Face Recognition Techniques: A Survey for Forensic Applications

Ritika Lohiya, Pooja Shah

***Abstract*— A wide variety of systems requires reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. Face Recognition is the most natural process for human beings and therefore it is the most natural biometric technique. It is widely used in the areas like information security, access management, law enforcement, criminal investigation etc.**

The goal of the paper is to have a comprehensive analysis of facial recognition in forensic investigation. It discusses the role of face recognition in investigation and law enforcement applications, with a brief background and origin of face recognition in forensic investigation. The paper also describes the applications of face recognition, a generic automated face recognition system and face recognition algorithms.

***Index Terms*— face recognition, forensic investigation, applications, face recognition algorithms.**

I. Introduction

Human brain and eyes are the best existing face recognition approaches, for obvious reasons. Moreover, according to science there is a specific area of the human brain which is called as fusiform face area (FFA) which is a part of the human visual system, it is speculated and is specialized for face recognition.
Face recognition is the most natural thing for human beings and therefore it is the most natural biometric technique. Face recognition is widely used for information security, access management, biometrics, law enforcement and personal security.
There are many applications of face recognition but automatic face recognition is actually used in civil spheres like identification of documents or in law enforcement applications like tracking the suspect from the recorded CCTV footage. The wide use of face recognition in these fields is only because all these applications demands uniqueness. The table below gives the details about the various areas and the respective applications where face recognition plays an important role.

Area	Applications
Information Security	Access Security, Data Privacy, User Authentication
Access Management	Secure Access Authentication, Permission based System, Access log or audit trails.
Biometrics	Person identification, Automated identity verification.
Law enforcement	Suspect Tracking, Video Surveillance, forensic reconstruction of faces from remains.
Personal Security	Home video surveillance systems, Expression interpretation.

Table 1: Applications of Face Recognition

Since the trend of photography has started, government firms as well as private companies have kept the photos of the people to ensure security and for record keeping purpose. Also now photographs have a great impact on personal identifications from passport to identity cards of any firm. Facial recognition was already a great deal of research before the computerized method of identifying faces came into picture. For instance:
- Earlier suspect identification techniques were used in which witness was con-fronted to a group of similar individuals and he/she had to identify the suspect among them.
- Also there were methods of making sketches and portraying the persons who were arrested repeatedly.
This paper discusses about face recognition in general, face recognition in forensic science investigation, various approaches of face recognition.

II. Origin and Background

With the commencement of photography, both government organizations and private agencies have started keeping collections of photo ID for personal identification. Photo ID are used not only for official passports at airports but also for the informal membership cards which are used for sports clubs.
Before the use of computer vision for identification and tracking human faces began, face recognition was already a subject to great deal of research. For instance:

In 1853, a criminologist developed a judicial anthropometry in France, which was referred to as “line-up” techniques in

***Manuscript received April, 2015*.**
Ritika Lohiya, *Computer Science and Engineering, Nirma University, ., Ahmedabad, India, +91-8866438698*
Pooja Shah, *Computer Science and Engineering, Nirma University, Ahmedabad, India, +91-9825434173.*

United Kingdom, in which witness is confronted to a group physically look alike people from which one of whom is suspect. By observing the group of people the witness has to decide whether one of the person from the group was present at the crime scene or not.

A famous criminologist introduced a system based on face classification which was widely accepted by the American law. This system was invented in 1879 and was known as bertillonage and quickly gained wide acceptance as a reliable and scientific method for criminal investigation. In order to recognize the suspects who were repeatedly arrested, portraits were stored sorted by common morphological characteristics that are the specific shapes of the different parts of the face. These stored portraits are now referred to as 'mug shots' and are even used today.

During 1964 and 1965, pioneers of automated facial recognition Woody Bledsoe, Helen Chan Wolf and Charles Bisson used computer for face recognition. They used a large set of databases of images and a photograph to be recognized among the image records.

On his work, Bledsoe quoted that, " The recognition problem is made difficult by a great variability in head rotation, illumination, angle, facial expression, aging etc. Some other attempts at facial recognition by machine have allowed for little or no variability in these quantities. Yet the method of pattern matching of unprocessed optical data, which is often used by some researchers is certain to fall in cases where the variability is great. In particular, the correlation is very low between two pictures of the same person with two different head rotations."

A. Background

The first attempt to automate the face recognition started in 1960 it was basically a semi automatic face recognition system. The system checked the coherence of measurements between different characteristics points of the face like distance between the eyes, hairline etc. But the system did not responded well with different orientation of the face.
And some of the researchers like A. Jay Goldstein, Leon D. Harmon [1] used subjective face features like between the eye distance etc to recognition the face. They defined a vector containing a set of 40 subjective features as the basis to recognition faces using pattern classification techniques. Some researchers like Fischler and Elschanger also worked on the similar feature based approach. In 1973 a researcher Kenade,[3] proposed a fully automated face recognition system. The program extracted sixteen facial parameters with small difference.

Early in 1980, Mark Nixon[4] implemented a method following the basis of earlier research and used eye-spacing as a prominent feature for recognition. Also, in the decade ahead many approaches were developed using artificial neural networks. In 1986, L Sirovich and M Kirley[5] proposed the technique of "eigenfaces" which led a remarkable impact in research ahead and even today. In this approach, they used principle component analysis to represent the image dataset in lower dimension and then reconstruct it. The eigenface approach laid the foundation for many new algorithms in face recognition.

Researchers found that the overall issue of facial recognition was complex but could be solved by taking into consideration images that are coherent in terms of pose, illumination, expression and resolution.

III. Applications of face recognition

Face is a complex multidimensional structure which needs good computing approaches for recognition. The face is our primary way, playing an important role in identifying an individual in the society. We as humans are capable of identifying a number of faces learned throughout our life span and identify that faces even after years. There may be variations in the faces of individual due to ageing, distractions like lighting, pose, or due to physical changes like beard, glasses, change in hairstyles etc. Face recognition is an important part of biometrics. In biometrics basic traits of human is matched to the already available data and depending on the result obtained identification and matching is done. Facial features are extracted and different algorithms are applied which are efficient and face recognition systems are build which are used for detection and identification.

Criminal Investigation and Identity checks

Face recognition just like fingerprint recognition helps the investigation team to manage files of the criminals and keeps track of the suspects that have different records over the years.

Fingerprint recognition and analysis has been performed over the years but facial recognition provides more benefits: Face recognition enables identification of individuals covering large population whose fingerprints cannot be acquired for various reasons.

By using face recognition with fingerprint analysis, will provide superior performance reducing the workload in verification process.

Also in face recognition no human participation is required. While in the case of Identity checks a suitable camera is needed to capture a person's identity. Most of the police officers are equipped with computers and tablets provided by the government, so they can make a quick search of an individual against any previous criminal records. Also these can also be presented for law enforcement.

Information Retrieval

Images are often available for inquires through surveillance videos, cameras, internet sites and also from the police records. The images can help in the identification process of the suspect.

Information about the suspect identity could be extracted from the available evidence. For instance, hundreds of hours of video footage from the CCTV cameras are analyzed manually in which suspect's faces are available. And to identify the face from the footage is time consuming and painstaking task. And this is the reason behind automated facial recognition.

Civil Applications

Face recognition can be used to check the uniqueness of the identity documents. In non-criminal context, it is quite easy to provide a photograph while to take fingerprints or image of an iris. For instance driving licence consists of the photograph of the holder rather than his fingerprint or any other biometric data. Other applications like travel documents, membership cards, election card, passport etc have face as the prime means for the person's identification.

Access Control

Access control means 'authorization' i.e to check anyone attempting to access a secure zone is entitled to do so. Face recognition can be used to check and verify a person's identity in highly control and sensitive environmental condition to grant access to the person.

IV. GENERIC FACE RECOGNITION PROCESS

Acquisition of the image.

In this step, the system may assess the resolution and quality of the image acquired. If the quality of the image acquired is not good, then re-acquisition of the portrait is done in order to obtain better quality image. This step is decisive, as the precision of the face recognition depends on the quality of the image acquired.

Localization and alignment

Before identification and verification of the images it is necessary to make sure that the images contain all sorts of information and all the images are aligned to same scale and positioned uniformly. This step is too simple with images that are still and with frontal view of the images. But it becomes too complex with the images in motion and distorted images.

Image Enhancement

Once the image has been properly calibrated, they need to be enhanced. For instance, the effects of compression can be reduced, illumination problem can be corrected etc. In this step, face images can be arranged with proper orientation, aging and expression effects.

Extraction of features

Face recognition algorithms use many mathematical transformations to compare the images and these transformations can highlight prominent features of the face.

Comparison

A template is extracted from the transformed image and then comparison function compares this template against the reference images stored in the database and scores each image. Higher the score, higher similarity with image of the wanted face.

V. FACE RECOGNITION ALGORITHMS

The face recognition algorithms includes procedural algorithms which carries out the steps decided by the operator and training algorithms through which mathematical logic is applied in order to define and use the criteria decided by the operator. These algorithms could be applied to images, videos and 3D images.

A. Training Algorithms

These algorithms are used not only in step 5 for comparison but also in step 2 to extract visible facial remarks and locate face images.

Principle Component Analysis

Faces in the images are viewed as high dimensional pixel arrays, and therefore in statistical approaches each image is considered as a point vector in d-dimensional space. Therefore if we have a 100*100 pixel image it would require 10000 pixel space which demands a high amount of storage space. This leads to apply some statistical approach to reduce the dimension of the space and get better result. One such method was introduced by Karl Pearson in 1901 called Principle Component Analysis. Principal Component Analysis is dimensionality reduction method which is used when there is a high scope of redundancy in the data available. By applying the principle component analysis reduction method to the data, the variables would be reduced to the small number of components called as principle components which will contain most of the information needed for recognition process which can also be referred to as variance. In principle component analysis the dataset of images available is simplified and arranged according to the order of variance available. For instance the dataset having the greatest variance are called the first principle components, the dataset having the second maximum variance are said to be the second principle components and so on. And so the main idea behind the PCA is to keep the dataset with the most variance i.e which gives the maximum amount of information available.

The input images which we get may contain a high amount of noise in terms of distractions like lighting, pose, occlusion, distortion etc thus these input images with noise are classified into signals and patterns are created where ever eyes, nose, mouth etc are detected. These are referred to as eigenfaces or principle components for the given set of input images. Now as eigenface for each image in the training data set is created, we can now use the eigenfaces to reconstruct the original image again.

However by considering only few eigenfaces an approximate image can be constructed. Thus the main idea behind PCA if to reduce the high dimensional face space into low dimensional feature space. Consider for instance we have a large set of digital images which have same environmental conditions, same orientation and have same resolution m*n. Then by applying principle component analysis this m*n pixels images are converted to mn-dimensional vectors which whose components represent the pixels values of the images. Thus extraction is done by considering these vectors of the face images and hence by the statistical distribution of the covariance matrix of eigenvectors, eigenfaces are generated. Currently there face recognition algorithms which are used, use face representations generated by the unsupervised statistical methods. The main step of these methods lies in generating a set of basis images which are linear combination of original images. Principal Component Analysis (PCA) is one such type of algorithm. The basis set which is found by PCA is dependent on the relationship between the pixels of the images in the dataset. In face recognition important information lies in high-order relationships among pixel. Therefore it becomes necessary to expect that better basis images should be found by the methods sensitive to these high order statistics. Independent Component Analysis (ICA) which is generalization of PCA is one such method.

Linear Discriminant Analysis

The most famous example of dimensionality reduction is ”principal components analysis”. This technique searches for directions in the data that have largest variance and subsequently project the data onto it. In this way, we obtain a lower dimensional representation of the data, that removes some of the “noisy” directions.

Both Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are linear transformation techniques that are commonly used for dimensionality reduction. PCA can be described as an "unsupervised" algorithm, since it "ignores" class labels and its goal is to find the directions (the so-called principal components) that maximize the variance in a dataset. In contrast to PCA, LDA is "supervised" and computes the directions ("linear discriminants") that will represent the axes that maximize the separation between multiple classes[6].

Linear Discriminant Analysis (LDA) is an dimensionality reduction algorithm which is applied as a pre-processing step in many pattern recognition and machine learning applications for instance face recognition. The main aim of the algorithm is to create a dataset with a low dimensional space and good separability features in order to avoid overfitting ("curse of dimensionality") as well as to reduce computational costs.

Ronald A. Fisher formulated the *Linear Discriminant* in 1936 (The Use of Multiple Measurements in Taxonomic Problems)[7], and it also has some practical uses as classifier. The original Linear discriminant was described for a 2-class problem, and it was then later generalized as "multi-class Linear Discriminant Analysis" or "Multiple Discriminant Analysis" by C. R. Rao in 1948[8] (The utilization of multiple measurements in problems of biological classification)

The general LDA approach is very similar to a Principal Component Analysis (for more information about the PCA, see the previous article Implementing a Principal Component Analysis (PCA) in Python step by step), but in addition to finding the component axes that maximize the variance of our data (PCA), we are additionally interested in the axes that maximize the separation between multiple classes (LDA)[6]. So, in a nutshell, often the goal of an LDA is to project a feature space (a dataset n-dimensional samples) onto a smaller subspace k (where $k \leq n\text{-}1$) while maintaining the class-discriminatory information[6].

In computerised face recognition, each face is represented by a large number of pixel values. Linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template.

Neural Networks

Neural Networks (NN) have been widely applied for various pattern classification tasks, including face detection and recognition. The Multi-layer Perceptron (MLP) model was used for the first applications, such as the face detection systems by Vaillant, Monrocq, and Le Cun (1993)[9], Burel and Carel (1994)[10] and Juell and Marsh (1996)[11]. Rowley, Baluja, and Kanade (1998a)[12] suggested a more advanced system that employed a neural network to learn from 20 × 20 pixel image patches using a hidden layer with 26 units with retinal (spatially related) connections to smaller subregions (10 × 10, 5 × 5 and 20 × 5 pixels) of the image patch. The system also employed heuristics to cope with overlapping detections and multiple networks were combined using an arbitration strategy for improved performance. Rowley, Baluja, and Kanade (1998b)[13] further improved their system to deal with faces rotated on the image plane. Rapha¨el, Olivier, and Daniel (1997)[14] introduced a neural network classifier based on the Constrained Generative Model (CGM) that attempts to build a non-linear description of the face distribution, constrained by non-faces obtained from an iterative bootstrap training algorithm (where false positives from previous iterations become non-face training data for the following iterations). Similar to the PCA/eigenface technique, classification of a candidate image is based on the reconstruction error of the CGM. Another learning architecture, known as SNoW (Sparse Network of Winnows), came from the domain of natural language processing and was first applied to face detection by Yang, Roth, and Ahuja (2000). SNoW architecture is tailored to work with large numbers of features and consists of a sparse network of linear functions, where the target class labels are represented as linear functions over a common feature space. Yang et al. defined a binary feature space, where every index represents a unique location and pixel intensity combination: for 20 × 20 images with 256 possible pixel intensities, there are 400 × 256 = 102400 different features. Naturally, in any given image only 400 of those features will be active, and this sparseness ensures the efficiency of the algorithm. Two linear threshold units representing faces and non-faces respectively were trained with a simple Winnow update rule that promotes and demotes weights according to a mistake-driven policy.

Adaboost

AdaBoost, an Adaptive Boosting machine learning technique is considered one of the most successful object detection methods in computer vision. Introduced by Freund and

Schapire (1995)[15], AdaBoost is an improvement on the many previous boosting approaches. It is a meta-algorithm in a way that it uses some other weak classification algorithm to build a series of classifiers and combine them in a way to achieve strong classification. The series of weak classifiers is constructed in such a way that misclassified training examples are given more weight in the subsequent steps. Each classifier in the series also gets an importance weight based on how accurate it is. The final AdaBoost classifier is then an importance-weighted majority vote of the weak classifiers. AdaBoost has been successfully used to boost different classifiers, including perceptrons, PCA and LDA based classifiers, linear support vector machines and others. Viola and Jones (2002)[16] successfully applied AdaBoost to the object detection problem, combining it with a number of other refined techniques to achieve real-time performance in a face detection example application. Firstly, instead of dealing with pixel intensities in images directly, they used Haar-like features—differences between the sum of pixels in adjacent rectangular sub-areas of an image. Such features can be very rapidly calculated at different scales from an integral image, an array of sums of all pixel intensities to the top and to the left of a given pixel. While the number of potential Haar-like features in an image window is potentially very large (over 45K different features in the example 24 × 24 pixel window with 4 defined configurations of adjacent subregions), a very small set of such features is chosen to build the classifier using AdaBoost. The weak classifier for AdaBoost is a simple decision rule that finds a single feature and a threshold for its value that performs classification on the entire training set. The resulting series of classifiers then contain the most discriminative features, and as few as 200 features where sufficient for a classifier with 95% accuracy. Furthermore, Viola and Jones employed a classifier cascade technique to further reduce the run-time complexity of their method. Classifier cascading is an effective technique for object detection that targets the imbalance between the amount of object and non-object candidate images in a typical application, where an absolute majority of image windows contain background. A sequence of less accurate but more efficient classifiers are biased to have high confidence in discarding the non-object images, passing only the potential object windows to the next level. This way, most of background windows in a candidate image can be cheaply discarded, saving time for the hard candidate windows that reach the full classifier at the end of the cascade.

Support Vector Machine

First introduced by Vapnik, Boser, and Guyon (1992)[17], support vector machines (SVM) quickly became a widely used machine learning technique. It is the pioneering and arguably the most popular of the kernel methods, the class of algorithms that map data points into very high-dimensional feature spaces, where even complex object classes become highly separable. Direct operation on the mapped images of data points is avoided by the use of kernel functions that represent the dot-product operator in the feature space. This kernel trick can be used with different statistical methods; good examples are kernel PCA and kernel LDA. The core idea of SVMs is intuitive and simple: given a set of points belonging to two linearly separable classes, find a separating hyper-plane such that the margin (the distance from data points to the hyper-plane) is maximized. The kernel trick is then applied to this linear classifier: the points being linearly separated are actually images of original data points in the high-dimensional feature space. The motivation for this is the fact that while a set of N points can't generally be linearly separated in less than N dimensions, it can usually be separated if the number of dimensions is larger than N. Since with kernel methods little is usually known about the target space apart from the kernel function that induces it, arbitrary points cannot be calculated in this target space. Therefore the SVM algorithm limits the search for the best separating hyper-plane (whose orientation is defined by a vector in the target space) to only the linear combinations of the images of original data points. In other words, it finds a separating hyper-plane expressed in terms of the weighted combination of data point images it separates. Due to the non-linearity of the mapping, in the original data space the separating boundary can take a complex form to fit between the two classes of data points.

Osuna, Freund, and Girosi (1997b)[18] were first to apply the SVM technique to face detection. They used 19×19 pixel image patches containing faces and non-faces and a second degree polynomial kernel function to train SVM classifiers to achieve accuracies marginally exceeding performance of top systems at the time, with significant 30 fold increase in run-time speed achieved by a SVM approximation technique suggested by Burges (1996). More recently SVMs saw further improvements in attempts to improve their effi- ciency. Romdhani, Torr, and Sch¨olkopf (2001, 2004) built upon Burges (1996) approximation method to build efficient SVM classifier cascades using 20×20 pixel images and the Gaussian Radial Basis Function (RBF) kernel. This approach gained another 30 fold speed improvement and was comparably accurate even without any normalisation of images.

In the light of the very efficient AdaBoost face detection system of Viola and Jones, Ratsch, Romdhani, and Vetter (2004) attempted to further improve the SVM cascaded classifier. The run-time bottle-neck being the large 20 × 20 pixel vectors that the decision functions performed dot-product operations on, the new approach was to employ the integral image technique and force the approximated support vectors in the classifier cascade to have a Haar-like rectangular structure that allowed for fast dotproduct calculation and avoided the need for an image pyramid to deal with different scales. Overall, the Haar-like feature SVM cascade was 6 times faster than the raw pixel one and remains one of the state of the art face detection systems to date.

B. Procedural Algorithms

Number The main goal behind the procedural algorithms is to find out the visible facial landmarks for instance eyes, lips, lowest point of the chin or the colour of the skin etc. After detecting the marks on the face, the next step is to measure the coherence between the two faces. And all these steps are demonstrated by using models which are designed to know that how much a face is affected by expression, age, orientation and lighting.

These algorithms are mainly used in the step 5 discussed above to convert the images into templates for comparison.

Eigenface Approach

Most of the earlier work on facial recognition has ignored issues of what aspects of the face are important for identification. This led Mark and Pentland in 1991[19] to introduce an information theory approach of coding and decoding face images. The approach mainly focuses on the significant features of the face such as eyes, nose, lips etc. This means that their system wanted to extract relevant information from a face image, encode it as efficiently as possible and compare it against the database of faces.

Mathematically, eigenface approach was to find out the principle components or eigen vectors of the face distributed over a database. These eigen vectors can be thought of as a set of features that together characterize the variation between the face images. Every image in a database contributes more or less to each eigenvector which can then be used to display a ghostly face which is referred to as 'eigenface'. Every face can be represented in terms of the linear combination of the eigenfaces and a face can be identified using the best eigenfaces which means those that have largest eigenvaluesand which account for more variance in the set of face images.

FisherFace Approach

Eigenfaces approach uses Principal Component Analysis (PCA) to find the linear combination of features that generates the maximum variance in data. But this approach has a limitation, though it represents the data perfectly but it does not consider classes for discrimination. And therefore a lot of information is lost when throwing components away. For instance if the variance of dataset is generated by using an external source, say light. Then in that situation the components identified by a PCA do not necessarily contain any discriminative information. As a result the projected sample mix together and classification becomes impossible.

Sir R. A Fisher, the great statistician introduced Linear Discriminant Analysis which performs class-specific dimensionality reduction. This property was mainly used for classifying flowers in his paper *The use of multiple measurements in taxonomic problem* (1936)[7]. Linear Discriminant Analysis uses the ration of between class and within class scatter in order to find out the combination of features that separates best between the classes. By doing this it avoids maximizing the overall scatter. This means that same classes cluster together and different classes are as far away as possible from each other in the lower-dimensional presentation. This concept was also presented by Belhumeur, Hespanha and Kriegman and so they applied a Discriminant Analysis to face recognition in[20].

Fisherfaces method has an advantage over the Eigenfaces method. It uses the class-specific transformation matrix to overcome the limitation of illumination. On the other hand Discriminant Analysis finds the facial features to discriminate between the individuals. The most important thing to note here is that the performance of Fisherfaces heavily depends on the input data. This means that practically if we apply Fisherfaces for well illuminated images only and if we recognize faces in bad-illuminated scenes then method is likely to find the wrong components (just because those features may not be predominant on bad illuminated images).

Local Binary Pattern Histograms

Eigenfaces and Fisherfaces are holistic approaches to face recognition. In Eigenface method data is treated as a vector in a high-dimensional subspace. To overcome the limitations of high dimensional subspace, a lower dimensional subspace is used where probably useful information is preserved. In Eigenface approach total scatter is maximized which leads to problem if the variance is generated by an external source. This is because the components with maximum variance over all classes are not useful for classification.

Local Binary Pattern (LBP) is a texture operating method. In this method, pixel of the image are labelled by thresholding the neighbourhood of each pixel and generates the result as a binary number. LBP texture operator method has proved to be very efficient because of its discriminative power and computational simplicity. It has become a monopolized approach to the conventionally accepted divergent statistical and structured models of texture analysis. Moreover, in real-world applications, the most important property of LBP operator is its robustness to monotonic gray scale changes which are mostly caused by illumination variations.

For texture classification in LBP, the occurrences of the LBP codes in an images are stored into a histogram. The classification is then done by computing simple histogram similarities. Also, while considering this for face image representation it leads to loss of spatial information and therefore codifying the texture information while retaining their locations becomes necessary. One way to do this is to use LBP texture descriptors to build several local descriptors of the face and then combine them into global descriptors. These method of local feature based methods are more likely to be used than holistic methods because of their robustness against variations in pose and illumination.

LBP based face description method was proposed by Ahonen et al.(2006). The method includes dividing the face image into local regions and then LBP texture descriptors are extracted from each region independently. These local descriptors are then contracted to form a global descriptor of the face.

The description of the face is represented on three levels of locality in the histogram: first is the pixel level, which has LBP labels for the histogram that contain the information about the patterns. Second is the regional level which has labels that are summed over a small region. And lastly the regional histograms are concatenated to build a global descriptor of the face.

The biggest advantage of using histogram based method is that it is not necessary that the regions should be rectangular. Also it is irrespective of their shape and size and they need not cover the whole image. It can also be applied to partially overlapping regions.

VI. Conclusion

This paper has presented a survey about face recognition. It is not a trivial task and today there are many open issues like face expression, pose variation, aging etc which remains unresolved. Following are some current lines of research:
New or extended feature extraction methods. There is a lot of research being done about extending and improving well known algorithms like PCA, LDA etc. Also these feature extraction methods could be combined for instance LDA can be combined with Support Vector Machines to overcome the limitations of small sized samples.
The paper has also discussed about the applications of face recognition, a generic face recognition process and it algorithms.
Face recognition is also emerging in other areas as an important application like expression recognition and body motion recognition. Therefore it is not only and unresolved problem but also source of new applications and challenges.

Acknowledgment

This paper was taken as part of study for face recognition techniques and forensic science. I would like to express my sincere thanks to Head of Department Dr. Sanjay Garg and project guide Prof. Pooja Shah in Computer Science Engineering Department, Institute of technology, Nirma University for many fruitful discussions and constructive suggestions throughout.

References

[1] Goldstein, A. Jay, Leon D. Harmon, and Ann B. Lesk. "Identification of human faces." *Proceedings of the IEEE* 59.5 (1971): 748-760.
[2] Goldstein, A. Jay, Leon D. Harmon, and Ann B. Lesk. "Man-Machine Interaction in Human-Face Identification." *Bell System Technical Journal* 51.2 (1972): 399-427.
[3] Zhao, Wenyi, et al. "Face recognition: A literature survey." *Acm Computing Surveys (CSUR)* 35.4 (2003): 399-458.
[4] Nixon, Mark. *Feature extraction & image processing*. Academic Press, 2008.
[5] Everson, Richard, and Lawrence Sirovich. "Karhunen–Loeve procedure for gappy data." *JOSA A* 12.8 (1995): 1657-1664.
[6] http://sebastianraschka.com/Articles/2014_python_lda.html
[7] Fisher, Ronald A. "The use of multiple measurements in taxonomic problems." *Annals of eugenics* 7.2 (1936): 179-188.
[8] Rao, C. Radhakrishna. "The utilization of multiple measurements in problems of biological classification." *Journal of the Royal Statistical Society. Series B (Methodological)* 10.2 (1948):159-203.
[9] Osadchy, Margarita, Yann Le Cun, and Matthew L. Miller. "Synergistic face detection and pose estimation with energy-based models." *The Journal of Machine Learning Research* 8 (2007):1197-1215.
[10] Burel, Gilles, and Dominique Carel. "Detection and localization of faces on digital images." *Pattern Recognition Letters* 15.10 (1994): 963-967.
[11] Juell, Paul, and Ron Marsh. "A hierarchical neural network for human face detection." *Pattern Recognition* 29.5 (1996): 781-787.
[12] Rowley, Henry A., Shumeet Baluja, and Takeo Kanade. "Neural network-based face detection." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20.1 (1998): 23-38.
[13] Rowley, Henry A., Shumeet Baluja, and Takeo Kanade. "Rotation invariant neural network-based face detection." *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*. IEEE, 1998.
[14] Raphaël, Feraud, Bernier Olivier, and Collobert Daniel. "A constrained generative model applied to face detection." *Neural Processing Letters* 5.2 (1997): 11-19.
[15] Freund, Yoav, and Robert E. Schapire. "Experiments with a new boosting algorithm." *ICML*. Vol. 96. 1996.
[16] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, 2001.
[17] Smola, Alex J., and Bernhard Schölkopf. "A tutorial on support vector regression." *Statistics and computing* 14.3 (2004): 199-222.
[18] Osuna, Edgar, and Federico Girosi. "Reducing the run-time complexity of support vector machines." *International Conference on Pattern Recognition (submitted)*. 1998.
[19] Turk, Matthew, and Alex Pentland. "Eigenfaces for recognition." *Journal of cognitive neuroscience* 3.1 (1991): 71-86.
[20] Belhumeur, Peter N., Joao P. Hespanha, and David J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection." *Computer Vision—ECCV'96*. Springer Berlin Heidelberg, 1996. 43-58.

Ritika Lohiya is pursuing her M. Tech. degree in Computer Science and Engineering with specialization in Information and Network Security from Nirma University, India. Her area of interest includes Cryptography, Intrusion Detection System, Cyber Forensics, Face Recognition etc. She has published 2 articles in International Journals in this domain.

Prof. Pooja Shah is currently pursuing her Ph.D in the domain of image processing under Nirma University Ahmedabad. She has published 5 articles in International journals in this domain and in total 12 articles in various International journals and conferences. She is member of IEEE and life member of ISTE. Her research interests are inclined towards image/video based forensics and oceanographic image processing.

