Face Recognition based Attendance System

Mr. Vinay Sripathi[1], Ms. Neha Savakhande[2], Mr. Kiran Pote[3], Mr. Parth Shinde[4]
Guided By: Dr. Jayant Mahajan[5]

[1,2,3,4,5] *Department of Electronics and Telecommunication Engineering, Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, India*

Abstract - *In the field of image analysis and computer vision, one of the most arduous tasks presently considered is Face recognition. The biometric system which basically works on the principle of face recognition is used for the identification or verification of a person from a digitalized image preferably used in surveillance, security and attendance purpose. Diving into matter of concern, to preserve the record generated by attendance of the student and co-operating ordinary activities becomes a tedious task all together. The run-of-the-mill method of calling the name of each student is time ingesting and there is always a risk of proxy attendance. The defined system is based on real-time multiple face recognition to maintain the attendance record of students. The variation in illumination and posing as well as focus issues and blurring are important factor to be considered during face identification process in the classroom using real-time videography. This above-mentioned problem can be nullified to some extent by the proposed system. The execution is performed with the help of Histogram of Oriented Gradients algorithm for detecting the faces, which is done by its excellent performance in differentiating feature descriptor of individuals. The training data sets comprises of different qualities of images. These data sets are fed to Convolutional Neural Network, through which the result of recognition is thoroughly obtained. Thus, making the attendance process work uninterruptedly.*

Key Words: Face Detection, Haar cascade classifier, Face Recognition, HOG, Attendance System, CNN, Euclidean distance, Face identification, Embedding.

1. INTRODUCTION

The process through which we can recognize a face, involves certain steps that has to be mandatorily followed to get the desired output. The steps are as follows, firstly it begins with collecting the strong data sets for the students in the class, followed by training them with a good model, later on detecting the faces with a good model at the time of actual attendance, then feature extraction, classification of the data acquired and recognition of faces, ending it by transferring the faces recognized to a excel sheet through CSV. During the course of technological advancement in the world, plethora of algorithms were developed and improvised especially for detection and identification. The implementation of the overall project can subject to vary between real- time and digital photo frame. However, the process of feature extraction in both cases remains more or less the same. They might include eyes, nose, lips, ears etc. when we skim around through various algorithms, the process of training the faces and storing in database and then matching it with the input faces remains the same [1].

Albeit, face identification has diversified its popularity and applications, it stands challenging task to execute. In real-time facial recognition certainly, there are problems to be dealt with such as illumination, focus, blur, pose etc. The necessity of making the identification and recognition buttress under a varying environmental situation, where the captured images are likely to get deteriorated and also lose some of the essential details due to improper light conditions such as, under and over exposure of the image or video which is a major criterion for getting a functioning output. Due to these mentioned problems, the images or video look different at various time instant and leading it to a misclassification of input image to the trained dataset images. The mightier problem of all the problems that is pose variation, it makes same images to look different which occurs due to change in viewing angles. Due to this, it creates an obstacle while recognizing a face. In a face identification model, detection plays an important and crucial role when speaking overall. Various algorithms like histogram of oriented gradients, Local binary pattern histogram, Fisher faces eigenfaces. Our project inculcates histogram of oriented gradients (HOG). Because we found working with HOG more efficient and performance was, satisfactory HOG takes the input image and juxtapose the pixels with other pixels in terms of the darkness. Due to this, light problems are solved to some extent and detection was accurate when seen and observed in different conditions [2].

2. RELATED WORK

A) Haar Cascade

Haar classifier works through following steps:

1) Haar Features

As the training data is given to the model, the classifier starts taking out features from each face. Haar features are convolution kernels, which confirms that a particular feature is present on an image, or not.

2) Integral Image

Haar Cascade classifier makes use of 24 x 24 base window size, which produces more than 180,000 features. It is difficult to compute pixel difference for all the facial characteristics. By considering the four corner values of any rectangle, all the pixels are added which results in integral image.[12]

3) Adaboost

As 180,000 features are being generated in integral image step. However, not all features are required to detect a face. Therefore, to choose the best facial characteristics from the whole set, adaboost a ML algorithm is implemented. This decreases the features count from 180,000 to approximately 6000.

4) Cascade

Since it is difficult to keep all the 6000 extractions on a window, the features are combined into different steps of classifiers and applied one at a time. The window is removed if it fails in primary stage. If the window passes, apply the second step of facial characteristics and follow the process. The window which passes through all steps is recognized as a face region. [5]

B) Histogram of Oriented Gradients (HOG)

Histograms of Oriented Gradients (HOGs) are a feature descriptor that has been widely and successfully used for object detection. It represents objects as a single feature vector as opposed to a set of feature vectors where each represents a segment of the image.

In face identification steps, first we must find all of the faces in an image. This detection is accomplished by making use of HOG within the acquired photo. Before applying HOG, the input picture is converted from RGB into gray scale photography for simplicity. It is computed by sliding window detector over an image, where a HOG descriptor is computed for each position. By using an 8 x 8-pixel detection window or cell, gradient vector or edge orientations at each pixel are calculated. This generates 64 (8x 8) gradient vectors which are then represented as a histogram. Each cell is then split into angular bins, where each bin corresponds to a gradient direction (e.g. x, y). This effectively reduces 64 vectors to just 9 values. As it stores gradients magnitudes, it is relatively immune to deformations. Histogram counts are normalized so to curtail the loss due to illumination. This is accomplished by accumulating measure of local histogram gradient over the higher linked region and to support those, the outcomes are used to normalize all cells in the block. This imperturbable positioning of histogram characterizes the eventual hog descriptor. By withdrawing the descriptors through only primary key points within the scale space of the image using a rotation normalization, the scale and rotation invariance can be found.[4]

Change in x and y direction can be calculated as:

$G_x(y, x) = Y(y, x+1) - Y(y, x-1)$; $G_y(y, x) = Y(y+1, x) - Y(y-1, x)$

$Y(y, x)$ – pixel intensity

$G_x(y, x)$ – change in x- direction

$G_y(y, x)$ – change in y- direction

The values of change in x and y direction helps in finding a gradient vector. We can compute the magnitude and angle of a vector as shown below:

Consider gradient vector as: $\begin{bmatrix} 36 \\ 36 \end{bmatrix}$

Magnitude = $\sqrt{(36)^2+(36)^2}$ = 50.91

Angle = arctan $\left(\frac{36}{36}\right)$ = 0.785 rads

= 45 degrees

Working of HOG is shown below:

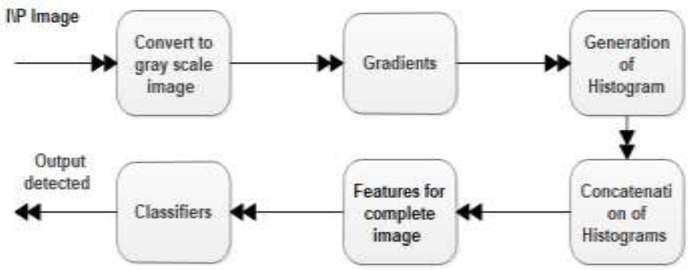

Fig 1: Block diagram for HOG.

C) CNN

A convolutional neural system (CNN) is a profound, feed-forward counterfeit neural system in which the neural system saves the various leveled structure by learning inner component portrayals and summing up the highlights in the basic picture issues like item acknowledgment and other computer vision issues. It isn't confined to pictures; it additionally accomplishes best in class brings about normal language preparing issues and discourse acknowledgment.

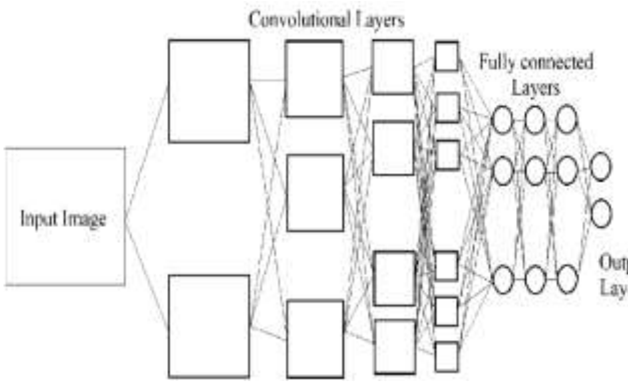

Fig 2: Layer Structure.

Convolution means to extricate highlights from the info picture, and thus it safeguards the spatial connection between pixels by learning picture highlights utilizing little squares of information. Rotational invariance, translation invariance, and scale invariance can be normal. For instance, a pivoted feline picture or rescaled feline picture can be effortlessly recognized by a CNN as a result of the convolution step. You slide the channel (square network) over your unique picture, and at each given position, you figure component insightful duplication (between the frameworks of the channel and the first picture) and add the increase yields to get the last whole number that shapes the components of the yield lattice.[2]

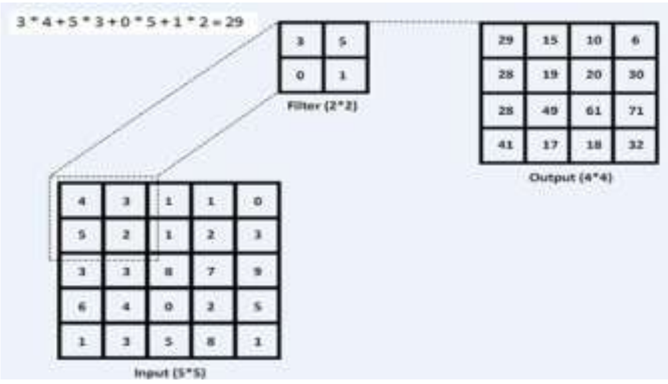

Fig 3: Subsampling

As appeared in Figure, channels have input loads and create a yield neuron. Suppose you characterize a convolutional layer with six channels and responsive fields that are 2 pixels wide and 2 pixels high and utilize a default walk width of 1, and the default cushioning is set to 0. Each channel gets contribution from 2� pixels, segment of picture. As it were, that is 4 pixels one after another. Henceforth, you can say it will require 4 + 1 (inclination) input loads. The information volume is 5�� (width �stature �number of channel), there are six channels of size 2� with walk 1 and cushion 0. Thus, the quantity of parameters right now each channel has 2*2*3 + 1 = 13 parameters (included +1 for inclination). Since there are six channels, you get 13*6 = 78 parameters.

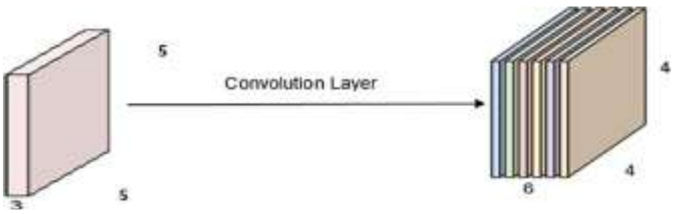

Fig 4: Input Volume

The pooling layers diminish the past layers' enactment maps. It is trailed by at least one convolutional layer and merges all the highlights that were found out in the past layers' enactment maps. This decreases the overfitting of the preparation information and sums up the highlights spoken to by the system. The responsive field size is quite often set to 2� and utilize a walk of 1 or 2 (or higher) to guarantee there is no cover. You will utilize a maximum activity for each open field with the goal that the initiation is the most extreme info esteem. Here, each four numbers guide to only one number. Along these lines, the quantity of pixels goes down to one-fourth of the first right now.[14]

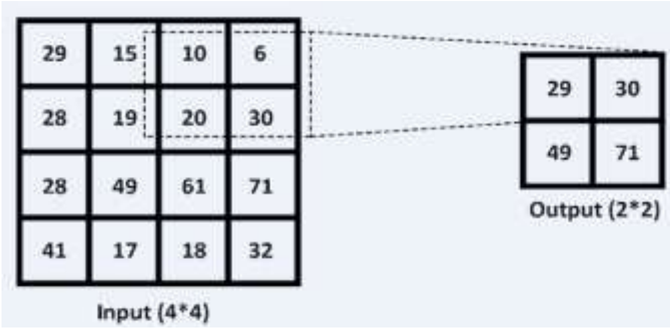

Fig 5: Maxpooling-reducing the number of pixels

A completely associated layer is a feed-forward fake neural system layer. These layers have a nonlinear initiation capacity to yield class forecast probabilities. They are utilized close to the end after all the highlights are recognized and separated by convolutional layers and have been solidified by the pooling layers in the system. Here, the covered up and yield layers are the completely associated layers.

D) Euclidean Distance

Euclidean distance is one of the most used distance metrics. It is calculated using Minkowski's Distance formula by setting p's value to 2. Thus, the distance 'd' formula is shown below:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Euclidean distance formula, which looks similar to the "Pythagoras Theorem", can be used to calculate the distance between two data points in a plane.

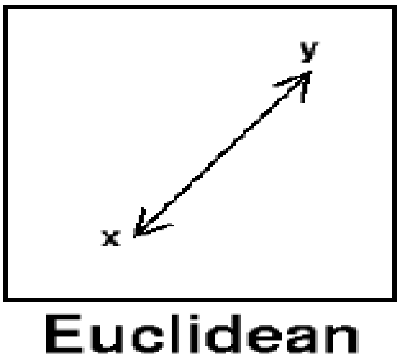

Fig 6: Euclidean Distance

It is shown in Figure, by using this formula of distance, Euclidean space becomes a metric space. The Euclidean distance between points x and y is the length of the line segment connecting them [xy]. In Cartesian coordinates, if x = (x1, x2, ... , xn) and y = (y1, y2, ... , yn) are two points in Euclidean n-space, then the distance from x to y, or from y to x is given by: d (x, y). In three-dimensional Euclidean space, the distance is:

d (x, y) = $(x1 - y1)^2 + (x2 - y2)^2 + (x3 - y3)^2$

The Euclidean distance between landmarks is used by most authors as a morphometric measure. Once facial feature points are obtained from a facial image or a two-dimensional face, they select some significant distances between them and compute the corresponding Euclidean distances. Then these distances are used to compare faces for face recognition systems. The Euclidean distances computed between all possible pairs the facial feature points constitute a vector of number of elements equal to the number of distances. This vector gives the human face features of 2D image and used as input of classification algorithm for face recognition system.

3. PROPOSED SYSTEM

The proposed system is to develop a face recognition-based attendance system with satisfactory rate of recognition from a classroom. First and the most important step in proposed system is to collect database of classroom students. Databased collection is to be done using Haar cascade classifiers to find face when webcam is turned on to collect database (images of face) of a student.

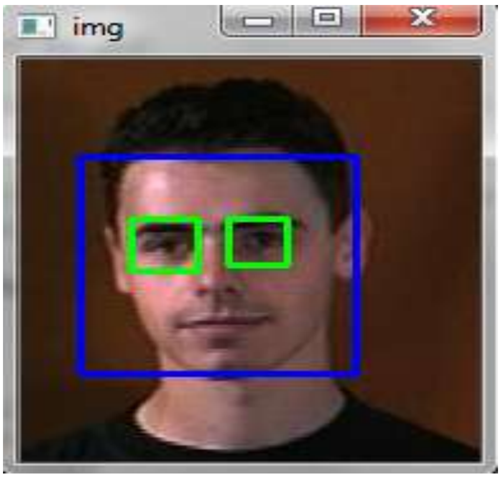

Fig 7: Detected Face

The above image is the result of Haar Cascade, where face is detected with the help of webcam and depending upon the value of n, represented in the program the number of snap shots is taken and stored as datasets.

Captured database images are stored in respective folder with respective name of student. After collecting database of all students, the encodings are to be extracted from those images using CNN model.

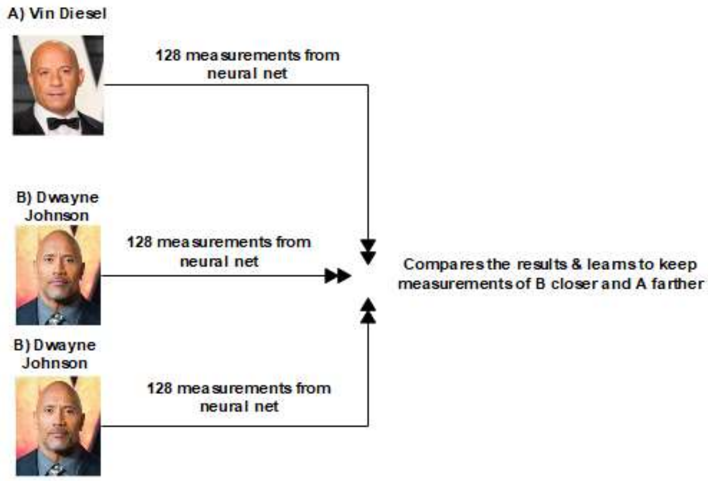

Fig 8: Triplet Training

Extracted encodings are then stored with respective student's name in a pickle file.

At the third step, webcam is turned on to capture the classroom. Faces are to be detected using HOG (Histogram of Gradient) method. Detected faces' encodings are extracted in real time and are to be compared with those stored previously with appropriate student names. Based on maximum counts gathered after comparing the extracted encodings and stored encodings, names identified in real time. Internally compare_faces function is used to compute the Euclidean distance between face in image and all faces in the dataset. If the current image is matched with the 85% threshold with the existing dataset, respective name is identified and marked in the attendance report being generated. Recognized names are then to be stored in a CSV file with time & date. Main aim of the proposed system is to develop a recognition algorithm that will identify maximum no. of faces correctly in real time with low false acceptance rate. The below block diagram refers to the complete proposed system.

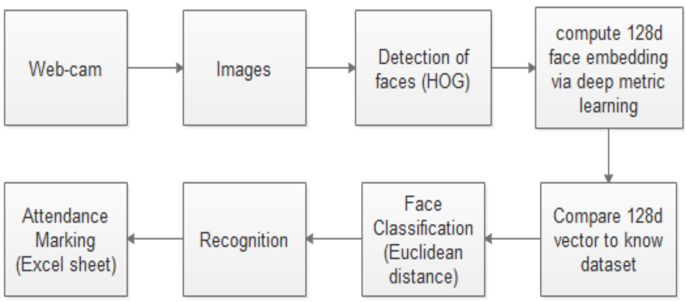

Fig 9: Block Diagram of Proposed System

4. Experimental Results

The results were generated in 3 different cases for understanding the impact of training the datasets, light conditions, pose variations.

Fig 10: Case-1

Case-1 was the first result through the proposed model. The arrangements of the model were done in a classroom in presence of little sun light and moderate classroom tube lights. Due to uneven lightening condition i.e. light coming from the window dominated the light which was in the room resulting in undersaturation on the face and oversaturation around the window thereby leading to drop in efficiency 30.77%.

Fig 11: Case-2

Case-2 was the second attempt which was performed in the Laboratory. This time around we made sure of the lightening conditions by balancing the light in front to the back side of the image. Efficiency increased by 7.69% compared to the case-1. But efficiency was not up to the mark due to the data set not containing the various pose variations of the students, meaning, the datasets in the database only contained frontal faces of the students.

Fig 12: Case-3

Case-3 was accomplished in the laboratory with just the tube lights. This case led us to the 100% efficiency due to even lightening condition and strong dataset generation. The dataset contained the frontal faces as well as side faces, resulting total faces for an individual about 50 images.

The below Table shows detailed result of all the cases.

Case no.	Present faces	Detected faces	Recognized faces	False recognition	Efficiency of recognition
1	13	11	10	1	69.23%
2	13	13	10	2	76.92%
3	8	8	8	0	100%

Table 1: Result

There are many papers based on algorithms like HOG and CNN combinations. For example [14]. The working of the model remains same but interpretation with respect to dataset and classification of images changes. By making use of Euclidean Distance for image Classification the hardware (laptop) is able to process the datasets as well as recognize more than 5 students. When using SVM as a Classifier the processing gets too complex for a decent laptop to perform the process. So, the maximum possible students which can be detected through the SVM classifier is 3 to 4 [14].

5. Conclusion

This system gives satisfactory results if conditions of capturing images is kept ideal i.e. sufficient illumination and enough head pose variation in dataset. The system uses HOG for face detection and CNN encodings to recognize multiple present faces in the frame, but the accuracy and efficiency of recognition process is hampered by factors like false acceptance due to mismatched encodings, poor lighting and occlusion. Therefore, the proposed system is useful in classroom attendance but its shortcomings cannot be ignored either.

6. Future scope

As the results state, the recognition efficiency drops as the number of students increase in a frame. Also, false acceptance rate is prone to increase if database is not prepared carefully i.e. using head-pose variation, illumination variation etc. The efficiency of CNN encodings which are used to identify each face, can be improved by fine tuning and training the Convolutional Neural Network algorithm according to generated database and target conditions of classroom. The proposed system is currently capable enough to being used in actual classrooms or laboratories or other places in organization which needs daily attendance report to be maintained. This can be further improved in terms of accuracy of CNN, User interface, connectivity and implementation of report in large systems which maintain organizational report.

REFERENCES

[1] Manisha M. Kasar1, Debnath Bhattacharyya1 and Tai-hoon Kim2 – "Face Recognition Using Neural Network: A Review" International Journal of Security and Its Applications Vol. 10, No. 3 (2016)

[2] Navneet Dalal and Bill Triggs - "Histograms of oriented gradients for human detection" Computer Vision and Pattern Recognition" 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE. 2005, pp. 886–893

[3] Radhika C. Damale, Bazeshree. V Pathak – "Face Recognition Based Attendance System Using Machine Learning Algorithms" 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)

[4] Mohsen Ghorbani, Alireza Tavakoli Targhi, and Mohammad Mahdi Dehshibi - "HOG and LBP: Towards a robust face recognition system" Digital Information Management (ICDIM), 2015 Tenth International Conference on. IEEE. 2015, pp. 138–141

[5] Neha Savakhande1, Vinay Sripathi2, Kiran Pote3, Parth Shinde4, Prof. Jayant Mahajan5 – "Face Recognition based Attendance System: Review" International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 02 | Feb 2020

[6] Vishakha Mehta, Sarika Khandelwal, Ashish Kumar Kumawat – "A Survey on Face Recognition Algorithm" 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 11-12 May 2018

[7] Vinay et al - "Effective Descriptors based Face Recognition Technique for Robotic Surveillance Systems" Procedia Computer Science 133 (2018), pp. 968–975

[8] Gang Zhang, Sikan Tang, and Jiaquan Li - "Face landmark point tracking using LK pyramid optical flow" Tenth International Conference on Machine Vision (ICMV 2017). Vol. 10696. International Society for Optics and Photonics. 2018, 106962B

[9] Yuqian Zhou,Ding Liu,Thomas Huang – "Survey of Face Detection on Low-Quality Images" 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),15-19 May 2018

[10] Anushka Waingankar1, Akash Upadhyay2, Ruchi Shah3, Nevil Pooniwala4, Prashant Kasambe5 – "Face Recognition based Attendance Management System using Machine Learning" International Research Journal of Engineering and Technology (IRJET)

[11] Thanh Tan Nguyen Thi, Khanh Nguyen Trong – "An Efficient Face Detection and Recognition" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-7 Issue-5, February, 2018

[12] Li Cui-mei, Qi Zhiliang – "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers"

[13] Mauzzam Siddiqui, Dr.Preetam Suman – "A Review on Face Image Recognition & Clustering by Using Various Types of Approaches in Deep Learning" International Journal of Computational Engineering Research (IJCER)

[14] R. Angeline1, Kavithvajen.K2, Toshita Balaji3, Malavika Saji4, Sushmitha.S.R5– "CNN Integrated with HOG for Efficient Face Recognition" International Journal of Recent Technology and Engineering (IJRTE) Volume: 07 Issue: 06 | March 2019

BIOGRAPHIES

VINAY SRIPATHI
Currently pursuing under graduation final year in the branch of EXTC at Rajiv Gandhi Institute of Technology, Versova.
Team Leader of the Project Group.

NEHA SAVAKHANDE
Currently pursuing under graduation final year in the branch of EXTC at Rajiv Gandhi Institute of Technology, Versova.
Team Member of the Project Group.

PARTH SHINDE
Currently pursuing under graduation final year in the branch of EXTC at Rajiv Gandhi Institute of Technology, Versova.
Team Member of the Project Group.

KIRAN POTE
Currently pursuing under graduation final year in the branch of EXTC at Rajiv Gandhi Institute of Technology, Versova.
Team Member of the Project Group.

Dr. JAYANT MAHAJAN (Ph.D.)
Project Guide at Rajiv Gandhi Institute of Technology, Versova.