Real Time Multiple Face detection from Live Camera, a Step towards Automatic Attendance System

Raikoti Sharanabasappa
Research Scholar
India.

Sanjaypande M. B
Prof. and Head of Dept.,
VVIET, Mysore, India.

ABSTRACT
Student attendance system is manual in most part of the world with essential Roll call and answering taking significant time. The objective of this work is to propose a model in openCV that captures live stream from camera and enables multiple face detection and segmentation. The segmented faces can further be used to recognize the students. As such the system leads towards the development of automatic attendance system, where the camera can be static and periodically can take the snap of the class. Further each image is processed to extract the faces. Haar cascade is used for face detection and Gaussian mixture model is used for face segmentation. A test over 1000 images reveals a result with 83% accuracy where accuracy is measured in terms of number of actual face detected v/s the number of faces present in a scene. The test are generated in various angles and light intensity.

Keywords
Face Detection, OpenCV, Haar Cascade, Gaussian Mixture Model

1. INTRODUCTION
Human Face conveys rich information about identity and emotional state of the person. Face detection in a complex scene and its segmentation from the background forms a prerequisite for any practical verification system using face as the main attribute. Therefore efficient face segmentation in real time environment is the most important step towards the goal. The system is primarily divided into Segmentation, Training and Testing. The role of segmentation process is to extract the face part from the rest of scenes.

Face segmentation is further a two step process. Firstly, the system is trained with several template faces. The template faces are arbitrary faces that are selected from the video frames, which may or may not include the faces of registered users. Some of the common back ground object templates are also stored which includes wall background, chairs and so on. As the system starts capturing the video, using a Normalized Cross Correlation technique the faces in the templates are matched with the scene and current faces are marked.

The marked faces are segmented and are stored in a face array. The size of the faces are normalized for faster and efficient detection. Each faces is recursively matched against the registerd database of faces and the instance is saved.

Considering that there would be some mis detection based on the angle that a student face is in a frame, the recognition is called through an asynchronous event at every T second where T must be large enough for successful detection and efficient recognition and must not be so large that the system can not correlate the initial background with the new one due to significant change in the intensity levels.

2. RELATED WORK
Skin color processing is easier than processing other facial components. Due to variations in lighting conditions, color appeared in the frames may differ but still remain invariant of orientation. The major difference between skin colors are Energy of the pixels [3]. Human skin color is different from the color of most other objects commonly appearing in the frames. Skin and Non Skin modeling is done in[4]. Color Space is an important statistical model for Segmentation of skin regions. The segmentation is efficient only if the chrominance is used in rather than the actual RGB color space for analysis. The process eliminates variations of luminance component with maximum possibility by choosing the CbCr plane (chrominance components) of the YCbCr color space to build the model. Various Research and Papers have demonstrated that skin color is grouped in a small region of the chrominance space [4]. unimodal Gaussian [3] model can effectively model such a clusterd color space. A huge database of skin pixels are needed to train the Gaussian model. The mean and the covariance of the database are the features which can be used for extraction of detection of further faces. [6] Presents the essential density mixture model relevant to the implementation.

3. METHODOLOGY
Contribution
There are several Face detection or segmentation and recognition algorithm already. With OpenCV and AForge.Net providing real time libraries for image processing and video processing, modeling the alogorithms practically has been easy. The main contribution of this work is GMM based face extraction. Note that several authors have urged that a face detection be followed by skin segmentation. But through our work we have proved that face pixels offers a unique color space which even though falls in the skin map, still offers a significant texture difference from the rest of the skin part. Hence detection is fast and efficient.

Secondly the recognition is performed over the normalized segmented faces. The normalization process adopts resampling with Gaussian interpolation which keeps is model intact even in the new space.

Mathematical Model

Let $c = [Cb\ Cr]^T$ denote the chrominance vector of an input pixel from an input image that may contain multiple faces. Then the probability that the given pixel lies in the face is given by:

$$p(c/Face) = \frac{\exp\left[-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)\right]}{2\pi\sqrt{|\Sigma_s|}} \quad (1)$$

where μ_s and Σ_s are the mean ovariance matrix respectively of the training data. Therefore the mean and the covariance needs to be estimated from the training data to characterize the skin color distribution as a unimodal Gaussian. This model is used to obtain the Skin Probability image of an input color image.

In the face extraction process with Gaussian model described above, the probability of each color value, given it is a skin color, is a linear combination of its probabilities calculated from the M Gaussian components. Hence the probability of a pixel

$$c = [Cb\ Cr]^T$$

given it is a skin pixel, is:

$$p(c/Face) = \sum_{j=1}^{M} p(c/j).P(j) \quad \textbf{(2)}$$

Where In a Mixture model ,M depicts the number of Gaussian, $P(j)$ is the probabilistic weight of the j^{th} component. This is also referred to as the prior probability of the data point having been generated from the component j of the model.

$$p(c/j) = \frac{\exp\left[-\frac{1}{2}(c-\mu_j)^T \Sigma_j^{-1}(c-\mu_j)\right]}{2\pi\sqrt{|\Sigma_j|}} \quad (3)$$

where μ_j is the mean and Σ_j is the covariance matrix of the j^{th} component

Note that the probabilities are chosen so that (4) is satisfied.

$$\sum_{j=1}^{M} P(j) = 1; 0 \le P(j) \le 1 \quad (4)$$

Therefore the parameters to be estimated from the given pixel matrix are M, μ_j , Σ_j ,and $P(j)$, j = 1 to M, i.e. for every M components. In this work, the number of components is decided automatically by a constructive algorithm [5] using the criteria of Maximum likelihood estimator. Once M is decided, the other parameters, of each component are calculated from the given matrix. One of the most adopted approach is to maximize a likelihood function of the parameters for the given set of data [8]. The negative log-likelihood for the dataset is given by:

$$E = -\ln L = -\sum_{n=1}^{N} \ln p(c_n) = -\sum_{n=1}^{N} \ln\left\{\sum_{j=1}^{M} p(c_n/j).P(j)\right\} \quad (5)$$

which can be regarded as an error function. Note that $\boldsymbol{L}$ is equivalent to minimizing the error function $\boldsymbol{E}$, N is the number of data points c_n. Maximizing the likelihood. A special case of Maximum Likelihood (ML) techniques is the Expectation Maximization (EM) algorithm. This algorithm has been used to determine the parameters of the mixture model that best fit the data in the ML sense.

3.1 Algorithm for Skin Modeling

Let i denote the iteration, M_i the number of components at the *ith* iteration and $\boldsymbol{L_i}$ the likelihood for the validation set w.r.t the model at iteration i. The initial number of components can be taken to be M_0 = 1. The algorithm [5] for model order selection can be outlined as follows:

1. Apply Expectation-Maximization for model with M_i components.
2. Compute $\boldsymbol{L_i}$ for validation set.
3. Save model
4. Find component j with the lowest total responsibility
5. Split component j
6. Restart from step 1 with $M_{i+1} = M_i+1$ and $i = i+1$.

The above sequences of steps are repeated until M_i reaches a desired value (10 in this implementation). The peak in the Likelihood function for the checking the data corresponding to the optimal model order. The parameters of this optimal order model are then used to estimate *p(cFace).* Once the Face texture is modeled using Gaussian, it is used to calculate the probability of whether an input pixel in the matrix is representing a face region ,i.e, $p(Face/c)$, where c is the input color value. The Gaussian model can be used to evaluate the probability of a color value given it is a skin color, i.e., *p(c/skin)*. This is again used to compute the required probability $p(skin/c)$ using the Bayes' formulation [3]:

$$p(skin/c) = \frac{p(c/Face)P(Face)}{p(c/Face).P(sFace) + p(c/non-NonFace).P(non-Face)} \quad (6)$$

To calculate the above probability for each input pixel, the skin and the non-skin classes are assumed to occur with equal probability [4]. Hence

$$P(skin) = P(non\text{-}skin) = 0.5 \quad (7)$$

which gives,

$$p(skin/c) = \frac{p(c/Face)}{p(c/Face) + p(c/non-Face)} \quad (8)$$

To obtain the probability, *p(c/non-skin)*, a non-skin or the background model is build with same Gaussian distribution.

Once the models are correctly obtained, face segmentation is performed using connected components. Connected components are mathematical operators that operates on results of binary structure post GMM. Once skin part is segmented from rest of the image, they need to be eliminated from being considered as probable face candidate regions. As a first step, gray level 'open' operation is performed. This operation involves gray level erosion followed by gray level dilation using the same structuring element. Erosion removes small and thin isolated noise-like regions that have very low probability of representing a face. Dilation preserves those regions that are not removed during erosion. Hence, the effect of using area open is removal of small but bright regions of the skin probability image. This is followed by gray level 'close' operation. Closing is dilation followed by erosion using the same structuring element. The dilation during close operation enhances small regions of low intensity that may lie within large regions of high intensity in the skin probability image. Hence, during the thresholding step that follows, holes are not created within large high probability regions with a small gray level depression inside their periphery. These depressions may be caused due to bad lighting conditions or the skin model may fail to give a high probability in those regions. The erosion (of close operation) removes the extra pixels that may be added, during the previous dilation operation, as high probability pixels around existing regions. A smaller structuring element is used for close operation so that a large area of pixels around existing regions is not enhanced [3].

This image is then thresholded into a binary image for further shape analysis. A threshold of 60 is chosen here so that large areas of relatively smaller gray levels that remain after open/close operation are not excluded from shape analysis. The connected components are labeled and isolated and shape analysis is done separately on each connected component. A hierarchy of 3 shape based connected operators [3] is used for deciding whether a component represents a face or not. These simple but effective operators rely on the combinations of the pixel area (A), perimeter (P) and the bounding box dimensions (Dx, Dy) of the connected components. Hence these have to be computed only once for the three operators. Finally a normalized area operator is used that rejects connected components that have face-like shape but have pixel area less than a certain fraction of that of the largest face component detected. The choice of this operator is based on observations made on a number of images containing multiple faces.

4. RESULTS

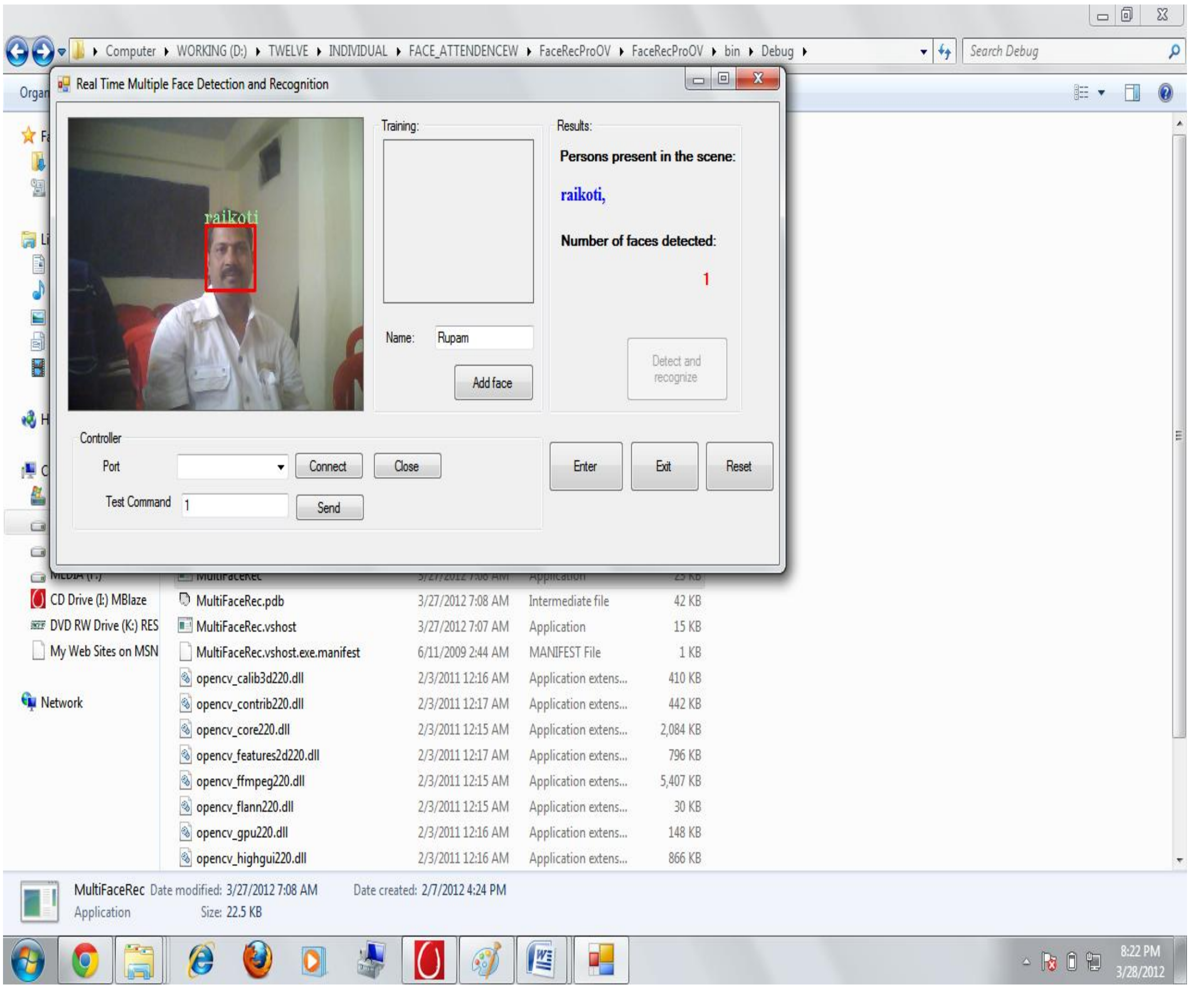


Fig 1 Detection of Faces at a distance from Camera

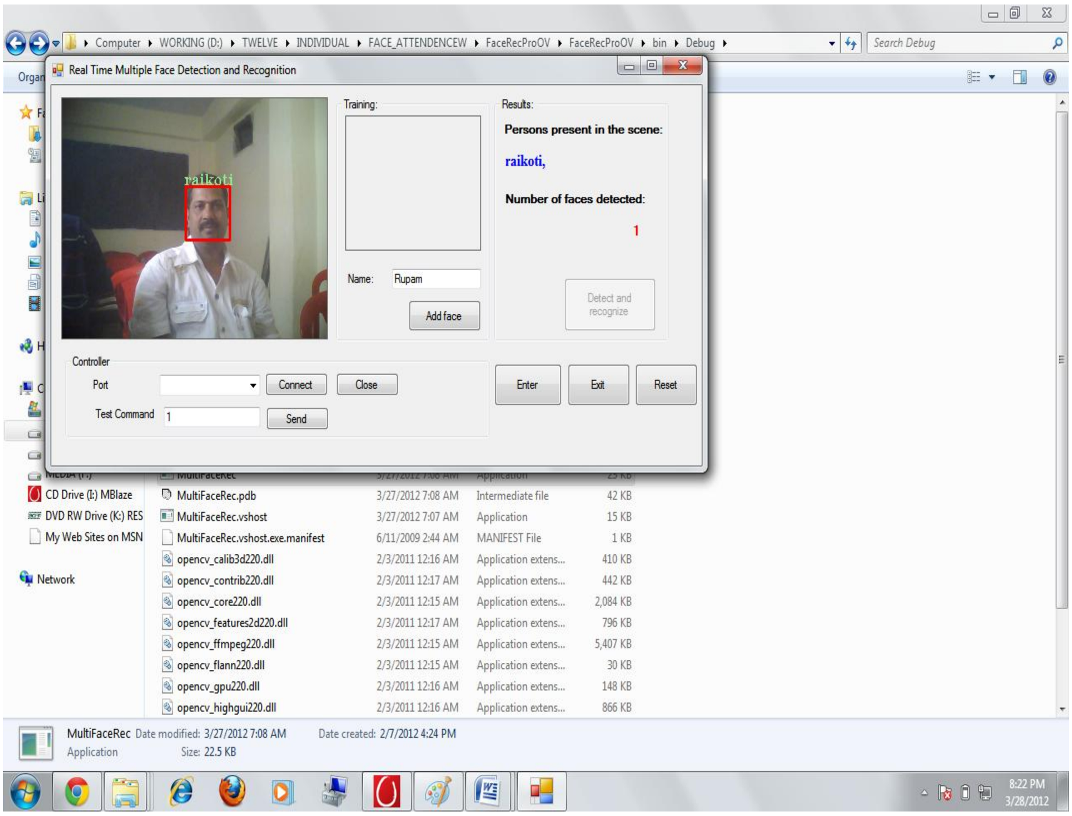


Fig 2 Detection of Single Close Face

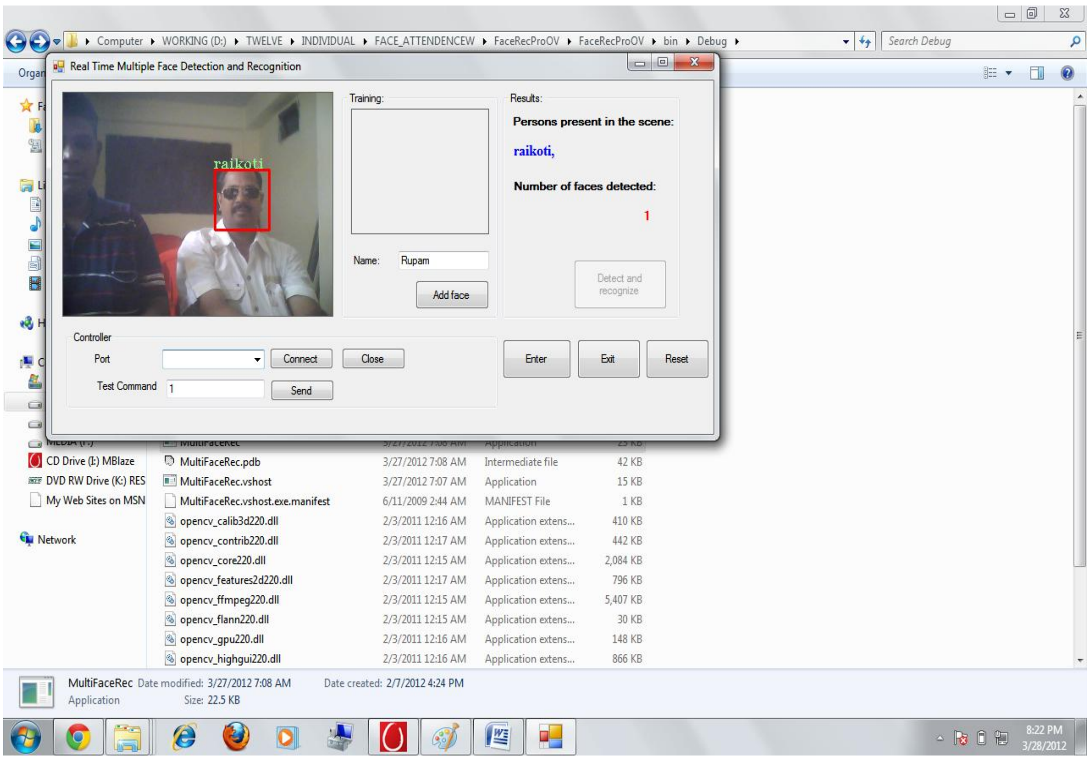
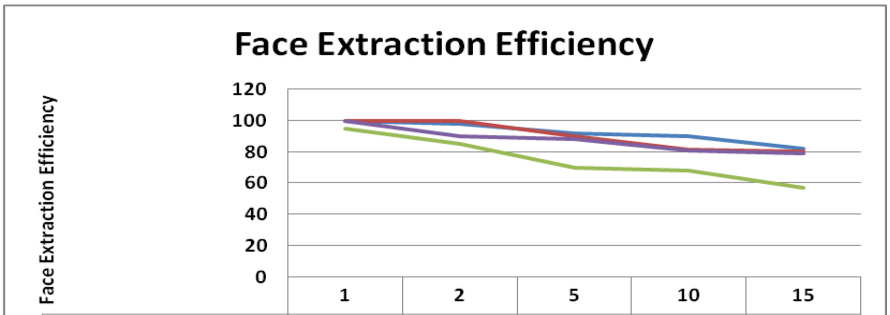
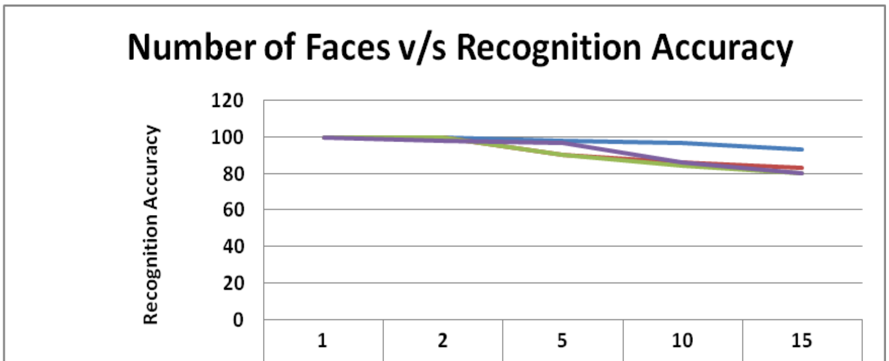


Fig 3 Detection of Face with dark glasses



	1	2	5	10	15
Proposecd Method	100	98	92	90	82
Haar Cascade Based Technique	100	100	90	81	80
Color Segmentation	95	85	70	68	57
PCA based Face Segmentation	100	90	88	81	79

Fig 4 Comparison of Face Extraction efficiency of different methods



	1	2	5	10	15
Proposecd Method	100	100	98	97	93
PCA	100	100	90	86	83
LBP	100	100	90	84	80
Gabor	100	98	97	86	80

Fig 5 Comparison of Number of faces versus Recognition Accuracy of proposed method with different techniques

5. CONCLUSION

Several face detection algorithms are being proposed over the years. Most of these algorithms are based on complex algorithmic process which makes them difficult to be detected from the live stream. Haar Cascade based solution provided with OpenCV suffers from lot of misdetection due to pose and intensity variations. The model and the technique proposed here is ideal for the id real time goal of achieving automatic face detection and recognition. Beside the algorithm provides a balance between speed and efficiency. It can be further improved by using better classifiers like support vector machines.

6. REFERENCES

[1] Chellappa, Wilson, Sirohey, "Human and Machine Recognition of Faces: A Survey", Proceedings of the IEEE, Vol. 83, No. 5, May 1995, pp. 705-740.

[2] M. Turk and A.P. Pentland, "Face recognition using eigenfaces", Proc. CVPR, 1991, pp 586-593.

[3] Menser, Wien, "Segmentation and Tracking of Facial Regions in Color Image Sequences", RWTH, Aachen, Germany, 1999.

[4] Jones, Rehg, "Statistical Color Models with Application to Skin Detection", Tech-Rep. CRL 98/11, Compaq Cambridge Research Lab 1998.

[5] Y. Raja, S. McKenna and S. Gong, "Colour Model Selection and Adaptation in Dynamic Scenes". In Proc. European Conference on Computer Vision, Freiburg, Germany, 2-6 June 1998.

[6] C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.

[7] M.-H. Yang and N. Ahuja, "Gaussian Mixture Modeling of Human Skin Color and Its Applications in Image and Video Databases", In the 1999 SPIE/EI&T Storage and Retrieval for Image and Video Databases, pp. 458-466, San Jose, January, 1999 .

[8] R.A. Redner, and H.F. Walker, "Mixture densities, maximum likelihood and the EM algorithm", SIAM Review, Vol. 26, no. 2, pp. 195-239, 1984.