International Journal of Engineering Sciences Paradigms and Researches: (Vol. 39, Issue 01) and (Publishing Month: February 2017)

# Component-based Face Recognition for Forensic Application using Deep CNN

Satyajit nayak[1], Yogo Maya Mahapatra[2] and Kamalkant Shaw [3]
[1,3]Associate Professor, Department of Computer Science Engineering, Gandhi Institute For Technology (GIFT), Bhubaneswar
[2]Assistant Professor, Department of Computer Science Engineering, Gandhi Engineering College, Bhubaneswar

**Publishing Date: February 27, 2017**

*Abstract*—Face recognition for forensic applications is a very challenging task. The challenges arise due to the unavailability of the mug shot of the culprit from a crime scene. The available face is either occluded or blurred. In both the cases, Componentbased face recognition is capable of recognizing an individual or the culprit through a part or component of the face, which is the major focus of the paper. Convolutional Neural Network is the base of the proposed framework for face recognition via components: nose, mouth and, eyes. The novelty of this framework is that it can single-handedly recognize an individual by full face or by anyone or all the components of the face using Deep Learning with high recognition rate. In a practical scenario, the collected images are corrupted with different noises that reduce the recognition rate. Robustness of the system is tested for different noises viz. White Gaussian noise, Speckle noise, and Salt Pepper noise. Results show the descending trend in the accuracy of the recognition as the intensity of the noise is increased. Faces94, which is the standard database of face, is used for the experiment.

*Keywords: Component Based Face Recognition, Deep Neu-ral Network.*

## I. INTRODUCTION

In today's world, human identification is of utmost importance and Face Recognition (FR) is a prominent method for this purpose. In this, people may not even know they are under surveillance, whereas other methods like a fingerprint scanner or iris scanner, are invasive [11].

When a human sights a new face, brain unconsciously extract the feature of the face which are commonly known as Nodal points. These features are used in FR for the identification of an individual. But in forensic applications, the face of the culprit is either occluded or only part of the face is available from the crime location. Hence FR becomes a tedious task to do. Therefore, recognition using only a small part of the face is essential.

Component- Based Face Recognition (CBFR) resolves the problem of occlusion. It can recognize an individual via components of face. Thus CBFR plays an important role in Forensic FR. Now, recognizing an individual only through a part of the face becomes difficult because the number of features involved is lesser. So there is a need for an algorithm that can extract features that humans can't extract through naked eyes. For this purpose, Convolutional Neural Network (CNN) is used. CNN is a part of deep learning wherein we don't have to input features, it is capable of performing the feature engineering task i.e. it can extract complex features on its own. This helps in recognizing task as many complex features are now involved in it.

The major contributions of this paper are • CBFR using deep learning method i.e. using CNN. The proposed system is competent in recognizing an individual by full face and also by the component of the face. The components considered in this paper are the Nose, Eyes and, Mouth.

- Also, the work covers the practical forensic application scenario wherein the images are corrupted by noise. Hence the recognition rate decreases with the increase in the variance of the noise. The noises covered in this paper are White Gaussian noise (Additive), Speckle noise (Multiplicative), and Salt Pepper noise (Impulse).

Organization of the paper is as follows: Section 2 gives a brief literature survey. Section 3 covers the background study required for the proposed work. Section 4 describes the proposed framework, followed by the experimentation and results illustrated in Section 5. Section 6 is about the robustness of the proposed system to different noises present in the FR and finally, Section 7 gives the conclusion and future scope of the work.

## II. RELATED WORK

Face Recognition technology depends totally on the features or landmarks of the face. FaceIt [10] defined 80 such landmarks that are to recognize a person. There are numerous methods that recognize full faces using various method. Few of them include Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA)[15, 19]. But for forensic applications, CBFR is the major focus.

Researchers used a probabilistic method for recognition using CBFR [17]. Also, they proved that the component system outperforms the holistic/global systems. The global system is sensitive to image variations such as rotations while component-based approach performs better for the same. This happens because, for small rotations, the change in the component is relatively small as compared to the changes in the whole face pattern [16]. Previous works also talk about the advantage of using multiple biometrics to improve the identification task [12, 2]. Also, using CNN network for holistic face recognition improves the accuracy of the recognition system. It gives better accuracy than the wellknown methods like PCA, Local Binary Patterns Histograms (LBPH) and K–Nearest Neighbor (KNN) [5]. Also, it says that increasing training samples increases the accuracy. From different layers of CNN, if batch normalization layer is added to the network is reduces the covariant shift problem that contributes in increasing the accuracy [9, 4]. [1] uses CNN only for feature extraction and later part is done by using transfer learning and KNN Classifier. Instead of KNN, SVM can also be used as a classifier [8].

All the images in practical forensic conditions are corrupted with noises like Gaussian and Speckle. These noises reduce the recognition rate. [18] considers Gaussian noise and use two cascaded Hit and Miss Transform that removes the noise. On similar lines, [7] states that noise removal using Gabor and Non-Negative Matrix Factorization gives a fair recognition rate. Robust deep FR technique can also be used. In this, a model trained on clean faces is used to detect the features from a noisy database [6]. The noisy images are removed using unsupervised learning. And instead of unsupervised learning, Self-Paced Neural Network can be used that gives better accuracy [3].

Previous work focuses majorly on holistic face recognition but for forensic applications component based recognition is more important. Hence CBFR becomes the major focus for the paper. Also, all the CNN based work is carried on the holistic face and CNN can single-handedly perform the feature extraction, regularization and classification. So, componentbased face recognition using CNN is a novel work. Moreover, all the algorithms focus on identifying and removing the noise from the corrupted images of full faces but none deals with the component of face. This motivates us to analyze the effect of noise on the components. Therefore, all the above reasons lead us to Component-Based Face Recognition using CNN and to analyze the effect of noise on it.

## III. Background Study

The major focus of the paper is to recognize a person via CBFR. But to the best of our knowledge, no such database is directly available, so it has to be created to perform the experiments. For this purpose, the Viola-Jones algorithm is used on full faces to separate the component from the face and form the component database.

### *A. Viola-Jones*

Viola-Jones is a basic object detection algorithm that is used to detect an object in real-time applications [13]. It comprises of four steps:

1) Haar filters are used to detect a face and they are feature specific. Few of them are shown in Figure 1.
2) Calculating the feature value is a tedious task and hence the image is converted into an integral image. This is done by adding all the top and the left pixel values at a considered pixel as shown in Figure 2.

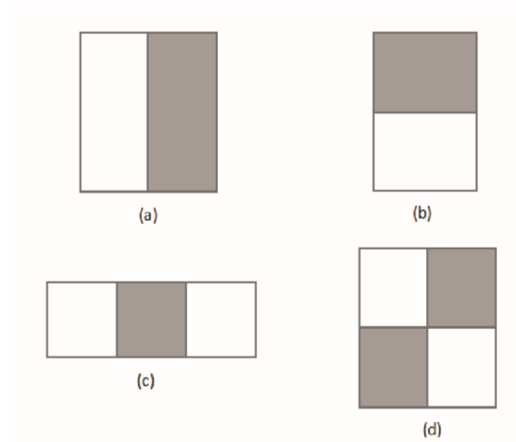



Figure 1. Haar Filters [14]

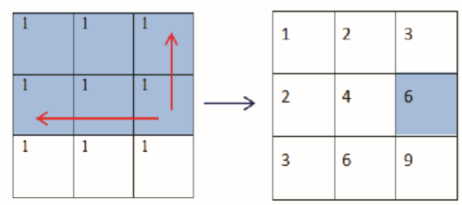



Figure 2. Integral Image

3) Adaboost is a machine learning algorithm which finds the best features among all. These features are a weak classifier, hence the Adaboost combines all to form a strong classifier.
4) To speed up this classification task, cascading is done. In these, if the first few features weights confirm that the sub-image does not have a face, it is directly discarded. If confirmation is not made, then the next set of features are used to confirm it. This is done until the classifier classifies whether the sub-image has a face or not.

This technique is extended to separate, the part of the face from the full face. In this, firstly the faces are detected and then components of the faces are separated using the above algorithm and those components are then used as a database for further experiments.

### *B. Convolutional Neural Network*

Convolutional Neural Network (CNN or ConvNet) is a category of deep learning that works as a spatial feature extractor and primarily used for structured data like image processing. The CNN consists of two parts: Convolutional Layer, Activation Layer (e.g. ReLU) and Pooling Layer (eg. Max Pooling) form the feature learning part; while Flattening, Fully Connected layer and Softmax layer form the classification part.

Convolutional layer (Conv layer) is an important part of CNN because it reduces the number of parameters to be trained. Various filters are used that extracts the features from the image. Several parameters like filter shape, filter size, filter stride, zero padding are involved in convolutional layers. Stride is the number of pixels shifts over the input matrix.

Activation layer is used to maps a set of inputs to set of outputs with a non-linear function like a hyperbolic tangent (tanh), Sigmoid or ReLU. Batch normalization layer is used to avoid the covariant shift problem that arises due to divergent images applied to the network.

Pooling layer is used for Dimension Reduction. It helps to keep only prominent features. For that, it takes the average or maximum value from a small group of pixels.

Fully connected layer converts the feature map into a single column vector. Lastly, The Softmax layer calculates the probability of the sample image. All the probability assigned to classes must add up to 1. And according to the probability assigned, the classification layer classifies the input sample.

## IV. PROPOSED FRAMEWORK

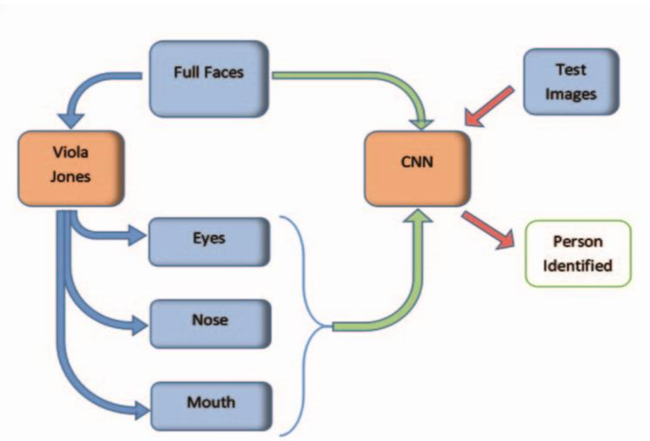



Figure 3. Schematic Diagram of the Proposed Framework

Figure 3 depicts the schematic diagram of the proposed framework. In this framework, full face database is considered and the component database is created from it using ViolaJones algorithm. The components that are considered for this work is Eyes, Nose, and Mouth. These components are now used to train the network. The whole database is segregated equally into two parts: training and testing. Hence 10 images per individual are used for training and 10 images for testing. Once the network is trained, it is used to recognize a person via a full face or component of the face. This becomes feasible as the network is trained to recognize the individual in either of the cases.

CNN network is used for this purpose. Figure 4 shows the proposed CNN architecture for this framework. The first layer of the network is the input layer wherein the images for training and testing is given. The input image is resized to [64x64x3] as the component size is smaller as compared to the holistic image. [64x64] size is considered as it is comparable to all the actual cropped component sizes from the database.

Next layer is the convolutional layer with 8 filters of size [3x3] with the stride of 1 and 'same' zero padding. Following

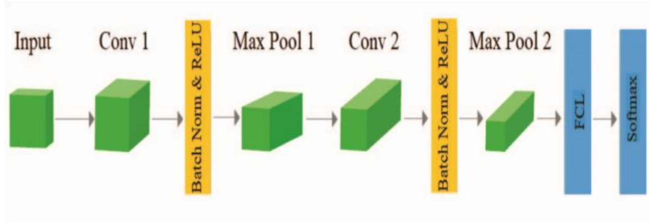



Figure 4. Proposed CNN Architecture

this is the batch norm and ReLU layer. Then comes the max pooling layer of size[2x2] with a stride of 2.

The above four layers are repeated one more time but the convolutional layer has 16 filters of size [3x3] with a stride of 1 and other layers remaining same. Next comes the fully connected layer that takes all the features from the previous layer and flattens it into a single column feature matrix. Lastly, Softmax layer and classification layer recognizes the person.

## V. EXPERIMENTS AND RESULTS

### A. Database

For CBFR, the required database is a component database. Hence as stated earlier, Viola-Jones algorithm is used on full faces to create the component database. For this purpose, Faces94, of Computer Vision Research Project, is used which is a standard database. It comprises 20 full face images of 153 individuals, totaling 3060 images. Using these, three sets of databases: eyes, mouth and nose database, are created. A melange of full face and the created database is shown in Figure 5. Faces94 database is selected as there are several variations in facial expression. Various lip and eye movements make a wide range of database that helps in making the proposed system robust towards the expression changes.

### B. Experimental Analysis

Various experiments are performed on the proposed framework and the accuracy of the system is analyzed using Equation 1. Error rate is also calculated using Equation 2. Besides that, the confusion matrix is computed to determine the number of misclassifications.

$$Accuracy = \frac{No.\ of\ testing\ samples\ classified\ correctly}{Total\ no.\ of\ testing\ samples} \quad (1)$$

$$ErrorRate = \frac{No.\ of\ misclassifications}{Total\ no.\ of\ testing\ samples} \quad (2)$$

Table I OVERALL ACCURACY FOR DIFFERENT NUMBER OF CONVOLUTIONAL LAYERS

| Network Trained by | One Conv Layer | Two Conv Layer | Three Conv Layer |
|---|---|---|---|
| Overall System Accuracy | 91.49% | 98.29% | 94.15% |


Figure 5. A Melange of Full Face and Component of face

Firstly, the number of layer for CNN is to be determined. The experiments are carried out to find out how many conv layers are sufficient to give good accuracy. Table I depicts the accuracies calculated when the number of convolutional layers is one, two, and three. It can be observed that the accuracy increases when two conv layers are used, but decreases as the layers are increased to three. This is because the hyperplane gets too complex, when three layers of convolution are used and leads to overfitting that decreases the overall accuracy. This implies that two conv layers are enough for good CBFR accuracy.

Secondly, all the face components and full face are used to train the proposed network simultaneously. This trained network is then tested by full face and all components individually as well as together. This is done to analyze, which of the component is more reliable in recognizing through CBFR. Table II shows the parameter calculation when the trained network is tested with the components and full faces individually and also when tested altogether.

Table II
PARAMETER CALCULATION FOR DIFFERENT TESTING IMAGES

| Parameters | Full Face | Eye | Nose | Mouth | Full Face + All components |
|---|---|---|---|---|---|
| Accuracy | 100 | 98.7805 | 99.3902 | 95 | 98.2927 |
| Misclassification | 0 | 10 | 5 | 41 | 56 |
| Error | 0 | 1.2195 | 0.6098 | 5 | 1.7073 |

Results clearly show that the nose has the highest accuracy among the components. This is because the nose features are more stable and are least affected by various facial expressions. Eyes and mouth component, as compared to the nose, have varying expressions. On a similar line, eye movements are less as compared to mouth. Hence the accuracy is even less for the mouth as compared to eyes. Therefore, the total system accuracy for recognition via both full face and component based is 98.29%. The full face recognition accuracy comes out to be 100%, as can be seen from Table II. This accuracy goes down to 99.9% when the proposed system is trained and tested only by the full face. Hence it can be stated that when the system is trained with full faces along with the components, the full face recognition accuracy increases. Thus, other than the forensic application, the proposed system can be applied to any full face recognition applications.

## VI. ROBUSTNESS OF THE PROPOSED SYSTEM

The proposed system recognizes an individual fairly via full face and the components. But in practical scenarios, the images that are collected are noisy. The reason for this being, the images captured in the forensic scenario is taken from a lowresolution camera and are mostly captured at night. And hence the chances of presence of noise is high.

The images are significantly corrupted with three common noises viz. White Gaussian, Speckle, and Salt Pepper noise. The experiments performed on the proposed system show that the system gives 98.29% accuracy when the images are not corrupted with any sort of noise. But this accuracy would reduce with the increase in the intensity of the noise. So to compare our results with the practical scenario, it is a necessary step to understand the effect of noise on the system performance.

### *A. Noisy Database*

For the experimentation purpose, the database having noise is to be created. For this, the White Gaussian noise, Speckle noise, and Salt Pepper noise are considered. These noises are added only to the test images i.e. 50% of the whole database of full face and all the components.

Figure 6 shows the effect of the noise on the database with

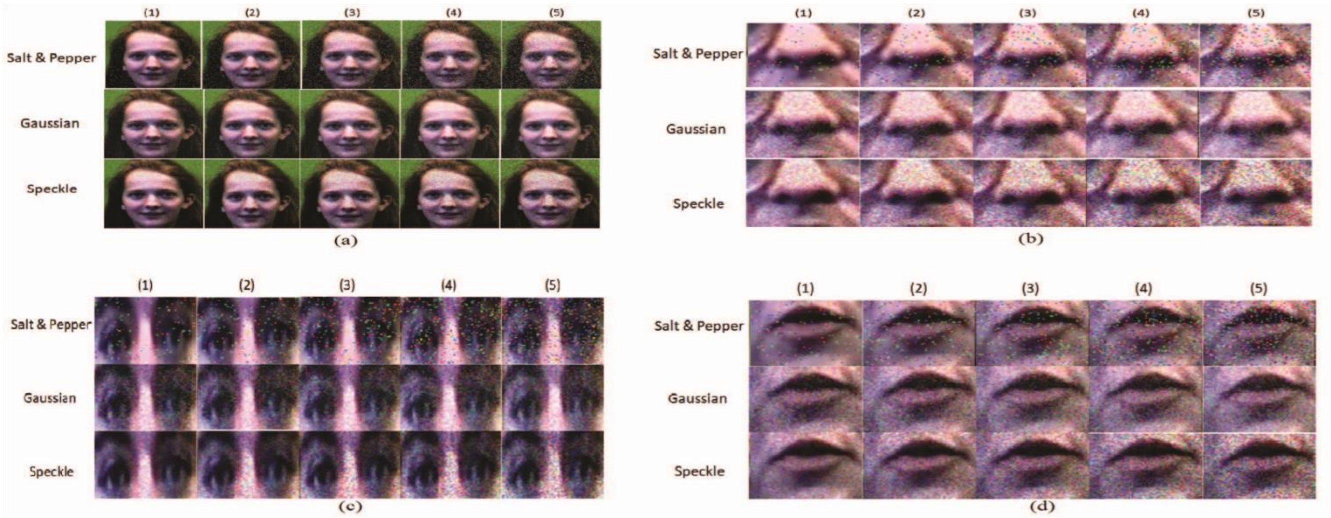


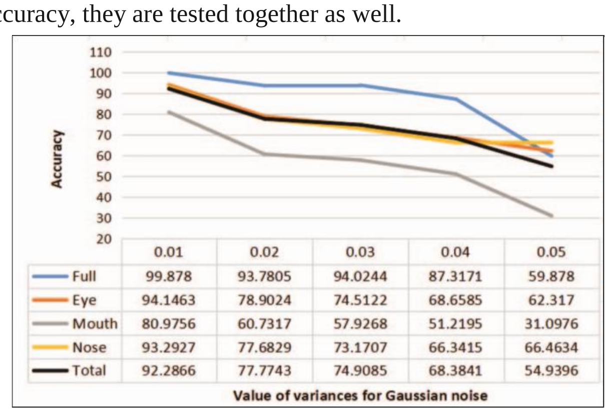

Figure 6. A Melange of Noisy Full Face and Component of face

5

varying intensity of noise. For Speckle and Salt Pepper noise, image (1) from the collage shows the image corrupted with zero mean and 0.02 variance. Similarly image (2) have variance 0.04, (3) have variance 0.06, (4) have 0.08 variance and finally (5) have 0.1 variance. For Gaussian noise, image (1)-(5) shows noise variance of 0.01-0.05 with a step size of 0.01. This range is defined by the values obtained from the results. The recognition rate drops to a very low value if the maximum limit of the given range is exceeded.

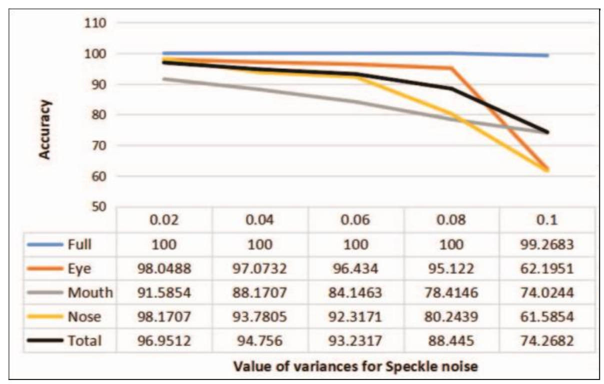
*B. Experimentation and Results*

The trained system is tested with corrupted images. Firstly, component-wise testing is carried out. This is done to understand which component among all, is more reliable for CBFR. Also to understand the descending trend in the overall accuracy, they are tested together as well.



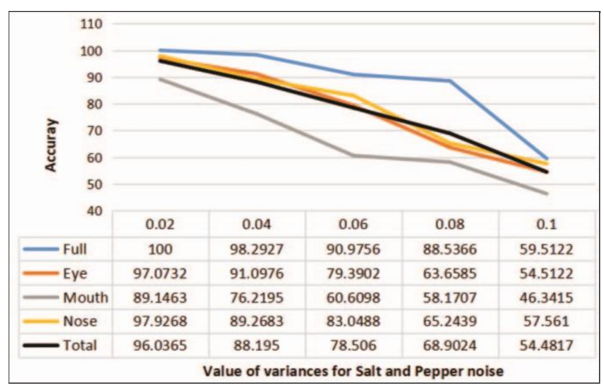

| | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 |
|---|---|---|---|---|---|
| Full | 99.878 | 93.7805 | 94.0244 | 87.3171 | 59.878 |
| Eye | 94.1463 | 78.9024 | 74.5122 | 68.6585 | 62.317 |
| Mouth | 80.9756 | 60.7317 | 57.9268 | 51.2195 | 31.0976 |
| Nose | 93.2927 | 77.6829 | 73.1707 | 66.3415 | 66.4634 |
| Total | 92.2866 | 77.7743 | 74.9085 | 68.3841 | 54.9396 |

Figure 7. Graph of accuracy for changing variances for Gaussian noise




| | 0.02 | 0.04 | 0.06 | 0.08 | 0.1 |
|---|---|---|---|---|---|
| Full | 100 | 100 | 100 | 100 | 99.2683 |
| Eye | 98.0488 | 97.0732 | 96.434 | 95.122 | 62.1951 |
| Mouth | 91.5854 | 88.1707 | 84.1463 | 78.4146 | 74.0244 |
| Nose | 98.1707 | 93.7805 | 92.3171 | 80.2439 | 61.5854 |
| Total | 96.9512 | 94.756 | 93.2317 | 88.445 | 74.2682 |

Figure 8. Graph of accuracy for changing variances for Speckle noise variances




| | 0.02 | 0.04 | 0.06 | 0.08 | 0.1 |
|---|---|---|---|---|---|
| Full | 100 | 98.2927 | 90.9756 | 88.5366 | 59.5122 |
| Eye | 97.0732 | 91.0976 | 79.3902 | 63.6585 | 54.5122 |
| Mouth | 89.1463 | 76.2195 | 60.6098 | 58.1707 | 46.3415 |
| Nose | 97.9268 | 89.2683 | 83.0488 | 65.2439 | 57.561 |
| Total | 96.0365 | 88.195 | 78.506 | 68.9024 | 54.4817 |

Figure 9. Graph of accuracy for changing variances for Salt Pepper noise

Figure 7, Figure 8 and Figure 9 show the accuracy calculation for all the components for various variances of Gaussian, Speckle. and Salt Pepper noise respectively. These Figures precisely depicts the descent in the accuracy for the increasing variance of all the noises considered. The black line in the above graphs portray the overall accuracy of the system which also conveys the decrease in the recognition rate.

Of all the noises, Speckle noise affects the least to the accuracy. For Gaussian and Salt Pepper noise, the slope for the decrease is much higher. Out of these, Gaussian noise makes the image smooth to the extent that the feature extraction becomes difficult. Also out of all, full face is least affected and mouth is affected the most by noise. The reason behind this is; face has ample of features, even if few of them are affected by noise, the rest of the features are capable of recognizing the person. But for mouth number of features that recognizes a person are very less and also the lip movement affects the accuracy.

## VII. Conclusion

This paper presents a novel approach for Component Based Face Recognition in a noisy environment. The novelty lies in two aspects: firstly, it can recognize a person via full face and/or by part of the face i.e. by eyes, nose or mouth and secondly, the framework is tested for the robustness against noisy mugshots. In this, instead of removing the noise, the system is tested on noisy images directly. So the overhead of identifying and removing the noise is avoided.

Our experiments conclude that the nose gives a better recognition rate than mouth and eyes. This is because the nose is stable towards pose and expression changes. The overall accuracy of the proposed system comes out to be 98.29%. Also, the accuracy of full face recognition becomes 100% when the system is trained on full face along with any or all of the components.

Additionally, the system is robust to the various noises viz. Gaussian, Speckle, and Salt Pepper noise, to a high level of variance. From the experiments, it is observed that the recognition is least affected by the Speckle noise and considerably affected by Gaussian and Salt Pepper noise. Moreover, as we increase the variance of the noise, accuracy reduces. Considering other face components like forehead, ears, right or left part of the face for recognition could be the future scope.

## References


[1] Rupali Sandip Kute, Vibha Vyas, and Alwin Anuse. “Component-based face recognition under transfer learning for forensic applications”. In: *Information Sciences* 476 (2019), pp. 176–191.

[2] Rupali Kute, Vibha Vyas, and Alwin Anuse. “Cross domain association using transfer subspace learning”. In: *Evolutionary Intelligence* 12 (Feb. 2019). DOI: 10. 1007/s12065-019-00211-y.

[3] Pengfei Zhu, Wenya Ma, and Qinghua Hu. “Self-paced Robust Deep Face Recognition with Label Noise”. In: Mar. 2019, pp. 425–435. ISBN: 978-3-030-16141-5. DOI: 10.1007/978-3-030-16142-2 _33.

[4] Musab Cos¸kun et al. “Face recognition based on convolutional neural network”. In: *2017 International Conference on Modern Electrical and Energy Systems (MEES)*. IEEE. 2017, pp. 376–379.

[5] Patrik Kamencay et al. “A new method for face recognition using convolutional neural network”. In: (2017).

[6] Jirui Yuan et al. “Robust Deep Face Recognition with Label Noise”. In: Oct. 2017, pp. 593–602. ISBN: 9783-319-70095-3. DOI: 10.1007/978-3-319-70096-0 61.

[7] Indra Budiman et al. “The effective noise removal techniques and illumination effect in face recognition using Gabor and Non-Negative Matrix Factorization”. In: *2016 International Conference on Informatics and Computing (ICIC)*. IEEE. 2016, pp. 32–36.

[8] Shanshan Guo, Shiyu Chen, and Yanjie Li. “Face recognition based on convolutional neural network and support vector machine”. In: *2016 IEEE International Conference on Information and Automation (ICIA)*. IEEE. 2016, pp. 1787–1792.

[9] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *arXiv preprint arXiv:1502.03167* (2015).

[10] S. B. Thorat, S. K. Nayak, and Jyoti P. Dandale. “Facial Recognition Technology: An analysis with scope in India”. en. In: (May 2010). URL: https://arxiv.org/abs/ 1005.4263v1 (visited on 05/26/2019).

[11] Chenggang Zhen and Yingmei Su. “Research about human face recognition technology”. In: *2009 International Conference on Test and Measurement*. Vol. 1. Dec. 2009, pp. 420–422. DOI: 10.1109/ICTM.2009. 5412901.

[12] Md Maruf Monwar and Marina L Gavrilova. “Multimodal biometric system using rank-level fusion approach”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39.4 (2009), pp. 867–878.

[13] Libor Spacek. *Facial Images: Faces94*. June 2009. URL: http://cmp.felk.cvut.cz/~spacelib/faces/faces94.html.

[14] *Viola–Jones object detection framework*. June 2009. URL: https://en.wikipedia.org/wiki/Viola%E2%80% 93Jones _object detection framework.

[15] Maha Sharkas and M Abou Elenien. “Eigenfaces vs. fisherfaces vs. ICA for face recognition; a comparative study”. In: *2008 9th International Conference on Signal Processing*. IEEE. 2008, pp. 914–919.

[16] Bernd Heisele et al. “Face recognition: componentbased versus global approaches”. In: *Computer vision and image understanding* 91.1-2 (2003), pp. 6–21.

[17] Aleix M Martınez. “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class”. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (2002), pp. 748–763.

[18] Rami Qahwaji and Roger Green. “Detecting faces in noisy images.” In: *Computers and Their Applications*. Citeseer. 2001, pp. 225–228.

[19] Peter N Belhumeur, Joao P Hespanha, and David J˜ Kriegman. “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 7 (1997), pp. 711–720.