

# Real-time imaging-based assessment model for improving teaching performance and student experience in e-learning

Muhammad Farhan<sup>1,2</sup> · Muhammad Aslam<sup>1</sup> · Sohail Jabbar<sup>2,3</sup> · Shehzad Khalid<sup>4</sup> · Mucheel Kim<sup>5</sup>

Received: 30 June 2016 / Accepted: 5 December 2016 / Published online: 11 January 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Multimedia is an essential and integral part of electronic learning (e-learning). In this study, teaching performance and student learning experience are measured using real-time multimedia processing tools and techniques for the e-learning paradigm. Visual attention and visual engagement analysis are performed using two developed algorithms. Video lectures are recorded and delivered to students in e-learning pedagogical setup, which are examined for the visual attention and visual engagement of the student and teacher, respectively. Proposed methodology integrates the assessment on both student and teacher ends. Multimedia processing of video lectures for teaching

performance produces scoring dataset. The same methodology on student end for visual attention is used to investigate student experience. These types of datasets then reduced to time-based datasets from the image-based dataset. Correlation and association of both datasets provide the opportunity to relate both student experience and teaching performance as well as to move forward to create content that is more useful. Computational performance of the developed algorithms is compared using different video lectures with their processed frames per second, which is analyzed as per their corresponding bins. Mean, max, and median of the processed frames of all the processed videos are also compared.

**Keywords** Video processing · Real-time computer vision · Real-time imaging · Visual attention · Visual engagement

✉ Mucheel Kim  
mucheel.kim@gmail.com

Muhammad Farhan  
farhansajid@gmail.com

Muhammad Aslam  
maslam@uet.edu.pk

Sohail Jabbar  
sjabbar.research@gmail.com

Shehzad Khalid  
shehzad\_khalid@hotmail.com

<sup>1</sup> Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Pakistan

<sup>2</sup> Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal, Pakistan

<sup>3</sup> Network Lab, School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea

<sup>4</sup> Department of Computer Engineering, Bahria University, Islamabad, Pakistan

<sup>5</sup> Division of Media Software, Sungkyul University, Anyang 430742, Korea

## 1 Introduction

This work is placed on the applications of real-time multimedia content analysis and, on the other hand, the analysis of human user. We present a real-time visual attention-scoring algorithm and a real-time visual engagement-scoring algorithm. Multimedia is an essential element of e-learning, but analysis of multimedia content for its visual engagement is not considered well before [1]. Visual engagement refers to the visual attractiveness of the video lecture for a student. This measure is useful to calculate the effectiveness of the video lecture. Student and teacher are the human users in this work, and they can take advantage of this analysis. Student's visual attention is measured with the attention-scoring algorithm. Both of the measures are useful for student and teacher to improve the learning experience and to improve the teaching performance [2].

We have demonstrated the analysis of video lectures by getting a video stream from the video file and extracting

images one by one. Each image is converted to grayscale image, image processing is performed to enhance it, and then face detection followed by eyes detection is performed. The same process is performed on the student side. A video stream is captured from the webcam mounted on student's laptop of computer [3]. Each image is enhanced, and detections are performed. This real-time recognition and tracking [4] of faces and eyes in the video produce two types of dataset.

Student–teacher interaction using real-time multimedia analytics has been nearly understudied. The aesthetic qualities of interaction design were not largely researched. Interaction design is a feature that is distinctive to the computer when the user is human. In case of e-learning, both the human users do not interact directly, so they have to use computers between [5]. Video lecture recording starts with the interaction of a video camera setup [6], whereas the student is not present. Recorded video lectures are delivered to the student. When a student takes lectures, no teacher is present in front. The student has to interact with a computer. Both of human users need to interact with computers. Elimination of face-to-face interaction and introduction of human–computer interaction by the user of multimedia content create an opportunity to analyze the situations with real-time analysis. Video services are becoming more pervasive due to the swift development of the next-generation communication and network systems. With the growth of these types of services, the numbers of users are increasing in exponential rate and numbers of devices are being used for it, e.g., wearable equipment, smartphones, tablets, personal computers or laptops, and smart televisions. Real-time object tracking is a challenging area of research, and an increasing number of good results are reported with respect to robustness and accuracy. Frequently, no clear statements about the tracking situation are made, when it comes to the application of these methods to real-world problems [4]. Human visual system (HVS) cannot give the same attention to the whole scene simultaneously especially when watching a video on a display device. Hence the overall perceptual quality of HVS from the subjective point of view can not be fully determined.

The objective of the study for an imaging system is to explore student-focused software applications of tracking and recognition of student and teacher in e-learning [2]. Perceptual inspired imaging methods have been widely used in many areas of visual information processing and multimedia tools and applications [7]. As the multimedia system is producing videos and other types of data, e.g., sounds and images that are eventually perceived by a human, it is important to note how HVS is using this information [8]. The HVS is only partially understood due to its complex nature, which exhibits many nonlinearities [9]. Hence, it is an opportunity, which can be exploited in multimedia systems with computational science, and aesthetics [10]. The HVS establishes communication

among eyes and the brain to respond external stimuli in the form of multimedia content [11]. Human vision is one of the senses that we depend on more in our daily life. Human vision works very fast to detect the objects and recognize the scene in front of the eyes [12]. Watching videos for learning purposes require the good visual content of the video lectures so that students can extract useful information [13]. The current study is about the measurement of video lectures for the visual engagement regardless of the teaching material or content covered in the lectures. Analysis of video lecture frame by frame is called visual engagement analysis. The frame-by-frame analysis of video lectures produces a dataset comprising a visual engagement score and an appropriate class label. In the same fashion, the student's attention is measured using a webcam or video camera setup [6, 14]. The measurement of student's attention using a webcam is called visual attention. Our vision is a fine-tuned process [13]. The brain interprets the images captured in the process of vision. The presence of a face, eyes and furthermore checking the eyes are open or close allow measuring visual attention and preparing a dataset to further analysis as well as the application of machine predictors to extract and predict the student learning experience [15]. Both datasets for visual engagement and visual attention are merged by transforming them into time-based data.

Student–teacher interaction is a critical part of the learning process at both ends. Effective lecture delivery is also essential for the students to learn. In the case of recorded video lectures, the visual attractiveness of the video content makes the students more attractive and engaged toward lectures [16]. e-learning refers to the type of learning where the student has to learn with own pace and no teacher is there to monitor [17, 18]. Automated tools can help such kind of scenarios as well as the visual engagement of the video lectures can also be measured using this type of tools [19]. During this research, a tool “Video Lecture Analysis” is developed. A video of standard quality usually has 16–20 frames per second, but this can be more if the video is of high quality (HD). The tool analyzes the video lecture or video presentation and produces a dataset. A video is processed frame by frame, and each frame is marked with the visual engagement of the teacher, e.g., presentation, partially engaging, partial eye contact with the camera, and fully engaging [6, 20]. The interactive video constructs are developed by time-based video constructs. This article presents a qualitative assessment methodology in an e-learning environment with the analysis of student attentiveness and teacher's performance measurements. A model is presented to demonstrate the association of attentiveness and engagement for both students and teachers, respectively. Furthermore, an introduction that describes the overall and background of the problem in learning and teaching proposed methodology to collect, analyze, and extract the information from student and teacher ends, results and discussion, literature review, and the conclusion of the study.

This paper is composed of five sections and organized as follows: Sect. 1 is the introduction of the paper, which illustrates the importance of the study and relationship between real-time imaging model and e-learning. Section 2 is the literature review, which is further divided into two sections. Section 2.1 is about the human visual system and applied visual perception; Sect. 2.2 is about the real-time computer vision and human–computer interaction. Section 3 is about proposed methodology, i.e., real-time imaging-based assessment model. Section 3 is further divided into two subsections. Section 3.1 is about tracking student’s visual attention, and an algorithm for real-time visual attention-scoring algorithm is presented. Section 3.2 is about tracking teacher’s visual engagement, and second algorithm about real-time visual engagement-scoring algorithm is presented. Section 4 gives results and discussion. Section 5 is the conclusion and future work of the current study.

## 2 Literature review

The goal of this section is to explore real-time models [4], algorithms, and technologies needed to enable multimedia systems. State-of-the-art research related to real-time imaging systems with computational analysis that can invoke the human emotional responses is the part of this section [21]. The real-time aspects of video processing systems and technologies including experimental prototypes, theoretical studies, emerging trends, and applications are also reviewed in this section [22, 23]. Content-based information retrieval from the video is very common phenomena in the literature [24]. A set of rules is applied to search the video content by extracting the sequence of the image [23]. At the time of the search, each image denotes a search state [25]. Comparison of different models is shown in Table 1.

### 2.1 Human visual system and applied visual perception

The ability to process visual details is due to the human visual system, which is the part of human central nervous system. The stimulus is received and interpreted to extract the information from the video [35]. The visual system performs a number of complex chores, including the identification and categorization of visual objects [36]. The HVS model is used by computer vision experts to video processing and image processing to deal with psychological processes as well as biological to get the full insight of human working for visuals and the better design of intelligent interfaces [37]. Such a model is used to simplify the behaviors of what is a very complex system [38]. HSV gives humans the ability to process visual detail [36, 39]. It interprets information after the detection using the biological phenomena to build a

representation of the environment [40]. Student receives the visual stimulus from the video lecture [41]. The presence of the teacher’s face encourages the student and visual attention is increased toward the lecture and same is the case if each image of the video is aesthetic and have a good contrast then the visual attention will be good [19]. Repetitive visual stimulus is a visual stimulus that has a characteristic [39]. The stimuli are simultaneously presented to the user when focusing his/her attention on the corresponding stimulus [39]. In typical, an entire video stream is initiated after it is divided into multilayered substreams to ensure the diversity of multiple devices and to satisfy the various demands from users [42]. Following the initiation, these substreams are distributed and transmitted through multiple subchannels in parallel to diverse end-users [43]. In order to enhance the recognition performance of the system, an innovative technique is presented [44]. The classification of segmented images is used to facilitate the search speed [45]. An efficient scheduling arrangement with the combination of a reinforcement algorithm significantly reduces time [46]. The massive amount of videos can be processed using cloud computing technologies [47]. Cloud computing-based solution for education institutes in proposed [48, 49]. Big data processing and analysis can also be useful for this purpose. Big data get problems when varieties of data sources are encountered. This problem can be resolved using transformation techniques; big data transformation algorithms are presented in [50]. Big data transformation solves the problems of heterogeneity, and selection of accurate and robust features is the central component of useful content-based image information retrieval using shape information of objects [51].

### 2.2 Real-time computer vision and human–computer interaction

The real-time aspect is critical for the analysis of video lecture files and visual inspection of the students. It is seen in the literature that usually that an image processing algorithm is developed and theoretically it is proven very sound, but its practical applications and the detailed steps, methodology, and trade-off analysis required to achieve its real-time performance are never fully explored [52]. This happens because it is presumably designed with a specific application in front. The current study is thus intended to bridge the gap between the theory and practice of image and video processing to design, implement, or utilize real-time image and video processing systems [23].

A good work has been done from the last decade that utilizes numerous kinds of visual features and also the combination of these features [38, 42, 44, 45, 51, 53]. Image segmentation based on skin color is also used by many researchers in the literature. At the same time, many color spaces have also been added in the literature [45].

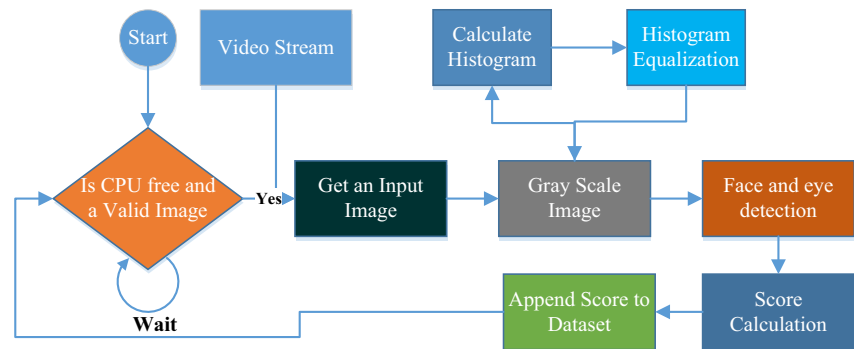
**Table 1** Feature set comparison used in different models for interaction assessment

	Features set extracted from models or frameworks or methodologies	ML for visual attention	Sensors used for eye data	Assessment using surveys	Face detection approach	Simulation-based attention	Visual attention calculation	Data collection and analysis	Engagement
1	Eye level clustering, simulation-based visual attention, eye movement for handwriting recognition [26]	✓	✓	✗	✗	✓	✓	✗	✗
2	Video sequence analysis, face segmentation, face tracking based on first frame [27]	✓	✗	✗	✓	✗	✓	✓	✗
3	Machine learning (SVM) for visual attention calculation, sensor-based eye tracking data collection and analysis [28]	✓	✓	✗	✗	✓	✓	✓	✗
4	Data mining integration with e-learning, data related to number of e-mails, messages and grading, etc. [29]	✓	✗	✓	✗	✗	✗	✓	✓
5	Eye movement datasets for visual attention measurement and analysis, sensor-based eye tracking data collection and analysis, prediction of human activities [30]	✓	✓	✗	✓	✗	✓	✓	✗
6	Teacher's pedagogical assessment in e-learning using questionnaire survey [31]	✗	✗	✓	✗	✗	✗	✓	✓
7	Student interaction in e-learning using surveys for data collection, LMS, distance learning, asynchronous online discussion [32]	✓	✗	✓	✗	✗	✗	✓	✓
8	Multiple object recognition used for the measurement of visual attention, Google street view images, image-based multiple object recognition [33]	✗	✗	✗	✓	✗	✓	✓	✗
9	Engagement analysis using social media and electronic devices for e-learning [34]	✓	✗	✓	✗	✗	✗	✓	✓

Some popular color spaces are YUV, YCrCb, HSV, RGB, etc. Color spaces [45] professionally separating the luminance components of the color from the chromaticity are usually understood well by the researchers [45]. In the literature, it is also common that contour extraction is used that is based on edge detection effects in a large number of edges that belong to the detecting objects but also to irrelevant background objects [44]. Significant work has been carried out on finding the objects in grayscale images based on their texture and appearance. Motion is an indicator, which is utilized by a few approaches to detection. Since it assumes that the only motion in the image is due to movement, the purpose is that motion-based detection demands a very controlled setup. The detection method can be used for tracking as well, because it is fast enough to

operate at image acquisition frame rate [54]. Methods for tracking and detection are very similar to each other [44]. These methods extremely limit the image search to keep the record of the object that was detected [55]. The implicit guess for the method to succeed due to frequent accusation of images [53]. Feature tracking is also another way that is found in the literature of computer vision [21]. This type of interaction enables the users of e-learning to interact with data in the way that is more natural [56, 57].

There are a number of software platforms and libraries available to the programmers to build the systems capable of real-time analysis and scoring systems for the video and image processing [52]. The list includes MATLAB [58], OpenCV [59], A Forge.NET Framework ([www.aforgenet.com/framework](http://www.aforgenet.com/framework)), and iGesture ([www.igesture.org](http://www.igesture.org)). OpenCV is widely

**Fig. 1** Real-time video processing model

used to develop the software applications for computer vision and human–computer interaction [60]. C++ is the natively supported language for it [61], while support for other programming languages is also available, e.g., Java, C# (EmguCV [62] is a wrapper library of OpenCV), and python. In the next section, the algorithms are presented for video processing to calculate visual engagement and visual attention. Face detection and tracking algorithms exist, but no intelligent algorithms to measure and analyze teaching performance and student experience. In the same way, eye movement datasets are available but no real-time measurement and analysis algorithms present in the literature. The principle goal of this methodology is to enhance student learning and teacher performance by improving student visual attention and learning experience and enhancing teacher visual engagement and teaching performance. In this study, we have developed real-time imaging-based assessment model. The assessment model tracks student’s visual attention by using our developed algorithm, i.e., real-time visual attention-scoring algorithm. In the same way, teacher’s visual engagement is tracked using our developed algorithm, i.e., real-time visual engagement-scoring algorithm.

### 3 Real-time imaging-based assessment model

The visual reception takes place at the retina, which gives an image, the color, and shadow. The visual cortex in the brain interprets the image to extract context, semantic, and form [63], while the image is transferred in the form of neural impulses and then transmitted through the optic nerve to the rest of the brain for processing. This study aims to assess the visual engagement of the video lectures. This analysis can be useful for the presenter and student to find out the overall visual attention of the videos [44]. For this purpose, a new algorithm and data collection module are developed. Videos can be transformed into a dataset with the help of data collection module. The dataset is prepared by extracting the image frames from the video and marking them with a number of faces [38], the number of eyes, the status of eyes, and the engagement score along with nominal values of

engagement level as shown in Fig. 1. In case of students, the video stream is taken from webcam, and in the case of the teacher, the video stream will be taken from video file. Frame-by-frame analysis is performed in real time. This process continues until the last frame in the video stream is present. Score of each frame is appended in the dataset.

Histogram equalization stretches out the range of grayscale value frequencies. This technique improves the contrast in an image, in order to spread out the intensity range, and detection is made easy for the computer vision algorithms [64].

#### 3.1 Tracking student’s visual attention

The first step in visual attention scoring is the detection of faces and eyes and the segmentation of the face and eyes from the image. The process of segmentation is central since it separates the scoring portion of the image for the calculation from the image background, so that it can pass it to the subsequent recognition and tracking phases as shown in Algorithm 1. Video processing in real time is done in this algorithm. Image is extracted one by one from the video. The extracted image is converted to a grayscale image, and this image is enhanced for the correct detection of face and eye. The detection of face and eyes is performed using multiscale cascade classifier. After the detection of faces and eyes, the process of scoring is started. If no face is detected, then zero is inserted for all attributes, and in the case of face detection, the score depends on the number of faces detected. The score calculation is performed using Eq. (1). Same is the way of score calculation for detected eyes. The number of faces and eyes depends upon the number of persons in the video. Face and eye ratio helps to determine the size of the eye in the image. This ratio further helps to determine either the eye is open or closed by calculating Hough circles. If no circle is found, that means the eye is closed, and in the case of any number of circles, the eye is open. Visual attention score is calculated using Eq. (5). Here constant 0.5 is taken because of the eye and face ratio, i.e., one face has two eyes. The constant 2 is taken because if one face is detected, it will normalize with respect to eyes detected. The level of attention is determined by the calculated score. The steps of the Algorithm 1 are self-explanatory.

**Algorithm 1** Real-Time Visual Attention Scoring Algorithm**Input:** A Video Stream from webcam while student is taking video lecture**Output:** A Dataset for Frame-based Visual Attentiveness Score and Attention Level i.e. "away", "sleepy", "satisfactory", "attentive"

1. **Begin**
2.     *Start the video capturing process*
3.     *Start Timer Tick*
4.     **If** Time Tick == 1000ms
5.         **Then** reset Timer and check the number of images processed i.e. PFPS
6. **If** Processor is entering IDLE state **AND** next is a valid image
7.     **Then** capture an image from the camera's video stream
8. **Else** exit
9.     *Convert the extracted color image to grayscale image*
10.    *Equalize the histogram of the grayscale image for light and brightness correction*
11.    *Detect multiscale face image using cascade classifier*
12.         **For Loop** Rectangle rect in detectFace
13.             *Draw rectangle around each face image*
14.             Copy imgOriginal to faceOnly
15.             Region of Interest of faceOnly = rect
16.             faceDetected = true
17.             
$$F(f) = y \begin{cases} 0 & \text{if no face} \\ \sum_{i=1}^n f_i & \text{on each face} \end{cases} \dots\dots\dots (1)$$
18.             *Append face detection score to the dataset*
19.         **End For Loop**
20.     *Crop and copy each face image*
21.     *Detect multiscale eye image using cascade classifier*
22.         **Loop For** Rectangle eyeRect in detecteye
23.             *Draw rectangle around each eye image*
24.             **If** faceDetected == true **Then**
25.                 
$$E(f) = x \begin{cases} 0 & \text{if no eye} \\ \sum_{i=1}^n e_i & \text{on each eye} \end{cases} \dots\dots\dots (2)$$
26.             *Append eye detection score to the dataset*
27.             **End If**
28.         **End For Loop**
29.     *Crop and Copy each detected Eye image*
30.     
$$\text{faceEyeRatio} = \frac{\text{dectctedFaceHeight} \times \text{dectctedFaceWidth}}{\text{dectctedEyeHeight} \times \text{dectctedEyeWidth}} \dots\dots\dots (3)$$
31.     
$$\text{eyeRadius} = \sqrt{\frac{\text{dectctedEyeHeight} \times \text{dectctedEyeWidth}}{\text{faceEyeRatio}}} \dots\dots\dots (4)$$
32.     Circles = Detect HoughCircles (canny Threshold = 100, circle Accumulator Threshold = 120, eyeRadius, eyeRadius × 2, eyeRadius – 10, eyeRadius + 20)
33.         **If** circles detected > 0 **Then**
34.             Take EOC score 1 for each opened eye
35.         **Else**
36.             Take score 0 for each closed eye
37.         **End If**
38.     
$$\text{VisualAttentivenessScore} = \frac{0.5 \times \text{numberOfFaces} \times 2}{\text{eyesDetected}} + \frac{\text{openedEyes}}{\text{eyesDetected} \times 0.5} \dots\dots\dots (5)$$
39.     *Append Visual Attentiveness Score to the dataset*
40.         **If** attentiveness < 0.49 **Then**
41.             Attention-level = "away"
42.         **Else-if** attentiveness > 0.49 && attentiveness <= 0.74
43.             Attention-level = "sleepy"
44.         **Else-if** attentiveness > 0.74 && attentiveness <= 0.86
45.             Attention-level = "satisfactory"
46.         **Else** Attention-level = "attentive"
47.         **End If**
48.     *Append Visual Attention level to the dataset*
49.     **Goto** Step 1
50. **End**

Teachers can use the data to observe students' activities and extract the patterns to recognize and categorize the problems on an individual basis as well as a whole. Although the results extracted from the data are usually postures server problems, they can be portrayed in a proper and suitable manner if student teacher comparison-based studies are conducted. The learning management systems usually display the information using very basic charts or in the text format [1]. The massive amount of data can be well organized to structure the semantics. So, the inference can be performed to extract the learning patterns and measure teaching performance. Moreover, very restricted interactivity is generally presented by these monitoring features, which can be overcome with the proposed methodology. Recorded video lectures and students attending those lectures remotely provide video files and video stream [65]. Both of the video sources can be analyzed to extract the patterns and to verify the effectiveness of the video lecture as well as attentiveness problems being faced by the students.

### 3.2 Tracking teacher's visual engagement

Video lecture has many instructive abundant potentials to make learning and teaching situation better if it is adopted with a proper way. Therefore there is a need to integrate technology and make novel prospects for learning and teaching. Video lecture development involves detailed implementation scheduling. The understanding philosophical perspective of pedagogy is a criterion for effective learning content. According to [66], the constant learning through the use of video lectures has the benefit that knowledge might get directed to the students. Preparing the video lectures and taking additional information from it will also seem to motivate medium's usability as shown in algorithm 2.

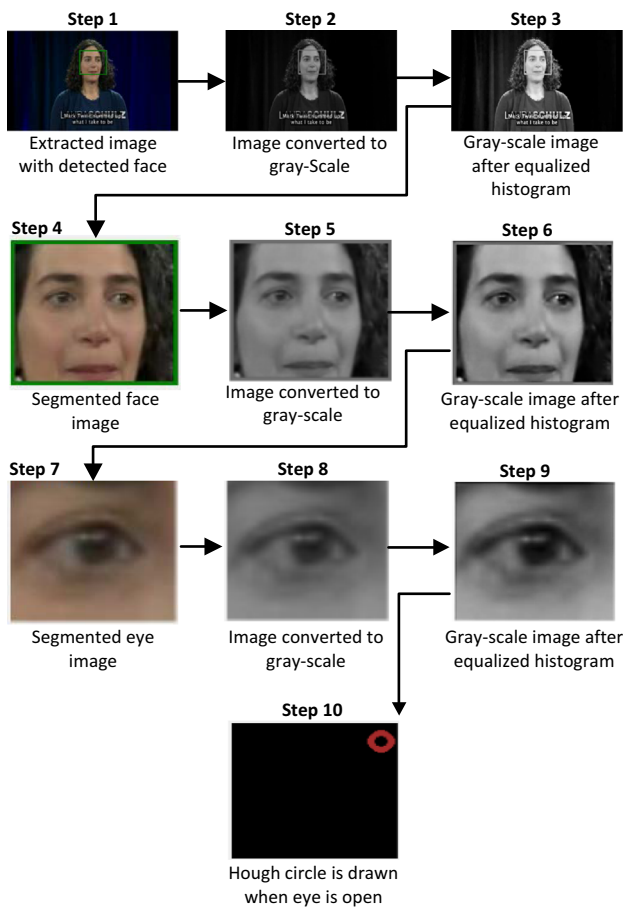
**Algorithm 2** Real-Time Visual Engagement Scoring Algorithm**Input:** A Video Stream from a recorded video lecture file**Output:** A Dataset for Frame based Visual Engagement Score and Engagement Level i.e. "presentation", "p-engaging", "p-eye-contact", "engaging"1. **Begin**2. *Start the video streaming process*3. *Start a Timer*4. *If one second is passed*5. *Then reset Timer and check the number of images processed i.e. PFPS*6. *If Processor is entering into IDLE state AND next is a valid image*7. *Then take the next image from video lecture file*8. *Else exit*9. *Convert the extracted color image to grayscale image*10. *Equalize the histogram of the grayscale image for light and brightness correction*11. *Detect multiscale face image using cascade classifier*12. **For Loop** Rectangle rect in detectFace13. *Draw rectangle around each face image*14. *Copy imgOriginal to faceOnly*15. *Region of Interest of faceOnly = rect*16. *faceDetected = true*17. *Calculate the score for detected face as in Eq. (1)*18. *Append face detection scores to the dataset*19. **End For Loop**20. *Crop and copy each face image*21. *Detect multiscale eye image using cascade classifier*22. **Loop For** Rectangle eyeRect in detecteye23. *Draw rectangle around each eye image*24. *If faceDetected == true Then*25. *Calculate the score for detected eyes as in Eq. (2)*26. *Append eye detection score to the dataset*27. **End If**28. **End For Loop**29. *Crop and Copy each detected Eye image*30. *Calculate face eye ratio as shown in Eq. (3)*31. *Calculate eye radius as shown in Eq. (4)*32. *Circles = Detect HoughCircles (canny Threshold = 100, circle Accumulator Threshold = 120, eyeRadius, eyeRadius × 2, eyeRadius − 10, eyeRadius + 20)*33. *If circles detected > 0 Then*34. *Take EOC score 1 for each opened eye*35. **Else**36. *Take score 0 for each closed eye*37. **End If**38. *Calculate visual engagement as calculated visual attention, shown in Eq. (5)*39. *Append Visual Engagement Score to the dataset*40. *If engagement < 0.49 Then*41. *Engagement-level = "presentation"*42. *Else-if engagement > 0.49 && engagement <= 0.74*43. *Engagement-level = "p-engaging"*44. *Else-if engagement > 0.74 && engagement <= 0.86*45. *Engagement-level = "p-eye-contact"*46. *Else Engagement-level = "engaging"*47. **End If**48. *Append Visual Engagement level to the dataset*49. **Goto** Step 150. **End**



Frame-by-frame analysis is performed, which are extracted from the video stream. The extracted images are converted to the grayscale image; histogram equalization is applied to correct the contrast and brightness of the image [64]. The enhanced image is used to detect the face in the first stage, and in the later stage, the eyes detection is performed. Histogram equalization is an essential part of the image correction due to the dynamic variations of light and environment. This also reduces the chances of

misdetections and false detections. The actual images used during the method are shown in Fig. 2. Image is extracted from the video stream of a video lecture, which is shown in the upper left-hand side. The image is converted to grayscale, and then histogram equalization is applied to make the contrast and brightness better. As the face is found, the face portion is segmented and converted to grayscale and histogram equalization is again applied. The eye is searched in the face image, it is cropped, and Hough circles are drawn on a blank image, which indicates that the eye is open. This process is repeated for each frame extracted from the video stream.

During the study seven videos, lectures were analyzed. The videos were of various lengths. Different frame rates were achieved depending upon the number of detections, i.e., number of faces or eyes [54]. The summary of the processed videos, the number of total frames processed, min, max and the average number of frames processed per second and number of detections are shown in Table 2.



**Fig. 2** Real-time image processing using computer vision approach along with processed frame results

### 4 Results and discussion

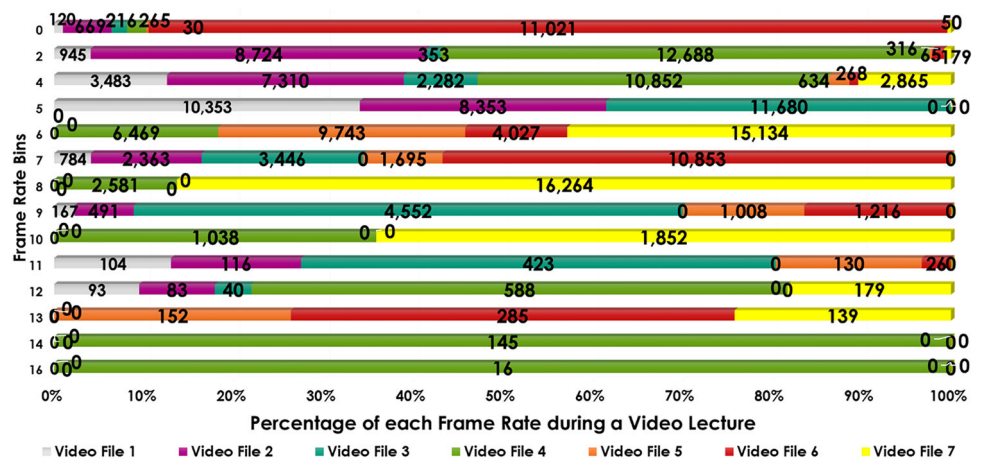
We evaluated the technique using a laptop machine. The specifications of the machine were HP ProBook 4540 s with an Intel Core i5 (3210 M, 2.50 GHz) third-generation processor, HP HD Webcam run at a resolution of  $1366 \times 768$ . The machine was having 5 GB of RAM. Microsoft Visual Studio 2013 along with OpenCV 2.4.10 and EmguCV Windows universal cuda 2.4.10 was used.

Learning analytics (LA) for the e-learning paradigm has many advantages for the students as well as for the teachers. This process is completed by following the steps such that, first of all, the data are captured, information is extracted, the prediction model is applied to the data, steps are taken after taking the recommendations from the expert system, and then the whole process is refined. The implications of such type of study can be further found in the signals project [1]. Video lecture files are processed using

**Table 2** Statistical summary for detections of faces and eyes with performance results

Video files	Resolution	Video length (mm:ss)	Processed frames per second (PFPS)			Average detections		Total frames processed
			Min	Max	Mean	Faces	Eyes	
Video file 1	854 × 480	08:53	0	13	6.03	0.11	0.09	16,047
Video file 2	854 × 480	15:33	0	13	4.73	0.47	0.04	28,107
Video file 3	854 × 480	12:30	0	13	7.09	0.35	0.04	22,990
Video file 4	854 × 480	19:14	0	17	5.81	0.35	0.01	34,640
Video file 5	854 × 480	07:28	0	14	7.22	0.26	0.01	13,455
Video file 6	854 × 480	15:34	0	14	5.28	0.50	0.20	28,010
Video file 7	854 × 480	20:22	0	15	7.49	0.35	0.18	36,660

**Fig. 3** Processed frames per second for each video lecture as per the frame rate



**Table 3** Summary of processed video files using Algorithm 1

PFPS (bin)	Video file 1	Video file 2	Video file 3	Video file 4	Video file 5	Video file 6	Video file 7
16	0	0	0	16	0	0	0
14	0	0	0	145	0	0	0
13	0	0	0	0	152	285	139
12	93	83	40	588	0	0	179
11	104	116	423	0	130	26	0
10	0	0	0	1038	0	0	1852
9	167	491	4552	0	1008	1216	0
8	0	0	0	2581	0	0	16,264
7	784	2363	3446	0	1695	10,853	0
6	0	0	0	6469	9743	4027	15,134
5	10,353	8353	11,680	0	0	0	0
4	3483	7310	2282	10,852	634	268	2865
2	945	8724	353	12,688	65	316	179
0	120	669	216	265	30	11,021	50

the real-time imaging-based assessment model [52], and results are shown in Table 3. The results are extracted using Algorithm 1 as shown in Fig. 3.

Aestheticism of video lecture for the student is the presence of teacher's face. Moreover, the eye contact of student and the teacher in the video plays an important role for the better interaction. The object of aesthetic evaluation could first be the program itself, which is available to the student for interaction, data model, and algorithm. Our model and technique are to enhance the teaching performance and to improve students' quality of experience by designing the software interface and human-computer interaction. The processed frames per second with an average number of frames per second and trend line and error bars are shown in Fig. 4. Summary of the performance results for each processed video files shows that minimum 5 frames per second are process. Thick black dotted line shows the trend of

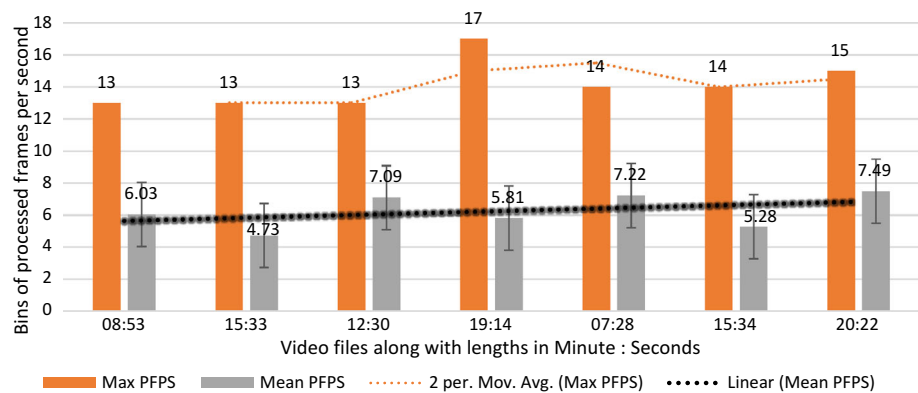
mean value of PFPS, i.e., 5 for each video. Upper dotted line shows mean of maximum processed frames per second, i.e., 14.

Real-time imaging is useful for the assessment and improvement of teaching performance and student experience in e-learning. The next section is the conclusion and future work of the study.

## 5 Conclusion and future work

The conclusions derived in this research article might produce a wider likeness on the student and teacher in the acceptance and all-time use of multimedia technologies. This is expected that the developments made during this study will motivate such future research efforts. Cross-sectional research designs are also needed to extend the validity of the findings. This type of multimedia real-time

**Fig. 4** Frames processed per second with max, mean, and error bars. Starting from the left is video file 1, and rightmost is the video file 7, respectively



analysis creates new opportunities to rethink about interaction design and instructional design. Due to an interactive system, the video constructs are designated within the computer by a set of rules [53]. Analyzing and recording the interactions in real time, the rules are applied in the system as it learns. The system can adapt its behavior in the light of its experience with students interacting with the video lectures. A video presentation that is focused on a specific topic can be more helpful for the students [23]. It can be used for the preparation of class for the deep exposure of the subject. Visual engagement measure earlier to attend the lecture in formal or informal teaching is more useful. This innovative method offers the student to prepare before time. It will also improve the student tendency toward the self-learning. It provides the opportunity to the learners to keep up their learning pace. Reusability of the learning material is another benefit of this method. Consolidating diverse information sources to make claims about student learning is well established and loaded with challenges in assessment.

This research has the following future directions. First, the analysis is performed on selected video lectures, and qualitative analysis is performed. Second, the selected number of students' data are analyzed qualitatively. Future research may be considered to collect the data, perform the quantitative analysis, and investigate the relationships between the teacher's engagement and student's attentiveness. The relationships among the possible states of students' visual attentiveness and teachers' visual engagement are required to be investigated by using quantitative research analysis methodology.

## References

- Gómez-Aguilar, D.A., et al.: Tap into visual analysis of customization of grouping of activities in eLearning. *Comput. Hum. Behav.* **47**, 60–67 (2015)
- Farhan, M.: A methodology to enrich student-teacher interaction in eLearning. In: *The 30th ACM/SIGAPP Symposium on Applied Computing* (2015)
- Zain-UI-Abidin, et al.: Analysis of video lecture's images for the compression formats support for the students in eLearning paradigm. *Sci. Int.* **27**(2), 1171–1176 (2015)
- Mörwald, T., et al.: Advances in real-time object tracking. *J. Real Time Image Proc.* **10**(4), 683–697 (2015)
- Farhan, M., et al.: Automated reply to students' queries in e-Learning environment using Web-BOT. In: *Artificial Intelligence (MICAI), 2012 11th Mexican International Conference on.* IEEE (2012)
- Wang, Y., et al.: Image acquisition for digital camera based on emgu CV. *Dianzi Keji Electron. Sci. Technol.* **25**(4), 31–32 (2012)
- Ponomarenko, N., et al.: Analysis of HVS-metrics' properties using color image database TID2013. In: *International Conference on Advanced Concepts for Intelligent Vision Systems.* Springer (2015)
- Khan, F.M.A., Masood, M.: The effectiveness of an interactive multimedia courseware with cooperative mastery approach in enhancing higher order thinking skills in learning cellular respiration. *Procedia Soc. Behav. Sci.* **176**, 977–984 (2015)
- Kriegeskorte, N.: Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015)
- Henriksen, S., Read, J.C.: Visual perception: a novel difference channel in binocular vision. *Curr. Biol.* **26**(12), R500–R503 (2016)
- Liu, J., et al.: Human visual system consistent quality assessment for remote sensing image fusion. *ISPRS J. Photogramm. Remote Sens.* **105**, 79–90 (2015)
- Rogowitz, B.E., Pappas T.N., de Ridder H.: Human vision and electronic imaging. *Hum. Vis. Electron. Imaging* **9394**, 586 (2015)
- Dickinson, S., Pizlo, Z.: *Shape perception in human and computer vision.* Springer (2015)
- Zarándy, Á., et al.: A real-time multi-camera vision system for UAV collision warning and navigation. *J. Real Time Image Process.* 1–16 (2014)
- Muhammad, M.I., et al.: Automated Web-bot implementation using machine learning techniques in eLearning paradigm. *J. Appl. Environ. Biol. Sci.* **4**(7S), 90–98 (2014)
- Fain, E., Smith, J.: Video lecture capture pedagogy context: does it matter and does it deliver?. In: *Society for Information Technology and Teacher Education International Conference* (2014)
- Aydin, M.K., Gürol, M., Vanderlinde, R.: Evaluating ICT integration in Turkish K-12 schools through teachers' views. *Eurasia J. Math. Sci. Technol. Educ.* **12**(4), 747–766 (2016)
- Seguin, R.A., Morgan, E.H., Connor, L.M., Garner, J.A., King, A.C., Sheats, J.L., et al.: Rural food and physical activity assessment using an electronic tablet-based application, New

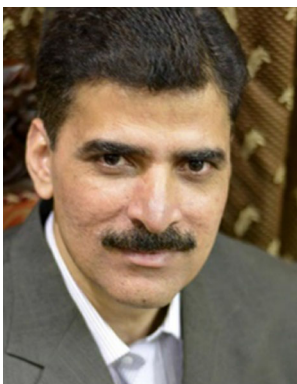
- York, 2013–2014. *Prev Chronic Dis.* **12**, 150147 (2015). doi:[10.5888/pcd12.150147](https://doi.org/10.5888/pcd12.150147)
19. Kizilcec, R.F., Papadopoulos, K., Sritanyaratana, L.: Showing face in video instruction: effects on information retention, visual attention, and affect. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014. ACM
  20. Wasserman, S., et al.: Assessing attention: relationship between circular visual-tracking and spatial span, and implication for electronic assessment. *Arch. Phys. Med. Rehabil.* **96**(12), e15 (2015)
  21. Selvakumar, K., et al.: Real-time vision based driver drowsiness detection using partial least squares analysis. *J. Signal Process. Syst.* 1–12 (2015)
  22. Plyer, A., Le Besnerais, G., Champagnat, F.: Massively parallel Lucas Kanade optical flow for real-time video processing applications. *J. Real Time Image Process.* **11**(4), 713–730 (2016)
  23. Saponara, S., et al.: Algorithmic and architectural design for real-time and power-efficient Retinex image/video processing. *J. Real Time Image Process.* **1**(4), 267–283 (2007)
  24. Sánchez, S., et al.: Real-time implementation of remotely sensed hyperspectral image unmixing on GPUs. *J. Real Time Image Process.* **10**(3), 469–483 (2015)
  25. Gomes, S.L., Rebouças, E.D.S., Neto, E.C., Papa, J.P., de Albuquerque, V.H., Rebouças Filho, P.P., Tavares, J.M.R.: Embedded real-time speed limit sign recognition using image processing and machine learning techniques. *Neural Comput. Appl.* 1–12 (2016)
  26. Salah, A.A., Alpaydin, E., Akarun, L.: A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(3), 420–425 (2002)
  27. Li, H., Ngan, K.N.: Saliency model-based face segmentation and tracking in head-and-shoulder video sequences. *J. Vis. Commun. Image Represent.* **19**(5), 320–333 (2008)
  28. Kienzle, W., et al.: Center-surround patterns emerge as optimal predictors for human saccade targets. *J. vis.* **9**(5), 7 (2009)
  29. Macfadyen, L.P., Dawson, S.: Mining LMS data to develop an “early warning system” for educators: a proof of concept. *Comput. Educ.* **54**(2), 588–599 (2010)
  30. Mathe, S., Sminchisescu, C.: Dynamic eye movement datasets and learnt saliency models for visual action recognition. In: *Computer Vision—ECCV 2012*, pp. 842–856. Springer (2012)
  31. Han, J.H., Finkelstein, A.: Understanding the effects of professors’ pedagogical development with Clicker Assessment and Feedback technologies and the impact on students’ engagement and learning in higher education. *Comput. Educ.* **65**, 64–76 (2013)
  32. Thoms, B., Eryilmaz, E.: How media choice affects learner interactions in distance learning classes. *Comput. Educ.* **75**, 112–126 (2014)
  33. Ba, J., Mnih, V., Kavukcuoglu, K.: Multiple object recognition with visual attention. arXiv preprint: [arXiv:1412.7755](https://arxiv.org/abs/1412.7755), 2014
  34. Gittins, H., et al.: Multilevel engagement: theory and practice integration. In: *Green Innovation and Future Technology: Engaging Regional SMEs in the Green Economy*, pp. 41–56. Springer (2015)
  35. Yamaguchi, Y.: Image synthesis in consideration of a human visual system. In: *The Visual Language of Technique*, pp. 65–79. Springer (2015)
  36. Li, M.S., et al.: Performance improvement of multi-view video coding based on geometric prediction and human visual system. *Int. J. Imaging Syst. Technol.* **25**(1), 41–49 (2015)
  37. Iqbal, M.M., Farhan, M., Saleem, Y.: Personalization with agent in eLearning framework. *Sci. Int.* **27**(1), 2541–2543 (2014)
  38. Liu, K.-Y., et al.: A new parallel particle filter face tracking method based on heterogeneous system. *J. Real Time Image Process.* **7**(3), 153–163 (2012)
  39. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. *Nature* **381**(6582), 520–522 (1996)
  40. Han, J., et al.: An infrared small target detecting algorithm based on human visual system. *IEEE Geosci. Remote Sens. Lett.* **13**(3), 452–456 (2016)
  41. Park, M.-C., Mun, S.: Overview of measurement methods for factors affecting the human visual system in 3d displays. *J. Disp. Technol.* **11**(11), 877–888 (2015)
  42. Ahmad, J., et al.: Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems. *J. Real Time Image Process.* 1–17 (2015). doi:[10.1007/s11554-015-0536-0](https://doi.org/10.1007/s11554-015-0536-0)
  43. Chen, B.-W., et al.: QoE-enabled big video streaming for large-scale heterogeneous clients and networks in smart cities. *IEEE Access* **4**, 97–107 (2016)
  44. Meng, H., et al.: Real-time human action recognition on an embedded, reconfigurable video processing architecture. *J. Real Time Image Process.* **3**(3), 163–176 (2008)
  45. Kapela, R., McGuinness, K., O’Connor, N.E.: Real-time field sports scene classification using colour and frequency space decompositions. *J. Real Time Image Process.* 1–13 (2014)
  46. Paul, A., et al.: Video search and indexing with reinforcement agent for interactive multimedia services. *ACM Trans. Embed. Comput. Syst. (TECS)* **12**(2), 25 (2013)
  47. Razaq Malik, K., et al.: Data security and privacy in cloud computing: threat level indications. *Sci. Int.* **26**(5), 1991–1996 (2014)
  48. Jabbar, S., et al.: Trust model at service layer of cloud computing for educational institutes. *J. Supercomput.* **72**(1), 58–83 (2016)
  49. Naseer, M.K., Jabbar, S., Zafar, I.: A novel trust model for selection of Cloud Service Provider. In: *Computer Applications and Research (WSCAR), 2014 World Symposium on.* (2014)
  50. Malik, K.R., Ahmad, T., Farhan, M., Aslam, M., Jabbar, S., Khalid, S., Kim, M.: Big-data: transformation from heterogeneous data to semantically-enriched simplified data. *Multimed. Tools Appl.* **75**(20), 12727–12747 (2015). doi:[10.1007/s11042-015-2918-5](https://doi.org/10.1007/s11042-015-2918-5)
  51. Khalid, S., et al.: Accurate and efficient shape matching approach using vocabularies of multi-feature space representations. *J. Real Time Image Process.* 1–17 (2015). doi:[10.1007/s11554-015-0545-z](https://doi.org/10.1007/s11554-015-0545-z)
  52. Chen, C., Liu, K., Kehtarnavaz, N.: Real-time human action recognition based on depth motion maps. *J. Real Time Image Process.* **12**(1), 155–163 (2016)
  53. Varfolomeiev, A., Lysenko, O.: An improved algorithm of median flow for visual object tracking and its implementation on ARM platform. *J. Real Time Image Process.* **11**(3), 527–534 (2016)
  54. Joseph Raj, A.N., Staunton, R.C.: Video-rate calculation of depth from defocus on a FPGA. *J. Real Time Image Process.* 1–12 (2014). doi:[10.1007/s11554-014-0480-4](https://doi.org/10.1007/s11554-014-0480-4)
  55. Rautaray, S.S., Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. *Artif. Intell. Rev.* **43**(1), 1–54 (2015)
  56. Farhan, M., et al.: Extracting parameters from e-feedback text data based on keyword based parsing in elearning environment. *Sci. Int.* **26**(3), 1395–1402 (2014)
  57. Poiesi, F., Cavallaro, A.: Predicting and recognizing human interactions in public spaces. *J. Real Time Image Process.* **10**(4), 785–803 (2015)
  58. Ohlsson, N., Ståhl, M.: Model-Based Approach to Computer Vision and Automatic Control using Matlab Simulink for an Autonomous Indoor Multirotor System. Chalmers University of Technology, Gothenburg (2013)
  59. Kaehler, A., Bradski, G.: *Learning OpenCV*. O’Reilly Media, Inc, Sebastopol (2014)

60. Anders, J., et al.: A hardware/software prototyping system for driving assistance investigations. *J. Real Time Image Process.* **11**(3), 559–569 (2016)
61. Li, C., et al.: Design and evaluation of a parallel and optimized light–tissue interaction-based method for fast skin lesion assessment. *J. Real Time Image Process.*, 1–14 (2015). doi:[10.1007/s11554-015-0494-6](https://doi.org/10.1007/s11554-015-0494-6)
62. Shi, S.: *Emgu CV Essentials*. Packt Publishing Ltd, Birmingham (2013)
63. Razzaq Malik, K., et al.: Multiagent Semantical Annotation Enhancement Model for IoT-Based Energy-Aware Data. *Int. J. Distrib. Sens. Netw.* **12**(6), 9103265 (2016). doi:[10.1155/2016/9103265](https://doi.org/10.1155/2016/9103265)
64. Chambers, O., Milenkovic, J., Tasic, J.F.: A pre-processing scheme for real-time registration of dynamic contrast-enhanced magnetic resonance images. *J. Real Time Image Process.* 1–10 (2014). doi:[10.1007/s11554-014-0468-0](https://doi.org/10.1007/s11554-014-0468-0)
65. Cano, A., et al.: Parallelization strategies for markerless human motion capture. *J. Real Time Image Process.* 1–15 (2014). doi:[10.1007/s11554-014-0467-1](https://doi.org/10.1007/s11554-014-0467-1)
66. Chorianopoulos, K., Giannakos, M.N.: Usability design for video lectures. In: *Proceedings of the 11th European Conference on Interactive TV and Video*. 2013. ACM



**Muhammad Farhan** is Assistant Professor at COMSATS Institute of Information Technology, Sahiwal Campus, Pakistan. He is a Ph.D. Scholar at Department of Computer Sciences and Engineering in *University of Engineering and Technology (UET), Lahore, Pakistan*. He has been working as instructor computer science in *Virtual University of Pakistan (VU)*. He obtained MSCS from *University of Management and Technology (UMT), Pakistan*.

He has received BSCS from *Virtual University of Pakistan (VU)*. Currently, he has 11+ years of teaching experience. His interests include digital image processing, computer vision, e-learning and computer programming. He has published about 30 research articles.



**Muhammad Aslam** is an Associate Professor at University of Engineering and Technology, Lahore, Pakistan. He has six years of experience in software architecture design, team leading, team building, and software project, fourteen years of experience in research and development as well as teaching at postgraduate level (supervising Ph.D. and M.Sc. theses). Completed his Ph.D. in Computer Sciences (CGPA: 8.9/10) session 2001–2005, Cin-

vestav-IPN, Mexico (Cultural Exchange Scholarships between Pakistan and Mexico). His research and teaching interests are knowledge-based systems, expert systems, intelligent agents, human–computer interaction, and computer-supported cooperative work, cooperative writing and authoring, communication, coordination, awareness, cooperative learning, and modern operating systems. He won award

of merit scholarship from Board of Intermediate and Secondary Education, Sargodha Division, Pakistan (1984–1986). He is also awarded merit scholarship during B.Sc. Agricultural Engineering from Faculty of Agricultural Engineering, University of Agricultural, Faisalabad, Pakistan (1987–1991). He was also awarded Cultural Exchange Scholarship between Pakistan and Mexico for higher studies (2000–2004) for Ph.D. studies. He has a number of publications.



**Sohail Jabbar** is a Post-Doctorate Researcher at Network Lab, Kyungpook National University, Daegu, South Korea, and Assistant Professor with the Department of Computer Science, COMSATS Institute of Information Technology (CIIT), Sahiwal. He is also the Head of Networks and Communications Research Group at CIIT, Sahiwal. He got MS Leading to Ph.D. Scholarship from Higher Education Commission, Islamabad, Pak-

istan, 2007–2012, won Magna Cum Laude honor (Silver Medalist) in MS (T&N), 2009, achieved Best Student Researcher Award of the Year, 2014, from Bahria University, and Research Productivity Award from COMSATS Institute of Information Technology in 2014 and 2015. His research work is published in various renowned journals of Springer, Elsevier, MDPI, Old City Publication, and Hindawi and conference proceedings of IEEE and ACM. He has also been the reviewer for leading journals (e.g., ACM TOSN, JoS, MTAP, AHSWN, ATECS) and conferences (ACM SAC 2016, ICACT 2016, ACM SAC 2015). His research interests are Internet of things, wireless sensor networks, and big data analytics.



**Shehzad Khalid** is a Professor and Head of Department at Department of Computer Engineering, Bahria University, Pakistan. He is a qualified academician and researcher with more than 60 international publications in various renowned journals and conference proceedings. Dr. Shehzad has graduated from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2000. He received the M.Sc. degree from National

University of Science and Technology, Pakistan, in 2003 and the Ph.D. degree from the University of Manchester, UK, in 2009. Dr. Shehzad is the Head of Computer Vision and Pattern Recognition (CVPR) research group which is a vibrant research group undertaking various funded research projects. His areas of research include, but are not limited to, shape analysis and recognition, motion-based data mining and behavior recognition, medical image analysis, ECG analysis for disease detection, biometrics using fingerprints, vessels patterns of hands/retina of eyes, ECG, Urdu stemmer development, short and long multi-lingual text mining, Urdu OCR. Dr. Shehzad has been the reviewer for various leading ISI indexed journals. He has received Best Researcher Award for the year 2014 from Bahria University. He has also been awarded Letter of Appreciation for Outstanding research contribution in year 2013 and outstanding performance award for the academic year 2013–2014.



**Mucheol Kim** is an Assistant Professor in a division of media software at Sungkyul University in Korea. His research interests include information retrieval, Web technology, social networks, and wireless sensor networks. He was a Senior Researcher in Korea Institute of Science and Technology Information (KISTI), Daejeon, Korea. He received the BS, MS, Ph.D. degrees from the School of Computer Science and Engineering at Chung-Ang University, Seoul, Korea, in 2005, 2007 and 2012, respectively.

Seoul, Korea, in 2005, 2007 and 2012, respectively.