Implementing A Novel Biometric Cryptosystem using Similarity Distance Measure Function Focusing on the Quantization Stage

R. Ramya¹* and T. Sasikala²

1Department of Computer Science and Engineering, Sathyabama University, Chennai – 600119, Tamil Nadu, India; ramya.rav@gmail.com
2SRR Engineering College, Chennai - 603103, Tamil Nadu, India; sasi_madhu2k2@yahoo.co.in

Abstract

Objectives: The essential requirement for a successful hashing method involves two distinct stages projection and quantization. In general, the projection stage is given much importance than the quantization stage. This stage has been concentrated in this paper which has equal importance as projection stage. Methods/Analysis: The using of Manhattan Distance method has been proposed in this paper instead of the widely used Hamming Distance, since it destroys the neighbourhood structure while measuring the similarity between points in the hashcode space. Findings: The problem of destroying the neighbourhood structure that existed in Hamming Distance is overcome by Manhattan hashing. Novelty/Improvement: The outperformance of Manhattan distance compared with Hamming distance has been shown and also, this paper has made an attempt to implement them in our Biocryptosystem to show its efficiency.

Keywords: Biocryptosystem, Biometrics, Hamming Distance, Manhattan Hashing, Similarity Measure Functions

1. Introduction

Human beings possess unique physical attributes that helps them to differentiate between each other. This feature has given a helping hand to the entry of the field of Biometrics. Analyzing human biological characteristics is known as the science of biometrics. The most common of these attributes are the fingerprint, hand, eye, face, and voice¹.

Biometrics is a field that recognizes individuals depending upon their behavioral or biological traits. Different types of biometrics involve DNA Matching (Chemical), Ears, Eyes (iris), Face Recognition, Fingerprint, Finger Geometry, Gait, Signature Recognition, Vein, Voice etc.

DNA Matching identifies a person based on the segments from DNA. Ear Biometrics uses the ear shape. Eye Biometrics uses the features of iris. Also Eye- Retina type uses patterns of veins at the eye back. Face Biometrics analyses using the facial recognition. They use eigenfaces in most of the systems or local feature analysis. Fingerprint which is most commonly used uses minutiae which can be found on the finger. Finger Geometry uses the finger’s 3D geometry. Gait which comes under behavioral biometrics uses the style of walking or any other gait. Signature Recognition uses the style of the handwriting of the individual. This Signature recognition are of two types, Static one and a Dynamic one. Static involves comparing the scanned signature with another scanned signature. Many algorithms are used to check their validity. Dynamic is much in fame as data is captured with different coordinates of the person who signs from the signing device. The data got from this kind is made lawful and used widely using the digital forensic tools for examination.

Vein identification biometrics uses the patterns of

* Author for correspondence
veins in a human body. Voice recognition\(^2\) uses the voice of the person who speaks. Finally, Multimodal biometrics involves two or more of the biological traits for enhanced security reasons.

1.1 An Introduction to Biocryptosystem
A Biometric template is a physical or behavioral snapshot captured from an individual. From the perspective on human behavior and its connection to society as a whole there is a need to protect Biometric Templates. Also from a perspective based on technology, there is also a need for the security of Biometric Templates.

This Biometric template can be one of the images of finger print, iris, etc. Unique features are taken from this template, and transformed into another file. Since such a file has the most essential credential information of a person it is also prone to hacking, failures, theft etc\(^3\).

The most risky areas where Biometric Templates are as follows:

- After template creation
- The biometric templates database.
- Transferring from the system where biometric template is stored to the server.

Cryptography paves a way to enhance the protection of Biometric Templates. Cryptography is the science that process data into an unintelligible form and reversing without data loss. Applying this process into Biometric has become feasible and is referred as “Bio-Cryptography”.

1.2 Security in Biocryptosystem
A recent study shows almost fortyone percentage of US customers prefer using any of their biometric data i.e. fingerprint or iris for logging in their credentials. It also says that today consumers are more conscious towards security and privacy than in the previous years, and the survey results show that people are taking steps to protect their personal information\(^4\). Hence we come to know that a growing number of people are towards the acceptance of a secured system.

A system using biometrics highly relies on the precision of the similarity measure functions. In the paper, by Yampolskiy et al.\(^5\) the authors have done the job of combining similarity measure functions for behavioural biometric systems which depends on the background knowledge of the data at the feature level.

In J. Zhe et al.\(^6\) paper, they have proposed a method for finger print template protection by using Randomized Graph-based Hamming Embedding. The binary template obtained here is secured against the problem of inversion. Angandi et al. in\(^7\), got the proposal of signature recognition system offline. The overall recognition precision obtained using this method ranges from 90% to 100%.

Here Belguechi et al.\(^8\), analyses the problem based on privacy in the current architecture of e-Passports for the storing and transferring of biometric data. A solution was proposed by using cryptographic protocols and cancelable biometrics. The Locality Sensitive Hashing (LSH)\(^9,10\) is included in representative data-independent methods its extensions\(^11–14\) generates hash data by applying a shifted cosine function. Hash data is generated by Shift invariant kernel hashing (SIKH) by using a shifted cosine function.

2. Distance Measures on Biocryptosystems

2.1 Similarity Distance Measure Functions
To measure how close a sample data resembles a template data a fresh sample has to be taken. A characteristic based on statistics of the data distribution is taken into account by a good similarity measure function\(^15\). Some of the well-established measures are, Euclidean Distance, Mahalanobis Distance, and Manhattan Distance.

2.1.1 Euclidean Distance
It is a known and a famous similarity distance function. It is the sum of the squared distances of two vector values \((x, y)\)\(^16\).

\[
E_{\text{iiid}}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}
\]

Euclidean Distance is a variant to the dimensionality of the vectors\(^17\).

2.1.2 Mahalanobis Distance
It can be given as:

\[
D_{\text{m}}(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}
\]

with \(\mu = (\mu_1, \mu_2, \mu_3, .., \mu_n)\) representing mean and \(\Sigma\) for covariance matrix for a multivariate vector \(x = (x_1, x_2, x_3, ..)\).
Mahalanobis distance is also a measure of dissimilarity between two random vectors of the same distribution with the covariance matrix \( \Sigma \):

\[
d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}
\]

(3)

If the covariance matrix is the identity matrix, then it is the same as Euclidean distance. If the covariance matrix is diagonal, then it is called normalized Euclidean distance:

\[
d_{xy} = \frac{\sum_{i=1}^{p} (x_i - y_i)^2}{\sigma_i^2}
\]

(4)

where, \( \sigma_i \) is the standard deviation of the \( x_i \) over the sample set. Mahalanobis distance is not dependent on the scale of measurements.

### 2.1.3 Manhattan Distance

It is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. Also, Manhattan distance is the sum of the absolute differences of the two vector values \((x, y)\).

\[
d_m = \sum_{i=1}^{n} |x_i - y_i|
\]

(5)

### 3. Proposed System

Most of the hashing methods available focus on the projection stage rather than the quantization stage. Here we concentrate on the effect of quantization. Quantization stage has equal importance like the projection stage. Manhattan Hashing is based on Manhattan distance that calculates the nearest neighbour search in the hash code space between points. The neighbourhood structure is preserved using MH in order to achieve the goal of hashing.

#### 3.1 Manhattan Distance based Quantization

In Hierarchical Quantization (HQ) neighbourhood structure is not preserved based on Hamming Distance. This drawback can be solved using Manhattan distance with natural binary code.

In order to use Manhattan distance for hashing, a q-bit quantization method is adopted. Once the real valued projection functions are learned, it is divided into 2q regions with each projected dimension.

When q value is 1, the output computed with Manhattan distance is almost same with Hamming distance. The Manhattan Quantization for Manhattan Hashing preserves the neighbourhood structure between points better comparatively with other methods.

### 3.2 Evaluation Metrics

The methodology has been widely used in 19–26. More specifically, Euclidean neighbours are defined using which recall, precision and the mean Average Precision (mAP) can be calculated. They are defined as follows:

\[
\text{Precision} = \frac{\text{the number of retrieved relevant points}}{\text{the number of all retrieved points}}
\]

\[
\text{mAP} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \sum_{k=1}^{R} \text{Precision}(R_{ik})
\]

where, \( q_i \in Q \) is a query, and the number of points relevant to \( q_i \) in the data set is \( n_i \), the related values are ordered as \( \{x_1, x_2, ..., x_{n_i}\} \), \( R_{ik} \) represents the set of ranked retrieval results.

### 4. Experimental Results

In this paper, experiments were made with sample fingerprint images as shown in Figure 1. We experimented initially with Hamming Distance and have got the output as given in Figure 2. First the sample fingerprint image was given as input and it was encrypted using the SHA-256 algorithm and thus we got the biocryptic template. Like this we took nearly 40 to 50 images and stored all of its corresponding biocryptic templates. A fresh input image after the conversion was compared against the stored templates using Hamming Distance Method. The image which matched the input image returned the value of 0, showing that the image matches it completely and exactly without showing 1 or 2 or any other value. This experiment has been done successfully for Hamming Distance method. The aim of showing it using Manhattan Distance focusing on quantization stage is under experimentation. The idea of using Manhattan Distance Method which can perform comparatively better was got from 27, which is referenced in Table 1.

![Figure 1. Extract features from an arbitrary fingerprint.](image-url)
Implementing A Novel Biometric Cryptosystem using Similarity Distance Measure Function Focusing on the Quantization Stage

From this we have got the output based on Hamming Distance. Based on Table 1.

Table 1. Comparison of Manhattan with other methods

| Method | SBQ | HQ  | 2-MQ | 3-MQ | 4-MQ |
|--------|-----|-----|------|------|------|
| ITQ    | 0.2771 | 0.3166 | 0.3537 | 0.286 | 0.27 |
| SIKH   | 0.0487 | 0.0512 | 0.0722 | 0.0457 | 0.0339 |
| LSH    | 0.1563 | 0.121 | 0.1382 | 0.0961 | 0.0684 |
| SH     | 0.0802 | 0.154 | 0.2207 | 0.2103 | 0.2026 |
| PCA    | 0.0503 | 0.1414 | 0.1913 | 0.2064 | 0.2092 |

From Figure 3 we can see that Manhattan hashing using Manhattan distance can outperform the other methods. This method is used in our biocryptosystem.

5. Conclusion

Most of the hashing methods do not concentrate on the quantization stage. As we have stated before, this paper has experimentally studied the results with Hamming Distance and have made an attempt to prove that Manhattan based Hashing can perform better in quantization stage. Hamming Distance as given in the results gives a solution, but it destroys the neighbourhood structure. The results have shown that Manhattan Distance provides a comparatively good result than the other distances methods and proves that it does not destroy the neighbourhood structure. This paper tries to give a novel solution, by applying Manhattan distance based hashing in biometric cryptosystem. Once the experiments which we are currently doing with Manhattan are done with, we will show its better working results in our upcoming work.
6. Acknowledgement

I would like to take the opportunity to thank the person who guided and supported me during this process. I have a great pleasure in expressing my deep sense of gratitude and indebtedness to Dr. T. Sasikala, my Supervisor for her continuous guidance and invaluable suggestion at all the times during the research work.

7. References

1. Mahalakshmi U, Sriram VS. An ECC based multimetric system for enhancing security. Indian Journal of Science and Technology. 2013 Apr 1; 6(4):4299–305. DOI: 10.17485/ijst/2013/v6i4/31857.
2. Shah HNM, Ab Rashid MZ, Abdallah MF, Kamarudin MN, Chow KL, Kamis Z. Biometric voice recognition in security system. Indian Journal of Science and Technology. 2014; 7(2):104–12. DOI: 10.17485/ijst/2014/v7i2/47673.
3. Saminathan K, Chakravarthy T, Devi MC. Comparative study on biometric iris recognition based on hamming distance and multi block local binary pattern. Indian Journal of Science and Technology. 2015 Jun 1; 8(11):1. DOI: 10.17485/ijst/2015/v8i11/71764.
4. Market wired: Press Release and News Wire Services [Internet]. [Cited 2015 Jun 15]. Available from: http://www.marketwired.com/press-release/gigya-study-reveals-increased-consumer-demand-data-privacy-personalized-marketing-2038324.htm.
5. Yampolskiy RV, Govindaraju V. Similarity measure functions for strategy-based biometrics. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering. 2008 Dec 20; 2(12):4254–9.
6. Jin Z, Lim MH, Teoh AB, Goi BM. A non-invertible Randomized Graph-based Hamming embedding for generating cancellable fingerprint template. Pattern Recognition Letters. 2014(42):137–47.
7. Angadi SA, Gour S. Euclidean distance based offline signature recognition system using global and local wavelet features. 2014 Fifth International Conference on Signal and Image Processing (ICSIP); 2014 Jan 8. p. 87–91.
8. Belguechi R, Lacharme P, Rosenberger C. Enhancing the privacy of electronic passports. International Journal of Information Technology and Management. 2012 Jan 1; 11(1–2):122–37.
9. Andoni A, Indyk P. Near-optimal hashing algorithms for approximate nearest neighbour in high dimensions. 47th Annual IEEE Symposium on Foundations of Computer Science, 2006. FOCS’06; 2006 Oct. p. 459–68.
10. Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing. VLDB ’99 Proceedings of the 25th International Conference on Very Large Data Bases; 1999. p. 518–29.
11. Datar M, Immorlica N, Indyk P, Mirrokni VS. Locality-sensitive hashing scheme based on p-stable distributions. Proceedings of the Twentieth Annual Symposium on Computational Geometry; 2004 Jun 8. p. 253–62.
12. Kulis B, Grauman K. Kernelized locality-sensitive hashing for scalable image search. 2009 IEEE 12th International Conference on Computer Vision; 2009 Sep 29. p. 2130–37.
13. Kulis B, Jain P, Grauman K. Fast similarity search for learned metrics. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009 Dec; 31(12):2143–57.
14. Mu Y, Yan S. Non-metric locality-sensitive hashing. AAAI; 2010 Jul 3.
15. Lee K, Park H. A new similarity measure based on intraclass statistics for biometric systems. ETRI Journal. 2003 Oct 14; 25(5):401–6.
16. Sturn A. Cluster analysis for large scale gene expression studies (Doctoral dissertation, Graz University of Technology).
17. Yampolskiy RV, Govindaraju V. Use of behavioral biometrics in intrusion detection and online gaming. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; 2006 Apr.
18. Mahalanobis distance [Internet]. [Cited 2015 Jul 25]. Available from: http://en.wikipedia.org/wiki/Mahalanobis_distance.
19. Gong Y, Lazebnik S. Iterative quantization: a procrastuate approach to learning binary codes. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2011 Jun 20. p. 817–24.
20. Raginsky M, Lazebnik S. Locality-sensitive binary codes from shift-invariant kernels. Advances in Neural Information Processing Systems; 2009. p. 1509–17.
21. Xu H, Wang J, Li Z, Xeng G, Li S, Yu N. Complementary hashing for approximate nearest neighbour search. 2011 IEEE International Conference on Computer Vision (ICCV); 2011 Nov 6. p. 1631–38.
22. Liu W, Wang J, Kumar S, Chang SF. Hashing with graphs. Proceedings of the 28th International Conference on Machine Learning (ICML-11); 2011. p. 1–8.
23. Zhen Y, Yeung DY. Active hashing and its application to image and text retrieval. Data Mining and Knowledge Discovery. 2013 Mar 1; 26(2):255–74.
24. Zhen Y, Yeung DY. A probabilistic model for multimodal hash function learning. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2012 Aug 12. p. 940–48.
25. Norouzi MM, Blei DM. Minimal loss hashing for compact binary codes. Proceedings of the 28th International Conference on Machine Learning (ICML-11); 2011. p. 353–60.
26. Zhang D, Wang F, Si LA. Composite hashing with multiple information sources. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2011 Jul 24. p. 225–34.
27. Kong W, Li WJ, Guo M. Manhattan hashing for large-scale image retrieval. Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2012 Aug 12. p. 45–54.