File Security Using Principle Component Analysis Induced Facial Recognition

[1]G.Kumaresan, [2]S.U.Abinandhanan, [3]P.Jai Ganesh
[1]Assistant professor, [2,3]Student
Department of Computer Science & Engineering,
Valliammai Engineering College, Chennai, India

Abstract - File security is a field which is less secure than what is needed for the current technological environment. The existing systems use normal encryption techniques using passwords, which can easily be bypassed. Even though the password is unknown, the security can be bypassed by using techniques like phishing and brute force. They can be protected by simple algorithms, but still, if the password is known by any third person, the security can still be breached. Our system encourages the concept of security for a file which belongs to, and can be accessed by only authenticated people . The system also proposes a higher security level ring that compliments the security of password protected encryption. It also acts as a monitoring system in addition to the security system.

***Index Terms* - File Security, Face Recognition, Bio-metrics, Neural Networks**

I. INTRODUCTION

Current day file security system use minimal amount of security as the importance of security for files are rather underestimated. This was due to the fact that it was true a few decades back when file systems were used to store normal carry-around data and it did not carry anything which amounts to the value of that which has to be protected by an advanced means of security system. But in modern day technological environment, files carry data which account to intellectual property rights and information regarding national and international security.

File security in the maximum level is based only on the encryption algorithms and it is only as safe as the key. Intrusion detection may contribute to the locking up of files when an intruder is detected, but as we say, the vault is only as safe as the lock, and here it is the encryption key.

We propose this system, where the file is protected in a second level of protection which is the facial recognition layer, where the file will be decrypted only when the face of the user matches that of the owner. Multi algorithm systems use many algorithms for single biometric. The technique such as rank level, score level and feature level matching technique are used to integrate them. It is a cheaper system as there is no extra device required. Comparison of parts based versus holistic image, set based facial recognition versus frame based is done by this method. The system is complicated because of use of secular algorithms.

Multi instance systems employ multiple occurrence of a single biometric. For Ex:- if a person's iris is to be recognized then both the right and left irises are scanned. This system is cost effective, since a single sensor can be used to obtain the images. Multi sample systems capture samples of single biometric feature. For Ex:- In capturing the image of front face, this system captures both the right and left profiles too. The scope of this project extends from single user Home use desktops to multi user shared Mainframe workstations. This can be used by individuals who simply have confidential data in files, and then in multi user systems where in corporate only specific people can access specific files. This can solve the use of different operating system logins for different users, as the users can simply lock it with their face. And also any forms of intrusion can be detected as anybody who tries to open the file will be identified by the owner from their mail.

The increased accuracy in output is far more important than the space and time complexity compromised. The modified component analysis incorporates forced elimination of the training images with higher Euclidean distance values leading o an increase in the accuracy of the recognition of the face. This forced elimination accounts for an increase in the detection of accuracy of the face of the subject. This is due to the virtual decrease in the average of the Euclidean distance of the training image of the subject. So, as the accuracy is increased, this leads to a promising reduction in the false recognition rate.

II. REFERENCES:

Though there are variations in illumination [1] and facial expression, from the recognition rates, it can be found that MSPCA outperforms the other principal component analysis algorithms. SPCA usually performs better than PCA, and MSPCA also performs better than MPCA. This shows that MSPCA enhances its robustness by introducing the L1 and L2 norms penalty regression to select the most important factors/variables for feature extraction. Another reason why MSPCA and MPCA perform better than SPCA and PCA, respectively, is that the structure information embedded in the higher order tensors does enhance the feature extraction abilities of one method. Thus, the higher order tensor extension of a feature extraction method is a tractable way to enhance the performance for feature extraction. In Data Uncertainty in facial recognition [2], the second step devises a good scheme to determine the most useful training samples from the set of all the original and synthesized virtual training samples. The key of this step is to make the determined useful training samples helpful for correctly classifying the test sample. The third step applies the 2-norm based representation algorithm to classify the test sample. We assume that $X_i \in m \times l$ represents the matrix of the

training samples of the i^{th} class, and each column of Xi is a training sample of the i^{th} class. Suppose that there are c classes and n training samples in total. The multi view face recognition [3] is considered as a challenging pattern recognition problem due to its non-linearity. From this point of view, a face image is supposed as a data point in a high-dimensional space. In order to simplify the data presentation and avoid high computations in high dimension, dimension reduction techniques are used to learn the low-dimensional subspaces. We refer to homogeneous illumination [4] as those sets of images with frontal face illumination. In the same manner, we refer to non-homogeneous illumination as those sets of images with asymmetric face illumination generating shadows in different parts of the face. Existing illumination compensation methods are not robust since they may improve face recognition in images with non-homogeneous illumination, but worsen recognition rates in images with homogeneous illumination. In Reference Face Graph framework (RFG)[5] is presented for face recognition. We focus on modeling the task of face recognition as a graph analysis problem. An early work by Wiskott et al. proposed to represent each face by a bunch graph based on a Gabor Wavelet transform. In our approach, the identity of an unknown face is described by its similarity to the reference faces in the constructed graph.

III. SYSTEM ARCHITECTURE

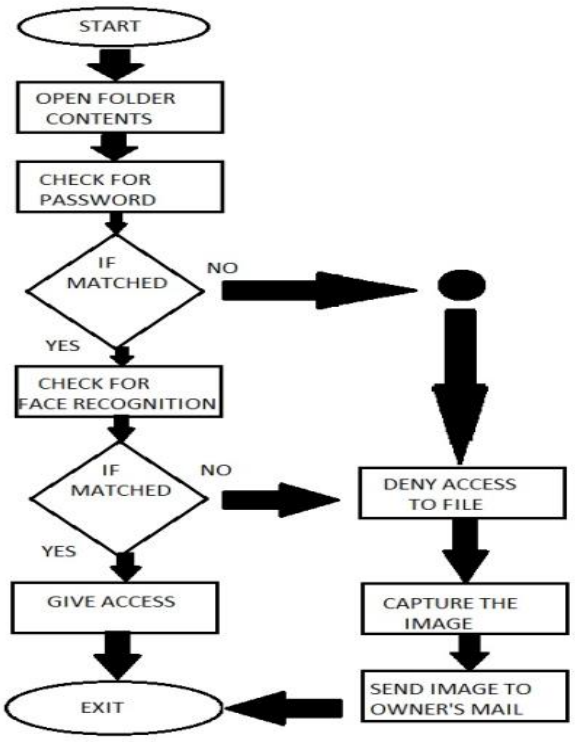

Figure1. System Architecture

When a user needs to access a protected file, the user enters the username and password. When the password does not match, any sort of read or write access to the file is denied. If the passwords match, then the process runs in the foreground, meanwhile in the background the system recognizes the face of the user through the web-cam. If the face matches that of the owner, then the program grants access to the user. If in case the faces do not match, the access is simply denied, and the picture of the user is sent via a server of the software to the mail of the owner of the file, and also an SMS prompt is issued to the phone number of the owner of the file.

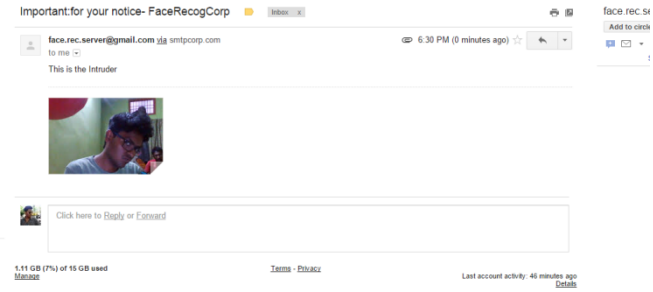

Figure2 E-mail Notifications to Owner of File

Face Image Preprocessing

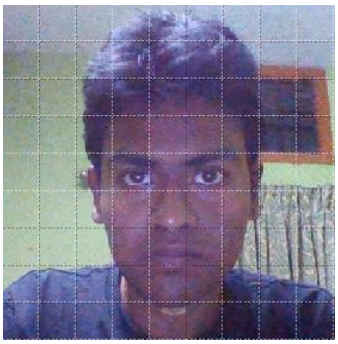

Figure3 Face Image Processing

The image with the face is preprocessed to an optimal level to achieve a smooth and uniform face image. Lighting defects are removed by converting to gray-scale. The size of the image is reduced as per the uniformity constraints specified by the algorithm which is used. The lighting affects the determination of the facial shape, so filters are applied so that the face shape is perfectly extractable.

Binarization

Figure4 Binarization

The preprocessed face image is subjected to binarization, which extracts the features of the face by converting the face into a feature-specific image. Binarization is done by specifying thresholds to specific localities in the image of the face and assigning binary values to the pixels by determining the histogram variations when compared to the threshold.

Face Extraction

The binarized image is subjected to face extraction by making the processed image as a matrix. The matrix values correspond to the pixel intensity values in the coordinate. The size of the matrix is proportinal to the accuracy of the extraction, but it also increases the space and time complexity. The matrix is stored as the extracted face.

The feature extraction is done using the PCA algorithm. Let there be R face images in the training set. Each image Xi is a 2-D array of any size m× n of intensity values. The image can be converted into avector D where D = m× n pixels, and, Xi = (xi1, xi2, ….,xiD). The pixel rows of the image are arranged as per the vector formation. The training set image is defined by X = (X1, X2, …, XR) ⊂ D×R ℜ .

The covariance matrix is defined as follows:

which is the mean image of the training set. The dimension of the covariance matrix Γ is D × D . Then, the eigenvalues and eigenvectors are calculated from the covariance matrix Γ. Let Q = (Q1, Q2, …, Qr) ⊂ D×R ℜ (r

The proposed system eliminates the training image from the training set, whenever an image is found to be matching (euclidean distance is lower than the threshold). The training set is analyzed and the training image with the greatest euclidean distance is flagged everytime. It is eliminated whenever a new image is found to be lower than that of the threshold Euclidean distance.

$$\Gamma = \frac{1}{R}\sum_{i=1}^{R}(X_i - \bar{X})(X_i - \bar{X})^T$$

$$= \phi\phi^T$$

Where $\phi = (\phi_1, \phi_2, \ldots, \phi_R) \subset R^{D*R}$

$$\bar{X} = \frac{1}{R}\sum_{i=1}^{R} X_i - \frac{1}{R}\sum_{i=1}^{R} X_m$$

Neural network architecture

Figure5 Neural Network Architecture

There are three layers in the neural network, namely, input layer, output layer, and a hidden layer in the middle. Each of these layers is composed of neurons and they have weights which can be assigned and modified. The path in which they communicate with each other when a sample face is encountered determines the recognition of the face. A training set consisting of the face values of the owner of the file is submitted to the algorithm. The accuracy grows with the increase in training set.

Face recognition

When a user tries to access the file, the face is compared to the training set using the extracted featured. They are both compared and if the euclidean distance between the neuron is found to be lower than that of the threshold, the face is determined to be the same.

1) For every person, there are N images in the training set. The unknown K (K< N) has to be found which are the subclusters from the image space spanned by the N training images.
2) In the beginning, the whole of the training set is assigned to be N distinct clusters. Let k = N.
3) The inter cluster distance is computed d(i, j) with the help of the equation:
   a. $d(i,j) = ||c_i - c_j||$;I,j=1,2,…..,k
   b. C_i and C_j are the i^{th} and j^{th} clusters which are the Euclidean standard.
4) The two nearest clusters C_i and C_j are computed by the equation:
   a. $d_{min}(i,j) = \arg \quad min\{d(i,j)\}; i,j = 1,2,\ldots\ldots,N, i \neq j$
5) The average of two clusters is found and a new cluster is formed and the value k is set as k = k-1.
6) Steps 3 and 5 are repeated until k becomes K.
7) All the steps are repeated for all the other subjects in the training set.

IV. Comparison With Traditional Principle Component Analysis

Analysis, the Modified component analysis showed significant levels of increase in the recognition accuracy over a training set of ten people with random training images and test images varying from Straight, Tilted, Turned and Low light conditions. Excepting low light conditions where the accuracy level matches with that of the PCA, all other conditions showed an average increase of accuracy. The straight face condition saw an average increase from 8.8 – 9.1, tilted faces 6.8 – 7.8 and turned faces from 8.7 – 8.9. Low light faces remained 7 for both the algorithms. When compared with traditional principle Component

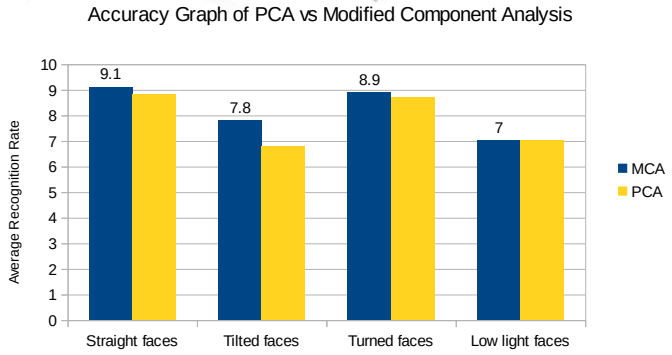

Figure 5 Comparison Of MCA With PCA

V. Scope

The scope of this project extends from single user Home use desktops to multi user shared Mainframe workstations. This can be used by individuals who simply have confidential data in files, and then in multi user systems where in corporate only specific

people can access specific files. This can solve the use of different operating system logins for different users, as the users can simply lock it with their face. And also any forms of intrusion can be detected as anybody who tries to open the file will be identified by the owner from their mail.

VI. CONCLUSION

The existing machine learning approach for facial recognition fails to settle for more, as intelligence incorporated does not allow forced memory loss, as it is architecture like our natural neural network. The increased accuracy in output is far more important than the space and time complexity compromised. The modified component analysis incorporates forced elimination of the training images with higher Euclidean distance values leading o an increase in the accuracy of the recognition of the face.

VII. REFERENCES

[1]Zhihui Lai, Yong Xu, Qingcai Chen, Jian Yang, Member, IEEE, and David Zhang, Fellow, IEEE "Multilinear Sparse Principal Component Analysis" IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, vol. 25, no. 10, october 2014.

[2] Yong Xu, Member, IEEE, Xiaozhao Fang, Xuelong Li, Fellow, IEEE, Jiang Yang, Member, IEEE,Jane You, Hong Liu, and ShaohuaTeng "Data Uncertainty in Face Recognition" IEEE transactions on cybernetics, vol. 44, no. 10, october 2014.

[3]HadisMohseniTakallou, ShohrehKasaei "Multiview Face Recognition Based On Multilinear Decomposition And Pose Manifold" Published in IET Image Processing Accepted on 14th February 2013.

[4]L.E. Castillo, L.A. Cament, F.J. Galdames and C.A. Perez "Illumination Normalisation Method Using Kolmogorov-Nagumo-Based Statistics For Face Recognition" ELECTRONICS LETTERS 19th June 2014 Vol. 50 No. 13 pp. 940–942.

[5]MehranKafai, Member, IEEE, Le An, Student Member, IEEE, and BirBhanu, Fellow, IEEE "Reference Face Graph For Face Recognition" IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, vol. 9, no. 12, december 2014.