

Random Forest and SVM Based Face Recognition Using Subspace Methods

Ali Asghar Khosravi^{1,a}, Ali Pourmohammad^{2,b,*} and Amir Amini^{1,c}

¹Department of Electrical Engineering, Faculty of Engineering and Technology, West Tehran Branch – Islamic Azad University, Tehran, IRAN

²Electrical Engineering Department, Amirkabir University of Technology, Tehran, IRAN

a. Ali.kh1358@gmail.com, b. pourmohammad@aut.ac.ir, c. amini@wtiau.ac.ir

*corresponding author

Keywords: Support Vector Machine, Random Forest, Subspace Methods, PCA, LDA.

Abstract: Face recognition is one of the methods used to identifying the people. Due to its ease of use, this method has been used in recent decades. This method is more user friendly than other people identifying methods through iris, retina, fingerprint, etc. Because of the lack of cooperation of the person being examined, the face recognition method is more acceptable than others. In this paper, using one of the Subspace Methods as the face attribute extractor and applying its pre-processing technique as the initial stage of the face recognition system, has been investigated for increasing the face recognition rate, and extraction of a certain aspect of the face has led to improvement in the correctness of the diagnosis. The subspace algorithm, by highlighting the features needs to identify and remove unnecessary information, reduces the volume of computations and increases the speed of detection. Then 80% of the data is trained by support vector machine and random forests, both of which are classifiers, and tested on 20% of the data. In this paper, due to changes in its pre-processing and changes in the way of extracting the characteristics, the results are obtained efficient.

1. Introduction

In recent years, face recognition has become one of the most important pattern recognition applications. Various scholars in many parts of the world are engaged in research in this field [1][2]. Engineers continue their research on facial recognition machines and how to implement them from a computational perspective [3]. Facial recognition is important because of its use in ATMs, smart door buildings, law enforcement in finding offenders, matching images with people at airports, etc. After the twin towers in the United States, governments around the world paying more attention to the level of security at their airports and outlets, the country's annual budget has intensified to reach the latest technology in detecting, identifying and tracking suspects. Increasing demand for these applications has helped researchers reap the cost of their research projects. There are many biometric methods for automatic processing and identification of individuals by their physiological or behavioural characteristics [4]. Some of these validation technologies include face recognition,

fingerprint identification, handwriting recognition, eye iris detection, voice recognition, signature recognition, retina detection, DNA matching among individuals, and facial recognition [2]. Although facial recognition is not as accurate as other methods of detecting, such as fingerprints [2]. The attention of researchers has still attracted many researchers in the field of machine vision. In fact, the main reason for this interest is that the face is a common way for people to identify each other and that facial recognition systems do not need to interfere with each other [5].

In the pre-processing section of this study, the first step is to remove the darkness of the image as much as possible by using the gamma correction and using the Gaussian function to try to highlight the edges and remove the local darkness. In the extraction section, the output image of the pre-processing section is reduced by using the combination of LDA + PCA Dimensions are given and in the final section by using two methods of random forage and support vector machine.

2. The Algorithm

2.1. Pre-Processing

Facial recognition is a vital role in many applications, such as face detection, which is very useful, cost-effective and also inexpensive to identify an offender. A face recognition system should recognize the face in different light and bright conditions in the image. An optimal human face detection system typically uses several integrated techniques. These techniques include normalization, feature extraction, and classification. The input image is pre-processed in this way, and during this pre-processing of exposure, shadows, lighting effects using pre-processing techniques apply without affecting the facial features. In the next step, the attributes are extracted from the prepared image [2].

2.2. Converting a Colour Image to Grayscale

A colour image has been converted to a grayscale image using the density level, so pixels can be from 0 to 255. The reason why the colour image is not intended to be detected is that there is a clutter of processing to identify the pixel location in this image. Also, the size of this image is higher than the grayscale image, and it also prevents information saturation [1][2].

2.3. Gamma Correction

Gamma correction is a gray-gray non-linear transformation used to modify power law. Applying a modified gamma turns the image into an empty image of the dark. This is done by compressing dark areas in bright areas. This relationship is defined as follows.

$$S = i^{\text{gamma}} \quad (1)$$

In relation (1), i represents the brightness of the input image pixels and S is the brightness of the image output pixels. This action converts the gray level i to i^{gamma} , assuming that the gamma is greater than zero, but its range is between 0 and 1 [1][2]. Figure (1) shows an example of gamma correction. The main and general rule of gamma correction is that the light reflection density of an object is the product of the input brightness and superficial reflection. Here the image obtained after gamma correction, the image is almost dark. The gamma value, which is considered between zero and one, is based on the results of the gamma value in this study. Figure 2 shows the image and histogram before and after gamma correction [1][2].

2.4. Gaussian Difference

Gamma correction is used to remove darkness, but gamma correction cannot completely eliminate darkness. Local shadows can be removed using a high pass filter, which helps to simplify detection. The high pass filter weakens the low frequencies while passing through the high frequencies. So the edges will be highlighted in the image and the local shadows will be deleted. Gaussian Filter is a special analytic tool that is easily customizable. Using a Gaussian function of the output image, an image is useful. And, using the difference between two Gaussian filters, the

contrast of the image improves according to local contrast information. The Gaussian filters used with the variances of Sigma 1 and Sigma 2 are considered to be small Sigma 1 than Sigma 2. Therefore, gamma correction and DOG filters produce a useful image [6][7]. This relationship is as follows.

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x^2}{2\sigma^2}\right] \quad (2)$$

By using equation (2), the edges of the image are highlighted, and the use of the difference in the two Gaussian filters makes the edges of the image larger and the local shadows less.



Figure 1: Modified gamma image [1]

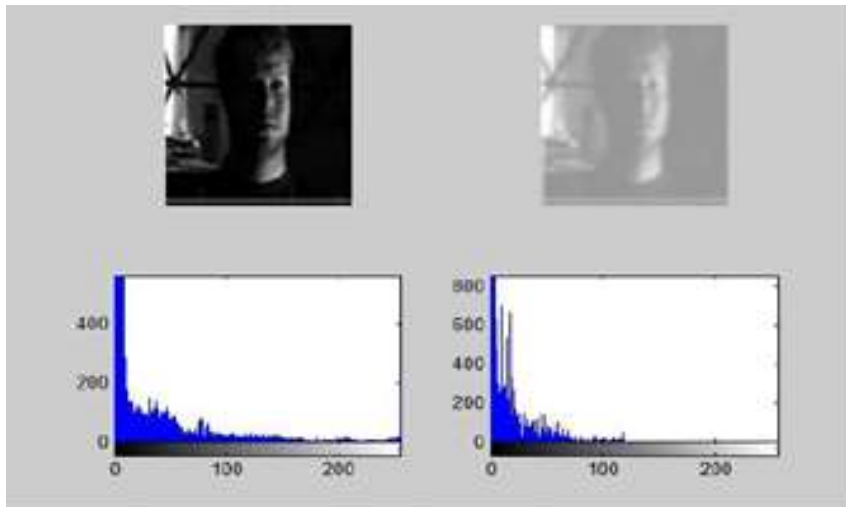


Figure 2: The difference between the histogram of the image of the input face before and after the proposed pre-processing stage [1]

2.5. Normalization

The final step in the normalization pre-processing chain is to scale the image density, thus highlighting most of the information inside the image while maintaining the essential visual elements. This is achieved using the mean absolute signal value [1].

2.6. Histogram and Calculation Time

The difference between the histogram of the image of the input face before and after the proposed pre-processing stage is shown in Fig. 2. In this form, it is specifically explained that its pre-processing stage is critical to reducing the unwanted noise or the difference in brightness. With the help of this step, the useful information and the main features of the photo are easily extracted, in addition, the execution time of the program is also considered very important [2].

2.7. Principal Component Analysis

The basic component analysis is a statistical method that has many uses and reduces later, including the uses that are commonly used in image processing. In the PCA component, the features that are distinguishable are considered. The purpose of the analysis is to analyse the basic components of the transfer of the data set X with dimensions M to the data of Y with dimensions N. Extracting the attribute using the PCA method is the same as reducing image dimensions. For example: if there is a 50x50 image, this image has 2500 unique dimensions, each self-contained pixel alone being a feature if the dimensions of this image are reduced without using the PCA or any other method. And these images are categorized. It is necessary to compare this image with the reference image every 2500th, which is a lot of calculations and time-consuming, and is usually not practical, therefore, using this technique or similar techniques, it decreases later. And in this later decline, pixels that are distinguished from other pixels Ned (Fig. 3).

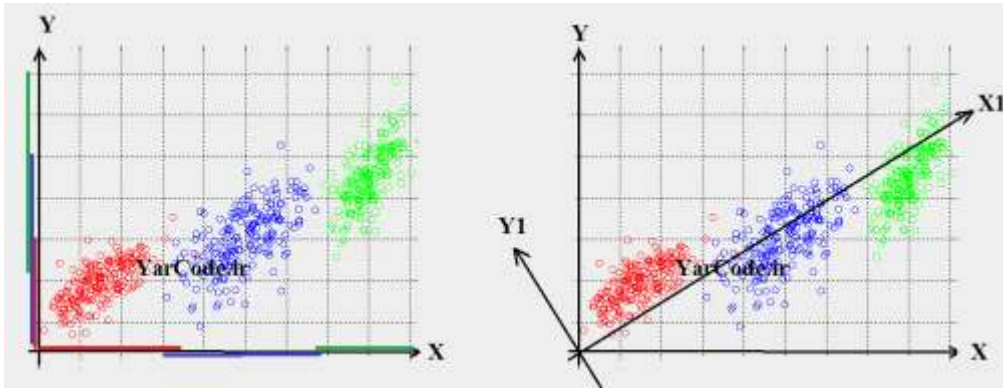


Figure 3: Selecting the X1 axis for data separation is a priority in a direction that is more dispersed than data [3]

2.8. Liner Discriminant Analysis

Linear analysis is a type of statistical method. Which is used in machine learning to find a linear combination of features that best differentiate between classes of objects. The LDA algorithm is one of the ways to reduce the dimensionality and feature selection in machine learning. This algorithm has two features:

1- Calculating Inside Class Characteristics:

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \quad (3)$$

$$S_w = \frac{1}{N-1} \sum_{i=1}^M \sum_{j=1}^{N_i} (x_{ij} - m_i)(x_{ij} - m_i)^t \quad (4)$$

2- Calculating Between Class Characteristics:

$$m_k = \sum_{k=1}^M N_k = \frac{1}{N} \mu \quad (5)$$

$$S_B = \frac{1}{N-1} \sum_{i=1}^M \sum_{j=1}^{N_i} (m_i - m)(m_i - m)^t \quad (6)$$

Inside the class property means that features that are supposed to be in a class. The same features should have the same characteristics: the LDA algorithm can only be applied to databases that have enough data in each class or group, in other words, if in each class or A group is an image, this algorithm will not work.

2.9. Comparison between PCA and LDA

In general, it has been believed from the past that the LDA is doing better than the PCA. Indeed, under certain conditions, but there are exceptions, the PCA surpasses the LDA's small face-to-face banks, and this method, unlike the LDA, is not sensitive to face-to-face banks and training

collections. It is clear that the root of the LDA's superiority relates to the mapping of images in it. LDA explores the internal differences between classes directly, while the PCA works with large-scale data (all faces). That's why the PCA encounters problems when it comes to expanding the face-to-face database. On the other hand, because the LDA needs to create facial expressions and their faces, if each individual's educational samples are not enough in the face-to-face bank, it will face a significant drop. So if the size of the data bank is small, the PCA works better than the LDA. Studies in this field prove theoretical issues. On the other hand, it can be expected that the LDA will outperform the PCA for a large image bank for each person. In the previous section, we mentioned the inadequacy of the LDA for face banks, which have fewer educational imagery than facial measurements. The solution proposed is to combine the two methods of PCA and LDA to obtain an optimal solution. By doing so, we can both use the two methods together and benefit from both. In Figure 4, a comparison is made between the two methods and their composition.

2.10. Support Vector Machine

A support vector machine is a type of categorization that is part of the kernel methods branch in the machine learning. SVM was introduced in 1992 by vapnik and is based on the statistical learning theory. The SVM algorithm is used to distinguish and distinguish complex patterns in data. Support vector machine is one of the similar methods of neural network, which is very useful in model recognition of the information we teach it, and it has two linear and nonlinear modes. SVM is used to recognize pattern from geometry and cloud-concept concept he does. Also, SVM is inherently dual or binary, that is, we can use two groups to teach the information in this way as an input so that the cloud computes the boundary between the two groups, and this is called the data classification. After classifying the data, we take the input from us and according to what our input information is in the class, it performs its own pattern recognition, which is very accurate. [1]

If the categories are linearly separable, we will obtain super-marginal pages for the separation of categories. In the case that the data are not linearly separated, the data is mapped to a more dimensional space so that it can be linearly separated in the new space. According to learning theory, in the case of linear separation, the separation that maximizes the margin of educational data will minimize the generalization error. The closest educational data is called the backup vector separator cloud. If the proper use of the SVM algorithm is used, this algorithm can have good generalizations and, in spite of its large size, it prevents the edges from being positioned, and this property is due to the optimization of this algorithm. It also uses support vectors instead of educational materials, which compresses the information. Figure (5) shows the support vectors.

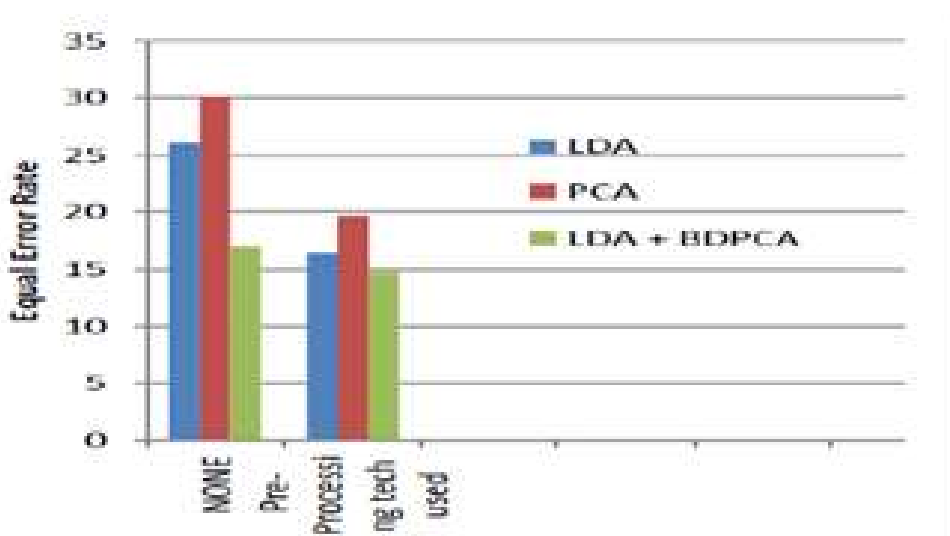


Figure 4: ERR diagram for the mentioned methods [1]

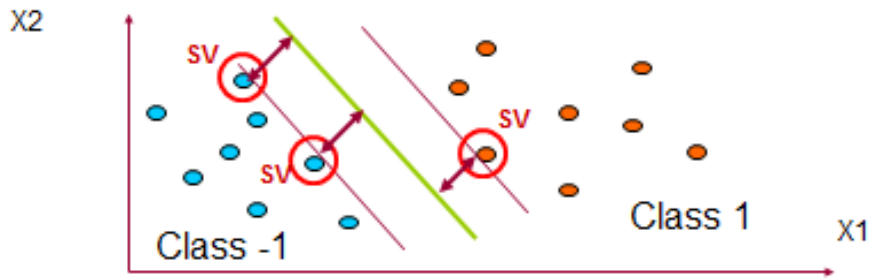


Figure 5: Support Viewer [1]

2.11. Random Forest Algorithm

The Random Forest algorithm, like SVM, is a classifier algorithm. The basic idea of the random forest is to find the mean value of the noise, using this method, the volume of input data is reduced, and complex calculations are calculated in a simpler space; other benefits of random forest include easy training, parallel work and rapid training speed] 2 [. Each learning method involves some random methods, and the goal is to select the best option in each step. RF method is very famous. RF generates a random decision path for each repetition of the algorithm and is often a great predictor [2].

The main idea of RF is to find the mean value of RF, the RF is capable of computing many complicated commands, and it can calculate the complex input space into the same space. RF directs decision-making routines, so reduced data is available [2]. One of the methods of learning a car is the random forest, which is used in various sciences. As the name suggests, a collection or forest of trees is used. Each tree is composed of roots, nodes and leaves. The random tree of a tree produces a great deal of decision, each tree in a class or class, and tells the story of which tree's vote belongs to which class, the class that has the highest vote among the forest trees is known as the forest class. The decision tree, with its simplicity, can easily analyse and categorize the data.

3. The Simulations and Discussion on the Results

In the pre-processing section in Gamma correction, different values for gamma are considered and the simulation results are shown in Table (1). Considering the results obtained hereafter, gamma 0.9 is simulated. Further studies are being carried out on other variables. In the Gaussian filter section, different values are considered for Sigma 1 and the simulation results are shown in Table (2).

The results show that Sigma has 0.5 times the best accuracy for both RF and SVM methods. So, the simulation of Sigma is considered to be 1 to 0.5. In the next step, different values are considered for Sigma 2. Of course, it should be noted that Sigma 2 should be larger than Sigma 1. At this point, $\text{sigma2} = k * \text{sigma1}$ is considered. The simulation results are shown in Table (3). The results show that $K = 11$ has the best accuracy for both RF and SVM methods. The simulation is then considered $K = 11$. In the extraction field and in the PCA method, the values of the characteristic are obtained from the covariance matrix of the educational faces set. Each attribute value addresses one of the facial features. By sorting these values down, we can remove small values close to zero, reducing the size and reducing the size of the data, resulting in a reduction in the volume of computations and an increase in the computational speed. But by eliminating special values because of the fact that no special values are eliminated, there is the possibility of merging classes.

In LDA method, the vectors of the characteristic are derived from the covariance matrix of the educational set of faces. These vectors divide the classroom into the best of the faces of the people in the database. By sorting these vectors down, in fact, vectors with the most dispersion are

arranged to the least dispersion. Using the calculation of a linear transfer so that we have the greatest difference in the class. The first step in calculating this transition is to obtain the internal differences of each class (the class of internal dispersion matrix), and the next step is to obtain the difference between the class (the class of dispersion matrix).

Our goal is to obtain the highest scatter criterion for classroom dispersion, while minimal dispersion within the classroom. Regarding the fact that the PCA method in the data set is poorly performing than the low data set, and the LDA method, the more the data sets value, the better the performance and accuracy of this method. Therefore, combining these two methods leads to better performance in different databases, as well as the failure to integrate classes into each other and ultimately to extract features.

Table (1) results obtained from different Gamma

Accuracy of SVM	Accuracy of RF	Gamma value
89.8693	98.6928	0.2
97.7124	98.6928	0.7
99.3464	98.6928	0.8
99.3464	99.0196	0.9
97.549	99.0196	0.95
98.6928	97.7124	1

Table (2) shows the results obtained from different sigma

Accuracy of SVM	Accuracy of RF	Sigma 1
99.3464	98.6928	0.2
99.3464	99.0196	0.5
95.915	99.3464	0.55
99.0196	98.6928	0.7
96.0784	98.8562	0.75
86.4379	99.0196	0.9

Table (3) the results obtained from different K

Accuracy of SVM	Accuracy of RF	value of K
99.3464	99.0196	10
99.183	98.5294	10.5
99.5098	99.183	11
99.0196	98.369	11.5
97.549	98.8562	12
99.183	98.8562	15

The training data and test data were randomly selected for being similar to reality, this is due to the fact that the variables have been fixed, and the program has different responses in the output.

Because of this, the average output is 3 times the program running with $\text{Gamma} = 0.9$, $\text{Sigma1} = 0.5$, $\text{Sigma2} = 11\text{Sigma1}$. The results are 3 times implemented and their average can be seen in Table (4). With attention to the fact that in most articles, special attention has been paid to its pre-processing, in some articles it has been attempted to remove the darkness of the image, and in others it is attempting to highlight the edges in the image. In this research, using these two techniques for pre-processing the input image and extracting the property using the PCA + LDA combination technique and the use of two types of categorization, and comparing them, the two categorizers have obtained better results. The percentage accuracy in the proposed method for random forests is 99.196% and for support vector machine is 99.163%. Also, this research showed that the use of subfield with Random forest has been more accurate than SVM with subfield. Table (5) shows the comparison of the proposed method with other algorithms.

Table (4) shows the results obtained from 3 times and obtaining the average

Accuracy of SVM	Accuracy of RF	Variable values
99.5098	99.183	Gamma=0.9 Sigma1=0.5 Sigma2=11Sigma1
98.9628	98.5294	
99.3464	99.3464	
99.183	99.0196	Average

Table (5) Comparison of proposed algorithms with other algorithms

Comparison	Accuracy is achieved
Proposed SVM	99.183%
Proposed RF	99.0196%
SVM [1]	87.16%
RF [2]	97.16%
ACNN [6]	91.67%
CNN+SVM [6]	97.50%
MFA [7]	75.83%
SDA [8]	65.17%
SSDL [9]	95%

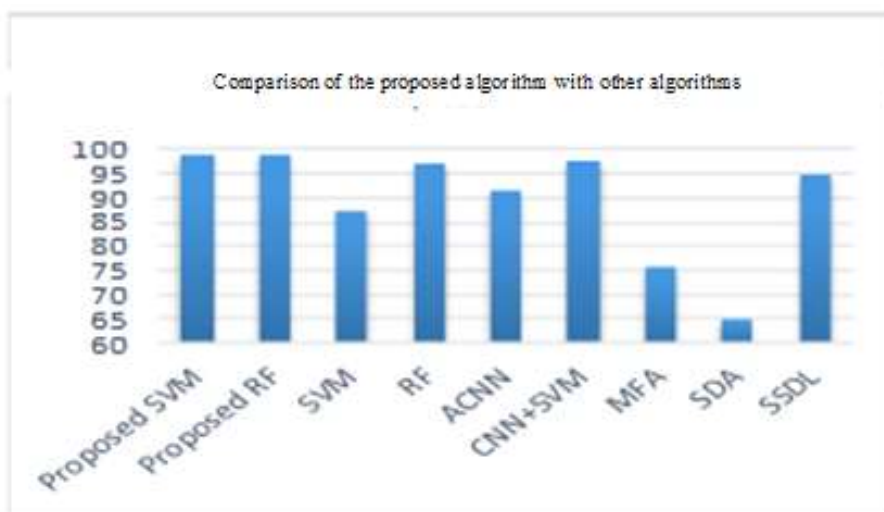


Figure 6: Comparison of the proposed algorithm with other algorithms.

4. Conclusion

The result shows that positive changes in the pre-processing can have positive effects at the end of the simulation. In this paper, with changes to the pre-processing (removal of darkness, and also highlighting the edges of the image), better and significantly better results than the similar ones. Of course, the positive or negative changes in the pre-processing are empirically obtained and still have a precise relationship not found for it.

Image processing algorithms are important in face detection applications as in ATMs, in buildings with smart doors, in matching of photos with the person at the airports and etc.. Considering the fact that its pre-processing is very important in image processing, in this paper we tried to change the pre-processing used algorithms, obtained the images with high and dark edges in the preproduction output, combined extractor input of PCA + LDA and classified by Improved random forest and backup machines. Further research on this topic can be found, and it examines the relationship between changes in pre-processing and changes in its pre-processing processes with face recognition rates through different methods of compilation.

References

- [1] Teja, G. Prabhu, and S. Ravi. "Face recognition using Subspace techniques." *Recent Trends In Information Technology (ICRTIT), 2012 International Conference on. IEEE, 2012.*
- [2] Kremic, Emir, and Abdulhamit Subasi. "Performance of Random Forest and SVM in Face Recognition." *The International Arab Journal of Information Technology (2015).*
- [3] Hyeongoon moon, P. Jonathon Philips, "Computational and Prformance aspects of PCA-based Face Recognition Algorithms", *IEEE, 2001.*
- [4] Shanshan Guo, Shiyu Chen, Yanjie Li, " Face Recignition Based on Convolutional Neural Network and Support Vector Machine ", *Proceedings of the IEEE, Ningbo, China, August 2016.*
- [5] Shih-Hsuan Yang, Yu-Xiang Zhan, "Improved Face Recognition by Incorporating Local Color Information into the Active Shape Model". *Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 2016*
- [6] Shih-Hsuan Yang, Yu-Xiang Zhan, "Improved Face Recognition by Incorporating Local Color Information into the Active Shape Model". *Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 2016*
- [7] Pratibha Sukhija, Sunny Behal, Pritpal Singh, "Face Recognition System Using Genetic Algorithm", *Procedia*