An Interactive Assessment Framework for Visual Engagement: Statistical Analysis of a TEDx Video

Muhammad Farhan
Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, PAKISTAN
Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal, PAKISTAN

Muhammad Aslam
Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, PAKISTAN

ABSTRACT

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. 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, the number of eyes, the status of eyes and the engagement score along with nominal values of engagement level. This data is transformed into time-based data items by using the attribute number of frames processed per second (PFPS). A case study for the assessment of TEDx video (length 8 minutes and 53 seconds) is included to validate the results and to extract statistical information from the dataset. Frames in the video are 16047 and they are transformed into 2675 keyframes. Machine learning classifiers are applied for the analysis of the dataset. The findings of this analysis help the presenter and the student to measure the quality of the visual content of the videos without actually going through it.

Keywords: interactive assessment, electronic learning, visual engagement, visual attention, video lecture, MOOCs

INTRODUCTION

Online learning video lectures are growing in number and variety, but usefulness and effectiveness in terms of learning and usability are not well understood. The assessment of these video lectures is necessary for better pedagogy and instructional design (Meade, Ball, & Brandsness, 2012). A significant research literature is found within the each domain of video learning and usability, but their unified design is not well accepting (Chorianopoulos & Giannakos, 2013). Above all, very less research is found on strategies for effective video lecture design, for instance, navigation support through the video, statistical analysis and the manifestation of humans in the video. According to (Chorianopoulos & Giannakos, 2013), the students take advantage from well-structured learning material, but it is feasible to manually

Correspondence: Muhammad Farhan, Department of Computer Science and Engineering, UET, Lahore, Pakistan.
✉ farhansajid@gmail.com

State of the literature

- The visual attention of the students can be measured using the biological measures
- The user of the system can be tracked by using special purpose devices like eye tracker to check the attentiveness by analyzing the eye movement data
- Intelligent tutoring systems are used to analyze the learning requirements and deficiencies of the students to predict and support them

Contribution of this paper to the literature

- Framework for video analysis to assess the effectiveness and visual engagement of a video lecture
- This paper presents a novel approach to measure the visual engagement of a video lecture by determining the activity of the teacher
- Computer Software and Machine Learning techniques used to facilitate the learning process and make it effective for students and teachers

edit the video for most teachers and institutes. A statistical analysis method is drawn from the research literature on video interaction and educational technologies for design principles. Moreover, a comprehensive approach to the design of the content and effective video lecture structure is provided. It is also suggested that teachers and institutes should spend further energy in such video lecture systems that take care of a unified methodology to controlling of video lectures, their editing, and sharing (Chorianopoulos & Giannakos, 2013).

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 towards lectures (Fain & Smith, 2014). Electronic learning (eLearning) refers to the type of learning where the student has to learn with own pace and no teacher is there to monitor (Aydin, Gürol, & Vanderlinde, 2016; Seguin et al., 2015). 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 (Kizilcec, Papadopoulos, & Sritanyaratana, 2014). During this research, a tool "Video Lecture Analysis" is developed. A video of standard quality usually has 16 to 20 frames per second but this can be more provided that 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 visual engagement of the teacher e.g. presentation, partially engaging, partial eye contact with the camera, and fully engaging (Wasserman et al., 2015).

Electronic Learning (E-learning) is becoming a principal way of learning in the higher education industry. In this medium of instruction delivery video lectures are the main teaching tool. Various video lecture delivery designs and structures are used to deliver course content among students (Fain & Smith, 2014; Farhan, Iqbal, & Naeem, 2015). It is also noteworthy that there are a very few numbers of the manuscripts found that examine how diverse style of

video lecture delivery formats affect the way a student feels engaged with the video lecture (Gilardi, Holroyd, Newbury, & Watten, 2015; Lyons, Reysen, & Pierce, 2012).

Another style of teaching allows engaging both sides to get what they want. In this type of model, the video lectures are played back by students on their individually selected time and lectures are pre-recorded. In principle, educational experience and the whole classroom environment currently being provided and that can be acquired online. Each student completes the assignment outside of class. This does not mean that the institutions will be out of the educational industry. Rather, active learning activities take place within the classroom while each student can complete the assignment and lecture outside of class. Therefore, it may no longer be a suggestion, with the assignment and active learning on the one hand while lectures on the other (Chen & Wu, 2015).

According to the (Gilardi et al., 2015) Massive Open Online Course (MOOC) is another style introduced for the online teaching. In fact, the student enrollments are upsetting: UDACITYi, one of the three principle performers in the MOOC collaborative, has enlisted 300,000 learners in a course called "Introduction to Computer Science". This is a record-breaking number for an MOOC-based course. Somewhere in the range of 20 million students in more than 200 countries have selected in an MOOC, and the pattern is rising strongly (Sinha, Jermann, Li, & Dillenbourg, 2014). Video lectures are a more efficient way of delivering content, compared to live lectures (McCammon & Parker, 2014). Many engineering graduates does not have the skill to solve the scenario-based problems in the workforce; but, these identical students are proficient in solving the problems in their textbook which involve a little effort than to determine the appropriate method for the application (Bishop, 2013; Nabiyev, Çakiroğlu, Karal, Erümit, & Çebi, 2016).

MATERIAL AND METHODS

The TEDx Program is planned to support groups, societies and those to initiate the discussion, exchange of ideas via TED-like practices (Nicolle, Britton, Janakiram, & Robichaud, 2014). In these proceedings, a transmission and recording of TED Talks videos initiate the detailed discussion (Cettolo, Niehues, Stüker, Bentivogli, & Federico, 2013). A case study is conducted by selecting a TEDx video for the interactive assessment and statistical analysis (Fostier, Patel, Clarke, & Prokop, 2015). The study is divided into the following stages for statistical and behavioral analysis (Liu, Chang, Huang, & Chang, 2016) of the video lectures:

Data Collection Module

The selected TEDx video of Alison Killingii is loaded into the software tool i.e. data collection module. This module starts extracting the frames from the TEDx video. The extracted frames are used for the data collection about the presenters. First of all, faces are identifying by searching multi-scale face images. The frames are extracted at real-time. The frequency of the frames is recorded automatically. Each detected face is cropped from the extracted image. Each cropped image contains only faces portion. Then each face image is

searched and checked for the eyes. The eyes are extracted and cropped from the face image. The count is recorded for the further processing. The number of the detected eyes are now checked for the either open or closed. Engagement score is calculated based on the detected number of faces, eyes, and status of eyes. Score data with following attributes is stored: a number of faces in the image, the actual number of eyes i.e. double the number faces detected in the last image, detected the number of eyes by the tool, opened a number of eyes, and the closed a number of eyes. Eyes detection error is calculated as shown in Eq. (1):

$$eE = dE - (oE + cE) \tag{1}$$

Eyes detection error is calculated as shown in Eq. (2):

$$EDE = aE - dE \tag{2}$$

Engagement is calculated as shown in Eq. (3):

$$engagement = \frac{pA \times nF \times fE}{aE} + \frac{oE}{aE \times pE} \tag{3}$$

where

nF stands for number of faces in each frame of the video lecture

aE stands for actual eyes in each frame in the eye image

oE stands for opened eyes in each frame in the face image

EDE is eyes detection error

eE is eyes error

cE is the number of closed eyes

dE is the number of detected eye

pA is the probability of attentiveness, which is taken 0.5 because if the video is being processed that means that teacher is present. Hence, the initial engagement score is considered 0.5

pE is the probability of the eye. If the eye is open that means it is an actual eye so the value is taken 0.5, e.g. if the oE is 1 and aE is 1 then the score will be taken 0.5. If oE is 2 and aE is 2 then the result of this part of equation will be 1

fE is the face to eye ratio, which is always 2 because a normal person has two eyes on a face.

Visual engagement is related to teacher's face and eye contact. In a video lecture, if the teacher is only presenting slides all the time, the student will most probably, lose attention. The face of the teacher in the video lecture makes the student engaging. The number of times teacher's face is detected, which is used to calculate the score of visual engagement, which means the video lecture, will be more engaging for the student. Actual eyes are directly related to a number of faces and this value is calculated by taking the double of the number of faces

in an image. If opened eyes of the teacher are detected, then the students will most probably be engaged in the video lecture attentively. Closed eyes help to calculate the teacher's visual engagement, which means if the teacher has closed the eyes during the recording of video lecture then the student will not be much engaged. The process of detecting faces and eyes is done by using computer vision and machine learning classifiers. Eye detection error is the part of the dataset. This error helps to calculate the engagement score correctly.

Processed Frames per Second (PFPS) are calculated by checking the number of images processed in a second. Each time the counter is reset and counted the number of images in a second. The whole process is implemented and a tool is developed as shown in **Figure 1**. The green rectangle around the face of the presenter is drawn because the face is detected. Cropped image is shown in in lower box (on the left). The next box is showing the eye detected. The last box with a black background and two red circles shows that the eyes are open. On the right side, PFPS and series of events detected by the tool are displayed.
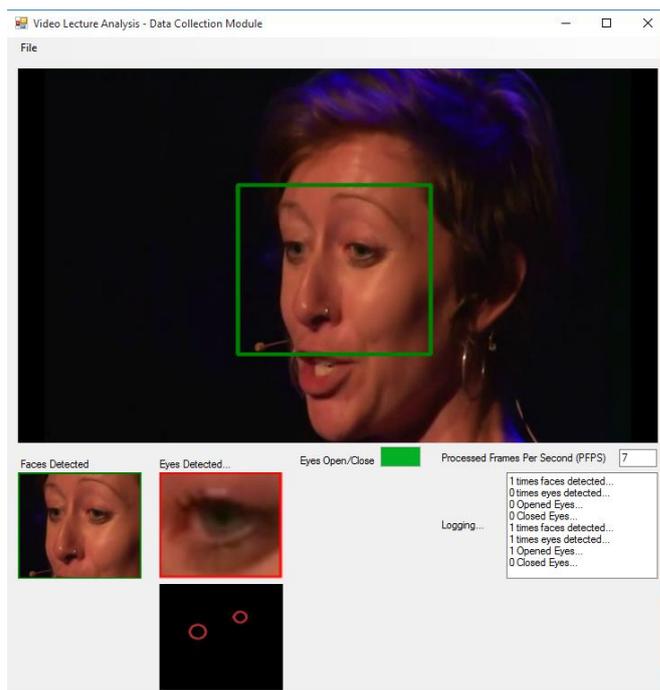
Figure 1. Data collection module for video lecture analysis

Transformation from Frame-based data to Time-based Data

Arithmetic mean from the frame-based data is calculated for the data items for PFPS attribute and mean frames per second (MPFPS) as shown in Eq. (4):

$$MPFPS = \frac{1}{n}\sum_{i=1}^{n} pfps_i \tag{4}$$

The calculated MPFPS is used to define the row differences and the time based total number of rows (TbR) are calculated as shown in Eq. (5):

$$TbR = \frac{\sum_{i=1}^{n} FbR_i}{MPFPS} \tag{5}$$

where *FbR* stands for frame based total number of rows.

Each attribute is averaged and converted to time-based frames except the nominal attributes. Keyframes can be defined as the non-duplicate tuples of the dataset. Keyframes are extracted from the time-based frame dataset. The value of MPFPS is calculated by using Eq. (4) and the calculated value is 6. Time based total number of rows (TbR) are calculated by using Eq. (5) and the calculated value is 2675. Analytical results of the video lecture analysis are then more reduced by eliminating the duplicate tuples and only key tuples are left, which are 54. Five minutes and 53 seconds video has been reduced to 54 unique tuples. The distribution of the key frames is shown in **Figure 2(a & b)**.
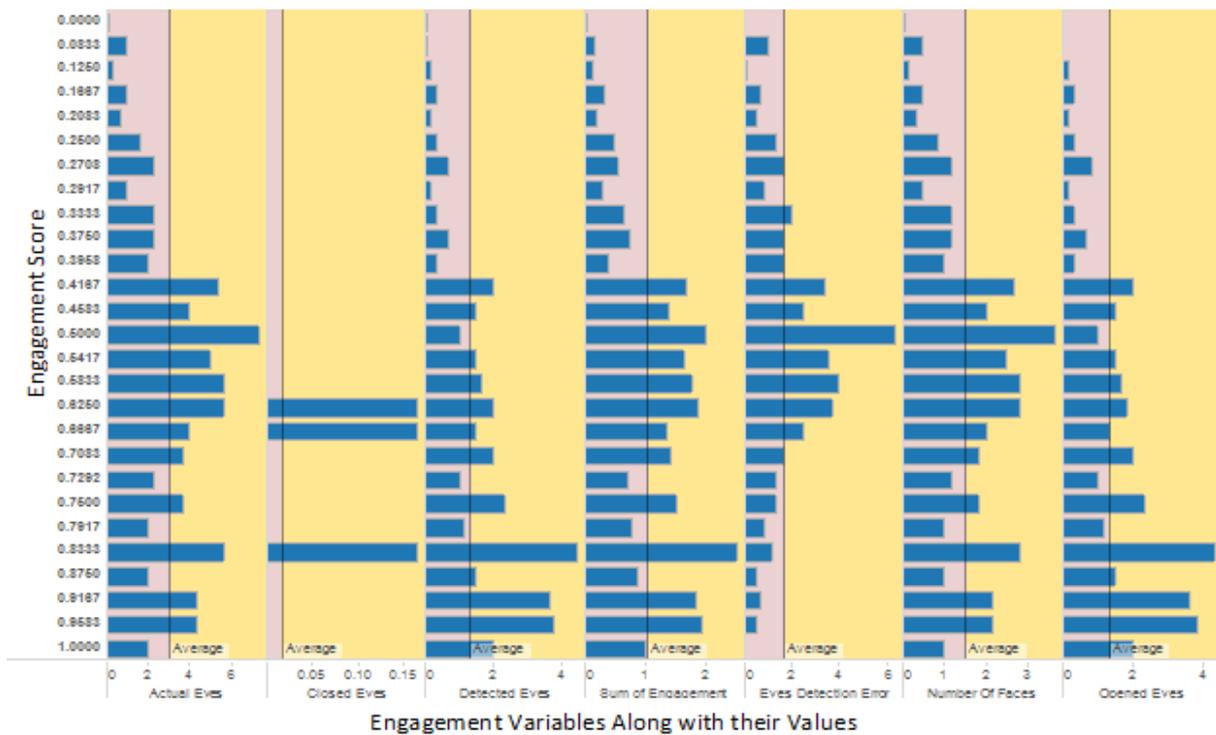
Figure 2. (a). Sum of actual eyes, sum of closed eyes, sum of detected eyes, sum of engagement, sum of number of faces and sum of opened eyes for each engagement
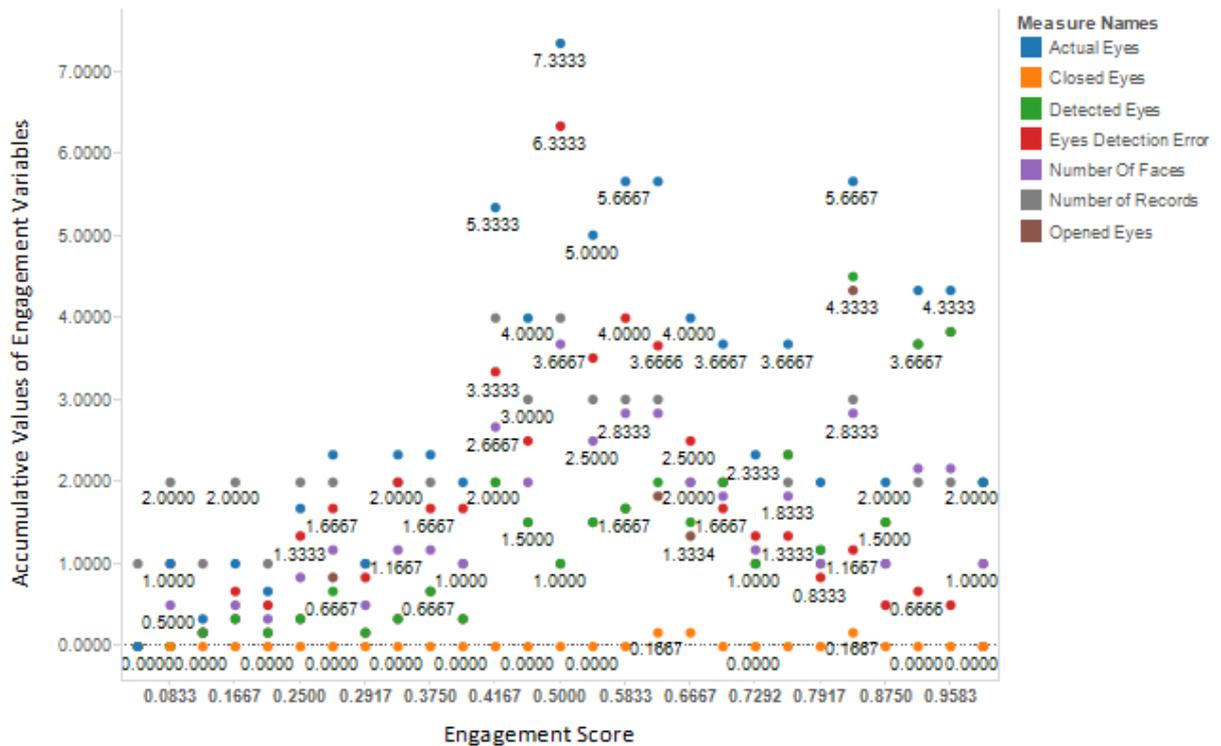
Figure 2. (b). Distribution of data for the sum of all variables

Video lecture has many instructive abundant potentials to make learning and teaching situation better if it is adopted with a proper way. Therefore, the thing is that the usage of technology to modify and update the educational-subjects and make novel prospects for the learning while the problem is not whether teachers would blend technology in their pedagogy. Video lecture development involves detailed implementation scheduling. The understanding philosophical perspective of pedagogy is a criterion for effective learning content. According to (Chorianopoulos & Giannakos, 2013), the constant learning through the use of video lectures has the benefit that knowledge might get direct to the students. Rehearing the video lectures and taking additional information from it will also seem to motivate medium's usability. Time-based keyframes attribute values and their delta difference is shown in **Figure 3.**
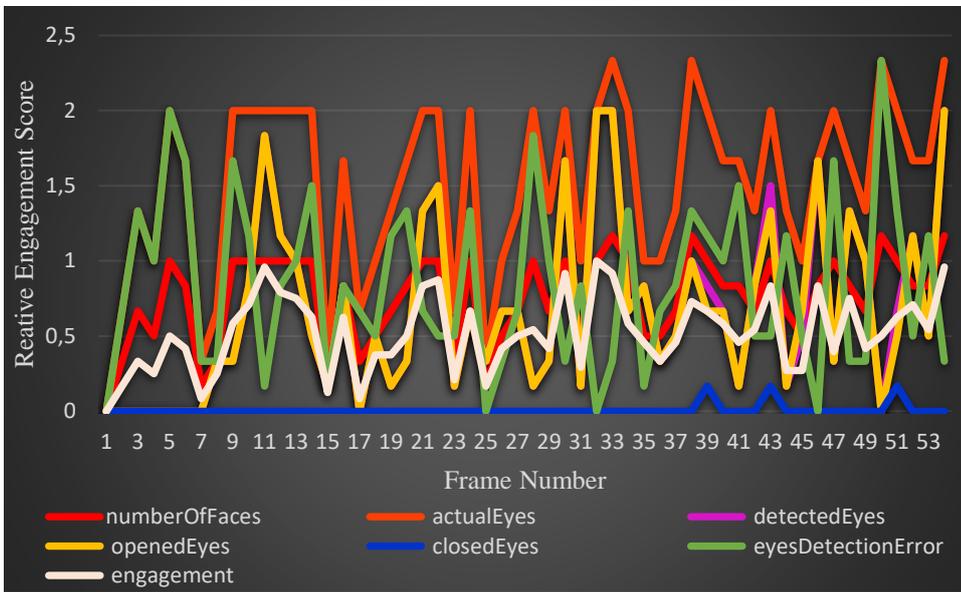
Figure 3. (a). Time based non-repeating attribute-values for the number of faces, actual eyes, detected eyes, opened eyes, closed eyes, eyes detection error which is calculated by subtracting the detected eyes from actual eyes and engagement level.
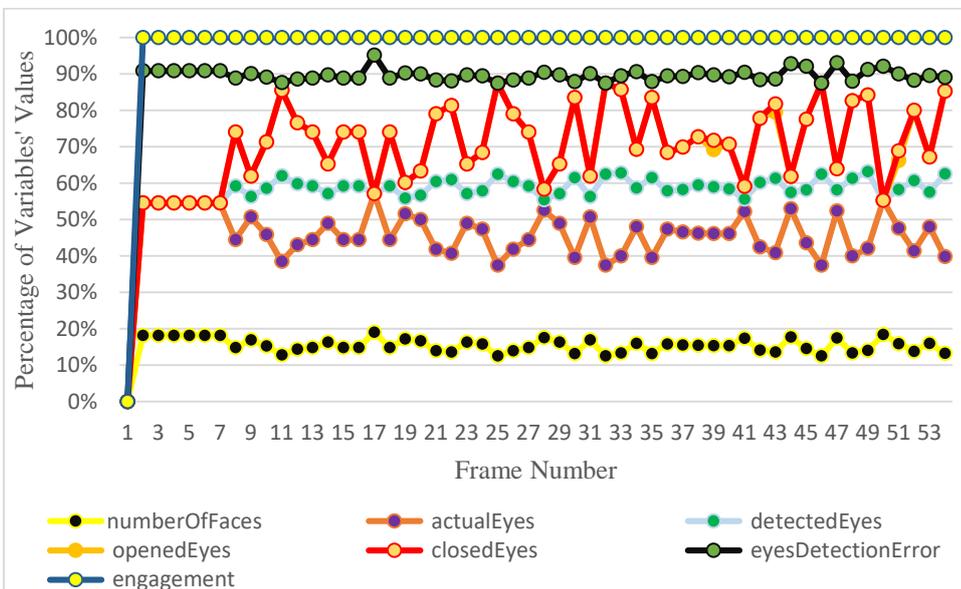
Figure 3. (b). Time-based delta-difference between the variables calculated during the analysis.

1114
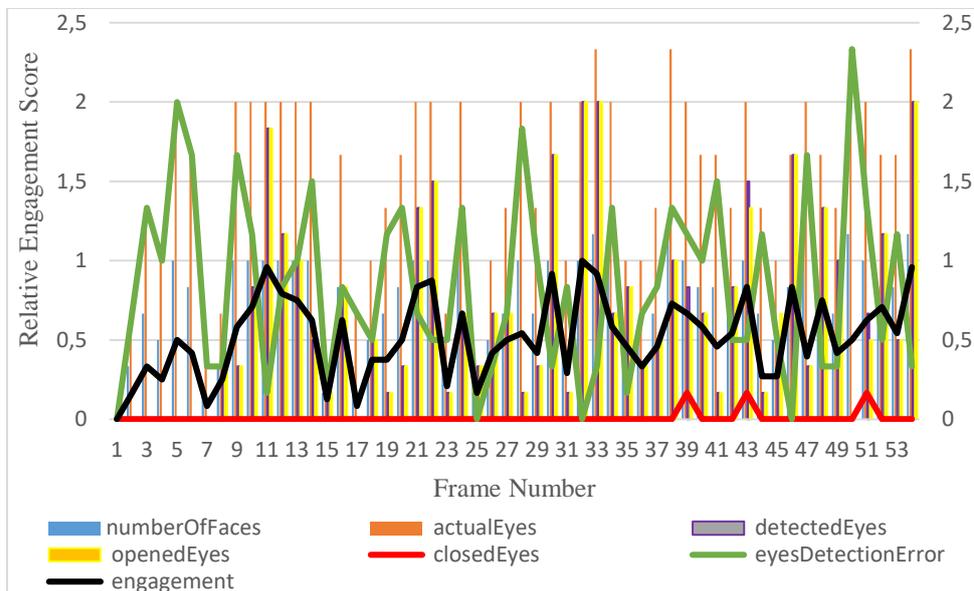
Figure 3. (c). Time-based progressive change for the engagement, other influencing variables are also shown.

The flipped classroom model appears to be a likely technique to develop student learning and understanding of concepts, it is unlucky that the few proof of its success (Yirci, Karakose, Uygun, & Ozdemir, 2016). The latest study of the flipped classroom is initiated and numerous research articles have been published, nearly none of the research findings described results of this type of arrangements on objective assessment procedures or described in the articles used controlled research designs (Chorianopoulos & Giannakos, 2013). So, it is the prime factor for the motivation for this research (Chen & Wu, 2015).

Inferential statistics is a type of statistic to deduce from the given information that what the overall population represents. It is used to make decisions of the likelihood that an experimental variance among groups is a reliable result or the one, which may have occurred by chance in the study. Frame-based data is used for rules creation and time-based data is used to test those rules. K-mean clustering is used to predict the cluster for time-based data and four clusters are formed as shown in **Table 1**. The clusters show corresponding engagement levels e.g. cluster 0 is created around presentation. Cluster 1 is created around engaging. Cluster 2 is created around p-eye-contact, which means that the teacher has partial eye contact with the camera. Cluster 3 is created around p-engaging, which means that the video lecture is partially engaging.

Table 1. Clustering is performed using K-mean clustering and Euclidean distance method, which is used to calculate the centers of the clusters.

Attributes	Full Data (16046) 100%	Cluster 0 (14244) ~89%	Cluster 1 (278) ~2%	Cluster 2 (942) ~6%	Cluster 3 (582) ~4%
Number of Faces	0.1130	0	1.0108	1.0011	1.0137
actual Eyes	0.2261	0	2.0216	2.0021	2.0275
detected Eyes	0.0941	0	2.0144	1.0011	0.0120
opened Eyes	0.0940	0	2.0180	1.0011	0.0069
closed Eyes	0.0002	0	0	0	0.0052
engagement	0.0795	0	0.9996	0.7500	0.5009
class		presentation	engaging	p-eye-contact	p-engaging

DISCUSSION AND CONCLUSION

The developed tool extract video analytics based dataset from the video file. This dataset is frame by frame score calculation. This dataset is transformed into time-based data by rounding off the arithmetic means of the variable PFPS, that is 6 in this case study. The tuples of frame based dataset were reduced were reduced, which resulted from 16047 tuples to 2675 tuples. Analytical results of the video lecture analysis are then more reduced by eliminating the duplicate tuples and only key tuples are left, which are 54. Five minutes and 53 seconds video has been reduced to 54 unique tuples. This reduced dataset is very easy to explore and visualize as compared to watch video lecture to analyze the effectiveness of the video lecture. Engagement score is between zero and one and higher the value of engagement score means the lecture is more effective and engaging.

Other than the style and structure of the video, it is also necessary to offer navigational support in the content of the video to the overall instructional design for peer-support on the learning management system. Mostly the video lecture arrangements facilitate the simple video navigation, such as random seek, play and pause. It depends on the accessibility of video fragments that a video lecturing system might also provide an extra user interface to connect the video to the remaining part of the instructional design. For the instance, Khan Academyiii and UDACITY provide a highly segmented video capturing approach (Chorianopoulos & Giannakos, 2013). Video lecture analysis facilitate the teacher and provides an opportunity to redesign the lecture. The major purpose of lecture's redesign is to make it effective and more useful for the students.

Lengthy video lectures cause ineffective content delivery. Video presentation up to 10 minutes that is focused on a specific topic can be more helpful for the students. 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

towards 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.

ACKNOWLEDGEMENTS

REFERENCES

Aydin, M. K., Gürol, M., & Vanderlinde, R. (2016). Evaluating ICT Integration in Turkish K-12 Schools through Teachers' Views. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(4), 747-766.

Bishop, J. (2013). Testing the flipped classroom with model-eliciting activities and video lectures in a mid-level undergraduate engineering course. Paper presented at the Frontiers in Education Conference, 2013 IEEE.

Cettolo, M., Niehues, J., Stüker, S., Bentivogli, L., & Federico, M. (2013). Report on the 10th IWSLT evaluation campaign. Paper presented at the Proceedings of the International Workshop on Spoken Language Translation, Heidelberg, Germany.

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108-121.

Chorianopoulos, K., & Giannakos, M. N. (2013). Usability design for video lectures. Paper presented at the Proceedings of the 11th european conference on Interactive TV and video.

Fain, E., & Smith, J. (2014). Video Lecture Capture Pedagogy Context: Does it matter and does it Deliver? Paper presented at the Society for Information Technology & Teacher Education International Conference.

Farhan, M., Iqbal, M. M., & Naeem, M. R. (2015). Analysis of Video Lecture's Images for the Compression Formats Support for the Students in eLearning Paradigm. *Science Education International, 27*(2), 1171-1176.

Fostier, M., Patel, S., Clarke, S., & Prokop, A. (2015). A Novel Electronic Assessment Strategy to Support Applied Drosophila Genetics Training in University Courses. G3: Genes| Genomes| Genetics, 5(5), 689-698.

Gilardi, M., Holroyd, P., Newbury, P., & Watten, P. (2015). The effects of video lecture delivery formats on student engagement. Paper presented at the Science and Information Conference (SAI), 2015.

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: effects on information retention, visual attention, and affect. Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Liu, L.-W., Chang, C.-M., Huang, H.-C., & Chang, Y.-L. (2016). Verification of social network site use behavior of the university physical education students. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(4), 793-805.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior, 28*(1), 181-186.

McCammon, L., & Parker, B. (2014). Live Lecture versus Video Lecture.

Meade, D. P., Ball, A. D., & Brandsness, M. (2012). The Evolution of an Electronic Assessment for Evaluating Trip Risk, Compliance, Approval Level and Tracking. Paper presented at the

International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production.

Nabiyev, V. V., Çakiroğlu, Ü., Karal, H., Erümit, A. K., & Çebi, A. (2016). Application of Graph Theory in an Intelligent Tutoring System for Solving Mathematical Word Problems. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(4), 687-701.

Nicolle, E., Britton, E., Janakiram, P., & Robichaud, P.-M. (2014). Using TED Talks to teach social determinants of health Maximize the message with a modern medium. *Canadian Family Physician, 60*(9), 777-778.

Seguin, R. A., Morgan, E. H., Connor, L. M., Garner, J. A., King, A. C., Sheats, J. L., . . . Buman, M. P. (2015). Peer Reviewed: Rural Food and Physical Activity Assessment Using an Electronic Tablet-Based Application, New York, 2013–2014. Preventing chronic disease, 12.

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. arXiv preprint arXiv:1407.7131.

Wasserman, S., Larsen, M., Asiedu, N., Tseretopoulos, I., Spielman, L., & Ghajar, J. (2015). Assessing Attention: Relationship between Circular Visual-Tracking and Spatial Span, and Implication for Electronic Assessment. *Archives of Physical Medicine and Rehabilitation, 96*(12), e15.

Yirci, R., Karakose, T., Uygun, H., & Ozdemir, T. Y. (2016). Turkish Adaptation of the Mentorship Effectiveness Scale: A validity and Reliability Study. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(4), 821-832.

http://iserjournals.com/journals/eurasia

APPENDIX A: A Few Screenshots of the Alison Killing TEDx Video

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4. (a). Presentation mode because the presenter is too far **(b)** Engaging because the presenter is clear in the video **(c)** partially engaging because the presenter's face is visible but not the eyes **(d) (e) (f)** Presentation mode because the face is not detected of the presenter.

---

[i] Udacity - Free Online Classes & Nanodegrees, Retrieved from https://www.udacity.com

[ii] Alison Killing, What happens when a city runs out of room for its dead, Retrieved from https://www.ted.com/talks/alison_killing_what_happens_when_a_city_runs_out_of_room_for_its_dead

[iii] Learn for free about math, art, computer programming, economics, physics, chemistry, biology, medicine, finance, history, and more, Retrieved from https://www.khanacademy.org