A Video Analytics System for Class Room Surveillance Applications

Saeed Ahmed, Nirmal Krishnnan, Thanmay Ganta, Gurusamy Jeyakumar

Abstract--- Using video analytics to give insights about events happening in classroom is a very important task in classroom surveillance systems. This paper proposes a new algorithmic frameworkto identify abrupt changes in a class room video and thenevaluate the attention level of students.The proposed algorithm is implemented with and without video key frame extraction approaches. The SSIM (Structural Similarity Index) approach for key frame extraction is used in this study. After extracting the key frames, the detection of face and upper body of the students to evaluate their attention level is performed on the key frames. The results comparing thealgorithms with and without SSIM reveals that the SSIM based algorithm gives better results. The algorithmic design of the proposed approach, the results obtained and sample cases are presented in this paper.

Keywords--- Video Analytics, Key Frame Extraction,SSIM values,Computer vision, Face Detection, Upper Body Detection.

I. INTRODUCTION

The growing need for information and high-quality video cameras has led to the proliferation of video based systems that perform tasks such as surveillance, traffic monitoring, etc. This is necessary because such tasks are time consuming given the amount of data involved. Hence manual detection has become almost impossible.A variety of domains (classroom surveillance, content based video retrieval, etc.) require optimized and robust mechanisms to detect an event from the video. Hence automatic event detection and decision making by the machine is of an urgent need to support the growing need of surveillance systems around the world which will reduce the chances of error and the amount of time spend in analysing the video if it was to be done manually by humans.

In a video, key frames are those which represents the most significant events occurring in that video. It highlights most of the key contents of the video. A video can be naively summarized with the help of key frames. This is an area that has gained the attraction of many researchers. A lot of research has been done to propose differentapproaches for key frame extraction.Selecting suitable approach is very much application specific. Using *SSIM* to extract key frames is one of the most conventional approachesto summarize videos. This is because it best corresponds to human perceived measurement.

Attention level of students can be evaluated by many approaches. One such approach is by using upper body and frontal face detection.This is a very robust idea and has the ability to work in real time. This idea is very flexible and add-ons can be made to make it more accurate.

In this paper we propose an algorithm to detect the attention level of a student in class room during the lecture hours by extracting key frames. Conventional *SSIM* approach is unified with frontal face and upper body detection to evaluate a student's attention level.The performance of the algorithm is validated by implementing it for three stored videos. The processing time of this approach in detecting attention level of a student is compared with and without key frames. The proposed algorithm shows robust performance with good customization possibilities.

II. RELATED WORKS

The research works highlighting the key frame extraction approaches and student behaviour analysis are summarized in this section, with a note on the idea for the proposed algorithmic framework.

The key frame extraction methods are largely used for change detection and summarization. An improved histogram based approach for key frame extraction is proposed in [1]. This approach uses the "shots" and frames of a compressed video for its process. A Genetic Algorithm (*GA*) based approach which uses human motion to detect the key frames is presented in [2]. The methods presented in [3] are to detect the key frames using the optimized key frame difference. The difference between adjacent frames are denoted as the inter-frame similarity. A chi-square histogram algorithm for extracting key frame is proposed in [4]. An approach to optimize the frame differences is proposed in [5]. Among different approaches (as proposed in [6], [7] and [8]) for solving key frame extraction problem, the*SSIM* based approach (as mentioned in [9]) has become popular due to its simplicity and its similarity with human perception.

Insights on how *SSIM*based algorithm can be used for video summarization and other processing is presented in [10].In this work, key frames are generated using visual features like *SSIM*, colour histogram, moment measure and correlation measure.The algorithm generated summaries that *SSIM*were close to human perception.There are few interesting attempts in the literature to integrate SSIM method with other algorithms.

In [11], authors have presented a novel approach to integrate the *SSIM* method

Manuscript received February 01, 2019

Saeed Ahmed, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Tamil Nadu, Amrita Vishwa Vidyapeetham, India. (e-mail: cb.en.u4cse15542@cb.amrita.students.edu)

Nirmal Krishnnan, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Tamil Nadu, Amrita Vishwa Vidyapeetham, India. (e-mail:cb.en.u4cse15431@cb.amrita.students.edu

Thanmay Ganta, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Tamil Nadu, Amrita Vishwa Vidyapeetham, India. (e-mail:cb.en.u4cse15355@ cb.amrita.students.edu

Gurusamy Jeyakumar, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Tamil Nadu, Amrita Vishwa Vidyapeetham, India. (e-mail:g_jeyakumar@cb.amrita.edu)

with an Evolutionary Algorithm (*EA*) called Differential Evolution (*DE*). The proposed *DE_SSIM* outperforms the *SSIM* method. This work is extended and presented in [12], tointegrate *DE* with other two key frame extraction methods viz Euclidean method and Entropy methods. This work presents a comprehensive comparative analysis of *DE* based methods with the classical methods of key frame extraction. The results showed that the *DE* based methods outperform the classical methods of key frame extraction methods.

Though there exists many key frame extraction approaches for video analytics, their efficiency can be validated by using such approaches for suitable real world applications where quick video analytics is needed. Some of the interesting works providing video analytics based solution for the real world scenarios are presented below.

In [13], the author aims at detecting the suspicious activities in an examination hall when students are writing examinations. The framework presented in this work monitors the activities of students during examination. It is done in three parts. First, detection of face region of the students is done and its orientation is monitored. The Haar feature extraction method is used for monitoring the orientation of the faces. Second, by using a grid formation the hand contact is detected and analysed to identify the exchange of papers between the students. Third, the hand signalling for the students is monitored and alert message is given to the invigilator.

In [14], authors combine the concepts of IoT with Computer Vision. The work proposes a method that takes student attendance and detects student motion and behaviour.The modules are: Face Detection module, Motion Analysis module and Behaviour Analysis module. In Face Detection module, a combination of Eigen Face and Fisher Face approach is used. In Motion Analysis module, Upper Body Detection is done using Cascade Classifier and Full Body using HOG. Motion Tracking is done using Lucas and Kanade's Optical Flow algorithm. The Behaviour Analysis module in [14], is referred by the present work to get theinsights about the rules to determine the behaviour of the student.

In [15],the "Viola Jones" algorithm is used for face detection. The algorithm looks for specific "Haar" feature of a face and the integral image concept to find out the area of rectangle. The summation of pixel values of the original image is used to define the integral image. In integral image, the value at pixel (*x*,*y*) is the sum of pixel values above and left to the pixel at (*x*,*y*). Using these two concepts and the "Viola Jones" algorithm it is possible to detect both face and upper body of a person.

Though there are many approaches for key frame extraction and student behaviour analyses, none of them combines the two for evaluating the attention of a student in classroom. Our algorithm focuses on analysing the attention levels of a student in a classroom room. Frontal Face and Upper Body Detection was performed to find out whether the student is paying attention to the class. The effectiveness of this approach isvalidated with three sample videos and the results are discussed in the later section.

III. EXPERIMENTAL DESIGN

This section explains the architecture and approach of the proposed algorithm. The architecture diagram is shown in Fig. 1. The main components of the algorithm are the video acquisition, change detection, faceand upper body detection and report generations.

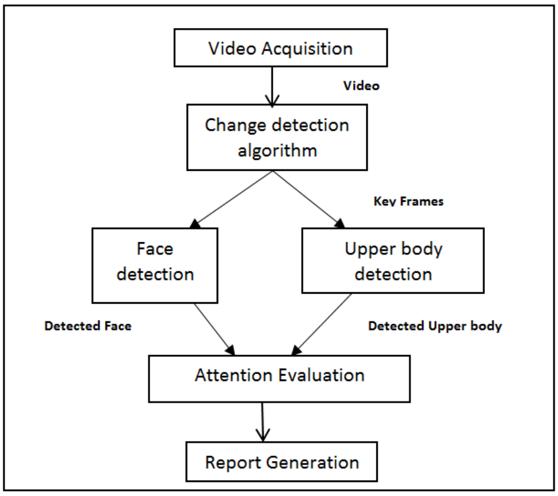

Fig 1: Architecture Diagram

The working of proposed algorithm is explained below, mentioning the functionality of each of the component.

- Video acquisition - The video of the classroom where the students are sitting is acquired. Care must be taken while taking the video so as to avoid illumination changes, shadow etc.
- Change Detection Algorithm -In this component, the relevant frames are extracted from the video collected at the video acquisition phase. *SSIM* values are used to compare frames and check whether they are key frames.*SSIM* says that a measure of structural information change can provide a good approximation to perceived image distortion.It assesses the visual impact of changes in luminance, contrast and structure in an image. The *SSIM* compares two images *x* and *y*, and finds the luminance *l*(*x*,*y*), contrast *c*(*x*,*y*) and structure *s*(*x*,*y*). The *SSIM* is calculated as given in equation 4.1. The *l*(*x*,*y*), *c*(*x*,*y*) and *s*(*x*,*y*) are calculated using the equations 4.2, 4.3 and 4.4, respectively (as given in [9]).

$$SSIM(x,y) = l(x,y).c(x,y).s(x,y) \quad (4.1)$$

where

$$l(x,y) = \frac{2\mu_x\mu_y+c_1}{\mu^2x+\mu^2y+c_1} \quad (4.2)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y+c_2}{\sigma^2x+\sigma^2y+c_2} \quad (4.3)$$

$$s(x,y) = \frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3} \quad (4.4)$$

- Face Detectionand Upper Body Detection - This module takes the key frames from the previous component and performs frontal face detection on those key frames. For this Viola-Jones algorithm has been used. The 4 stages in the algorithm are Haar feature selection, creating an integral image, Adaboost training and Cascading Classifiers.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- Checking whether the student is attentive or not - From the previous componentthe images showing the detected face and upper bodyis extracted. They areanalyzed to make the following conclusions.
    - If frontal face and upper body are both detected thenit is concluded that the student is attentive.
    - If only upper body is detectedthen it is concluded that the student is not paying attention.
- Report generation - In this component, the count of frames in which the student is attentive and not attentiveis measure. From these counts the percentage of attentiveness of the student is calculated and reported to class the teacher and/or the student.This feedback about the attentiveness of a student will enable him/her to realize his level of attention during the lecture hours.

IV. RESULTS AND DISCUSSIONS

The working of proposed algorithm is tested with three short sample videos taken in different class room environment. The video is converted into frames,each of the frames is analyzed and key frames are identified. The face and upper-body detection is done on the key frames,the percentage of attention of the student is calculated and reported.

The Figure 2 and 3 shows the results on attention detection for theVideo #1.Figure 2 shows a frame in which the student is paying attention in class. In this frame both the upper body and frontal face are detected.Figure 3 shows a frame in which the student is not paying attention in class. In this frame only the upper bodyis detected.Frontal face is not detected.Table 1 shows the attention percentage of the student in each video. It is seen from the result that in each video the student has shown different percentage of attention. In Video #2, the student is very less attentive with only 47% of attention. These statistics about students attentiveness can be used by the student advisor to advise the student for his/her betterment.

Table 1: Student's attention percentage in each video

Video Number	Attention Percentage
Video #1	65 %
Video #2	47 %
Video #3	70 %

The experiment is extended to analyze the importance of using the key frame detection approach for this application. A comparison is made between the response times of the proposed algorithm with *SSIM* and without *SSIM* approach. Table 2 depicts the response time comparison. As it is seen in Table 2, the algorithm's response time is better in calculating the attention percentage after detecting the key frames (using *SSIM*) from the videos.

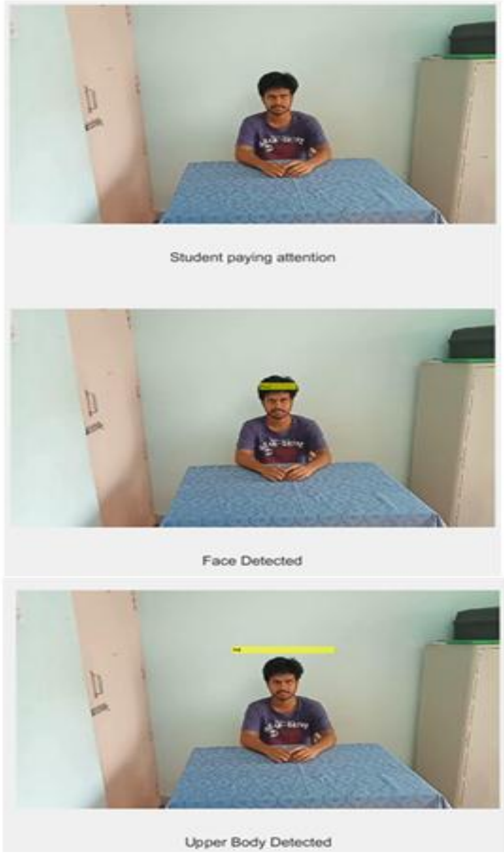

Fig. 2: KeyFrame in which student is paying attention

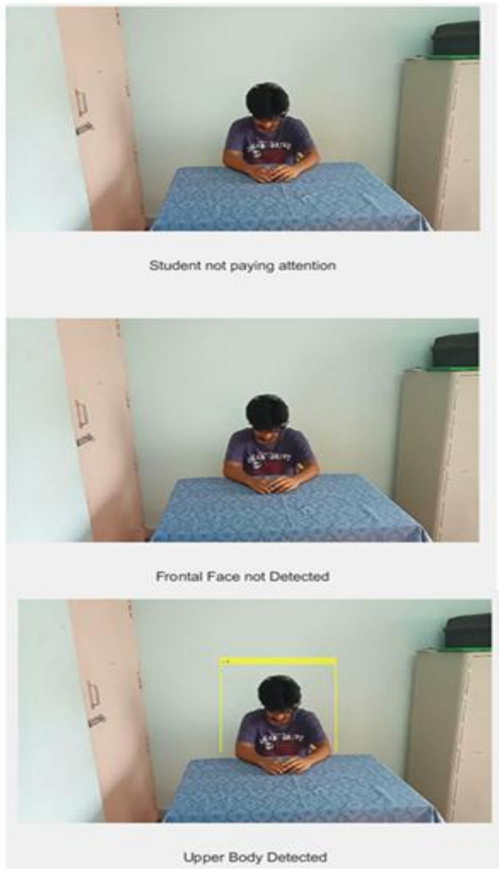

Fig. 3: Key Frame in which student is not paying attention

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 2: Response time comparison

Video Number	Response Time (in Minutes)	
	with *SSIM*	without *SSIM*
Video #1	00:33	00:56
Video #2	01:09	01:22
Video #3	01:47	02:20

Though this proposed algorithm is successful in implementing an integrated approach which combines key frame detection, face and upper body detection for an class room surveillance application, it has certainly many limitation and leaves lot of scope for further research. The major limitations are highlighted below.

- The usual problems associated with video analytics such as dynamic background, illumination change, shadow, camouflage, occlusion and negative recognition are present in the proposed algorithm.
- Different detections of the exact face due to sub window overlapping.
- The algorithm works only for a single student.
- The algorithm doesn't work for real time videos.
- In case the video is not collected in a controlled space, then it gives rise to many falsedetections. A sample case of multiple false detections of face is shown in Fig.4, where other than the face of the students many other similar structured objects also detected as face.

Fig. 4: False detection of faces

V. CONCLUSIONS

As an attempt to evaluate the attention level of a student in classroom, this paper has proposed an algorithm which combinesa key frame extraction approach with face and upper body detection methods. The proposed algorithm is verified using 3 videos. In each of the video the algorithm calculated the percentage of attention of a student in a class room. The algorithm is implemented with and without finding key frames from the given videos. The response times are compared and found that the algorithm gives the results faster when it is used with a key frame extraction approach. The *SSIM* approach of key frame detection is used in this study.

The proposed approach can be improved further to implementit for multiple students in a classroom. Also the proposed algorithm assumes that the students are not writing anything during the lecture. The algorithm needs to be improved to ignore this assumption. Inclusion of optimization algorithms viz Evolutionary Algorithms(*EAs*) can be verified to get more optimized key frames during the process of attention level detection.

REFERENCES

1. Guozhu Liu, and Junming Zhao, "Key Frame Extraction from MPEG Video Stream", *In Proceedings of Second Symposium International Computer Science and Computational Technology*, 2009.
2. Qiang Zhang, Shulu Zhang and Dongsheng Zhou, "Keyframe Extraction from Human Motion Capture Data Based on a Multiple Population Genetic Algorithm", *Symmetry Open Access Journal*, Vol. 6, No. 4, 2014.
3. Huayong Liu and Wenting MengZhi Liu, "Key Frame Extraction of Online Video Based on Optimized Frame Difference", *In Proceedings 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 2012.
4. Ramender, G, MovvaPavani and Kishore Kumar, G, "Evolving optimized video processing and wireless transmission system based on arm-cortex-a8 and gsm", International journal of computer networking, wireless and mobile communications, Vol. 3, No. 5, 2013.
5. Huayong Liu, LingyunPan and WentingMeng, "Key Frame Extraction from Online Video Based on Improved Frame Difference Optimization", *In Proceedings of 14th International Conference on Communication Technology (ICCT)*, 2012.
6. Ran Zheng ,Chuanwei Yao ,Hai Jin ,Lei Zhu,Qin Zhang and Wei Deng, "Parallel Key Frame Extraction for Surveillance Video Service in a Smart City" *Plos One*, Vol. 10 No.8, 2015.
7. Sheena C.V and N,K.Narayanan, "Key Frame Extraction by Analysis of Histograms of Video Frames using Statsistical Videos", *Procedia Computer Science*, Vol. 70, pp. 36-40, 2015.
8. Zhang Ronghua and Liu Changzheng, "The Key Frame Extraction Algorithm Based on the Indigenous Disturbance Variation Difference Video", *The Open Cybernetics &Systemics Journal*, Vol. 9, pp. 36-40, 2015.
9. Z. Wang and A. C. Bovik, "A universal image quality index", *IEEE Signal Processing Letters*, Vol. 9, pp. 81–84, 2002.
10. SharanjeetKaurSandhu and AnupamAgarwal "Summarizing Videos by Key frame extraction using SSIM and other Visual Features", *In Proceedings of - ICCCT '15 - Sixth International Conference on Computer and Communication Technology*, 2015.
11. Kevin Thomas Abraham., Ashwin M., DharsakSundar., TharicAshoor and Jeyakumar G, "An Evolutionary Computing Approach for Solving Key Frame Extraction Problem in Video Analytics", *In Proceedings of – ICCSP'17 - International Conference on Communication and Signal Processing*, 2017.
12. Abraham K.T., Ashwin M., Sundar D., Ashoor T and Jeyakumar G, "Empirical Comparison of Different Key Frame Extraction Approaches with Differential Evolution Based Algorithms", *In Advances in Intelligent Systems and Computing*, Vol 683, *Thampi S., Mitra S., Mukhopadhyay J., Li KC., James A., Berretti S.* (Eds), 2017.
13. Senthil Kumar T and Narmatha G, "Video Analysis for Malpractice Detection in Classroom Examination". In: Suresh L., Panigrahi B. (eds) *Proceedings of the International Conference on Soft Computing Systems. Advances in Intelligent Systems and Computing*, Vol 397, 2016.
14. Jian Han Lim., EngYeow The., Ming Han Geh and Chern Hong Lim, "Automated classroom monitoring with connected visioning system", *In Proceedings of*

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2017.

15. Mehul K Dabhi and Bhavna K Pancholi, "Face Detection System Based on Viola - Jones Algorithm", *International Journal of Science and Research*, Vol. 5., No. 4., 2016.