

A Novel Face Parts Detection Mechanism for Biometric Recognition

Satish Kumar Singh^{1*} Narendra Kumar Gupta²

¹Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad

²Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad

Abstract – In surveillance, the role of Closed-Circuit Television (CCTV) is found more significant in recent days for security reasons in different places such as town centres, airports, public transport, evidence in court etc. Human face detection is the most important part in the body that helps to identify the person during identification. This paper has proposed a new technique that accurately detects different parts of the face such as nose, eyes and mouth. During detection, this technique also considers the illumination and background changes of the facial parts. This work is composed of three phases. The first is the computation of an integral image representation which allows the features to be computed quickly. The second is a learning algorithm based on modified AdaBoost which selects a small number of critical visual features from a larger set and provides an efficient classifier. The third phase is a method for combining increasingly more complex classifiers. The proposed technique provides good results maintaining a low computational cost. Results obtained from experiments shows that the proposed technique gives better results than other state-of-the-art techniques.

Keywords: Face Detection, Surveillance, Security, CCTV.

-----X-----

1. INTRODUCTION

It is found that the use of CCTV cameras has grown immensely in recent years for video surveillance due to security threats. Terrorist activities are one of the major threats found in recent years after 9/11 and 26/11 attacks in New York and Mumbai respectively. A huge number of CCTV cameras have been installed worldwide in public and prime locations to cover all the criminal and terrorist activities. There are many instances have been found regarding the installation of CCTV cameras in a very large quantity such as in London (Yu-Lung and Chi-Jui Chang, 2016). However, many techniques have been developed to manage CCTV network. But, still they are found insufficient to manage such a growing huge number of CCTV network. In current scenario, humans are directly involved to monitor screens or review stored videos. This is unreliable and inefficient way that makes proactive surveillance impractical (Brands et. al., 2016).

Security personals only see and do analysis of the footage obtained from the recorded videos of the incident. These video footages depend on human monitoring, which seems to be unreliable, invalid and less effective as compared to automatic and intelligent CCTV system. So, developing an intelligent and fully

automatic CCTV system is demand for the current scenario and it is increasing day by day.

To track the criminal activities, human face detection is one of the most important content in the making of an intelligent CCTV system (Wilber et. al., 2016. Yang et. al., 2016).

Face detection is the process of using computing technology to determine the exact place and size of human facial parts in the form of digital image. During this process, other objects (background images) except facial parts are totally ignored. In all face processing techniques, face detection is the very first important step and its accuracy affects the overall performance of the face recognition process.

However, most of the research on face detection focuses on high quality full face detection based on different facial parts separately for better accuracy is still on-going research. Many problems still need to be solved before facial parts detection can approach the capability of the human perception system. Face parts detection is a challenging task due to the variances in background, facial expression and illumination. To handle these challenges, we propose a technique that performs advanced detection of different facial parts such as Eye Pair, Nose and Mouth and can achieve fast, accurate detection that

is robust to changes in illumination and background. This work is composed of three phases. The first is the introduction of a new integral image representation which allows the features to be computed quickly. The second is a learning algorithm based on modified AdaBoost which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers. The third contribution is a method for combining increasingly more complex classifiers in a cascade which allows background regions of the image to be quickly discarded while spending more computation on significant face-like regions. The detection stage provides good results maintaining a low computational cost. After the integration of the three stages, several improvements are proposed which increase the facial parts detection rate and the overall performance of the system. The experimental results demonstrate the significant performance improvement using the proposed approach over other existing techniques. It can be seen that the proposed method is very efficient and has significant value in application.

This paper is further organized in to 4 sections. A brief literature survey is given in section 2, section 3 describes the proposed algorithm, experiments and results are discussed in section 4 and finally the paper is concluded in section 5.

2. LITERATURE SURVEY

To detect the facial part of any image generally two types of approaches have been used i.e. feature based and image based. The first approach matches the extracted feature of an image against knowledge of face feature. While getting best match between training and testing image is used under image based approach.

2.1 Feature Based Approach

It contains three techniques such as active shape models, low level analysis and feature analysis.

Active shape model is classified into three groups i.e. snakes, Point distribution model, deformable templates.

The first type uses a generic active contour called snakes, first introduced by Kass et al. in 1987 (Crowley and Berard, 2016). Snakes are used to identify head boundaries (Low and Ibrahim, 1997. Heisele et. al., 2000. Sharif et. al., 2011). In order to achieve the task, a snake is first initialized at the proximity around a head boundary. It then locks onto nearby edges and subsequently assumes the shape of the head. The evolution of a snake is achieved by minimizing an energy function, E_{snake} , denoted as

$$E_{snake} = E_{internal} + E_{external} \quad (1)$$

Where $E_{internal}$ and $E_{external}$ are internal and external energy functions.

Ideal Point Distribution Models can deform in ways that are characteristic of the object. Cootes and his colleagues were seeking models which do exactly that so if a beard, say, covers the chin, the shape model can override the image to approximate the position of the chin under the beard. It was therefore natural (but perhaps only in retrospect) to adopt Point Distribution Models. This synthesis of ideas from image processing and statistical shape modelling led to the active shape model. The first parametric statistical shape model for image analysis based on principal components of inter-landmark distance was presented by Cootes and Taylor in (Rahman and Bhuiyan, 2008).

On this approach, Cootes, Taylor, and their colleagues, then released a series of papers that cumulated in what we call the classical active shape model (Burl et. al., 1955. Huang et. al., 1998).

Deformable templates were then introduced by Yuille et al. (Viola and Michael, 2004). to take into account the a priori of facial features and to better the performance of snakes. Locating a facial feature boundary is not an easy task because the local evidence of facial edges is difficult to organize into a sensible global entity using generic contours. The low brightness contrast around some of these features also makes the edge detection process problematic.

Low level analysis is based on low level visual features such as color, intensity, edges, motion, etc.

Color is a vital feature of human faces. Using skin-color as a feature for tracking a face has several advantages. Color processing is much faster than processing other facial features. Under certain lighting conditions, color is orientation invariant. This property makes motion estimation much easier because only a translation model is needed for motion estimation (Krizhevsky et. al. 2012).

When use of video sequence is available, motion information can be used to locate moving objects. Moving silhouettes like face and body parts can be extracted by simply thresholding accumulated frame differences (Ratsch et. al., 2004). Besides face regions, facial features can be located by frame differences (Wang and Ji, 2004).

Face detection based on edges was introduced based on analyzing line drawings of the faces from photographs, aiming to locate facial features. A hierarchical framework based on to trace a human head outline was proposed in (Sharif et. al., 2011). Remarkable works were carried out by many researchers in this specific area. Method suggested by Anila and Devarajan (Anila and Devarajan, 2010). was very simple and fast.

Feature analysis algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate

faces. These methods are designed mainly for face localization.

Paul Viola and Michael Jones, proposed a fast and robust method for face detection which is 15 times quicker than any technique at the time of release with 95% accuracy at around 17 f/s. The technique relies on the use of simple Haar-like features that are evaluated quickly through the use of a new image representation based on the concept of an-Integral Image.

Authors in (Muhammad et. al., 2011. Rahman and Bhuiyan, 2008) proposed elastic bunch graph map (EBGM) algorithm that successfully implements face detection using Gabor filters. The proposed system applies 40 different Gabor filters on an image. As a result of which 40 images with different angles and orientation are received. Next, maximum intensity points in each filtered image are calculated and mark them as fiducial points.

2.2 Image Based Approach

Different approaches have been found under image based approach like Neural Network Support Vector Machine (SVM) and Principle Component Analysis (PCA).

A method based on neural networks for face detection is given in (Krizhevsky et. al., 2008). Their network consists of layers with 1,024 input units, 256 units in the first hidden layer, eight units in the second hidden layer, and two output units. Method using auto associative neural networks is based on (Zhang and Zhang, 2014). which show an auto associative network with five layers is able to perform a nonlinear principal component analysis. Chen et al. presented a face detection system using probabilistic decision-based neural network (PDBNN) (Chen et. al., 2009). The architecture of PDBNN is similar to a radial basis function (RBF) network with modified learning rules and probabilistic interpretation.

An efficient method to train an SVM for large scale problems and applied it to face detection is developed in (Ratsch and Romdhani, 2004). Based on two test sets of 10,000,000 test patterns of 19×19 pixels, their system has slightly lower error rates and runs approximately 30 times faster than the system in (Wang and Ji, 2004).

PCA is a technique based on the concept of Eigen faces (Shah et. al., 2013). It is also known as Hotelling transform. Turk and Pentland proposed PCA to face recognition and detection (Susheel, 2011).

3. PROPOSED METHOD IMAGE BASED APPROACHES

Proposed algorithm is based on scanning a particular area that captures face parts across a given input image. The technique of standard image processing is to rescale the input image in to different sizes and after that apply fixed size detection on these images. We have used scale invariant technique that is a combination of integral image and simple rectangular features. Details of the proposed method are as given below:

3.1 Feature selection through scale invariant technique

This is first step of the proposed algorithm. In this step we tune the input image into an integral image. This process is done by making each pixel to sum of all concerned pixels above and sum of all concerned pixels to the left of each pixel which is shown in figure 1.

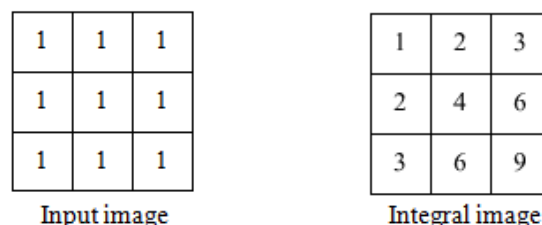


Fig. 1: Input and Integral Image.

So, four values can be used to find the sum of all pixels inside a given rectangle. Calculation of the sum of all pixels is as follows.

$$\text{Sum of rectangle (grey)} = A + D - (B + C) \quad (2)$$

It has now been shown that how the sum of pixels within rectangles of arbitrary size can be calculated in constant time. The proposed method analyzes a given sub-window using features consisting of two or more rectangles. Different types of features are shown in Fig. 2.

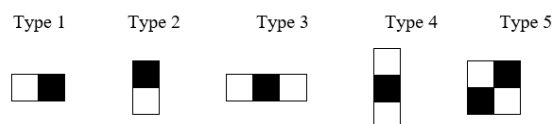


Fig. 2: The different types of features.

Each feature results in a single value which is calculated by subtracting the sum of the white rectangle(s) from the sum of the black rectangle(s). The proposed algorithm has empirically found that a detection approach with a base resolution of 24×24 pixels gives satisfactory results. When allowing for all possible sizes and positions of the features in Figure

4 a total of approximately 160,000 different features can then be constructed. Thus, the amount of possible features vastly outnumbers the 576 pixels contained in the detector at base resolution. These features may seem overly simple to perform such an advanced task as face detection, but what the features lack in complexity they most certainly have in computational efficiency.

3.2 The modified AdaBoost algorithm

To calculate the features as stated above a modified version of AdaBoost algorithms has been used. AdaBoost is a kind of machine learning boosting algorithm which constructs a strong classifier with the use of weighted combination of weak classifiers.

The proposed modified AdaBoost algorithm is as follows:

Step 1: Take sample images from (x_1, y_1) to (x_n, y_n) where for negative samples the value of $y_i = 0$ and for positive samples the values of $y_i = 1$.

Step 2: Initialize weights for negative and positive samples:

(a) For negative samples: weight $(w_i) = 1/3m$

Where $y_i = 0$

(b) For positive samples: weight $(w_i) = 1/3l$

Where $y_i = 1$

Here m denotes the number of negative samples and l denotes the number of positive samples.

Step 3: For $k = 1, \dots, n$

a) Normalize the weights

$$w_{t,i} \approx \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

b) Perform selection of the best weak classifier among all available weak classifiers with respect to weighted error:

$$\mu_t = \min_{f,p,\theta} \sum_i w_i |g(x_i, f, p, \theta) - y_i|$$

c) Definition of $g_t(x) = g(x, f_t, p_t, \theta_t)$ where f_t, p_t and θ_t are the minimizers of μ_t .

d) Update the weights: $w_{t+1,i} = w_t, i \beta^{1-e_i}$. Here the value of $e_i = 0$ if classification of i is correct sample x_i and otherwise value of $e_i = 1$ and $\beta_t = \mu_{t/(1-\mu_t)}$

Step 4: The strong classifier is defined as follows:

$$Classifier(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t g_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

Here $\alpha_t = \log(1/\beta_t)$

Modified AdaBoost algorithm is based on three things such as selection of best features with polarity and threshold. In order to find the best feature with high performance capability involves evaluation of each feature of all training examples.

An important part of the modified AdaBoost algorithm is the determination of the best feature, polarity and threshold. There seems to be no smart solution to this problem and the proposed method suggest a simple brute force method. This means that the determination of each new weak classifier involves evaluating each feature on all the training examples in order to find the best performing feature. This is expected to be the most time consuming part of the training procedure.

The best performing feature is chosen based on the weighted error it produces. This weighted error is a function of the weights belonging to the training samples. The weight to correctly classified sample is decreased and the weight of a misclassified sample is kept constant. As a result it is more expensive for the second feature (in the final classifier) to misclassify as ample also misclassified by the first feature, than as ample classified correctly. The point being that the weights area is the vital part of the mechanics of the AdaBoost algorithm.

3.3 Classification of Images

The basic principle behind the proposed algorithmic program is to scan the detector repeatedly through a constant image size. Although a picture ought to contain one or additional faces, it's obvious that associate degree excessive great amount of the evaluated sub-windows would still be negatives (non-faces). This realization ends up in a special formulation of the problem:

Instead of finding facial parts, the algorithm ought to discard non-facial parts.

The thought behind this statement is that it's quicker to discard non-facial parts than to search out facial parts. With this in mind a detector consisting of just one (strong) classifier suddenly appears inefficient since the analysis time is constant in spite of the input. Thus the necessity for a cascaded classifier arises.

This paper presents a number of the concerns relating to the implementation of the proposed methodology beside final results.

4. EXPERIMENTAL WORK AND RESULT ANALYSIS

This section presents some of the considerations regarding the implementation of the proposed method along with final results.

4.1 Creation of Positive Samples

The face training set consisted of roughly 800 hands tagged face pictures scaled and aligned to a base resolution of twenty four × twenty four pixels. Fig 3 shows a sample of positive pictures.



Fig. 3: Sample Positive Images.

The variance normalization is suggested as a mean of reducing the effect of different lighting conditions. Sample of normalized image is shown in Fig 4.



Fig. 4: Normalized Image.

4.2 Creation of Negative Samples

For the training of the completely different stages within the sequence classifier negative samples area unit are conjointly needed. A negative sample is essentially simply a picture not containing a face. This criterion might initially appear straightforward to fulfil, however ideally the negative examples ought to represent all kinds of non-facial textures that the planned technique are often expected to satisfy. Fig 5 shows the sample negative images.

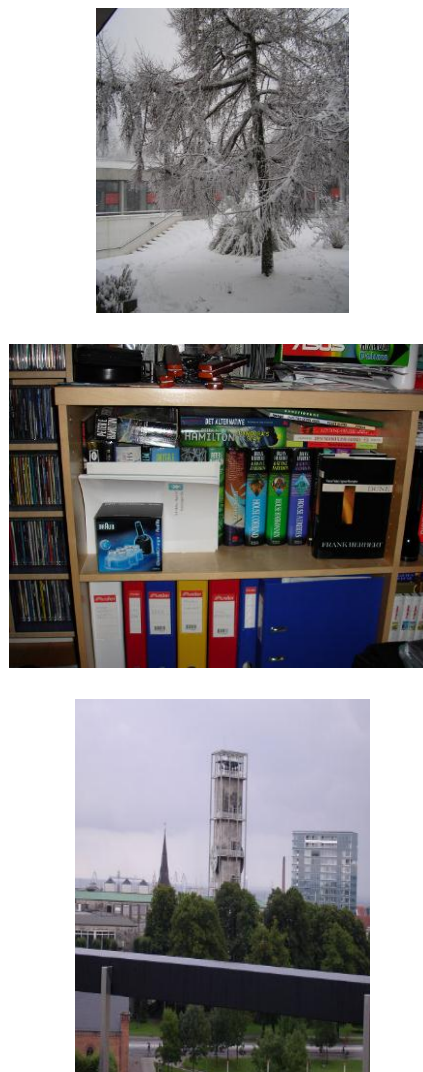


Fig. 5: Sample negative images.

4.3 Training the classifier

Each stage within the sequence classifier was trained employing a positive set, a negative set and an evaluation set. For every stage the positive set and therefore the analysis set was identical whereas the negative set was particularly designed for precisely that stage. As delineate earlier false positives square measure most well-liked over false negatives and since the AdaBoost algorithm aims at minimizing false negatives it desires a bit tweaking. A procedure

that bit by bit lowers the classifier threshold of a stage till a given performance demand is employed.

Since the evaluation set during this project solely consists of positive samples true positive rate became the key measuring for a stage's performance. Secondary, the false positive rate was calculable by holding the existing sequential classifier appraise the current negative examples. Un-remarkably it's not suggested to use the training data as evaluation data, however owing to the successive structure it will be allowed during this case. Particularly for the upper stages the negative coaching sets were generated by running through virtually innumerable examples and so making certain training sets nearly a twin of the real world knowledge the classifier is anticipated to encounter.

4.4 Visual Results

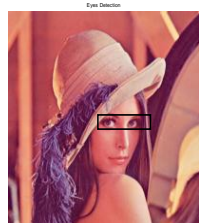
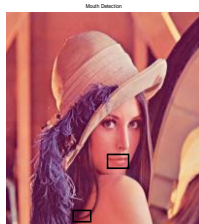


Fig. 6 (a) is standard 'Lena' image for which the proposed method is capable of recognizing different facial parts nose, mouth and eyes accurately.

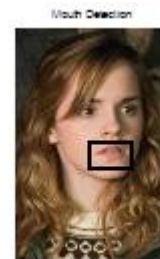
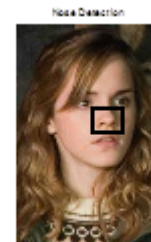


Fig. 6 (b) contains frontal-face for which the proposed method is capable of recognizing different facial parts nose, mouth and eyes accurately





Fig. 6 (c) contains a smiling face for which the proposed method is capable of recognizing different facial parts nose, mouth and eyes accurately.



Fig. 6 (d) shows that the proposed method correctly recognizes different facial parts nose, mouth and eyes accurately in presence of upper body part.

From all the above figures shown in Fig. 6, we can see that there are several challenges present such as change in illumination conditions, variation in face pose, variation in facial expressions, variation in age of persons and complex background. In spite of these challenges we have obtained correct detection results.

4.5 RESULTS

For performance evaluation of the proposed technique, four parameters have been considered. These parameters are True Positive Rate (TPR), False Negative Rate (FNR), False Positive Rate (FPR) and True Negative Rate (TNR).

On observing Table 1, it can be seen that the value of TPR is very high (almost near to 1) and FNR is very less (almost near to 0). Similarly, TNR is high and FPR is low. Hence, it is clear that the proposed technique performs well in terms of True Positive Rate (TPR), False Negative Rate (FNR), False Positive Rate (FPR) and True Negative Rate (TNR).

All experimental results have been obtained in a very less computation time. Therefore, time complexity is also very less for the proposed method.

Table 1: Evaluation Results for Images.

Sr. No.	Images	Evaluation set (Positive)		Evaluation set (Negative)	
		TPR	FNR	FPR	TNR
1	50	0.9750	0.0258	0.3814	0.6195
2	100	0.9750	0.0258	0.4154	0.5844
3	150	0.9750	0.0258	0.4917	0.5082
4	200	0.9750	0.0258	0.4854	0.5126
5	250	0.9750	0.0258	0.2262	0.7727
6	300	0.9750	0.0258	0.3311	0.6669
7	350	0.9750	0.0258	0.3968	0.6040
8	400	0.9722	0.0270	0.3537	0.6464
9	450	0.9710	0.0290	0.3933	0.6058
10	500	0.9710	0.0290	0.4132	0.5877
11	550	0.9681	0.0310	0.2801	0.7197
12	600	0.9681	0.0310	0.3043	0.6957
13	650	0.9681	0.0310	0.4323	0.5677
14	700	0.9681	0.0310	0.4871	0.5129
15	750	0.9681	0.0310	0.2355	0.7654
16	800	0.9681	0.0310	0.4382	0.5617

5. CONCLUSION

In the proposed work, an efficient approach has been applied to detect the facial parts of different images with minimum computation time and high detection accuracy. Techniques applied in the proposed work are very generic in nature and may have widely applicable in image processing.

Main contribution of the proposed work is to compute a rich set of image features by applying integral image. Using the integral image, face detection is completed in almost the same time as it takes for an image pyramid to be computed. Our proposed approach is very efficient in detecting face parts in the presence of wide range of conditions like illumination, scale, pose, and camera variation. On analysis of experiments it is found that our approach is highly efficient for face images.

The proposed approach is efficiently tested on face images. However, it may be extended to other various real life applications like object detection, face detection, and pedestrian detection, etc. for biometric recognition in surveillance systems.

REFERENCES

- A. Krizhevsky, I. Sutskever, G. E. Hinton (2008). Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.
- A. Krizhevsky, I. Sutskever, G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.
- B. Heisele, T. Poggio, M. Pontil (2000). Face detection in still gray images, Tech. rep., Center for Biological and Computational Learning, MIT, A.I. Memo 1687.
- B. K. Low, M. K. Ibrahim (1997). "A Fast and Accurate Algorithm Facial Feature Segmentation", Brands, Jelle, Tim Schwanen, and Irina Van Aalst (2016). "What are you looking at? Visitors' perspectives on CCTV in the night-time economy." European Urban and Regional Studies, vol. 23, no. 1, pp. 23-39.
- C. Zhang, Z. Zhang (2014). Improving multiview face detection with multi-task deep convolutional neural networks, in: Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on, IEEE, 2014, pp. 1036–1041.
- J. L. Crowley, F. Berard (2016). "Multi-model Tracking of faces for Video Communications".
- M. C. Burl, T. K. Leung, and P. Perona (1995). Face localization via shape statistics, in Int. Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland.
- M. Ratsch, S. Romdhani, T. (2004). Vetter, Efficient face detection by a cascaded support vector machine using haar-like features, in: Pattern Recognition Symposium.
- M. Ratsch, S. Romdhani, T. Vetter (2004). Efficient face detection by a cascaded support vector machine using haar-like features, in: Pattern Recognition Symposium.
- Muhammad SHARIF, Adeel KHALID, Mudassar RAZA, Sajjad MOHSIN (2011). Face Recognition using Gabor Filters, Journal of Applied Computer Science & Mathematics, no. 11(5)/2011, Suceavapp, pp. 53-57.
- P. Wang, Q. Ji (2004). Multi-view face detection under complex scene based on combined svms, in: Proc. of ICPR.
- P. Wang, Q. Ji (2004). Multi-view face detection under complex scene based on combined svms, in: Proc. of ICPR.
- Paul Viola, Michael Jones (2004). Robust Real-Time Face Detection, International Journal of Computer Vision 57(2), pp. 137–154.
- Paul Viola, Michael Jones (2004). Robust Real-Time Face Detection, International Journal of Computer Vision 57(2), pp. 137–154.
- Rahman M. T. and Bhuiyan M. A. (2008). Face Recognition using Gabor Filters, 11th IEEE International Conference on Computer and Information Technology, 2008. ICCIT 2008, pp. 510-515, 33.
- Rahman M. T. and Bhuiyan M. A. (2008). Face Recognition using Gabor Filters, 11th IEEE International Conference on Computer and Information Technology, ICCIT 2008, pp. 510-515.

S. Anila, N. Devarajan (2010). Simple and Fast Face Detection System Based on Edges, International Journal of Universal Computer Sciences, vol. 1, no. 2, pp. 54-58.

Sam Higginbottom University of Agriculture, Technology & Sciences, Allahabad

E-Mail – satishsinghcs@gmail.com

Shah J., Sharif M., Raza M. and Azeem A. (2013). A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques, The International Arab journal of Information technology, 10(6).

Sharif M., Raza M. and Mohsin S. (2011). Face Recognition Using Edge Information and DCT, Sindh Univ. Res. Jour. (Sci.Ser.), 43(2), pp. 209-214, 29.

Sharif M., Raza M. and Mohsin S. (2011). Face Recognition Using Edge Information and DCT, Sindh Univ. Res. Jour. (Sci.Ser.), 43(2), pp. 209-214.

Susheel Kumar K. (2011). Bhaskar Semwal V. and Tripathi R., Real time face recognition using adaboost improved fast PCA algorithm.

W. Huang, Q. Sun, C. P. Lam, and J. K. Wu (1998). A robust approach to face and eyes detection from images with cluttered background, in Proc. Of International Conference on Pattern Recognition.

Wilber, Michael J., Vitaly Shmatikov, and Serge Belongie (2016). "Can we still avoid automatic face detection?" In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-9.

Wu, Yu-Lung, and Chi-Jui Chang (2016). "Surveillance of public space: CCTV, privacy and sense of safety." Global Journal for Research Analysis, vol. 5, no. 4.

Y.-N. Chen, C.-C. Han, C.-T. Wang, B.-S. Jeng, K.-C. Fan (2009). A cnn-based face detector with a simple feature map and a coarse-to-fine classifier, Pattern Analysis and Machine Intelligence, IEEE Transactions on (99) pp. 1–1.

Yang, Shuo, Ping Luo, Chen-Change Loy, and Xiaoou Tang (2016). "Wider face: A face detection benchmark." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5525-5533.

Corresponding Author

Satish Kumar Singh*

Satish Kumar Singh^{1*} Narendra Kumar Gupta²