A Unique Document Security Technique using Face Biometric Template

Raikoti SharanaBasappa\textsuperscript{1} and Sanjaypande M. B.\textsuperscript{2}

\textsuperscript{1}Research Scholar, India
\textsuperscript{2}Prof. and Head of Dept., VVIET, Mysore, India
sr.raikoti@gmail.com, rkroop99@gmail.com

Abstract

Human face is generally viewed as most flexible model when it comes to the field of biometric applications. The reason being it is a cost effective one due to its easy data acquisition fundamentals as well as it being invariable when considered over a fixed time period. There have been various algorithms that have been invented for the purpose of facial recognition as well as demonstrated over last three decades which are mainly categorized into kernel based techniques, Vector Methods and so on. A unique binary string generated out of facial features of a person is called a template. This template is used in several applications like network security, public key cryptography and so on. The main challenge in this regard is that no two instances of the faces acquired at different time instance can be same. The templates are similar in regard to the hamming distance but are not same. Therefore it is important to propose an algorithm that presents a face recognition system with the invariant template generation. In this work we propose a unique security architecture where face features and subsequently the generated templates are used as the key for document security. A framework needs users to register with the system along with their face instances. The instances are used for training samples. Once a user selects a folder for encryption, all the files of the folders are encrypted with the templates generated from users face data. A decryption request needs to be authenticated through the face data and the template generated at the time of decryption is used for decrypting the encrypted files. Rinjdal method is used for the cryptographic framework. Frames are acquired in real time from camera and face part is segmented based on skin segmentation. Further Eigen face based template generation and matching is used for face recognition. Results show significant low FAR in comparison to FRR and improved performance in encryption process and recognition rate.

Keywords: Public Key Cryptography, Hamming Distance, Biometric, Template, Key, Cryptographic Framework, Eigen Face

1. Introduction

Several techniques are developed over the years for document security. One of the common techniques to achieve security is through encryption and authentication. Encryption is a method where the actual contents of any digital document is changed using a password called a key. Authentication is a system where by the persons authenticity is verified first before giving him an access of the system. Biometric system is one where either the generated
key is from user’s biometric features and is verified through his biometric patterns. Face of each person is unique and therefore biometric system based on faces are unique and of high utility.

Therefore in this Work, we combine the technique of face recognition based authentication system and generate a key from the facial features of a person. This key is used as the key for encryption and only after the same user is authenticated, his documents are decrypted.

It’s just a normal thing that a human can easily identify any person’s face whom he had seen in the past. Humans have always had the innate ability to recognize different faces. Here the human brain simply compares the characteristics of the face with his known face and recognize the person. This is the common example of face recognition in day to day life.

Computers recently have shown the ability of recognizing the face. With a lot of innovations in technology happening around us primarily centered on the computer aided technology, the goal of creating a practical and fine face recognition program that can be applied in the real time is being strategically achieved.

By combining the tools of image processing, computer aided vision as well as pattern recognition effective human face recognition system which is fully automatic is being developed. The most unique feature is that this system differentiates between various people based on a comparison of past records in the databases.

The proposed topic of this Work deals with facial recognition under different lighting conditions.

The most unique feature of biometrics regarding the human face is that it makes use of the striking features of the human face to differentiate between various users. Human face recognition technology using biometrics mainly focuses on the following techniques of Neural Networking, Eigen Faces, Feature Analysis and Automatic face processing. If we take the example of a Neural Network we see that it works on a very distinguished procedure of using the template of the users face and checking it against its database to see if it matches with that of the original template of the user. In case the pattern doesn’t match or the image is rejected then the program modifies itself to avoid the mistake in future searches. The Eigen faces approach is by far the most innovative method, wherein a database is set to contain about 60-120 template faces. When the user enters his face pattern it is immediately mixed up with the available patterns to give in to produce an exact image of the user’s face which can be used for future identification and verifications.

A. Enrollment

Enrollment signifies that the user information is taken in for future processing of data in case of checking the authenticity of the searches.

During the first phase, features are calculated from the input face given by the user by a process called Feature extraction, and are modeled as a template. The modeling is a process of enrolling user to the verification system by constructing a model of his/her face, based on the features extracted from his/her facial sample. After the features are extracted and all the signal processing is done the system checks for the quality of the templates that are extracted from face, if sufficient quality of features are extracted then the templates are kept in the database else the system needs to acquire new facial image. The collection of all such enrolled models is called facial database.
B. Authentication/Verification

It is of vital importance to capture and process the vital details of biometric data to get an authentic picture of the templates to match with those of the database.

In the second phase, known as the verification phase we can see that certain striking features of the human face are taken in and compared with previous features preserved in the database. Comparison between these two is carried out by software loaded in the computer called as the Feature Matching.

Based on this comparison the final decision is made about the user identity.

Both these phases include Feature extraction, which is used to extract user dependent characteristics from face. The reason we use this feature is the reduction in the data taken for testing and at the same time with hold the striking information of the user. The complete
system is as shown below in schematic flow diagram which combines both the enrollment and verification phases.

C. Fundamentals Issues in Face Recognition

It is a known fact that the human face is not always constant and given a period of time there are a lot of variations in the patterns of the face. It is said that the human face is a myriad combinations of 3D images wherein the face is lighted up by the various sources of light. If the human face is captured at an instant we find that it contains certain background images as well. Hence when a 3D image is converted into a 2D image the resultant picture can vary a lot. Hence while developing a system for face recognition one must keep in mind that an articulate system with accurate recognition features despite disturbances must be developed. It has indeed been found out that the 3D patterns, lighting effects, aspects apart from the human face do play a major role in creating these facial recognition systems.

The fact that matters most in developing accurate recognition and detection systems is that the variations and changes in the pattern of the human face must be made tolerable. The human face can in fact be distorted into a million possible images which are unique in their own way. The variations are further classified into the Inter Personal and Intra Personal variations. The former will be mostly according to the genetics, race and identity of the person while the latter is mostly due to the age factor, facial changes, growth of hair on face, changes in facial expressions etc.

The basic working of a facial recognition system has to be precise as it has to give a name or identification of a particular face by comparing it with the myriad images located in its database. Furthermore it has to bear with the otherwise unassociated disturbances which one might come across while taking a picture like problems with the resolution of images, noise factor etc. hence the facial recognition in fact is posed with multi-dimensional problems of pattern detection and recognizing the image.

Also the system must be flexible, i.e. it must be compatible with other computational devices. In short the capturing and processing of the images must be quick and precise keeping in mind the run-time and storage space.

2. Related Work

[1] elaborates the various techniques for biometric data security by proposing a matching environment based on smart cards. It also elaborates the mechanism of generating a hash from the fingerprint biometric data and encrypting data through this hash.

Authors in [2] deal with the variability in the biometric templates. According to the finding of the author, every matching for authentication must also measure the quality of the just generated template and periodically must update the template in order to maintain high accuracy in biometric authentication process. The paper also presents matrices for biometric template quality analysis which is used in the proposed benchmark analysis work.

Authors in [3] elaborate the main problem with the conventional password based security techniques and emphasizes on the facts as to why the biometric keys are better than conventional keys. The work defines the technical steps associated with the keys and also discusses the properties of good biometric keys.

For biometric key generation from biometric features like faces and fingerprints, first image specific features must be extracted which is called templates. A quantization is applied on the template to generate the key. As real number range is infinity, if a key is generated without quantization, matching becomes a difficult process. Hence authors in [4] elaborate the mechanism for the quantization process for key generation. They also suggest a log likelihood based technique for the same.
Even though biometric keys are strong technique for cryptographic key generation, because they are stored in the database, there remains a chance for the keys to be eavesdropped by unauthenticated users which makes the system vulnerable. Therefore techniques must also be adopted for such key generation. Hence authors in [5] propose the techniques for securing the biometric key itself.


The strength of a biometric key is defined from the inability of a proper guessing of a key using brute force technique. A biometric key appears random to any intruder. Therefore how he guesses the key depends upon the entropy information of a random variable that generates the key. [7] defines a mathematical relationship of probability of successful guess of a key with the entropy information of the templates and hence quantifies the fact that entropy analysis of any template is an important step in deciding the strength of the key.

The closeness of a template with the stored template depends upon the distance between the templates. This distance can be represented in various mathematical forms as proposed by the authors of [8]. The authors also prove experimentally that log likelihood measure is one of the better ways of representing the closeness of two templates.

Whenever a mechanism is selected for biometric template matching for authenticating purpose, it invariably presents a false rejection and false acceptance on the biometric data and the mechanism itself. The authors in [9] present a unique way to select the appropriate tradeoff between the rejection and acceptance tradeoff so that the adopted technique is acceptable and efficient. The author also presents a benchmark analysis for optimality for any recognition technique in [10] and illustrates the proposed theory with the help of character recognition system.

Gabor based techniques are widely adopted for biometric feature representation or generation of templates. But the size of such initial vectors is so high that it presents a practical problem of storage. Thus biometric template reduction becomes an important aspect for biometric key generation or authentication technique. [12] presents a technique for minimizing the number of feature vectors for template generation.

Out of all the possible attacks on biometric keys the most severe attack is on the stored keys. This finding of [13] forms a base for our assumption that if a system can be devised without the necessity for the key to be saved, a biometric system can be made un-attackable. Even though the authors present multi biometric model for hardening the security of a biometric system, the system still remains vulnerable against the attacks.

For any biometric like face or fingerprints or iris, templates will differ from one instance to the other instance of acquisition. Therefore out of N number of acquisition of the templates of a single feature type of a person there would certainly exist variability amongst the templates. Therefore authors in [14] present a systematic way of extracting the best template out of all the templates. This paper also presents an important step for biometric feature benchmark analysis by proposing a clustering technique for segregating the biometric features by their average distance measures.

As against most widely used Gabor convolve methods, authors in [15] proves that bi orthogonal wavelets achieve higher accuracy with low FAR and high FRR. Thus Bi orthogonal wavelet based template generation is selected as one of the methods for testing the quality of a biometric template for experiments in this paper.

[16] classifies the type of attack on the stored biometric systems and proposes a unique mechanism for minimizing the risk emerging from such attack. But the authors presents an important finding that encrypting the biometric keys or appending more security layer for
protecting the biometric data leads to more complicated processing in the recognition phase which delays the overall time for recognition. This finding also strengthens the need for a system which can empirically analyze the quality of a biometric system before adopting it.

[17] presents a technique to remove the need of a centralize storage for biometric data with the help of smart cards. But the authors also emphasis that even such a card based system cannot guarantee a perfect un attackable system as the user information can be eavesdropped from the card itself.

[18] discusses various techniques by means of which the biometric data itself (not the templates) can be acquired, forged and can be used. It also presents a wide verity of attacks and describes at which instance and which data sets the attacks can affect.

In [19] authors describe a technique for generation of digital signature from biometric template. Such signatures can be used in public key cryptography for encryption and decryption of digital documents shared across a network.

[20] analyzes the technique proposed by [11] to rectify the errors in generated template by using Reed Solmon code. The author also proposes a unique mechanism for protection of biometric key by storing the checksum rather than the key in the database. The authors also show that a key generation from biometric template and subsequent correction using the checksum improves the system security to a great deal. This work has adopted the elaborated model for error correction in biometric template.

In [21] authors have used irreversibility and revocability of the templates as a measurement for biometric key security and show the technique of biometric matching in the secured environment. The authors present another significant observation here in terms of feature selection. [21] finds that lower order features like means and standard deviation can not reveal the actual information of any feature set and higher order spectra is better suited for the representation of the features. Hence in our benchmark analysis, we have used higher order spectra for determining the closeness of the templates.

[22] proposes an alternative method for biometric security through enthronization. It is a process of adding extra feature systematically in the biometric template to increase the randomness of the template and also demonstrates that even after enthronization, the achieved results are satisfactory.

[23] discusses one of the most little talked possible attacks on the biometric templates. The author here finds out a technique through which the image itself can be interpolated or reconstructed from leaked template. Therefore this image can be subsequently used to obtain un authorized information through hill climbing attack. Thus the findings demonstrates that not only the randomness of the templates are enough to provide security to the biometric system but at the same time it is also important to ensure that the inverse process of reconstruction of images should not be possible from the templates.

[24] elaborates a new technique called biometric template transformation be means of which direct mathematical transformation can be achieved on the templates to extract more information from the templates. This important transformation is the answer for the following severe problem. Whenever a biometric key is encrypted, it cannot be used for direct matching at the time of verification. The key needs to be decrypted before matching which invariably exposes the key. But with the help of direct transformation on the biometric templates, key matching can be performed over the encrypted keys, minimizing the requirement for a decryption stage prior to verification.

In [25] authors prove that presence of noise in the biometric data actually improves the security of the data itself. This claim is supported through entropy analysis of the generated key from the template and the relationship of energy and the entropy of the information.
[26] discusses various mathematical transformation related to guessing of a biometric key and presents a fish-bone model for categorization of the attacks on biometric data. The model helps in identifying various areas which must be considered for ensuring the security of biometric data.

[27] proposes a mechanism for mixing a biometric template with biometric key for biometric key protection.

[28] discusses about possible technique for face biometric and is used for selection of the techniques for the current work.

Merely ensuring a strong key generation technique is not sufficient for biometric technique adaptation. It must be checked for the feasibility of adaptation. [29] presents various means of feasibility check for biometric techniques.

[30] proposes an indexing scheme through binning to index the biometric images in a large database into groups for better classification and representation.

Authors in [31] discuss a unique technique for secured biometric mechanism by introduction of N-template system. The authors claim and proves that if more than one template is generated from the same biometric feature like iris and a similarity measurement function can be devised to find the similarity between the templates than, the empirical value of similarity can be used to authenticating the users rather than the template itself.

The method of selecting best templates or function that generates best templates is always result driven tests. Thus it is quite difficult to select the best templates out of available templates. [32] accesses this problem partially and presents a score based mechanism for template selection. In this work, the prototype of [32] is extended for selecting not only the best templates but for selecting the mechanism itself which can generate most acceptable templates.

[33] discusses various non technical issues alongside the technical issues for accepting a biometric access control technique.

Improving the biometric recognition with 0% FAR and 0% FRR is considered to be ideal theoretical biometric system which is not yet achieved. Authors in [34] observe that most widely used technique for enhancing the security of a biometric system is through combining more than one modality like combining face and iris. But this needs the user to expose his features twice before two different sensors.

3. Methodology

The main Modules of the work are as Explained Below:

1. Face capturing: This is essentially a camera interface written in C#.Net. The objective of this element is to acquire face in real time from web cam for processing

2. Preprocessing: This is an image processing part. The main role is to resize the images and convert the images into gray scale image.

3. Feature extraction: It is the process whereby Fourier magnitude features are extracted from the faces. Totally 128 features are used. The algorithm used is FFT( butterfly structure)

4. Classification: This is a support vector machine based technique. The input is a face of unknown class and the output is the detected person’s name.

5. Template generation: This generates a unique password called template from features.

6. Authentication: This module takes input as name of the person and the template and authenticates if that face belongs to the person or not.
7. Encryption and Decryption: This technique takes a folder as input, extracts all the files from the folder and applies Rinjdal technique for encryption and decryption. Along with the password, it takes a random key called salt.

4. Proposed System

The proposed face recognition technique is based on Eigen features extracted from the face. These vectors are used as feature vectors.

![Figure 3. Proposed System](image)

Nearest Neighbor Classifier is used for classification of faces. Once a face is recognized a hash function is generated from the facial features and the support vectors. These features are used to encrypt and decrypt the document in a standalone system and over a network. Overall process is described in following figure.

The system is explained in short as below:

1. First users register to the system by giving two face instances called face*a.jpg and face*b.jpg which are obtained from a live web cam
2. These images are extracted by a VB module which trains the faces by extracting Eigen face feature vectors.
3. Once the user wants to encrypt any document, he has to input his face. The face is first matched with the database face and if matched then a hash is generated. This has is used as the key for password.
4. Rinjdal algorithm encrypts the document using this key.
5. The same process is followed for decryption.
The overall process of data security can be elaborated in Figure 4.

When the user is registered for the first time, biometric template is extracted and stored in a database. When user receives an encrypted document which he needs to decrypt, he needs to get authenticated first. In authentication process biometric template of the user is regenerated and is compared with the features of the database and if they match the private key is generated from the stored features because every time new features are generated, they will not be 100% same as that of the stored templates. In key generation even deviation of a symbol out of a matrix of size say 1024, will result in generating wrong key. Hence the best method is to generate the key from the stored vector rather than the new vector. Further the stored vectors are protected using either watermarking techniques or are encrypted by a common domain key and are stored. Therefore if an unauthenticated intruder can track the domain key, he can easily temper with the stored key. Therefore the strength of the technique is considered to be dependent on the techniques of protecting the templates themselves, which interns minimizes the strength of the biometric templates itself.

\[\text{Figure 4. Biometric Template Based Public Key Cryptography}\]

A possible alternative to this problem is to device a mechanism by means of which it is not required to store the key in the database. Every time a document arrives, user’s biometric template is generated and a private key is generated from recently generated template rather than the existing stored template.
The solution is depicted in simple block diagram in Figure 5.

![Proposed Solution for Biometric Security of the Digital Document](image)

Figure 5. Proposed Solution for Biometric Security of the Digital Document

Figure 5 clearly elaborates that there are no stored element to temper with. Therefore theoretically the key generated out of this technique is temper proof.

The main challenge with this technique is the variability of the features. Every time biometric data is acquired, it must be unique and the selected function must be such that it produces the exact key at every instance. Thus the first step is to find out the most invariant biometric features and template generation methods.

### A. Encryption

Encryption is a technique by means of which the sender scrambles and changes the message in a way that only the receiver is able to unscramble it.

Consider a Node B with public key and private key are UB and PB respectively. Since UB is public, everyone will be able to get it. For anybody who needs to send the message ‘Msg’ in a secured way to device B, will Scramble the data using B’s public key to obtain the cipher text ‘Ctx’. A cipher is the base of cryptography that changes the original message. The encrypted and the cipher text can only be decrypted using the private key of B’s. Upon receiving the Msg, B decrypt it with the help of its private key PB. Since only B has the knowledge of its private key PB, nobody else including A can decrypt Msg. The entire process is as given bellow. It needs to be noted here that the cryptography is a mathematical function where as the keys and the messages are binary or numerical values. The cryptographic function is designed in such a way that the message cannot be predicted without having the actual keys. Even the variation of a single bit should lead to unsuccessful decryption.
B. Eigen Face for Face Recognition

Face recognition process is explained in detail in earlier chapters. As discussed there are several techniques which are generally classified into supervised and unsupervised classification. An Eigen face approach is one of the most used and efficient approaches of face recognition.

Now let us discuss the entire technique of face recognition based on Eigen face technique.

We consider that there are N users whose faces are to be recognized. For Eigen face technique to work properly, it is important that there should be training images available of each user in the database. More the number of training instances the better the efficiency of the system will be.

Now our face consists of various distinct components like eyes, mouth, forehead, chick and so on. Each of these affects the way a face is recognized. Therefore if the smaller details of faces are extracted, then the classification stage can compare each of these features of a face with the other face images.

All the faces in the training database can be represented as linear combination of Eigen faces. An Eigen face is essentially the principal components extracted from the covariance matrix of the faces in the database. Any face can be reconstructed by correctly selecting the weight vectors for combining the Eigen faces. Maximum number of Eigen faces is equal to total numbers of faces present in the database.

In general rather than representing a face with many Eigen faces, generally the face is represented or approximated by most significant Eigen faces or the Eigen faces that can reproduce the actual face image. Therefore they are the most variance out of all other images from the set of training images.

Therefore the Eigen face concept can be considered as face decomposition into M-sub space or M sub face components. It may be argued that more Eigen face per image will result in better efficiency. Though the argument is true to some extent but it results in a lot of computational complexity. Therefore best faces are used instead of all Eigen faces.

The concept of Eigen faces is derived from the fact that faces can be efficiently represented by set of standard face images and some weights corresponding to the facial features. Therefore, if a Number of images of the faces of the users are re constructed by combining the weighted sum of a minimum numbers of characteristic images, then an effective way to train and recognize the faces can be build.
Figure 7. Concept of Eigen Faces

Figure 7 shows the concept described above. The face of a person is divided into many small faces and these faces can be used to reconstruct the original face by adding with a weight factor.

The algorithm of the entire process is explained in detail in Figure 8. The Technique for face recognition based on Eigen faces are performed with the help of following steps:

1. First capture two instances of each person as training images. The set of all people’s to be registered are captured first.
2. Eigen face of all the training set images is calculated.
3. Calculate the principal components of the faces and find out M best faces. Keep the best M faces which are having highest eigenvalues. These M main face images describe the “face space”. When new faces are registered, the Eigen face space needs to be updated.
4. The weight of each face is calculated by Working the faces to the face space.

Once the system is initialized with the previous steps, followings are performed.

1. Given a Face to be Identified, Obtain a set of weight-vectors of the M Eigen face-images by Working the Eigen face onto each of the Eigen faces.
2. Using an appropriate threshold determines if the given image is a face image or not.
3. If it is a face, classify the weight vectors as one of the persons registered or unregistered.
Figure 8. Overall Algorithm for Eigen Based Face Recognition
5. Results

Figure 9. Detection and Recognition of only Trained face from Group of Faces

Figure 10. Detection of Face without Dark Glasses and at a Distance from Camera

Figure 11. Generated Password as Face Template from User Face
Figure 12. When a Folder is Selected, Files will be Displayed. All the Files will be Encrypted at a Stretch.

Figure 13. Encrypted File View of the Folder

Figure 14. Number of Registered Faces versus %FAR and %FRR
Figure 15. Number of Persons in Frame versus %FAR and %FRR

Figure 16. Number of Files versus Encryption Time

Figure 17. Number of Registered Faces versus Recognition Accuracy
6. Conclusion

Document security is one of the most important aspects of the computer security. Digital document security is generally done through either cryptography or authentication. Both of the mechanisms have their limitations. With advancement of Biometric technologies, more and more applications are using biometric solutions. However such a solution is yet to be introduced in windows application for providing security to the documents or folders. In this work we have proposed a real time face recognition based on Eigen principle, extracted the features post authentication and have generated templates. These templates are than used as the key to protect documents.

In this Work we combine both authentication and template based techniques to develop a much secured system for document security. Results show that that without appropriate authentication, it is not impossible to encrypt or decrypt any document. Misdetection rate for face recognition is very low. The Work can be further improved by Rapid face detection using Haar like classifier and applying face recognition over only detected face objects.

References


Authors

S. R Raikoti is M.Tech and M. Phill in Computer Science. He is currently pursuing research in Biometric Imaging. He has guided over 50 M. Phill students in their research work. His area of interest are Image Processing, Pattern Recognition, Biometric templates and Template security.

Dr. Sanjay Pande M. B. is professor and Head of Department of Computer Science Department in VVIET, Mysore. He received his P.H.D from Mysore University in 2005. His area of interest is Imaging studies for cognition and Recognition in Pattern Recognition.