DOI: http://dx.doi.org/10.26483/ijarcs.v10i4.6445

ISSN No. 0976-5697

Volume 10, No. 4, July – August 2019

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

REAL-TIME STUDENT SURVEILLANCE SYSTEM USING MACHINE LEARNING AND COMPUTER VISION

Rajat Mehta
Department of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India

Rashi Bhardwaj
Department of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India

Prakash Ramani
Department of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India

Abstract: In a classroom full of students, it is practically not possible for a single teacher to give personal attention to every student. Continuous monitoring and identification of students who show signs of lethargy, sadness or anger in classrooms can help management counsel them and advice some preventive measures – which if kept unchecked can lead the students to take adverse steps. Potential students who need help in some form can be identified. There are some unproductive activities that take place in a classroom which can be easily automated and the time saved can be devoted to productive activities. Manual attendance calling is one of them which can be automated using facial recognition algorithms. In this paper, a robust surveillance system using Machine Learning and Computer Vision algorithms is presented that can take on the above challenges. For facial recognition, Local Binary Pattern Histograms have been used and for emotion recognition, Deep Learning model has been used.

Keywords: Computer Vision; Deep Learning; Facial Recognition; Emotion Recognition

I. INTRODUCTION

Every 40 seconds one person commits suicide globally. Suicide is the second leading cause of death after road injury for the ages 15 - 29 years [1]. Lack of personal attention to the students makes these signs of suicidal tendencies go unnoticed. This is a matter of concern and needs quick involvement of the appropriate authorities for students. Giving personal attention to every student will definitely decrease these dangerous numbers. Personal attention means to analyze every activity of the student and to counsel them whenever needed. For a teacher to identify a student facing any issue is not an easy task. It becomes more difficult as the number of students increases.

The solution to this challenge is using surveillance cameras in the classrooms which can monitor the student activities and identify the ones in dire need of guidance. Such a surveillance system will also decrease the time devoted to unproductive activities such as manual attendance registering by roll call. The model created also involves a facial recognition algorithm to register attendance of the students present in the class. The time saved here can be devoted to productive activities such as discussions, problem-solving and question-answer sessions. The model can also store the activity data in a database for data-driven insights and decision making.

For detecting whether a face is present in camera frame, Haar Cascade for Frontal Face have been used [2]. For facial recognition, Local Binary Pattern Histogram (LBPH) is used [3]. For sleep and yawn detection, dlib – a modern toolkit that contains tools and algorithms to solve real world problems – is used [4]. Using dlib's 68 face landmarks, important points on a face are detected which help in performing various operations.

FER2013 dataset is used to train the emotion recognition module [5]. It contains 48 x 48 pixel images in grayscale. There are 28,709 images in the training set and 3,589 images in the test set. It contains the emotion value in the first column and the pixel values in the second column. It contains 7 basic emotion images i.e. happy, surprise, neutral, sad, angry, disgust and fear. A Convolutional Neural Network is used to train the model. This paper illustrates how a robust student surveillance system has been researched and developed.

II. LITERATURE REVIEW

For facial detection, multiple techniques are available [6]. One of the most popular and accurate classifiers is the Haar feature based cascade which is used in this model [7].

Facial recognition has been a crucial topic of research in Machine Learning and Computer Vision studies. Research has been conducted on it since the 1960s with initial work conducted by extracting features manually by hand [8]. Albeit the low computational power at that time limited the accuracy, but during 1991, a major breakthrough was achieved by Eigenfaces approach [9].

Local Binary Pattern Histogram (LBPH) is another technique for facial recognition [3]. LBPH is widely used for image analysis related to faces. It takes a central pixel value, compares it with the nearby remaining 8 pixels in a 3 x 3 window size. Any value greater than the central value is replaced with 1 and any value smaller is replaced by 0. The resulting binary number is used for labeling. This process is

known as Local Binary Patterns [10]. This process is repeated for every pixel and their local structure is encoded. To train the model, LBPH has been used.

Deep Learning models which use Convolutional Neural Networks also show promising results on benchmark datasets [11].

For sleep detection, previous works include using an infrared lens and a thermal sensor which monitor and evaluate temperatures around the nose and mouth – which change when a person is about to sleep [12]. Albeit such systems use multiple hardware systems to find whether a person is sleeping or not, it can also be performed using cameras [13]. Use of Haar cascades for eye detection failed in the cases when the eyes were kept closed. Yawn detection is another factor which helps detect drowsiness. Edge detection has been used to determine if the distance between the lips is sufficient to call it a yawn [14].

Facial emotion recognition or facial expression analysis is also an interesting area which uses various Computer Vision and Machine Learning algorithms to identify human emotions from their facial landmarks. FER2013 is a benchmark dataset released as a Kaggle Competition [15]. The highest categorization accuracy achieved in the competition for public test set was 69.4% and private test set was 71.2%. Deep CNN architectures like BKVGG12 have achieved slightly better accuracy on the same dataset [16].

III. METHODOLOGY

The model is divided into 3 different modules:

A. *Facial Recognition Module*

For facial recognition, this module creates a dataset, trains the model and saves the weights for future use. Figure 1 shows the working of facial recognition module.

1) Dataset Creation: Custom dataset has been used for this module. Per person, 50 images are taken via a web cam. In this model frontal face is captured using the Frontal Face Haar Cascades [2]. This removes the redundant regions like background and torso. While image capturing, multiple image augmentation techniques like rotation, reflection and converting to grayscale are performed. This not only increases the training images but also induces variations in the images to match real-world scenario. The registration details of the student are also stored in the database with registration number as the primary key. The images captured per person is stored with a name corresponding to registration number.

2) Model Training: LBPH is used to train the model. The images are fed in grayscale which reduces computations.

3) Using the Weights: Once the training is complete, model weights are saved in a YML file. Using the YML file, faces can be recognized. When a familiar face is encountered, the event is logged into the database with the timestamp of the occurrence. Here, a face is recognized based on confidence percentage.

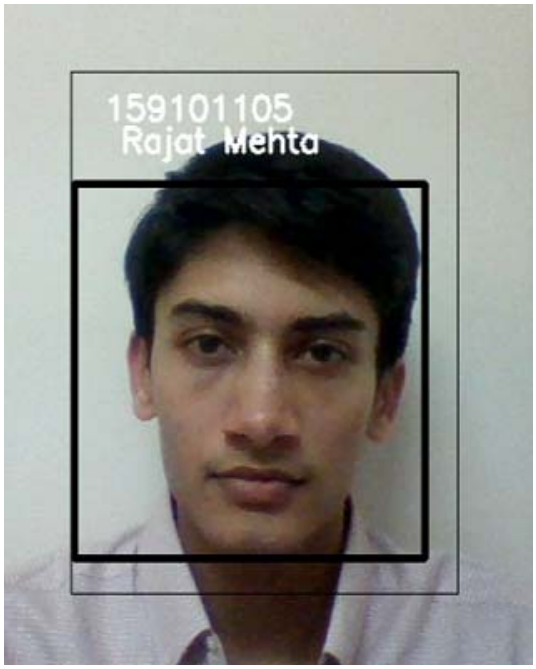

Figure 1. Working of Facial Recognition Module

B. *Sleep and Yawn Detection Module*

Once the weights have been saved, faces can be recognized. Now, various image processing techniques can be applied to get insights. In this module, sleep and yawn detection processes have been described. Figures 2 and 3 show the working of sleep and yawn detection algorithms respectively.

1) Recognizing Faces: Sleep and yawn detection operations are performed only on the faces that are recognized by the model. So, the weights saved from the previous module are first used to identify the faces and then further operations are performed.

2) Locating Region of Interest: Using shape predictor 68 face landmarks, contours are drawn on both the eyes and the lips. These contours help in computing the changes in distances between eyes and lips and visualizing the process.

3) Computing Changes: Eye Aspect Ratio (EAR) is calculated to determine whether the eyes are closed or not.

$$EAR = \frac{| e_2 - e_6 | + | e_3 - e_5 |}{2 | e_1 - e_4 |}$$

The values e1, . . . , e6 are eye landmark points. When EAR value decreases from a particular threshold (0.20 in this case), a counter is initiated. If this counter is running for more than a particular time duration (48 frames in this case), the information about the event is logged into the database. It stores the total time duration for which the eyes have been kept closed.

For yawn detection, the mean of all the top lip landmark points is subtracted from the mean of all the bottom lip landmark points. If the absolute value of the difference exceeds a particular threshold (25 in this case), the event is logged into the database. The threshold values in both the cases can be changed as per requirement. Contours on the eyes and lips with EAR is displayed so that it does not look like a black-box algorithm.

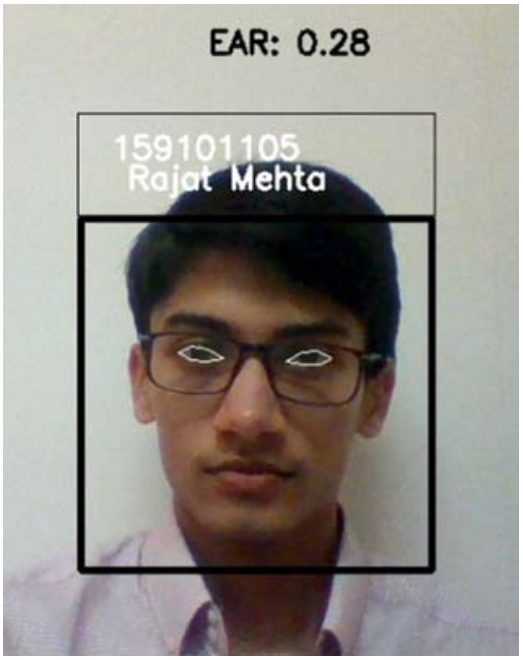

Figure 2. Working of Sleep Detection Module

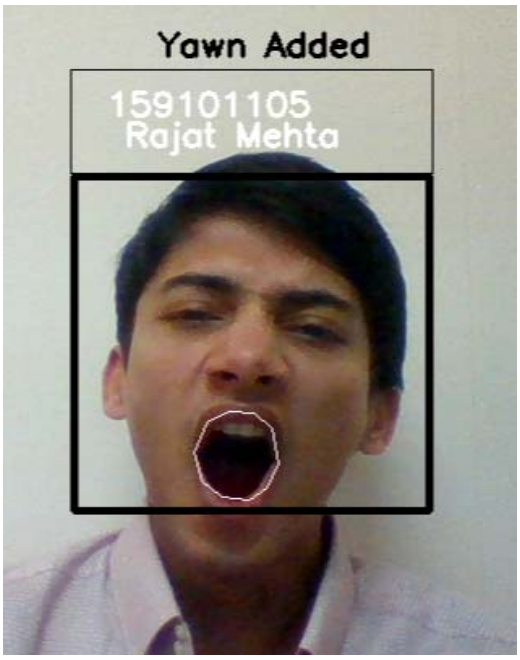

Figure 3. Working of Yawn Detection Module

C. *Emotion Recognition Module*

This module uses Convolutional Neural Networks (CNN) to classify facial expressions into 7 basic emotions i.e. happy, surprise, neutral, angry, sad, disgust and fear. Figures 4, 5 and 6 show 3 out of 7 emotions that can be detected by the model.

1) Model Architecture: The CNN architecture consists of 17 layers with Rectified Linear Units and Softmax as the activation function, Adam as the optimizer.

2) Model Training: The model is trained for 50 epochs with 256 as the batch size. The weights are then saved in a JavaScript Object Notation (JSON) format file. This JSON file is later used for emotion recognition of faces on a live video feed.

3) Using the Weights: Once the weights are saved, it can be used on a live feed. Initially, Haar cascades are used to detect a face. If a face is detected, saved weights are used to find what type of emotion is being expressed. Frame count for every emotion is separately added to the database.

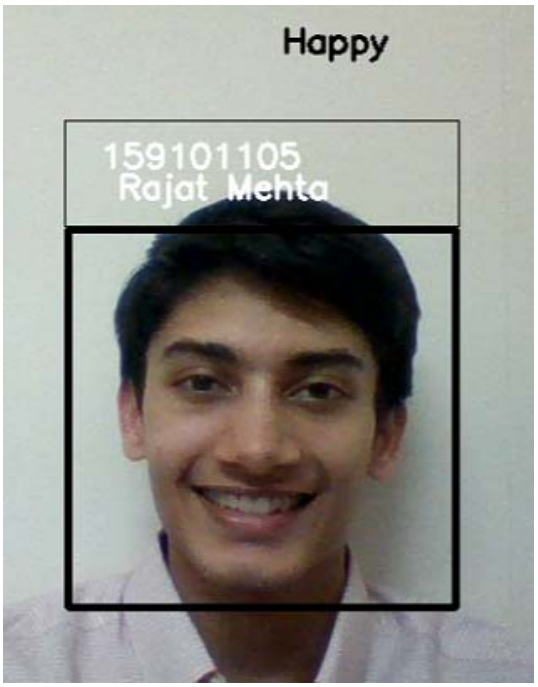

Figure 4. Detection of Emotion - Happy

Figure 5. Detection of Emotion - Surprise

Figure 6. Detection of Emotion - Sad

IV. RESULTS AND DISCUSSION

For facial recognition, confidence is used as a metric to match a live face with those saved in the database. Only when the confidence level of a live face is more than 85%, the information will be logged into the database. The various details of the subject can be seen when the confidence level is greater than 85%.

The sleep detection algorithm is robust to subjects wearing spectacles with transparent glasses. If only one eye is closed even then the EAR will not go beyond the threshold.

CNN for emotion recognition has achieved the highest accuracy of 61.3%. This accuracy is close to the highest accuracy achieved by the winner of the Kaggle competition i.e. 71.2%. Once a face is detected accurately, sleep, yawn and emotion recognition work without any faults.

The model was also tested on live video stream. Data of 20 students was stored and the model was trained. For the testing, a 720p camera was used. In a single frame there were at-most 4 students. Table I shows the results received during the live testing.

Table I. Live Testing Results for Facial Recognition

Day Number	***Total Students***	***Correctly Identified***	***Incorrectly Identified***
Day 1	18	16	2
Day 2	20	17	3
Day 3	20	18	2

V. CONCLUSION AND FUTURE WORK

Using these Machine Learning and Computer Vision algorithms, real-time surveillance of students can be conducted and the data-driven decision-making process by the management will help in decreasing depression and suicide rates of the students.

Deep Learning algorithms can be applied to the facial recognition module to achieve even higher accuracies. Images for other expressions like crying, annoyance, boredom etc. can be added for a better understanding of facial expressions.

VI. REFERENCES

[1] Mental Health, Suicide Data, World Health Organization. https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/, 2016 (Accessed 3 January 2019).

[2] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 2001, pp. I511-I518.

[3] T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7), pp. 971-987.

[4] D.E. King, “Dlib-ml: a machine learning toolkit”, Journal of Machine Learning Research, 2009, vol. 10, pp. 1755-1758.

[5] I.J. Goodfellow, D. Erhan, P.L. Carrier, A. Courville, M. Mirza, B. Hamner, et al. “Challenges in representation learning: a report on three machine learning contests”, In Lee, M., Hirose, A., Hou, Z.G., Kil, R.M. (Eds.) Neural Information Processing: ICONIP 2013, Part III, Lecture Notes in Computer Science, Berlin, Heidelberg, 2013, 8228, pp. 117–124.

[6] G. Hemalatha and C.P. Sumathi, “A study of techniques for facial detection and expression classification”, International Journal of Computer Science and Engineering Survey, 2014, 5(2), pp. 27-37, https://doi.org/10.5121/ijcses.2014.5203.

[7] P. Viola and M.J. Jones, “Robust real-time face detection”, International Journal of Computer Vision, 2004, 57(2), pp. 137-154.

[8] W.W. Bledsoe, “The model method in facial recognition”, Technical Report PRI 15, Panoramic Research, Inc., Palo Alto, California, 1964.

[9] M. Turk and A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, 1991, 3(1), pp. 71-86.

[10] D. Huang, C. Shan, M. Ardebilian, Y. Wang and L. Chen, “Local binary patterns and its application to facial image analysis: a survey”, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2011, 41(6), pp. 765-781.

[11] Y. Sun, D. Liang, X. Wang and X. Tang, “DeepID3: face recognition with very deep neural networks”, arXiv:1502.00873 [cs.CV], 2015, https://arxiv.org/abs/1502.00873

[12] J.R. Clarke Sr. and P.M. Clarke, “Sleep detection and driver alert apparatus”, United States Patent, Patent No. 5,689,241, https://patents.google.com/patent/US5689241A/en, 1997 (Accessed 25 December 2018).

[13] A.B. Albu, B. Widsten, T. Wang, J. Lan and J. Mah, “A computer vision-based system for real-time detection of sleep onset in fatigued drivers”, In Intelligent Vehicles Symposium, IEEE, Eindhoven, The Netherlands, 2008, pp. 25-30.

[14] V.B. Hemadri and U.P. Kulkarni, “Detection of drowsiness using fusion of yawning and eyelid movements”, In Unnikrishnan, S., Surve, S., Bhoir, D. (Eds.) Advances in Computing, Communication, and Control: ICAC3 2013, Communications in Computer and Information Science, Berlin, Heidelberg, 2013, 361, pp. 583–594.

[15] P.L. Carrier and A. Courville, “Challenges in representation learning: facial expression recognition challenge”, Kaggle Competition, https://www.kaggle.com/c/challenges-in-

representation-learning-facial-expression-recognition-challenge/data, 2013 (Accessed 29 September 2018).

[16] D.V. Sang, N.V. Dat and D.P. Thuan, "Facial expression recognition using deep convolutional neural networks", 9th International Conference on Knowledge and Systems Engineering(KSE), IEEE, 2017, pp. 130-135.