Revolution in Teaching/Moocs by the use of Real-Time Face Detection

Bharadwaj Vishanth Thiyagarajan
Computer Engineering Department, P.I.C.T (University Of Pune), Pune, India
Vishanthbharadwaj@gmail.Com

***Abstract* -- The use of computer has become essential and easy for students. Using the computer to perform teaching for the students can be a great advantage. Here the focus is made on the working of the Haar-feature computer vision algorithm and the connectivism {Massive Open Online Courses (MOOCs)}, which are then collaborated to aggrandize the student learning capabilities. The use of these methods can overcome the absence of a human tutor where computer can communicate to teach the students. The principles/Algorithms of FACE-DETECTION, FACE-TRACKING, CONNECTIVISM and FOCUS-CAPTURE are mentioned in this paper. Hence a new creative approach for MOOCs can be used to overcome the age-old study methods that were monotonous. This novel method can also increase the speed and accuracy of the learning.**

Key Words—Object Detection; face detection; MOOCs; Viola-Jones; Real-time; KLT.

I. INTRODUCTION

In this paper an attempt is made to combine the research of pattern detection and Connectivism (MOOCs) to enhance the student learning capability.

The field of computer vision has an expansive scope with the availability of augmented reality, virtual reality, Optical Character Recognition, Optical Brail Recognition etc. The multicore/hybrid computer systems can take an advantage of fast processing and great accuracy, which humans cannot perceive. In case of Object/Pattern detection the humans have a very high accuracy but continuous working creates a monotonous environment, whereas computers can perform the task continuously without interruption but coding the pattern recognition becomes tedious. The computer can be used to examine the real time experiments or the activities of the students. According to the result, the next move can be executed.

The face detection algorithm used here was proposed by Viola-Jones in 2001. Thus the computer can scan and recognize the faces in the real-time and collaborate the output with the Connectivism method of MOOCs. Connectivism can be defined as a method where computer can contain the source of knowledge and deliver it to the humans for flexible. The use of pdfs and videos are increasing in MOOCs. However if the student is not present in front of the E-learning device, there is a possibility that they may miss something important. Hence to solve this kind of problem the presence of the student and their concentration is to be checked and the teaching would be paused if the student were not present. Hence the potential can be augmented by these kinds of intelligent methods, which can increase the output capacity and overcome the absence of human – human interaction in certain situation where there is absence of human tutor.

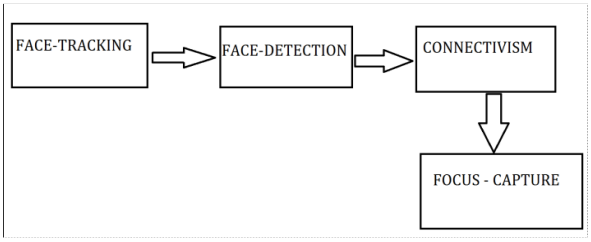

Fig. 1. Flow Chart of methodology

II. RELATED WORK

The age-old method of personal study becomes very monotonous. It is found that the understanding of the subject becomes difficult using the regular method. If the student does not have any interest in the subject he/she may tend to ignore some important points, which may cause a downfall in results. Hence an interactive method can overcome this problem. Thus the software is designed using the proposed algorithms in this paper to create an interactive learning environment in order to clear the boredom and increase the efficiency of learning.

III. METHODOLOGY OF THE PROPOSED WORK

The paper presented consists of 4 phases Methodology namely Face detection, Face tracking, Connectivism, Focus capture. Face detection needs to be performed to check the presence of the subject. While performing face detection it is important for scanning of the surrounding to grab correct frames containing a face for which the Face-Tracking algorithm is used. Hence the accuracy of the webcam in the real-time increases the face detection.

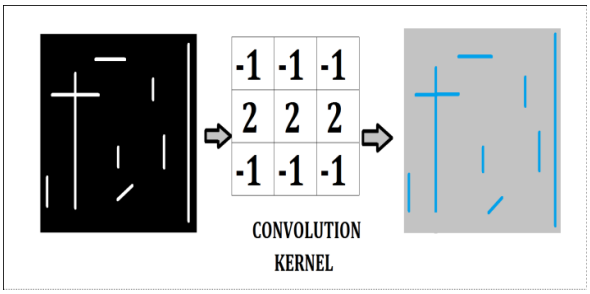

Fig. 2. Example of convolution kernel

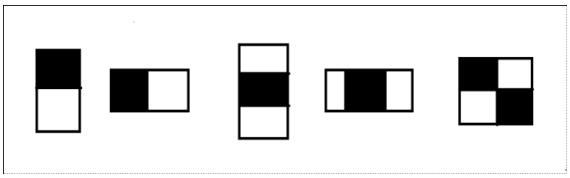

Fig. 3. Five basic types of Haar Features

Connectivism is the algorithm, which can be used to impart knowledge using this software. Focus-Capture block is the implementation of the above-mentioned blocks. The main parts and their sub-parts are listed below.

1. FACE-DETECTION algorithm.
a. HAAR Features.
b. Integral Image.
c. ADA – Boost for the images.
d. Cascading the features of classifiers.
2. FACE-TRACKING algorithm.
a. Mean -Shift.
b. Histogram back-projection.
c. Continues adaptive mean (CAM) – shift.
3. CONNECTIVISM.
4. FOCUS-CAPTURE using innovative images and text to speech engine.

IV. FACE DETECTION ALGORITHM

Haar Features:
Here the following image shows a basic introduction to edge detection. Now the above output image contains some high (blue) intensity pixels, which perfectly matches the convolution kernel pixels, which are derived from the white pixel in the black background of the input image. Viola-Jones algorithm also depends on the same principle explained above. Here any face detection performance includes scanning of the face and filling it in the form of black and white pixels. Then the calculation is made for every feature where the sums of the black pixels are calculated and the white pixels are calculated. Then the subtraction is done where the while pixels' sum is subtracted from the black pixels' sum. There are 5 types of Haar features.

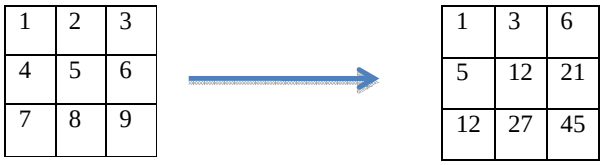

Fig. 4. Theoretical Example

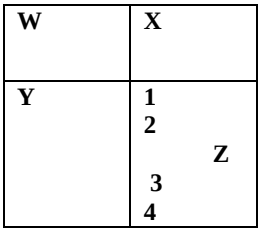

Fig. 5. Practical example

The algorithm uses a 24x24 window for training purpose and includes the features like position, scale and type so the total features go beyond 160,000+ features. Now for the calculation of the sum of the pixels a new method has been introduced which is called Integral Image. For efficient classifiers we need at least 5000 positive images and negative images of 9000.

Integral Image:
Integral image is a solution found out to encounter the problem of calculating the area of the rectangle, which was mentioned in Haar Features above. According to this algorithm by knowing, the co-ordinates of the four corners of the rectangle, the area of the rectangle can be calculated. We need to perform addition in $\Sigma(a) + \Sigma(L)$. Where 'a' is for co-ordinates above and 'L' is for co – ordinates left.
Here the L.H.S is the normal image co-ordinates and the R.H.S is the co-ordinates generated using Integral Image.Here the numbers 1,2,3,4 represent the co-ordinates of pixels of the block Z. Now to calculate the area of block Z, we write Area(Z) = 1+4 – (2+3). In other terms as per the example that is given in Theoretical Region, we can write it as

Area(Z) = W +(W+X+Y+Z) –((W+X)+(W+Y))

So using this technique, it becomes easy to calculate the values of the area of the desired region and aggrandize the performance of the HAAR features.

Ada-Boost:
ADA-BOOST is a machine-learning algorithm, which is used to track the best features in the above-mentioned 160,000+ features. Post findings of these features, a weighted compilation of all these features are taken into account where it can be found out whether the given image is a face or not.

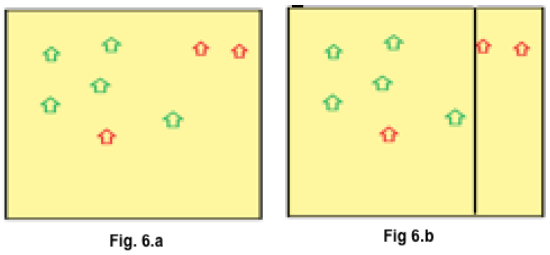

Fig. 6. Weights of the images

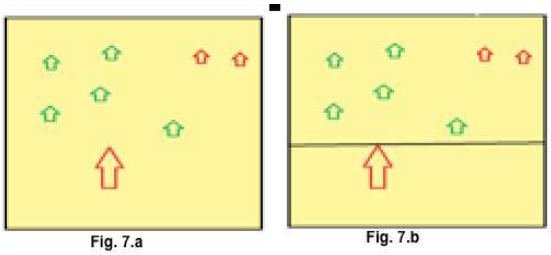

Fig. 7. Weights of the images

A simple formula is demonstrated below:

$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \alpha_4 f_4(x)\ldots. \quad (1)$$

α = assigned weight.
$f_N(x)$= weak classifier (features tracked).
F(x) = strong classifier generated from the tracked features.
Every weak classifier can have two outputs namely '1' or '0'.

The output '1' is generated when a successful detection is done and the output '0' is generated when detection fails. There is always a chance that the weak classifiers can detect more than 50% faces in the given set of images. The remaining classifiers carry this chain where each classifier can detect more than 50% of the remaining faces. All these detected features are combined to get a strong classifier, which can detect the faces. This concept is applicable not only to faces but also to many other objects.
Consider the following figure:
The green arrows show the positive images or faces. The red arrows show the negative images. Initially we provide same weight for them in Fig. (6.a). Now to separate the positive and negative images a black line is drawn in Fig. (6.b). This is the output of the first classifier. But there is a false detection in the Fig (6.b). Hence we increase the weight of the falsely detected image and the next out put is shown below:

Now the weights of the red arrows have increased (refer Fig. 7.a) .The second classifier (refer Fig. 7.b) detects the positive images but there are 2 false detections so the weights of these images must be increased.

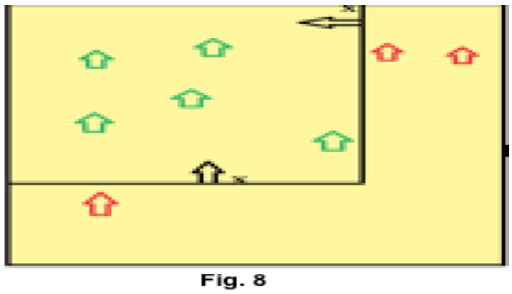

Fig. 8. Weights of the images

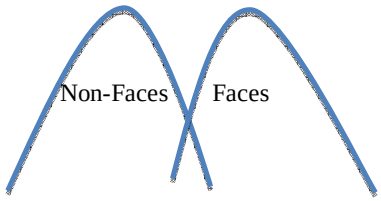

Fig. 9. Graph of Gaussian weak classifier

Similarly the results are calculated by increasing the weights of the false detections and then the final classifier is done which is the strong classifier. The third classifier results are shown below:
'X' is the co-ordinate system in the Fig.8. This creates an efficient classifier.ADABOOST is used to find a single rectangular feature with threshold that separates the positive and negative training samples, in terms of weighted error.

Ada boost algorithm:
For n rounds we consider:
1. Weighted error should be evaluated for each feature and the best one is chosen.
2. Re-weight the example taken:
   Incorrect examples: - higher weight.
   Correct examples: - lower weight.

Finally a good classifier is generated from the weak classifiers classified according to the error they had.

Cascading classifiers the features of classifiers:
There exists a basic problem with the above approach. Every time the algorithm has to scan the same image with new window size and this becomes prohibitive because of the time and energy cost that will incur. Even if an image, consider that it will have n faces. But there may be negative images >= n. Hence scanning these negative images puts an extra burden on the processor costing us energy and time. Hence the algorithm must be so effective that it quickly rejects the negative images (non-faces in this case) and spend more time on analyzing the positive images (faces in this case). Hence the use of cascade of classifiers came to existence because forming a classifier out of linear combination of all these classifiers becomes quite tedious. Hence

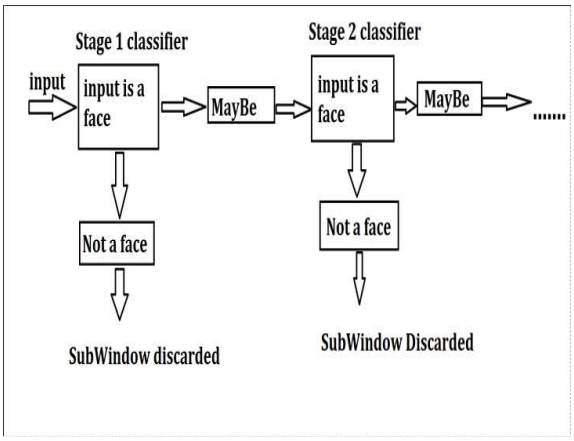

Fig. 10. Schematic representation of face detection

acascade of classifier is made which contains each strong classifier. Thus all features are grouped into stages. Every stage present determines whether the given classifier is a face or not. If it is a face it sends it to the next classifier and if it is not a face it rejects it. Since the input is in the form of sub-window the current sub window is discarded if it is not a face.

There are 3 main steps in designing a classifier:
1.) Total number of stages
2.) Total number of features
3.) Threshold of each strong classifier.

V. FACE TRACKING ALGORITHM

a) Mean Shift Algorithm:
Mean shift algorithm is a real time algorithm to process the face in a convenient image for tracking of the object (face in this condition). The steps for algorithm are given as follows:

1. Convert image into HSV.
2. Variable H of the image is selected.
3. Size of the search window is chosen.
4. Initial location of search window must be chosen.
5. Center of mass of the object must be found.
6. Center search the window at the Center of mass.
7. If convergence occurs got to step. 8 or else go to step no. 5
8. Compute the parameters.

b) Histogram Back Projection:
Initially a window is chosen to track the face. All the hue values are extracted from the window and a histogram is formed with the help of those values. This histogram is used for reference and whenever an image is captured in the frame then their pixel values are converted into histogram pixel values and then used for further processing, this is called the histogram back projection.

Fig. 11. Example of Histogram Back-projection

CAM shift:
This algorithm has the facility to adapt to the screen size of the object that is being tracked. Every iteration when occurs the window size is adjusted based on the first moment calculated within the previous window.
The size can be S and M is the moment and the formula is given by:

$$S = 2*\sqrt{M_{00}} / 256 \qquad (2)$$

Window width=S
Window length=1.2S
M_{00} = Moment of order zero

VI. CONNECTIVISM AS A TOOL TO LEARNING

When it comes to education, the human-human interaction is at its best. Here we make use of connectivism as a method for communication between human and computer to make it more effective. Connectivism describes that knowledge or learning materials can also exist in non-human environment.

Before seeing connectivism we have to see 3 main definitions:

1. Behaviorism
2. Cognitivism.
3. Constructivism.

These 3 methodologies have a major disadvantage which is the intrapersonal view of training, failure of utilizing the available resources within the knowledge and contribution to the valid judgments to the real life environment. Hence Downes and Siemens introduced a methodology, which uses the network as a medium for spreading useful knowledge for education known as connectivism.

principles of connectivism:

1. Wide range of opinions for learning and knowledge.
2. Learning as a tool connects information services and special nodes.
3. Non-human appliances may contain knowledge for learning purpose.
4. Capacity of acquiring knowledge is more critical than the currently present knowledge.

5. Cultivating and maintenance of connectivity is important aspect where the knowledge can be efficiently shared.
6. Connections must be established between ideas, concepts and fields and they must be maintained.
7. The intent of all the activities of connectivism is currency (accurate, up-to-date knowledge).
8. Decision-making is an important learning process. There are two sub-types of decision-making.
9. Waiting to take the right decision.
10. Taking the decision and making it right.

Both these techniques are efficient ones depending upon our implementation skills. The first technique requires knowledge whereas the second technique requires experience.

In brief the connectivism is not a theory but an ideology or a method that harnesses the age-old human-human interaction system and paves a way for the machines to perform this task. In this way artificially intelligent software can be created using pattern detection as a tool or machine learning algorithms where the computer adapts itself to the environment. Thus the human-computer interaction methodology is explained in the following section with necessary proof.

VII. FOCUS-CAPTURE

It is pertinent to mention that the human brain resonates to the vision, sound and smell to a great extent. Human intelligence is designed on these three basic factors. The Human brain is nothing but a high capability organ, which can perform heavy IF-ELSE cases in the real time. The involuntary organs like spinal chord also perform thoughtless actions in the jeopardized situations can be considered as a high performing IF-ELSE machine. Now the motto of this project is to capture the human intelligence and to focus them in a real time learning environment. Human brain can be attracted with the help of sound and providing vision easily. 66.66% of the intelligence providing organs can be made to focus on the computer by providing relevant and novel images along with some good language at which the subject is comfortable with.

Even though these features are provided, the computer must be able to interact with the student in the form of a human tutor. This is possible only when the computer is able to see what the student is doing. The computer must pause when the student is not present. It must also tell the student to concentrate when he/she is not looking towards the computer and if his/her mind weavers. With these methods the computer can monitor the student's ction and give the output accordingly.

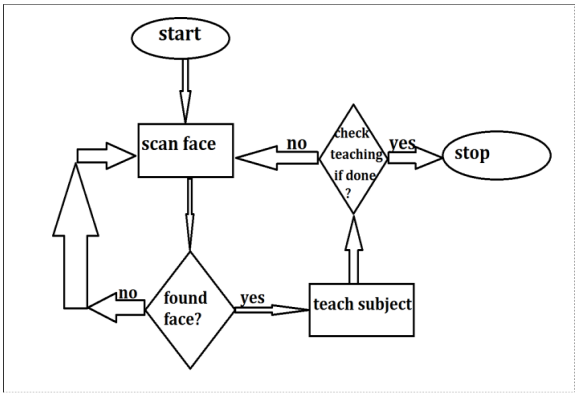

Fig. 12. Flow-chart representing the working model of teaching

TABLE 1. EXPERIMENTAL RESULTS ON PROPOSED METHOD

Subject	Average Regular Learn Time	Learning Time On Using The Proposed Technique	Result
Algebra formulae (student age 7)	210 minutes	150 minutes	Success
Learning Trigonometry (student age 14-> class 9)	240 minutes	195 minutes	Success
Learning Human Body Functions (student age 11 -> class 6)	150 minutes	120 minutes	Success
Learning History of World War I(student age 12 -> class 7)	180 minutes	135 minutes	Success
Learning Construction of Figures in Geometry (student age 14:- class 9)	180 minutes	120 minutes	Success
Learning Nature of soil in geographical regions of Geography (student age 14:- class 9)	210 minutes	165 minutes	Success

Using the steps given in the flow chart and with rich set of relevant images we can capture the focus of the student. By providing the vocal teaching, it is ensured that student will have a better understanding of the concepts. The technologies used for MOOCs are redefining the terms of education by providing faster means of learning methodologies. This project since can be used in the absence of human tutor and

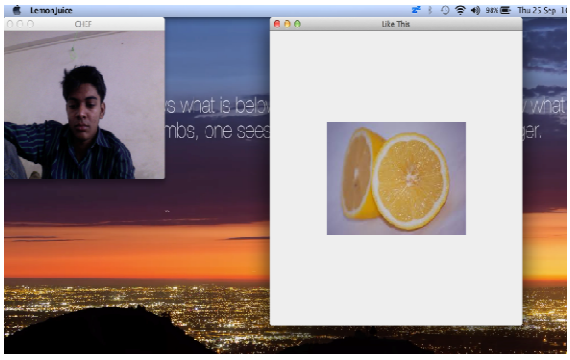

Fig.13. Student concentrating towards the computer.

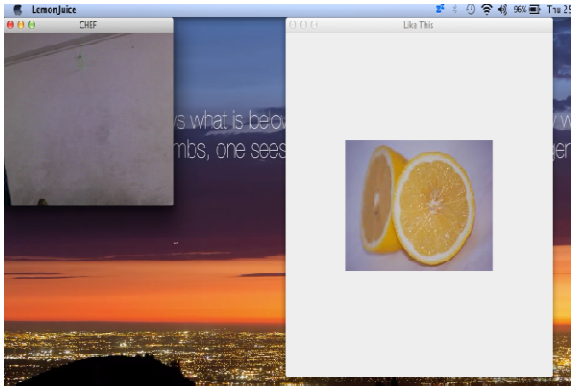

Fig.14. Student not present in front of the computer.

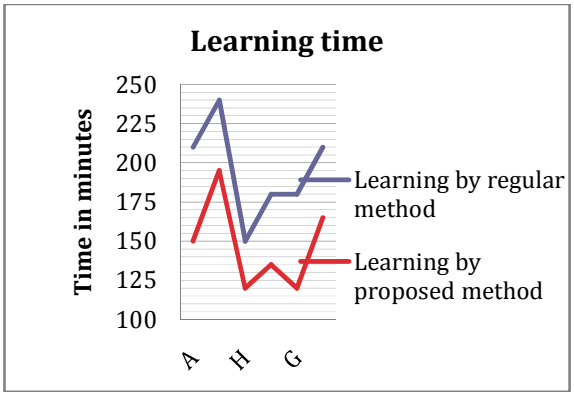

Fig. 15.Graph representing the output as per learning time (time in minutes)

revolutionized and aggrandized the grasping power of the human brain and thereby achieving a greater results.

VIII. RESULTS

Experiments conducted on the students' (age :- 7-14) self study using book vs students' self study using this software and their learning time is shown below:

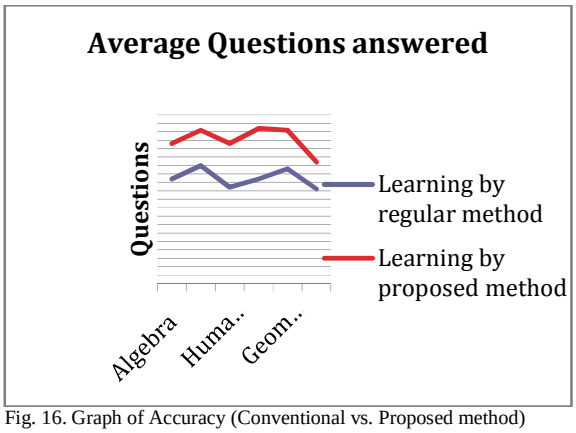

Fig. 16. Graph of Accuracy (Conventional vs. Proposed method)

TEST CONDUCTED ON 200 STUDENTS.
ACCURACY IN FACE DETECTION :- 100%.

IX. DISCUSSIONS

The above result consists of 4 figures and a table representing a comparison between the regular and the newly proposed theory. Here in the figure 1 the computer detects a face with an angle facing towards the webcam. Hence it realises that the subject is concentrating and hence it provides the knowledge in the form of pictures and voice. In the figure 2 the subject is not looking towards the computer due to some disturbance. Hence the computer is not able to recognize a face, which is trained in the Haar file. So it does not proceed further and gives an alarm of "YOU HAVE AN INCOMPLETE TASK SIR" continuously for every 5 seconds. Hence it is accepted that if the subject is concentrating, the teaching will be continued. A proper time interval is kept between every note shown so that the student can implement it in a proper way.

The table described above is the test conducted on 200 students. In every group 40 students were taken and were split into a random group of 20 each namely Group A and Group B. The students of Group A of every set were made to study from the Book and the average time for them was calculated. The Group B students were educated with the help of the software and the readings were observed and finally a comparison was made in the learning time and it was found that the students who learnt from this software had a greater accuracy and greater learning speed. All the students who were taken for this experiment were having the academics percentage in 75 – 80. Hence the intelligence level of the students on which the tests were conducted was almost same level. Hence the software had the similar impact as that of a personal tutor teaching the students independently. The Bar Graph 1 helps to describe the learning time.

In the next part we have asked 10 questions to each of the students. On basis of the answer, we calculated the average accuracy of the students and found out the results, which are seen in Graph 2. The blue bar indicates the accuracy using the regular method. The red bar indicates the results after using the new method.

X. CONCLUSION

The principle of Connectivism with human-human interaction is applied on human-computer interface. Computer vision algorithm is used where a face is detected in the real time using the Face Detection algorithm. It consists of Haar features, Integral image, Ada-Boost and Cascading the features. The Face Tracking algorithm is used to filter out the frames using Mean-Shift, Histogram Back-Propagation and CAM-shift used respectively. Now the output of these algorithms is a detected face, which is then combined with the principles of Connectivism to aggrandize the accuracy and speed of learning. Connectivism consists of principle where the computer can use the human-human interaction policies to perform teaching. The output of all these methods is then used in focus capture, which utilizes the vocal teaching and image popping techniques to perform efficient teaching. According to the results, we can say that the students can learn with a speed of 1.2 – 1.5 times faster with a higher accuracy than their usual speed. It is also found that the students can study in a creative way rather than the traditional method, which is very monotonous. Adding the character recognition or emotion detection to this object/face detection can provide teaching in a more sophisticated format. Hence the learning capacity of the students increases.

ACKNOWLEDGMENT

I would sincerely thank my college mates, Teachers and my mentors T. Ananthi (Mathematician) who helped me to think about this idea and make it a huge success, which is now a startup.

REFERENCES

[1]. B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” Proceedings of the DARPA imaging understanding workshop, pp. 121–130, 1981.

[2]. J. Quinlan, “Induction of decision trees,” Machine Learning, Journal, vol. 1, no. 1, pp. 81-106, 1986.

[3]. P. Menezes, J. C. Barreto, and J. Dias, “face tracking based on haar like features and eigen faces,” 5th IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, Portugal, pp. 1-6, July 5-7, 2004.

[4]. Frances Bell, “Connectivism: Its Place in Theory-Informed Research and Innovation in Technology-Enabled Learning”, IRRODL journal, March-2014.

[5]. Siemens, G(2005), “Connectivism: A learning theory for digital age”, Retrieved from www.elearnspace.org.

[6]. Smith M.K. (1999), “Learning Theory: The encyclopedia of informal education.” retrieved from www.infred.org.

[7]. E. Hjelmas and B.K. Low, “Face detection: a survey,” Computer vision and image understanding, vol. 83, pp. 236-274, 2001.

[8]. P. Viola and M. Jones, “Robust Real-time Object Detection,” 2nd international workshop on statistical and computational theories of vision - modeling, learning, computing, and sampling, pp. 1-25, 2001.

[9]. J. Shi and C. Tomasi, “Good Features to Track,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600, 1994.

[10]. M. H Yang, “Detecting faces images, A survey,” IEEE Transations on Pattern Analysis and Machine Inteligence vol. 24, no. 1, pp. 34–58, 2002.

[11]. Bharadwaj Vishanth Thiyagarajan, “Facial Detection and facial feature points’ detection with the help of KLT algorithm”, IJARCSMS, vol. 2,issue 8, august 2014.