we discussed the application of biometric hashing to facial recognition, iris recognition, and fingerprinting. We reviewed the current applications of facial and iris recognition as well as their current strengths and weaknesses. These two forms of biometric security were then reviewed after an implementation of a hashing algorithm where they both benefited in security, error reduction, and efficiency. In fingerprinting we discussed methods of symmetric hashing and biohashing algorithms. As this technology is relatively new, there is an enormity of advances that could be made to further improve effectiveness in hashing. We hope that these hashing algorithms can be improved on in the future and further generalized so that they can be applied to other forms of biometric security thus improving the security of future users and systems.

8. REFERENCES

[1] Daugman, John. "How iris recognition works." IEEE Transactions on circuits and systems for video technology 14, no. 1 (2004): 21-30.

[2] Daugman, John. "New methods in iris recognition." IEEE Transactions on Systems, Man, and Cybernetics, Part B 37, no. 5 (2007): 1167-1175.

[3] Jin, Andrew Teoh Beng, David Ngo Chek Ling, and Alwyn Goh. "Biohashing: two factor authentication featuring fingerprint data and tokenised random number." Pattern recognition 37, no. 11 (2004): 2245-2255.

[4] Laurenz Wiskott et al. "Elastic Bunch Graph Matching." Scholarpedia, 9(3):10587.

[5] Lu, Xiaoguang. "Image analysis for face recognition." personal notes, May 5 (2003).

[6] Lumini, Alessandra, and Loris Nanni. "An improved BioHashing for human authentication." Pattern recognition 40, no. 3 (2007): 1057-1065.

[7] Ma, Li, Tieniu Tan, Yunhong Wang, and Dexin Zhang. "Efficient iris recognition by characterizing key local variations." IEEE Transactions on image processing 13, no. 6 (2004): 739-750.

[8] Nandakumar, Karthik, Anil K. Jain, and Sharath Pankanti. "Fingerprint-based fuzzy vault: Implementation and performance." IEEE transactions on information forensics and security 2, no. 4 (2007): 744-757.

[9] Radha, N., and S. Karthikeyan. "An evaluation of fingerprint security using noninvertible biohash." International Journal of Network Security & Its Applications 3, no. 4 (2011).

[10] Rathgeb, Christian, and Andreas Uhl. "Iris-biometric hash generation for biometric database indexing." In Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 2848-2851. IEEE, 2010.

[11] Sharif, Muhammad, Kamran Ayub, Danish Sattar, Mudassar Raza, and Sajjad Mohsin. "Enhanced and fast face recognition by hashing algorithm." *Journal of applied research and technology* 10, no. 4 (2012): 607-617.

[12] Symmetric Hash Functions For Secure Fingerprint Biometric Systems. 1st ed. State University of New York at Buffalo: J.A.Robinson, 2007. Web. 17 Dec. 2016.

[13] Turk, Matthew A., and Alex P. Pentland. "Face recognition using eigenfaces." In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, pp. 586-591. IEEE, 1991.

[14] Wikipedia contributors, "Eigenface," *Wikipedia, The Free Encyclopedia,* https://en.wikipedia.org/w/index.php?title=Eigenface&oldid=751828722 (accessed November 28, 2016).

[15] Wikipedia contributors, "Linear discriminant analysis," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Linear_discriminant_analysis&oldid=754849680 (accessed December 14, 2016).