I.J. Image, Graphics and Signal Processing, **2020, 2, 19-29**
Published Online April 2020 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2020.02.03

Human Distraction Detection from Video Stream Using Artificial Emotional Intelligence

Rafflesia Khan*
Computer Science and Engineering discipline, Khulna University, Khulna, Bangladesh
Email: rafflesiakhan.nw@gmail.com

Rameswar Debnath
Computer Science and Engineering discipline, Khulna University, Khulna, Bangladesh
Email: rdebnath@cseku.ac.bd

Received: 08 October 2019; Accepted: 28 October 2019; Published: 08 April 2020

***Abstract*—**This paper addresses the problem of identifying certain human behavior such as distraction and also predicting the pattern of it. This paper proposes an artificial emotional intelligent or emotional AI algorithm to detect any change in visual attention for individuals. Simply, this algorithm detects human's attentive and distracted periods from video stream. The algorithm uses deviation of normal facial alignment to identify any change in attentive and distractive activities, e.g., looking to a different direction, speaking, yawning, sleeping, attention deficit hyperactivity and so on. For detecting facial deviation we use facial landmarks but, not all landmarks are related to any change in human behavior. This paper proposes an attribute model to identify relevant attributes that best defines human's distraction using necessary facial landmark deviations. Once the change in those attributes is identified, the deviations are evaluated against a threshold based emotional AI model in order to detect any change in the corresponding behavior. These changes are then evaluated using time constraints to detect attention levels. Finally, another threshold model against the attention level is used to recognize inattentiveness. Our proposed algorithm is evaluated using video recording of human classroom learning activity to identify inattentive learners. Experimental results show that this algorithm can successfully identify the change in human attention which can be used as a learner or driver distraction detector. It can also be very useful for human distraction detection, adaptive learning and human computer interaction. This algorithm can also be used for early attention deficit hyperactivity disorder (ADHD) or dyslexia detection among patients.

***Index Terms*—**Distraction, E-learning, Facial Landmarks, Facial Alignment, Facial Movement, Artificial Emotional Intelligence.

I. Introduction

Computer vision has become extensively advanced and a number of complex problems can be solved using different aspects of computer vision. One of the aspects is facial landmark detection and compute the change, rotation and other features from facial landmarks over time. Impressive research works are ongoing in terms of facial recognition, landmark detection, emotion detection etc. However, there are not very works in the area of human interaction and psychological analysis based on the facial expression and body movement. For human computer interaction it is important to determine how behavior of individual changes over time. Also, it is important to find out the visual aspects of behavioral change. For example, if an individual is yawing frequently over time, it could be an indication of drowsiness, tiredness or boredom. Similarly, if a person is rotating head frequently, that could be an indication of different behavior for different contexts. If a learner is looking left or right, too frequently, instead of looking at the whiteboard in a classroom and the time duration of each rotation is comparatively too long or longer than a threshold, it is an indication of inattentiveness. Likewise, if a driver is looking left or right for a longer time duration that might cause a serious road accident and the driver needs an immediate alert.

If we could detect these inattentive behaviors, it would be beneficial to many sectors. Such as, if we can detect inattentive students in a classroom through intelligent surveillance system, we can help them for being active and also notify the teachers who have more inattentive students in his classroom to be more careful about students. Fig. 1. shows two scenarios where both active and distracted students are recognized by distraction detection system. Detecting the time period of activeness

and distraction for each students of a classroom and presenting a statistical analysis with those results will help both the students and the parents and teachers of those students. This will be a proficient contribution in case of ensuring effective learning as well as e-learning.

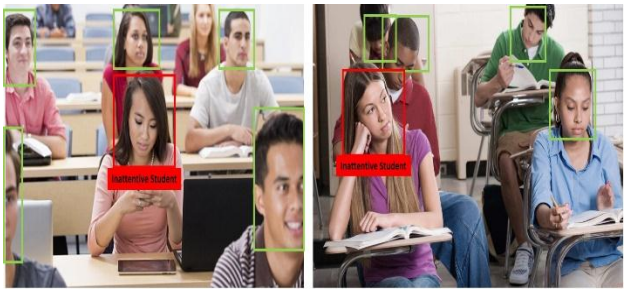


Fig. 1. Classroom scenarios with both active and distracted students.

Also, if we can detect distracted drivers timely, we might be able to reduce the amount of street accidents. Fig. 2. shows scenarios of continuously monitoring a driver. Where distraction detection model can continuously monitor a driver and detect whether he is active or distracted. In case of distraction the model can also be used for alarming the distracted driver and save him from any kind of accidents.

Fig. 2. Driver's activeness monitoring in a car.

Analyzing the importance of human distraction detection, in this paper, we propose a model for human attention and distraction detection using the correlation between facial attribute's deviation and behavioral change. In order to demonstrate the application of this work, in this paper, the use case of learning activity has been chosen.

In a classroom or in front of a computer, a learner has some common set of behavioral attributes. Facial landmark deviation helps to identify these attributes and then we detect the change in learner's behavior during any learning activity. In order to identify when there is a substantial change in their attention level, e.g., they are feeling sleepy or having difficulties to concentrate etc., Emotional AI [1, 2] based algorithm can be used to a great extent. Emotional AI is already being used successfully in game development in order to sensing and recognizing the players' emotions, and on tailoring the game responses to these emotions [3]. The similar concept can also help in learning platform development. It would also help teachers to keep peach with the students. Teachers will be able to recognize the inattentive students and their behavioral patterns and take necessary steps to help them. With this goal of improving the human interaction and understanding the pattern of change in human attention level, several works [4, 5, 6] are proposed to improve student learning activities inspection, guidance and assistance.

According to MOOCs platform analysis [5], understanding a learner's integration and assisting him accordingly is one of the core problems needs to be solved in any advanced e-learning [7] platform. Some of the main challenges in existing e-learning platform are as following

- Determine attendance
- Detection of distraction of learners
- Detection of difficulty in understanding of learners
- Detection of joint distraction where students close to each other and are both inattentive
- Detection of delayed response

An efficient e-learning platform needs to solve the above challenges. Also, classroom learning needs to be smarter as e-learning. So, this work proposes an artificially intelligent algorithm based on facial alignment and different facial attribute analysis. Our main objectives are: (1) to develop a distraction detection model for human attention level (i.e., attentive or distracted) detection, (2) to find the best facial attributes that defines attention and distraction, (3) to calculate the derivations within those attributes, (4) to monitor the deviation in facial alignment and related facial attributes based on a threshold model over time, (5) to use an emotional AI model in order to detect any change in human attention level, (6) to identify the change in attention level and continuously monitoring them, (7) to provide a statistical result of human attention and distraction period over time.

In our proposed model, we select the best attributes whose derivation defines distraction of human. All relevant facial attributes are extracted using facial landmarks from continuous video stream, therefore, this algorithm can detect changes in attention level in real time. The major contributions of this paper are as following,

- We define an emotional AI facial attribute model to determine the correlation between necessary facial landmark deviation and change in human behavior.
- We define a probabilistic emotional AI threshold model for identifying the change in human attention level.
- An emotional AI algorithm is developed to compute the deviation between the proposed emotional AI model's attributes and identify inattentiveness and inattentive events, e.g., yawning, sleeping, speaking, ADHD etc. from video stream using facial alignment and other facial landmark feature detection.

- We present quantitative and qualitative results to evaluate the performance of the proposed algorithm.

We have organized this paper in an order that some of the state of arts have been presented with their detail contribution in Section II. The proposed framework along with significant contributions are discussed in Section III. Section IV explains the system algorithm and work flow of our model. Section V demonstrates the significant improvement in performance and efficiency of our model along with comparative evaluation of performance. The final Section VI of this paper consists conclusion with some discussion and future plans of this work.

II. Leterature Review

Facial movement detection and human emotion recognition are the two popular research areas. Recently, these research methodologies are being used for distraction detection. Proper attention is a must for students at classroom or a driver while driving a car. So, most of the distraction detection models work with these two cases.

A. Models that works for learner's distraction detection

A recent work [4] has used OpenFace [8] to train a regression tree and SVM classifier to identify facial expression characteristics e.g., happiness, sadness etc. They have also detected the 68 facial landmarks and used head rotation, eye closure and mouth opening to estimate the attention label of a person. For a simple video frame, OpenFace API generates a 431 dimensional vector. The model described in [4] has used this large output vector of OpenFace to extract their required features e.g., head position, head posture, eyelid height, lip height, lip width, gaze direction, pupil position etc. Later the observed features difference between adjacent frames is computed. Then an SVM classifier is trained to obtain the criterion of classification.

However, this model constructs the learning samples by manual annotation. Therefore, the reference frame is selected manually. But, in real time video of a physical classroom, virtual classroom or car driver, it is not possible to annotate the video manually. Therefore, it is required to define an algorithm that can detect distraction in real time video without human interaction.

According to [9], the learning results can be classified and predicted by observing the student's learning activities participation, teaching objectives taxonomic ranks, attention level, interaction level and knowledge mapping analysis. So, considering the importance of attention level detection for learning another work proposed by Asteriadis, Stylianos, et al. [10] detected the interest of the attention of a person in learning. Model [10] works for human distraction detection in case of learning using head pose, gaze and hand tracking. They build information estimated from behavior-related features (e.g., level of interest, attention) of users reading documents in a computer screen. They detect the position of the irises to illustrate the direction of gaze and head pose and use the position and movement of prominent points around the eyes and head as vectors. Those vectors detect whether the particular user is looking into the screen or not and whether their eyes are fixed at a particular spot for long periods of time that is ultimately whether the learner is paying attention or not.

B. Models that works for driver's distraction detection

Alioua, Nawal, et al. [11] estimated driver attention level from monocular visible spectrum images using appearance based discrete head pose estimation. They have detected different head pose descriptors to find variations in driver head pose and proposed a fusion approach providing a good discrimination of pose variations. Then they evaluated the ability of those descriptors to represent pose variations by testing their efficiency using the SVM classifier. Model [11] calculates head pose to recognize driver's activeness and distraction. Their model works properly even if the facial features are not visible.

Choi, In-Ho, et al. [12] considered both head nodding and eye-blinking for driver's attention label detection. They have used Discriminative Bayesian—Active Shape Model (DB-ASM). The POSIT (Pose from Orthography and Scaling with Iterations) algorithm is employed to estimate the present head pose of the given face. When the driver's head pose crosses a certain threshold along a given direction extreme pose is detected. They also have used a Markov chain framework and then the driver's eye-blinking and head nodding are separately modelled based upon their visual features. After that, system makes a decision on whether the driver is distracted (e.g., drowsy) or not by combining those behavioral states.

Both model [11] and [12] work for driver's distraction detection but, none of them consider the case when a driver is talking over phone or not looking straight.

Analyzing the still existing challenges, in this proposed work, we propose an efficient model for human attention level detection where facial landmarks are used to detect attributes such as head pose, eye direction, lip movement and eye movement. We find these attributes to be directly relating to different distractive behavior such as talking, yawning, sleeping etc. Also, a threshold model is used to define the slandered value for each criterion described in Table 1. Instead of manually selecting the reference frame like [4], this work defines standard values for each of the behavior model attributes. The deviation from the standard values are controlled by a threshold model. If the deviation is less than the corresponding threshold value for a certain time, this proposed system detects a distraction event and a reduction in attention level. This proposed work can also identify distraction in real time video without any previous data processing.

III. Proposed Syatem Architecture

This proposed work focuses on sensing and recognition of the individuals' attention level and assists them to understand the pattern. For that, we first define the main

components for our proposed model. With an input video our model goes through these components one by one to recognize the proper behavioral status of a person. Fig. 3. shows the architecture of our proposed system with different components.

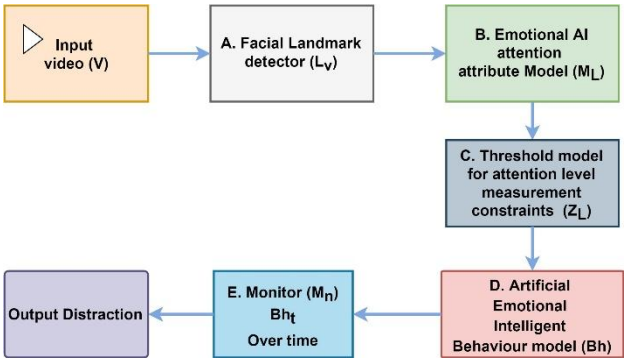


Fig. 3. Architecture for proposed system model.

The proposed system uses the following different components for attention level change detection:

- Facial landmark detection (L_V)
- Emotional AI attribute model (M_L)
- Threshold model for attention level measurement constraints (Z_L)
- Artificial Emotional Intelligent Behavior model (Bh)
- Monitor behavior continuously in video frame (Bh_t)

In the following section, we describe the major components for attention change detection of our model in brief.

A. Facial landmark detection (L_V)

Facial landmark detection component (L_V) detects different relevant facial landmarks from face. In this proposed model, instead of using open source tools like: OpenFace [13] and OpenVino [14] and getting large video analysis results, only relevant attributes for distraction detection are computed using facial landmarks [15]. A pre-trained landmark detection model of dlib library (i.e., implementation of [15]) is used to estimate the location of 68 (x, y)-landmark coordinates that map to facial structures as shown in Fig. 4(a). Not all landmarks on a face are essential for attention level change detection. Therefore, computing the deviation for all 68 landmarks computed by OpenFace is not often mandatory. Hence, after detecting landmark points we consider only those points that are required for our model's Emotional AI attribute model. Then, only the required points are used for one time computation. Fig. 4(b), (c) and (d). shows the selected landmark points used for detecting eye, head and lip correspondingly.

This work uses the ensemble of regression trees [15] based detection algorithm for facial landmark detection. Fig. 4(a). shows the 68 landmarks on a person's face detected by landmark detection model.

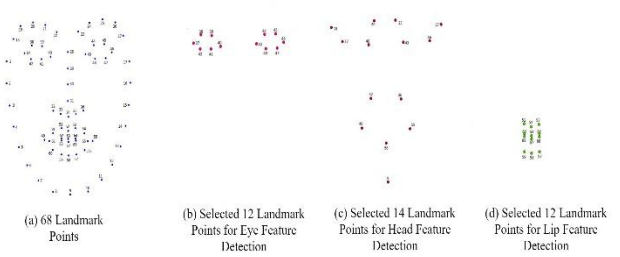


Fig. 4. Selected landmark points for different attribute detection.

We choose head pose, eye direction, lip movement and eye movement as the considerable attributes which best defines human's attention level (i.e., attentive or distracted). So, after detecting all landmarks 68 points, for our proposed model we select 12 landmarks shown in Fig. 4(b). for eyes detection, 14 landmarks shown in Fig. 4(c). for head detection and 12 landmarks shown in Fig. 4(d). for lips detection. So, instead of using all 68 landmarks our model works with 38 landmarks.

B. Emotional AI attribute model (M_L)

Among the facial landmarks, not all of them are relevant for distraction detection, therefore, this work proposes Emotional AI attention model to identify only relevant attributes for attention level change detection. For example, in a classroom, it is important to identify the attributes that influence the attention of students as well as to detect the change in their behavior and attention level in real time. Emotional AI can contribute significantly to generating effective behaviors for different game characters in gaming platform development [16]. Similarly, an effective emotional AI model can identify the relevant attributes that best relates to the human attention level. One of the major contributions of this work is that we develop an Emotional AI landmark model that identifies the efficient relevant facial attributes using facial landmarks for tracking individual's attention level. Table 1. describes the list of attributes that construct the Emotional AI attribute model for attention change detection. We have finalized 4 different attributes i.e., Head Posture alignment (H), Eye Aspect Ratio (E_r), Eye Direction (E_d) and Lip Distance (L_d) as relevant facial attributes for distraction detection. Using head (H) attribute we compute whether the person is not heading straight or heading at left, right, top or bottom. Using eye ratio (E_r) attribute we compute whether the person is frequently blinking and feeling drowsy or sleepy. Using eye direction (E_d) attribute we compute whether the person is looking straight or not. Using lip (L_d) attribute we compute whether the person is talking or yawning. All these essential attributes contribute to human distraction detection and make the model more efficient. Therefore, this Emotional AI attribute model can be used for so many sectors where distraction detection is necessary and extended for other kinds of behavior detection.

Table 1. Emotional AI attribute model (M_L)

Emotional AI Model Attribute Label	Emotional AI Model Attribute Name	Behavior
H	Head Posture Alignment	Distraction Looking right, left, top or bottom
E_r	Eye Aspect Ratio	Drowsiness, Sleeping
E_d	Eye Direction	Looking Left or right
L_d	Lip Distance	Talking, Yawning

At first, our proposed model detects 68 facial landmarks and then selects required landmarks for detecting the 4 efficient attributes of Emotional AI attribute model. So, this proposed work does not require huge computation, data pre-processing or individual trained data set for different use cases. Therefore, it is efficient enough to detect real time distraction for different use cases such as learning students, on street drivers, attention disorder patients etc.

After finalizing the 4 different attributes i.e., *H*, E_r, E_d and L_d as the relevant facial attributes for distraction detection, the next step is to compute the movements on them and classifying those movements as attentive or distracted behavior.

1) Head Posture Alignment (H)

Head posture has become one of the necessary attributes for attention decrease detection. Turning head posture is directly related to attention level change. If the head is straight or forward of an individual learner, the learner is paying attention. However, if the head is leaning back, rotating or translating the learner is potentially distracted.

In addition, alignment of head refers if the head is moving towards either left or right [4]. The proposed distraction detection model computes head posture using Euler angles (e.g., roll, pitch and yaw). Fig. 5. shows the roll, pitch and yaw angle's direction of a head.

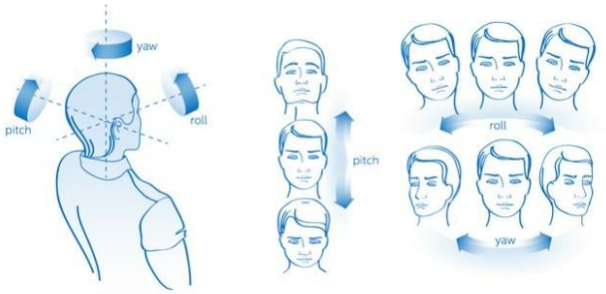


Fig. 5. Head movement at Euler angles (e.g., roll, pitch and yaw).

After detecting the Euler angles, using a proper threshold value our model estimates whether the head is moving upside-down or left to right or top to bottom. For pose estimation we follow the Perspective-n-Point [17] problem. In this problem, we find the pose of an object when we have a calibrated camera and we know the locations of some 3D points on an object and the corresponding 2D projections of the object in an image. Detail Explained in [18]. Using (1), for one object point say P, if we get the 3D coordinates (X, Y, Z) in the world coordinate space we can calculate the location of point P in the camera coordinate (C_x, C_y, C_z) when we have the rotation (3×3 matrix, r_{00} to r_{22}) and translation (a 3×1 vector t_x, t_y, t_z), of the world coordinates with respect to the camera coordinates.

$$\begin{bmatrix} C_X \\ C_Y \\ C_Z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

For head pose estimation, we select 14 face landmark points and we know the location of these 2D points (i.e., (x, y)). So, using (2) we can estimate the camera coordinates (C_x, C_y, C_z) of these points. Here in (2), (u_x, u_y) is a principal point that is usually the image center and (f_x, f_y) are the focal lengths in the x and y directions expressed in pixel units.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & 0 & u_x \\ 0 & f_y & u_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C_X \\ C_Y \\ C_Z \end{bmatrix} \quad (2)$$

After estimating the camera coordinates (C_x, C_y, C_z) of 2D points(x, y) of 14 face landmarks, the next step is to calculate the rotation and translation vector of the world coordinates with respect to the camera coordinates using (1). From the rotation vector the roll, pitch and yaw angle of the head is computed.

When the yaw angle stays between -20 to +20 threshold value, head remains at center. Yaw angle larger than 20 indicates head facing at left and smaller than 20 indicates head facing at right. Similarly, when the pitch angle stays between -20 to +20 threshold value, head remains at center. Pitch angle larger than 20 indicates head facing top and smaller than 20 indicates head facing at bottom. When the roll angle stays between -25 to +25 threshold value, head remains at center. Roll angle larger than 25 indicates head facing upside-down at left and smaller than 25 indicates head facing upside-down at right. When the head is not facing at the center for more than 40 seconds, a person is found distracted.

2) Eye Aspect Ratio (EAR) (E_r)

Scientific measuring believes that sleep deprivation leads to lower alertness and concentration [19]. So, we choose sleepiness or drowsiness as another vital attribute for our distraction detection model. To detect drowsiness we count the eye closure frequency of a person for every minute. Whenever this frequency is larger than a normal threshold, drowsiness is detected. For that, we simply store the 12 landmark points that represent the left eye (i.e., 37-42) and right eye (i.e., 43-48) from 68 points. Then, following [20] we calculate the aspect ratio (EAR) of each eye using (3).

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \quad (3)$$

Here, p_i, i= 1-6, represents the 6 points of each eye. the numerator of (3) computes the distance between the vertical eye landmarks and the denominator computes the distance between horizontal eye landmarks. Fig. 6(b). shows p_1 to p_6 points and Fig. 6(c). shows vertical and horizontal distance for a single eye.

(a) Landmark Points of left eye with actual labels. (b) Landmark Points of left eye referred as point (p1-p6) (c) Distnce between vertical and horizontal points

Fig. 6. Left eye with landmarks and distance between them.

The calculated average eye aspect ratio of both eyes remains almost constant when the eyes are open and it rapidly drops to zero while an eye-blink takes place. Whenever the average EAR is less than threshold (0.25) an eye-blink is counted. If the average EAR remains less than our model's threshold for about 5 seconds or more, the case is identified as drowsiness that indicates the person is distracted and feeling asleep. Also, if the average EAR remains less than our model's threshold for about 60 seconds or more, the case is identified as sleeping.

3) *Eye Direction (E_d)*

As another attribute of our distraction detection model, we select eye direction. In case of concentration, eye's pupil supposed to stay at center. A person is assumed distracted if his eye pupil directs to the left or right direction for certain period of time. In our model, we detect eyes direction using the same 12 points used for drowsiness detection. In Fig. 7. eye pupil direction for single eye and how eye pupil directs at left, center and right is shown.

(a) Looking LEFT (more white region at eye's 2nd half) (b) Looking CENTER (almost equal white region at both eye half) (c) Looking RIGHT (more white region at eye's 1st half)

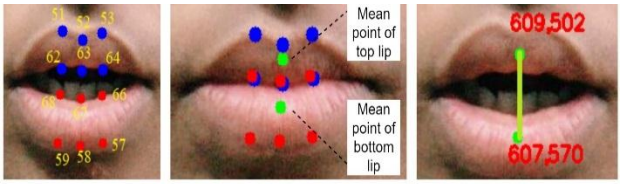
Fig. 7. Eye pupil direction for a single eye.

For direction detection, we simply create two images of eyes by cropping the two eye region using 6 landmark points for each. Then Binary-Thresholding [21] is applied to both eye images. For each eye, we divide the eye region into two equal parts. Next, we calculate the portion of white region by counting the number of non-zero pixels at each half part.

If the average white region of the first half is greater than the average white region of the second half for both eye regions, eyes are looking at right. If the average white region of second half is greater than the average white region of first half for both eye regions, eyes are looking at left. When the eyes remains directing at left or right for more than 30 seconds, the case is identified as not looking straight. When none of the above cases are found, eyes are looking at center. Fig. 7(a), (b) and (c). shows an eye with white region on different cases.

4) *Lip Distance (L_d)*

A common case of distraction is caused when a student is talking for a long time during class-time. So, our model detects the talking status of a person for recognizing distractive behavior. The talking status detection also helps to identify whether a person is yawning or not which is another symptom of distraction. For talking detection, the relevant landmark attribute is lip points, especially, the 6 top lip points and 6 bottom lip points as shown in Fig. 8 (a).



(a)Used landmark points of top and bottom lips (b) Mean point detected from used landmark points of top and bottom lips (c) Vertical distance detection of top and bottom mean point

Fig. 8. Lips with landmarks and distance between them.

Equation (4) computes the mean point for 6 individual top and bottom lip points.

$$MeanPoint = \frac{\sum LipPoint_{i,j}}{6} \quad (4)$$

Where, i, j represents the x, y coordinates respectively of the 6 points for both top and bottom lip.

After detecting mean point, the vertical distance between top and bottom lip's mean point is considered as lip distance. When the lip distance is more than threshold value 30 and it keeps increasing for about 3 to more seconds, the case is identified as yawn. On the other hand, when the lip distance is more than threshold value 30 and it remains consistent for about 20 or more consecutive frames, the case is identified as talking. Both talking and yawning decreases attention level.

C. *Threshold model for attention level measurement constraints (Z_L)*

Another major contribution of this work is that it has identified the threshold model that indicates the change in 4 major attributes selected for constructing Emotional AI attribute model. Table 2. describes the standard value for 4 different attributes when an individual is attentive. Whenever any variation from these standard values is detected the condition is considered as an influence to distraction or decrease in attention level. If these variations remains continuous for certain time constraints, the case is considered as distraction. Thus, these deviation from the standard threshold value can contribute to detect distracted behavior. For example, if head pose (H) angle is less than -20 degrees or greater than +20 degrees for 40 seconds, the individual student is considered looking at a different directions which is a potential distraction event. Any distraction event

monitored by model would reduce the attention level for the corresponding student.

Table 2. Threshold model for attention level measurement constraints (Z_L)

Emotional AI Model Attribute	Metrics	Threshold Value	Active Behavior
H	Head Posture alignment / Yaw Angle	-20 degree to +20 degree	Head at center or Heading straight
E_r	Eye Aspect Ratio	>0.25	Eyes Open
E_d	Eye Direction	Asymmetric average white region over time	Looking straight
L_d	Lip distance	<=20	not talking or yawning

D. Artificial Emotional Intelligent Behavior model(Bh)

Using the Threshold model and Emotional AI attribute model it is possible to detect behavior of any individual at any time. In this work, we have defined an Artificial Emotional Intelligent Behavior model to correlate human behavior and Emotional AI attribute model. Table 3. shows how our selected attributes, movements between attributes and threshold selected for them construct our Artificial Emotional Intelligent Behavior model.

Table 3. Artificial Emotional Intelligent Behavior model (Bh)

Emotional AI Model Attribute Label	Metrics	Standard Value	Time Constraints (For student in seconds)	Distracted Behavior
H	Head Posture Alignment / Yaw Angle	< - 20 degree or > + 20 degree	>40	Looking at different direction
	Pitch Angle	< - 20 degree or > + 20 degree		
	Roll Angle	< - 25 degree or > + 25 degree		
E_r	Eye Aspect Ratio	<0.25	>60	Sleeping
		>=0.25	>5	Drowsiness
E_d	Eye Direction	average white region of first half ≠ average white region of second half	>30	Looking at right or left
L_d	Lip Distance	>=30	>3	Yawning
		>=30 (Consistent)	>20	Talking

In Table 3. the Time Constraints (For student in seconds) define the threshold time periods that we consider as the threshold duration of variances to be recognized as distraction within a student's behavior. How these time constraints contributes to the model is briefly explained in Emotional AI attribute model (M_L). The time constraints for certain Emotional AI Model Attribute in Table 3. can be adjusted according to different situations. For example, the value for each of these thresholds in case of kindergarten students is different from the corresponding value for college students or drivers. This Artificial Emotional Intelligent Behavior model detects behaviors (e.g., Looking at different Direction, Sleeping, Drowsiness, Looking at right or left, Yawning and Talking) that best defines distraction of a person. And then the time constraints are set to consider those behaviors as distractive behavior.

E. Monitor (Mn) behavior continuously in video frame (Bht) over time

Our model continuously monitors the input video frame. Also, for a considered subject it detects required attributes and calculates the derivations. For each frame of a video our model detects and stores the estimated values for all attributes with the attention level estimated for those values. So, finally we receive all necessary majors with its corresponding attention level (i.e., whether the considered person is distracted or attentive). These values helps us creating statistical representation that statistically views the period of distraction and attention of a person over time.

IV. Proposed Algorithm

The proposed artificial emotional intelligent algorithm for detecting distraction is described in Algorithm 1.

The first step is to detect only the required landmarks where at any time t, or at any frame f, all the required landmarks for identifying distraction of subject is

detected based on the proposed Emotional AI attribute model (L_t).

Algorithm 1 Detection of distraction

```
procedure DETECTDISTRACTION(video, M_L, Z_L, Bh)
  while video ≠ null do
    Lt ← detectLandmarks(video frame)
    ML_t ← selectLandMarksRequiredforAttributeModel
           (L_t, M_L)
    A_t ← computeAIAttributeModel(M_Lt)
    A_t ← compareWithStandardThresholdOverTime
          (A_t, Z_L)
    if δA_t > 0 then
      NotifyDistraction
    else
      continue
    end if
  end while
end procedure
```

Unlike computing 431 vectors in case of OpenFace, only 68 landmarks are detected and then only the required landmarks are selected for attribute detection at this stage.

As unnecessary computation is avoided in this step, it is comparatively faster than [4]. The next step is to compute the essential attributes for Emotional AI attribute model (M_{Lt}) and then compute their derivations. The next step is to compute the difference between standard threshold value for any $l_i \in L$ defined in proposed threshold model Z_L using (5). Here n is the number of available landmarks (l_i) at any time t.

$$\delta A_t = \sqrt{\frac{\sum_{i=0}^{n}(l_i - Z_l)^2}{n}} \quad (5)$$

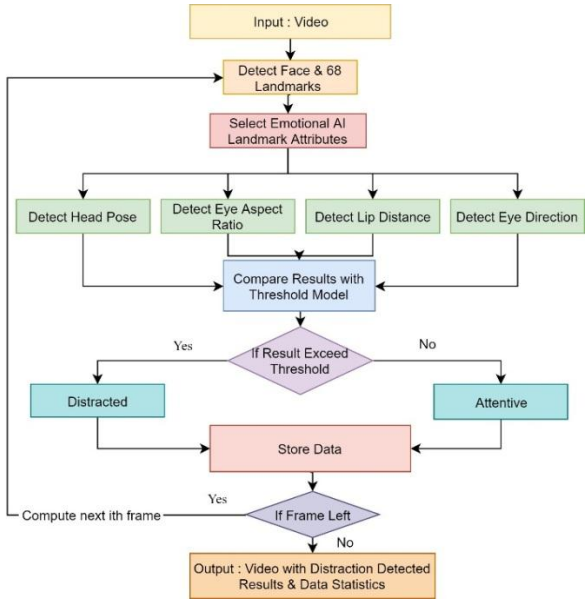


Fig. 9. Work-flow for proposed algorithm.

The current value for δA_t and the change in attribute model (M_L), is compared with the behavior model (Bh) time constraints and it is decided that the behavior is either a distraction or not, based on the behavior model described in Table 3. For a distraction event it is possible to generate an alert. After distraction and attention detection the result of distraction detection and the values of attributes derivation are stored for further processing. The stored data is used for statistical result representation. The procedure continues till any frame left in input video. Fig. 9. shows the work flow for proposed algorithm.

The process is continued for all the frames left in input video stream. Thus we can even calculate the duration of attentiveness and inattentiveness of a person throughout the input video stream.

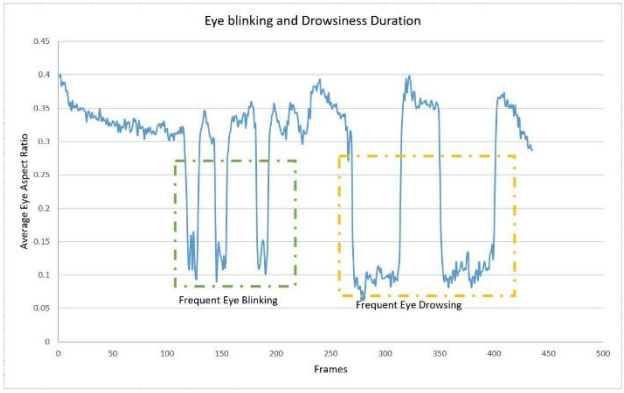
V. Performance and Compararive Analysis

For performance evaluation experiment we have used a video of 40 minutes long with a frame rate of 10fps (i.e., frame per second) and 640×480 resolution. In this video we have recorded one student's learning scenario within a classroom. We observe the scenario and enlist the active and distracted condition of the learner to create the known and actual status. Then we test the video with our model to find the model's estimated results.

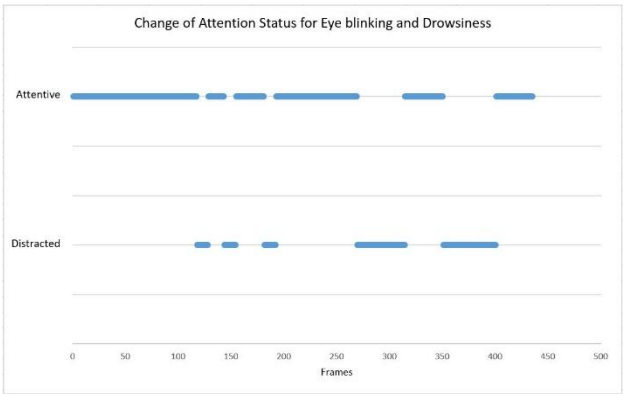
We have used face alignment landmarks [14] and our distraction detection attributes are estimated using those landmarks. No manual annotation was done. The threshold model is used to identify the difference between attentive behavior and any deviation from that. As a result, it is possible to evaluate any video in real time without any extra manual editing or video processing. Unlike [4], no additional classifier or training is used in this proposed algorithm. We have used an Emotional AI model as classifier and Threshold model to identify the distraction from standard behavior. This prediction result confirms 95.76% accuracy in case of detecting the actual behavior of a learner. We apply (6) to calculate the accuracy of distraction detection in percentage. Here, accuracy is calculated by estimating the ratio between total the number of frames where person is found distracted by the model with total number of frames where person is actually distracted.

$$Accuracy = \frac{Total\ number\ of\ frames\ where\ subject\ is\ found\ distracted\ by\ model}{Total\ number\ of\ frames\ where\ subject\ is\ actually\ distracted} \quad (6)$$

For our model's performance evaluation, we present some statistical features representations. Fig. 10. presents how attention and distraction status changes with frequent eye blinking and drowsiness detection over a video of 435 frames. Fig. 10. shows the average eye aspect ratio plotted against frames and how it changes over time and also how attention level changes with it.



(a) Average eye aspect ratio plotted against frames.

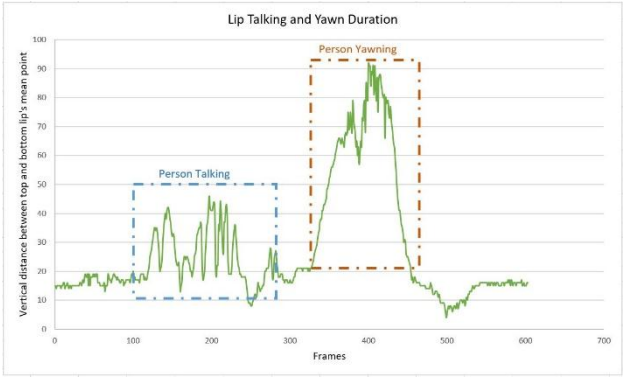


(b) Change in attention status because of frequent eye blinking and drowsiness.

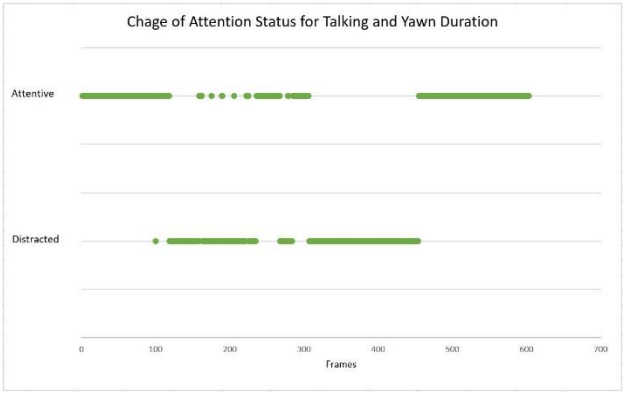
Fig. 10. Change in attention status with average eye aspect ratio.

Here in Fig. 10(a). average EAR is plotted against number of frames where we have also high lighten the eye blinking and drowsy situation. Fig. 10(b). shows how attention and distraction is detected with the derivation of average EAR. Fig. 10(b). also shows how eye blinking and drowsiness creates distraction.

Fig. 11. shows how attention and distraction status changes when a person talks or yawns over a video of 603 frames. Fig. 11. shows the vertical distance between two lips plotted against frames and how it changes over time also how attention level changes with it.



(a) Vertical distance between top and bottom lips plotted against frames.



(b) Change in attention status because of talking and yawning.

Fig. 11. Change in attention status with vertical distance of top and bottom lip.

Here in Fig. 11(a). vertical lip distance is plotted against number of frames where the talking and yawning scenarios are also high lighten. Fig. 11(b). shows how attention and distraction is detected with the derivation of distance between top and bottom lip.

Fig. 12. shows how attention and distraction status changes with head posture alignment over a video of 966 frames.

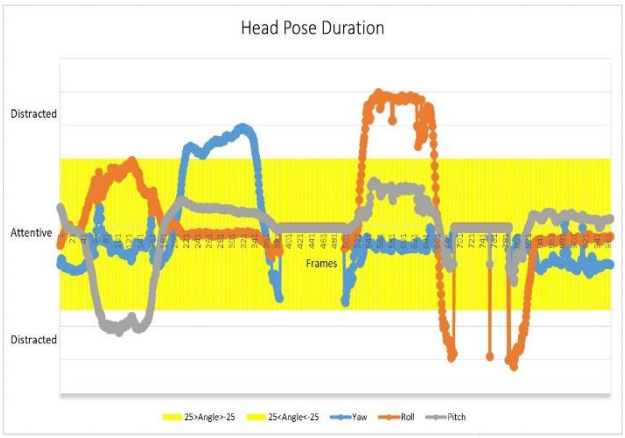


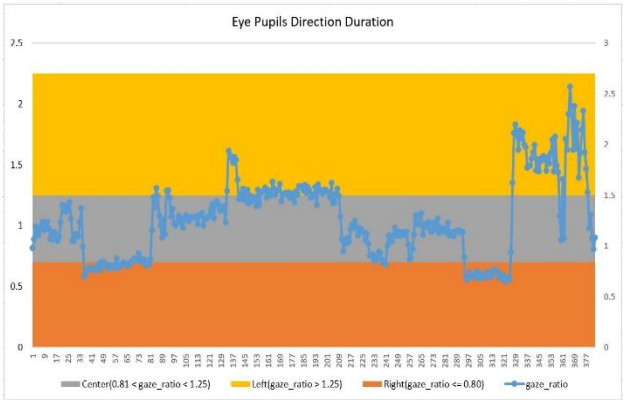
Fig. 12. Attentive and distracted result because of head pose (roll, yaw and pitch angle).

Our model detects distraction when a person head's roll, yaw and pitch angle exceeds the threshold value. In Fig. 12. the roll, yaw and pitch angle value of head is plotted against frame number. Fig. 12. also shows how attention level chances from attentive to distracted whenever head pose angle values exceeds the threshold.

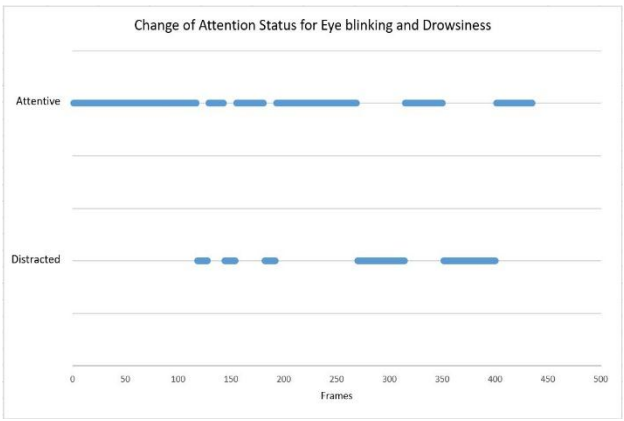
Fig. 13. shows how attention and distraction status changes when a person is not looking straight and his eye pupils are directed towards left or right over a video of 384 frames.

Here in Fig. 13(a). average eye gaze value (i.e., asymmetric average white region) is plotted against number of frames. Here for better understanding we marked (yellow as looking left, gray as looking straight

and orange as looking right) regions of Fig. 13(a). Fig. 13(b). shows how attention and distraction is detected with the derivation of average eye gaze value and also how the attention level changes whenever the eyes are not looking straight.



(a) Eye gaze ratio plotted against time in seconds where plotting area is segmented in regions indicating eyes pupil direction (left, right and center).



(b) Change in attention status because of eyes not looking at center.

Fig. 13. Change in attention status with eyes pupil direction.

Table 4. shows a comparative analysis of our work along with some existing models. This comparison shows that the proposed work is efficient in term of complex computation, data processing overhead, manual interaction and additional classifier. Unlike other work, this proposed work is not dependent on trained data set. For example in case of [11] different datasets need to be trained for different locality student. In this proposed work, the standard values are unique regardless of locality, race and other attributes. Table 4. shows what our model do no need for proper distraction detection and what other models do need.

Analyzing Table 4. we can conclude that our model works with real time data without any pre-processing. It does not require any previous dataset, training or classifier. Also, without additional computation, dataset or facial feature it ensures better result. Our model confirms an average 95.76% accuracy in case of distraction detection of single person from a real time video of 40 minutes with 10 fps frame rate and 640×480 image resolution. The rate is also comparatively better than model [4]'s 92% accuracy on a video with 15 fps frame rate and 1280×720 image resolution.

Table 4. Comparative analysis of proposed work with existing models.

	Our work	Nawal Alioua et al. [11]	In-HoC. et al. [12]	Liying W.et al. [4]	Stylians A. et al. [10]
Manual Annotation				√	
Data Pre-processing		√	√	√	
Additional classifier		√	√	√	√
Dataset dependency		√		√	
Real time without data pre-processing	√				
Additional computation		√	√	√	√
Additional Database			√		
Facial Feature Extraction		√	√		√

VI. Conclusion

In this paper, we propose a distraction detection model for human. Our model continuously monitors a video and recognizes attentive as well as distracted behavior of a person. We believe this distraction detection model can be helpful for so many sectors such as students learning, where human attention as well as distraction detection is necessary. The performance evaluation for the proposed system shows that the algorithm can effectively detect the change in individual learner's attention level. This algorithm can also detect any abnormal facial expression during learning activity based on the facial attribute's alignment deviation from standard value over time. This proposed work can detect distraction for real time video without any manual annotation. Due to our defined attribute model with selected landmarks, unnecessary computations are avoided and the system configuration is faster than state of art. Also, because of the novel behavior model, no additional machine learning or deep learning training is required for classifying distracted and attentive behavior.

Up to now we have been able to test our model with single person and detected the duration of his attentive and distracted periods. For a 40 minutes long video our model has achieved 95.76% accuracy in case of detecting the actual attention level of a person. In future we look forward to detect distraction of multiple persons within a classroom. Also, we are willing to work on the adaptability and robustness of the model with different camera angles or with the same camera under different distance.

ACKNOWLEDGEMENTS

We gratefully thank Fellowships, scholarships and grants for Innovative ICT-related sector, ICT Innovation fund, ICT division, Government of People's Republic of Bangladesh for financial support.

REFERENCES

[1] R.W. Picard, Affective computing, MIT press (2000).
[2] R. W. Picard, "Affective computing: challenges," International Journal of Human-Computer Studies 59(1-2), 55–64 (2003).
[3] E. Hudlicka, "Affective computing for game design," in Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation, 5–12, McGill University Montreal, Canada (2008).
[4] L. Wang, "Attention decrease detection based on video analysis in e-learning," in Transactions on Edutainment XIV, 166–179, Springer (2018).
[5] N. Hakami, S. White, and S. Chakaveh, "Motivational factors that influence the use of moocs: Learner's perspectives," in Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017), 323–331 (2017).
[6] W. Shunping, "An analysis of online learning behaviors and its influencing factors: A case study of students' learning process in online course" open education learning guide" in the open university of china [j]," open education research 4 (2012).
[7] R. H. Shea, "E-learning today," US News & World Report 28, 54–56 (2002).
[8] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 1–10, IEEE (2016).
[9] Wu, F., Mou, Z.: The design research of learning outcomes prediction based on the model of personalized behavior analysis for learners. E-educ. Res. 348(1), 41–48 (2016).
[10] S. Asteriadis, P. Tzouveli, K. Karpouzis, et al., "Estimation of behavioral user state based on eye gaze and head pose-application in an e-learning environment," Multimedia Tools and Applications 41(3), 469–493 (2009).
[11] N. Alioua, A. Amine, A. Rogozan, et al., "Driver head pose estimation using efficient descriptor fusion," EURASIP Journal on Image and Video Processing 2016(1), 2 (2016).
[12] I.-H. Choi, C.-H. Jeong, and Y.-G. Kim, "Tracking a driver's face against extreme head poses and inference of drowsiness using a hidden markov model," Applied Sciences 6(5), 137 (2016).
[13] B. Amos, B. Ludwiczuk, M. Satyanarayanan, et al., "Openface: A general-purpose face recognition library with mobile applications," CMU School of Computer Science 6 (2016).
[14] OpenVINO, "Openvino deep learning computer vision toolkit." Available link: https://software.intel.com/enus/openvino-toolkit (26 Sept. 2016).
[15] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in Proceedings of the IEEE conference on computer vision and pattern recognition, 1867–1874 (2014).
[16] C. Conati and X. Zhou, "Modeling student's emotions from cognitive appraisal in educational games," in International Conference on Intelligent Tutoring Systems, 944–954, Springer (2002).
[17] X. X. Lu, "A review of solutions for perspective-npoint problem in camera pose estimation," in Journal of Physics: Conference Series, 1087(5), 052009, IOP Publishing (2018).
[18] S. Mallick, ""home." learn opencv.." Available link: https://www.learnopencv.com/tag/solvepnp (26 Sept. 2016).
[19] P. Alhola and P. Polo-Kantola, "Sleep deprivation: Impact on cognitive performance," Neuropsychiatric disease and treatment (2007).
[20] T. Soukupova and J. Cech, "Eye blink detection using facial landmarks," in 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia, (2016).
[21] I. Thresholding, "Opencv." Available link: https://docs.opencv.org/ref/master/d7/d4d/tutorial_py _thresholding.html (Accessed September 27, 2019.).

Authors' Profiles

Rafflesia Khan is a student of M.Sc. at Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh. She has completed her Bachelor's degree from Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh in 2017. Her research areas of interest are image and video processing, visual object detection & recognition, facial behavior recognition, machine learning, pattern recognition, and internet of things security.

Rameswar Debnath is a Professor of Computer Science and Engineering Discipline at Khulna University, Khulna, Bangladesh. He has completed his Bachelor degree from Computer Science and Engineering discipline, Khulna University, Khulna, Bangladesh in 1997. He has received his Masters in Engineering degree and PhD degree in Computer Science and Engineering from the University of Electro-Communications, Tokyo, Japan in 2002 and 2005 respectively. His research areas of interest are image processing, statistical machine learning and its applications to pattern recognition, visual object detection and natural language processing.

How to cite this paper: Rafflesia Khan, Rameswar Debnath, " Human Distraction Detection from Video Stream Using Artificial Emotional Intelligence", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.12, No.2, pp. 19-29, 2020.DOI: 10.5815/ijigsp.2020.02.03

