IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 10, 2018, accepted August 8, 2018, date of publication August 16, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2865754

A Complete System for Analysis of Video Lecture Based on Eye Tracking

XUEBAI ZHANG[2,3], SHYAN-MING YUAN[3], MING-DAO CHEN[3], AND XIAOLONG LIU[1]
[1]College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]Xiamen University, Xiamen 361005, China
[3]Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan

Corresponding authors: Shyan-Ming Yuan (smyuan@cs.nctu.edu.tw) and Xiaolong Liu (xlliu@fafu.edu.cn)

This work was supported in part by the Fund of Ministry of Education in China (MOE) Project of Humanities and Social Sciences under Grant 17YJC880076, in part by the Educational Commission of Fujian Province, China, under Grant JAS170171, and in part by the MOST of Taiwan under Grant 105-2511-S-009-007-MY3.

ABSTRACT An eye tracking technology provides a faster and more intuitive interface to explore the learner psychology. In this paper, with the aim of assisting researcher to expand the knowledge of video lectures, we present an efficient system named VLEYE based on eye tracking. The system allows gathering the dynamic areas of interest (AOIs) and combining them with eye movement data. There are mainly three modules in the proposed system, including "eye movement recorder," "dynamic AOI module," and "video analyzer." The "eye movement recorder" records the eye movement data by eye tracking device, where mouse cursor control function is applied to support the use of low-cost devices. The AOIs could be drawn manually or detected and tracked by the provided algorithm automatically in the "dynamic AOI module." The "video analyzer" provides various visualization modes to view and compare the eye movement information for single or multiple subject, as well as statistical charts and detail fixation information. In addition to performance testing, a case study was conducted to verify the applicability of the proposed system. We demonstrate that this system works well and is an essential step forward in efficiently helping studies to better understand learner cognitive processes for dynamic stimuli within video.

INDEX TERMS Dynamic area of interest (AOI), dynamic stimuli, eye tracking, eye movement, video lecture.

I. INTRODUCTION

Over the last years, the using of video lectures has significantly increased in online courses as a result of the improvements in the Internet infrastructure [1]. Literature suggests that video lectures are substantially appealing to many students and are perceived as effective for learning [2]. Video lectures are therefore being gradually adopted in digital learning contexts such as flipped classrooms, massive open online courses, and electronic textbooks [1]. Despite the growing number and variety of online video lectures available, there is a limited application to support the understanding of the nature and quality of their effectiveness [3]. Therefore, in this paper, we are making efforts to assist the researchers to extend knowledge around optimizing video as a learning tool.

Eye-tracking technology traces participants' real-time eye movements and provides details of the viewers' cognitive processes [4]. It provides unique information concerning perceptual and cognitive processes that underlie learning performances [5]. A previous study of Liu [6] claimed that eye-tracking technology added value in discovering the underlying mechanisms of effects on multimedia learning. Better understanding of the viewing behaviors of participants will help to improve the effectiveness of learning by using video. In general, eye movement analysis of eye-tracking technology can be conducted subjectively or objectively.

Subjective analysis emphasizes playback of gaze locations and viewing attention maps after task completion. Visualizations play an important role in subjective eye tracking analysis: they form a basic representation for question and hypothesis generation [7]. They offer a quick way to aggregate data across participants, but are limited in their ability to explain larger data trends. Heat or attention map [8] is popular eye tracking visualization that generally show all fixations within a period of time as either multi-colored or transparent areas that are superimposed on the stimulus. Heat map has become a very popular way to summarize eye tracking data, but provide limited objective, aggregated understanding of task strategies [9].

Objective analysis aims at more objective understanding of sequential factors and effects may be obtained through analysis of appropriate metrics. Objective analysis relies on eye movement metrics generated by the software, including fixation duration, fixation number, and many other metrics [10]. The data was then exported to analysis at a higher level of abstraction. Across all fields of research on eye tracking, area of interest (AOI) is the most important metric for objective analysis. AOIs are areas of a stimulus that are of high importance for a hypothesis. It used to link eye-movement measures to particular object of the stimulus (e.g., the time spent looking at a particular object in the stimulus). AOI statistics can make eye-movement data easier to interpret and are used in multiple fields of research [11]. Usually such analysis is performed manually and is mostly found in static stimuli, such as an image, webpage and slideshow.

To support both subjective analysis and objective analysis in dynamic stimuli, we propose a system for analysis video lectures based on eye tracking. The proposed system is the first system that is specially designed for supporting the researchers to determine and uncover students' learning process in video lecture. Our proposed system is named VLEYE, which provides different researchers with the opportunity to fulfill their own needs. VLEYE combines commonly known "state-of-the-art" gaze analyzing methods, such as AOI definition, heat map, real-time replay, along with an easy-to-use graphical user interface. Our design has therefore focused on ease of use, reliability, and minimizing cost. The main contributions of our proposed VLEYE system are the following:

1) Supporting low-cost eye trackers to create the possibility of widely used system to the research community.
2) The implementation of timesaving appropriate technologies, such as an automatic definition and tracking method for dynamic AOI definition.
3) An up-to-date natural and convenient operation mechanism for user to draw dynamic AOIs.
4) Visual-guided data mining through the use of different visual representations.
5) The implementation of eye tracking analysis functions to label dynamic AOIs which are meaningful to the subjects.

The remainder of this paper is organized as follows. Section 2 presents the related works in the area of eye tracking system for videos. The design and the implementation of the system are described in Section 3. Section 4 presents and discusses the experimental results. Finally, Section 5 draws conclusions.

II. RELATED WORKS

With the development of fast, high-end eye-trackers, Eye-tracking technology has now in widespread use within a number of disciplines, such as usability research, psychology, marketing research and so on [12]. While there is a bunch of eye tracking researches have been studied, most of them are based on analyzing of eye movement in static stimuli, such as an image, webpage and slideshow. There are limited researches about the application of the eye tracking technology in dynamic stimuli, such as video, game and interactive applications.

In earlier studies of the eye tracking technology in dynamic stimuli, eye movement data is collected and interpreted in a low-level form, as gaze-coordinates in the space of rendered visual stimuli that gazes were recorded for. Tien *et al.* [13] measured gaze overlaps of a video that showed a surgical task to compare experts' gaze with the gaze of trainees. Marchant *et al.* [14] described an approach to investigate the influence of directorial techniques on film viewers' experience. However, in their research, relating eye movement data to the semantic content of the stimuli is generally done offline by human analyst who inspects gaze heat maps visually, or defines AOIs manually. This process requires significant manual intervention and is especially difficult for studies of dynamic stimuli involving many moving objects [15].

To analysis players behaviors in video games efficiently, Sennersten et al. [16] developed a software interface between the eye tracking system (Tobii) and a game engine (HiFi) for use in automated logging of dynamic 3D objects of gaze attention. The eye movements are recorded and the gaze positions are matched with the presented objects in real time during the game. The static and moving objects within the virtual world themselves serve as AOIs, which are realized with the use of game engines to allow the real time visualization of virtual three-dimensional worlds. Unfortunately, their method is not possible to rerun the matching process after removing some objects [17]. Furthermore, it also has the disadvantage of not applying any method for correcting the eye-movement data [18]. To overcome these drawbacks, offline matching of dynamic AOIs is pursued in the field of eye movement analysis.

Dynamic AOIs are areas that represent the position and dimensions of important content over time in dynamic stimulus. Papenmeier and Huff [17] introduced a tool called DynAOI that allows for automatic definition of offline AOIs for animations that are based on three-dimensional models. The captured eye movement data on a two dimensional screen is converted into a three-dimensional representation of the events within the presented video files. The converted gaze vector then is used to determine all objects as AOIs on its path. Unlike the work of DynAOI, Alam and Jianu [19] proposed a work targeting data-rich information visualizations. They introduced data of interest-based analysis, which is a novel way of analyzing eye-tracking data in visualization and data space. This method leverages the known structure of an evaluated visualization to detect in real time which individual elements of the visualization are being viewed. In this approach, the data analysis method separately rates of every frame of the combined video manually. The method is very subjective because it can be different to interpret the eye position with related dynamic AOIs. Therefore, another approach for data analysis, which was depended on a synchronized and separate recording of eye movement and dynamic AOIs, was proposed by Zelinsky and Neider [20]. They used a shorted distant rule for assigning the eye-movement data to the presented dynamic AOIs. The distances between each eye movement point to each center of the visible dynamic AOI are calculated. The dynamic AOI that is closest to the eye movement is classified as "looked at." Though this approach assigned the eye movement data to dynamic objects automatically, it encounters problems when the subjects do not look at any object within the video. Further, Friedrich et al. [21] presented an application to automatic matching of eye movement and polygon AOIs. The application checks whether an eye-movement point is inside an AOI by taking the layer model into account the matching algorithm. The shape and position of the AOI defined by points are stored for every timestamp into a log file. The eye movement uses the same timestamp as the stored AOIs. The dynamic AOIs are then assigned to the eye movement. Although this method allows for the use of dynamic AOIs with any desired content, it is very time consuming to add the AOIs manually. Moreover, such a system is not appropriate to the situation that dynamic objects change their position or shape over the majority of time.

The related literatures mentioned above realized the analysis of eye movement in different dynamic stimuli. However, the objective analysis of eye-tracking data is still challenging and time-consuming, since some steps are still performed manually. To provide more automatically analysis, we present a complete system that not only provides manual drawing but also automatic method for dynamic AOI definition. Furthermore, most of the previous eye-tracking researches are depended on specific eye trackers, and most of them are quite expensive. It is difficult for research institutions with limited budgets to invest in the eye tracking research. The proposed system could support low-cost eye trackers as long as they provide mouse cursor controlling function. It provides different researchers with the opportunity to fulfill their own needs, which would cause considerable advance to the popularity of eye tracking-related research and applications.

III. SYSTEM DESIGN

A complete eye movement data analysis system VLEYE is provided for researchers to recode eye tracking data, dynamic AOIs setting and video lecture analyzing. The architecture of the proposed video analysis system VLEYE is shown in Fig. 1. The system contains three main modules: "Eye Movement Recorder," "Dynamic AOI Module" and "Video Analyzer." "Eye Movement Recorder" communicates with eye tracker device and records eye movement data of the participants when they are watching the video. "Dynamic AOI Module" creates and stores the AOI data of the video, which is independent to eye movement record and analysis. By loading the recorded eye movement data and stored AOI data, "Video Analyzer" analysis the eye movement information for the AOIs in the video. Output analysis results with chosen mode in Video Analyzer create the presentation patterns of the analysis results and stores the analysis results. These three modules can be performed separately, users can define AOIs

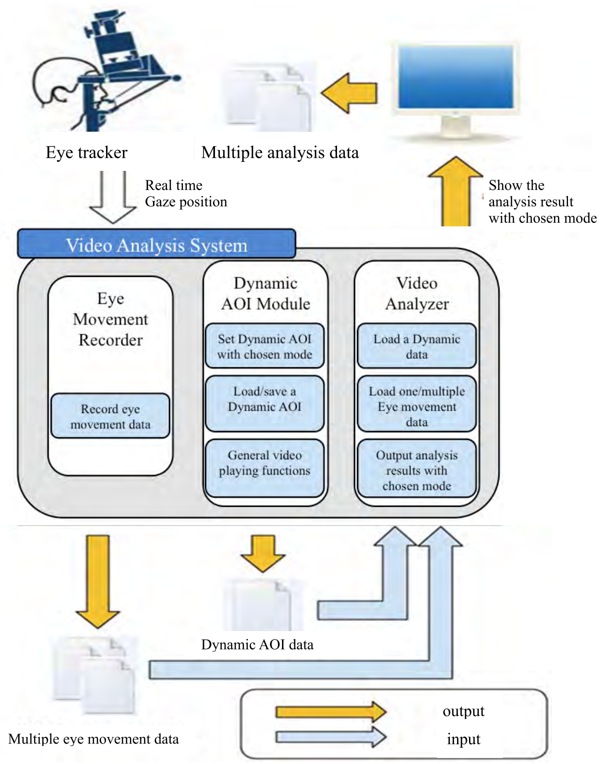


FIGURE 1. The architecture of the proposed video analysis system VLEYE.

at any time without affecting tracking and subsequent analysis of the eye movement data. The details of the proposed modules are elaborated in the following sub-sections.

A. EYE MOVEMENT RECORDER

The Eye Movement Recorder Module is designed to record the participant's eye movement information with dynamic video stimulus. To support a wide majority of low-cost eye trackers, such as Tobii EyeX and EyeTribe, we track the mouse cursor along with the eye movement. Mouse cursor control, which is supported by various eye tracking devices, allows users to redirect mouse cursor to the gaze position, so that they can track the position of the mouse cursor instead of the gaze. As long as the eye tracking device has the capability to control the mouse cursor, it can be used in our system. After connecting the eye tracker, user is required to start the eye movement recorder according to a series of stating, which are presented in system message to prompt user (as shown in Fig.2).

The stating process includes loading the video, resizing the playing window and playing the video. The loaded video is covered by a black block to prevent participants from predicting stimulus according to the first video frame. Resizing the window allows the accurately tracking of the gaze location corresponding to the video stimulus. After playing video, user can output the eye movement data and exit the recording by keyboard shortcuts. The output eye movement file in .txt format stores the timestamp (millisecond) and the corresponding

FIGURE 2. The user interface of the Eye Movement Recorder.

gaze coordinate (GazePosX, GazePoxY), which data will be used by the Video Analysis Module.

B. DYNAMIC AOI MODULE

In order to enable users to define dynamic objects efficiently, the Dynamic AOI Module provides two kinds of AOI setting mode: "Manual Setting Mode" and "Automatic Setting Mode." "Manual setting mode" provides user with using simple mouse operations to define dynamic AOIs in the video. "Automatic setting mode" provides user an automatic way to detect and track dynamic AOIs, which simplifies the defining process of AOIs in the video. Furthermore, this module also provides the possibility of importing a list of predefined AOIs from a text file for further modification and analysis of AOIs.

The AOI defining interface contains two main areas: AOI workspace and video timeline. The workspace is the area for video displaying and AOI creating. The Video timeline enables the user to control the playback progress of the media to create dynamic AOIs. We provide user with both mouse operation and Keyboard shortcuts to control the video replay. With mouse, user can drag the progress bar to any frame to create/modify the dynamic AOIs. The Keyboard shortcuts enables the user to precisely control the media playing, stepping forward/ back 1/10/100 frames of the video, as well as the basic video playing functions, such as Play, Pause, Speed up/down, and Replay.

In Manual Setting Mode, the dynamic AOIs can be created by simple mouse operations of user. To draw rectangles, which are available shapes to define the AOIs, click and drag until the shape is the size adapt to the researched object. Once the rectangle is drawn in the first frame that the object starts to play, a dynamic AOI is created around the object in the video. In order to automatically adjust the rectangle in each displaying frame to cover the dynamic object, we propose interpolation algorithm in the tracking process. The system will interpolate the shape and position of the AOI rectangles in-between the keyframes. The keyframe is a user defined shape and position of the AOI that corresponds to a certain point on the timeline of the video. In later processing, mouse operations are also capability to modify the dynamic AOIs. The supported operations including: moving the AOI, erasing rectangle frame by frame, deleting the AOI, adjusting the AOI rectangle and renaming the AOI.

In Automatic Setting Mode, in order to reduce the complexity of AOI setting procedures and improve the setting efficiency, we design an automatic dynamic AOI setting procedure for the human face, the key part of the human attention in the lecture video, according to the computer vision and dynamic problem solving techniques [22], [23]. This automatic setting procedure uses the face detection algorithms, object tracking methods and scene change detection algorithms to simplify the operations for users when defining the dynamic AOIs in the video. In this work, face detection is achieved using the Haar-based face detection [24] to automatically wrap the human face object with appropriate size box. This face detection algorithm locates faces in video and outputs a bounding box for every face detected in the imagery, with extremely rapidly image processing and high detection rates. The Kernelized Correlation Filter (KCF) object tracking method [25] is implemented to track objects with an interval of N frames and automatically record the position. Also, in order to automatically end the AOIs on last scene before scene change, traditional pixel-based scene change detection algorithm is used to detect the scene change. In the pixel-based method, the sum of mean absolute difference (MAD) of pixel values at corresponding location between two sequential frames is calculated. If the MAD is larger than a predefined threshold, it is considered for a scene change occurring.

The flowchart of the automatic dynamic AOI setting procedure for the video is shown in Fig.3. Once starting with the "Automatic Setting mode," in the first frame, the system will automatically create new AOIs of human object by face detection, or user could set AOIs of nonhuman object by manual operations. In the next frame, the system determines whether it reaches the end of the video. If so, the system ends the last AOI and outputs the AOIs data file. Otherwise, the system will detect whether the scene is changed in this frame. If scene changed, track the AOI in the last frame of the last period and redetect to create new AOIs at the current frame. If not, the system will determine whether this frame is with tracking interval of N on the tracking point for each AOI. In the Nth frame, the system will track AOI objects by tracking algorithm and record the tracking results. Otherwise, it calculates the position that AOI moves by interpolation method. Then the program goes on to the next frame of the sequence and repeats until the end of video is reached. The automatic setting mode also allowed set the AOIs manually at any frame. The tracking results will readjust and update based on the new manual setting of the user.

Finally, the defined dynamic AOI data are stored in a separate file in .txt format. It includes both the manual and automatic setting AOIs. For each AOI, AOI name, frame number, and coordinate of the AOI are stored in the .txt file. The output dynamic AOI file then can be used by the Video Analysis Module or exported to other experiments.

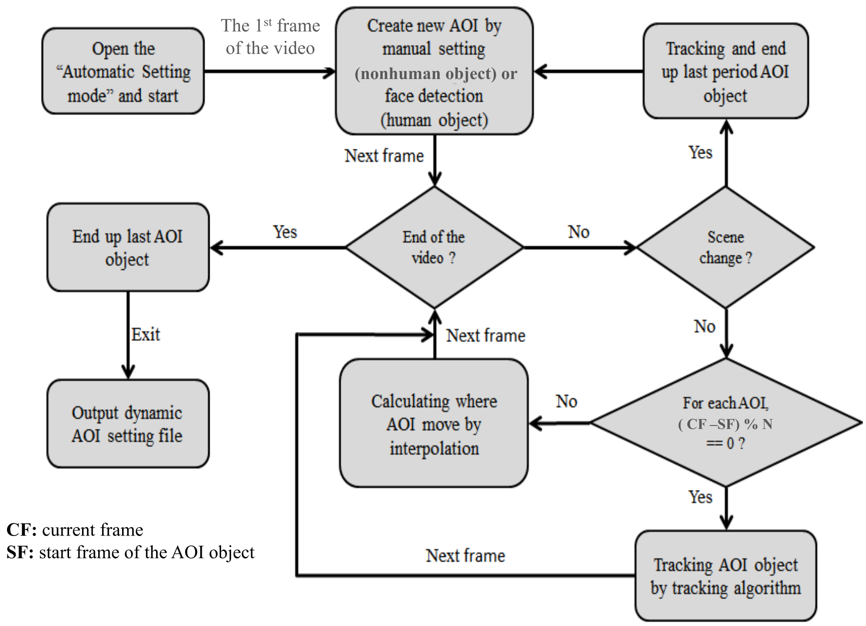


FIGURE 3. **The flowchart of the automatic dynamic AOI setting procedure.**

C. VIDEO ANALYZER

The Video Analyzer Module is design to calculate, store, and display the subjects' gaze states of each AOI in the video. The recorded eye movement data will be gathered in this module and the defined dynamic AOIs data are combined in this calculation, and the fixation of each AOI is calculated for all subjects.

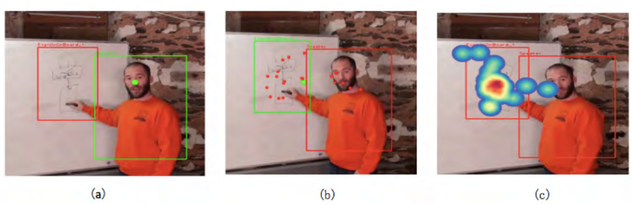


FIGURE 4. **Three display modes for video analysis: (a) Singular Spot; (b) Plural Spot; (c) Heat Map.**

To facilitate eye-movement research, this analysis module provides graphical representations and visualizations of gaze data and defined AOIs. Three display modes (as shown in Fig.4) are available in our system, including Singular Spot, Plural Spot and Heat Map. Singular Spot and Plural Spot modes provide user with combined results of eye positions and AOI information. Heat Map mode provides a dynamic representation of the gathered eye data for the selected recordings. The detail information of the three modes is as follows:

(1). Singular Spot (Fig. 4(a)): This mode allows the subject to look back at his individual eye movement after experiment. The purpose of this display mode is to show which dynamic AOI is exactly fixated by the subject. The green dot is the point where the subject gazed at during experiment. The rectangles are the defined dynamic AOIs in the video. If a dynamic AOI is gazed by the subject, this AOI will be shown as green rectangle. Otherwise it will be shown as red rectangle.

(2). Plural Spot (Fig. 4(b)): A group of subjects' eye movements are shown in this display mode. The purpose is to show the gaze difference between a selected subject and others subjects in this group. The selected subject's eye movement is shown as green dot, and other subject's eye movement is shown as red dot. The defined dynamic AOIs are also displayed throughout the timeline. The AOI gazed by the selected subject will be show as green rectangle.

(3). Heat Map (Fig. 4(c)): This mode allows researchers to analysis a group of selected subjects' eye movement data through different levels of observation intensity. Eye movement data of the group subjects are visually merged to identify regions of special attention in the stimulus video. The density function visualization [26] is used to represent the density of gazing point in the map. The heat map covers the regions where the group of subjects gaze at, the higher density of subjects' gazes in a region, the warmer the color is used to represent it.

After displaying, Singular Spot mode will output single subject's analysis data and Plural Spot mode will and output a group of the subjects' analysis data and the group average data. Two types of information are generated for single and multiple analyses, respectively, including detail information (store in .txt file) and final analysis results (store in .csv file). The detail information described the number of group member, the detail video information (video length, length of AOIs exist, amount of AOIs, average amount of AOIs watched, etc.), and the detail AOI fixation records (length of time, each member fixation time, etc.). The final analysis results show the calculation results of the three indicators of eye movement for each AOI: fixation time rate, fixation number and the fixation state. The fixation time rate for each person is calculated according to the following equation:

$$\textit{Fixation time rate} = \frac{\text{Sum of all the fixation time on this AOI}}{\text{Display time of this AOI}} \tag{1}$$

The fixation number is the total count of meaningful fixation whose duration value is larger than the fixation threshold. The fixation state (0/1) indicates whether the individual shows an impression on the AOI. We denote the AOI with no fixation number to be state 0 and any fixation number to be state 1. It is worth mentioning that our system supports user to set the threshold for a meaningful fixation duration. On average, the typical mean fixation duration is around 180-275ms during the eyes move in visual search [27]. Our default threshold values are defined in this range to be 250ms.

Overall, VLEYE is a complete system that provides eye movement recoding, AOI definition and video analyzing functions. By loading the recorded eye movement data and stored AOI data, the system provides researchers with the opportunity to fulfill their own needs. It is worth mentioning that VLEYE is just a fundamental system, significant investigations need to be undertaken by researchers to further analyzing the learner's learning process in video lecture. Here is an example of the possible learning investigation: The researcher could use our system to analyze what constitutes an effective video lecture to engage students and maintain their attention? Overlaying a pointer controlled by the teacher on the slides or including a video of the teachers face in the corner? Researcher could use VLEYE to define the pointer on the slides and teachers face as dynamic AOIs. The proposed system can help to collect the eye- tracking data and measure the amount of time the learner is paying attention to each AOI. The output result will provide an objective analysis for researcher to determine which element is more effective to engage student and maintain their attention. The investigation result would help researcher to enhance knowledge around optimizing video as a learning tool, and learner engagement with video can be maximized.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the experiment, three testing videos, i.e. video1,[1] video2[2] and video3,[3] selected from cloud computing online courses to represent typical video lectures with instructors on the screen are used. Experiment results will be presented to: (1) determine the appropriate tracking interval N for object tracking method of the Automatic Setting Mode; (2) prove the feasibility of the Automatic Setting Mode, comparing with the Manual Setting Mode; (3) evaluate system usability using questionnaires based on the System Usability Scale (SUS). Finally, a case study proves the feasibility of the proposed system for low cost eye tracking device in educational research, revealing students' impression effect by analyzing eye movement data on AOIs.

A. DETERMINE AUTOMATIC TRACKING INTERVAL N FOR VIDEO LECTURE

In "Automatic Setting mode," in order to speed up object tracking, we set the tracking interval to track the objects instead of tracking them at each frame. Appropriate tracking interval allows users to accurately locate the new position of the moving objects. To determine the appropriate tracking interval N of the moving object, two video clips of 200 frames are sampled from video1 and video2, respectively. The size of human object (instructor) in video2 is larger than that in video1. For video lectures these two sizes of instructor are displayed typically.

We define a dynamic AOI of human object for video2 from 4060th to 4160th frame, as well as for video1 from 2200th to 2400th frame. The human objects within the clips are presented continuously. Four automatic tracking intervals of 10, 15, 20, 25 frames between tracking are tested to discover the appropriate one. The tracking results with these intervals for each clip are shown in Fig. 5 and Fig. 6, where the results of interval 10, 15, 20 and 25 are shown as blue, green, purple, and cyan rectangles, respectively. According to these figures, we could observe that tracking results of intervals 10, 15,

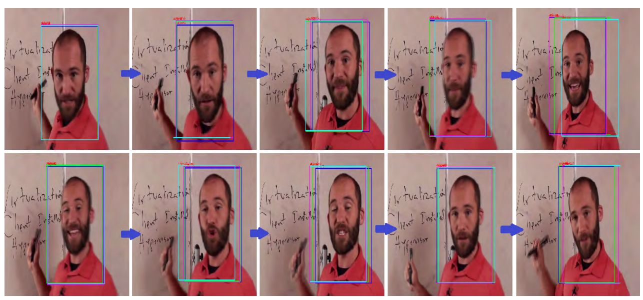


FIGURE 5. AOI tracking results from 4060th to 4160th frame with different interval N in video2.

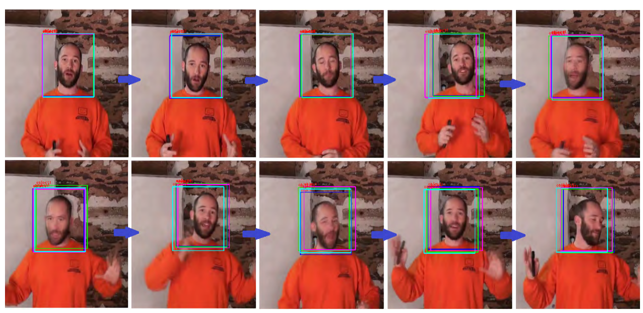


FIGURE 6. AOI tracking results from 2200th to 2300th frame with different interval N in video1.

[1]Video1: Introduction to Virtualization_Overview. (2017): https://www.youtube.com/watch?v=zLJbP6vBk2M

[2]Video2: Introduction to Cloud Computing_Virtualization. (2017): https://www.youtube.com/watch?v=QYzJl0Zrc4M

[3]Video3: Introduction to Cloud Computing_Three Levels of Cloud Computing. (2017): https://www.youtube.com/watch?v=10LFxehfWz4

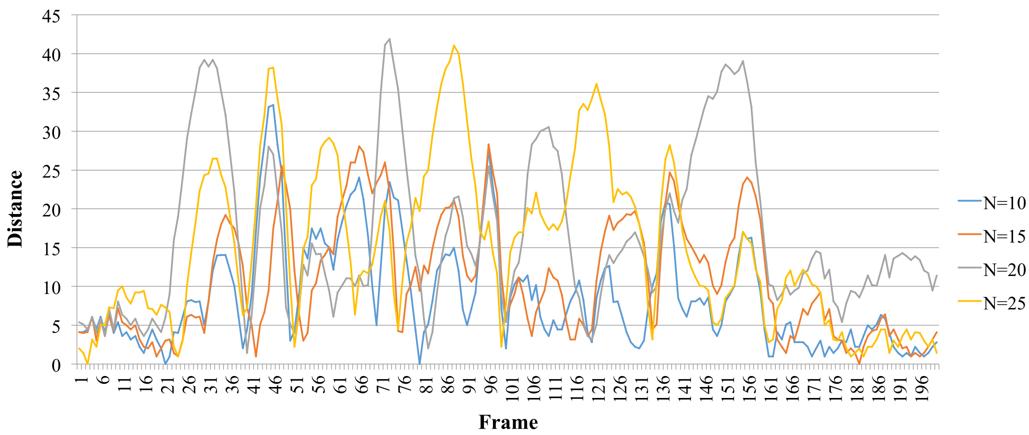


FIGURE 7. The Offset distance with different interval N compared to interval N=1 in video2.

20 and 25 are all acceptable for the human object with the video lectures. Moreover, if we want to pursue high accuracy of tracking results, intervals with 10, 15 or less than 10 are more appropriate.

To evaluate the typical suitable intervals for video lectures, we also calculate the offset distance of the AOI positions from these intervals to interval N=1. The formula for calculating offset distances is presented as Eq. (2):

$$Offset\ Distance = \sqrt{(X_{FN} - X_{F1})^2 + (Y_{FN} - Y_{F1})^2} \quad (2)$$

where X_{F1} and Y_{F1} refers to X and Y coordinate of AOI at F frame without tracking intervals, respectively. X_{FN} and Y_{FN} refers to X and Y coordinate of AOI at F frame with tracking interval of N respectively, which are the offset points. The results reveal that tracking interval of 10 frames achieves a nearest distance among the test intervals. Take video2 for example. The offset distance with different interval N= 10, 15, 20, 25 compared to interval N=1 in video2 is shown in Fig. 7. We could observe that tracking with interval N=10 and N=15 make less offset distance and more precise in video2. The average offset distances for each frame are 8.88, 10.86, 16.59 and 14.98 for intervals of 10, 15, 20 and 25, respectively.

B. ADVANTAGE OF TRACKING DYNAMIC AOI WITH "AUTOMATIC SETTING MODE"

In order to estimate how many operation times may be saved by using automatic AOI setting mode, number of mouse click and keystrokes were recorded for both automatic setting mode and manual setting mode. The experimental clip is produced from video1 from 2200 to 2400 frames, totally 200 frames. There are 5 participants involved in this experiment. Before the formal experiment, they were taught to use the manual setting mode of our system on the other video. In the experiment, they are required to use the manual setting mode to set the AOIs of human objects for this clip. The operation times of mouse click and keystroke are recorded during their manual setting process. For automatic setting mode, the operation times to set the same human objects as AOIs during this clip were recorded, and the tracking interval N was defined as 10 frames.

TABLE 1. AOI setting operation times by using "Manual Setting mode" and "Automatic Setting mode."

AOI definer	Keyboard operation times	Mouse operation times
Participant 1	33	17
Participant 2	43	22
Participant 3	64	18
Participant 4	29	14
Participant 5	34	17
Automatic (N=10)	2	3

Table 1 shows the comparing results of mouse click times and keystroke times using manual setting mode and automatic setting mode. It can be found that the average operation times with automatic mode were lower than those with manual mode both for mouse click and keystroke. The AOI setting results of video1 by using both manual setting mode and automatic setting mode with the interval N=10 frames were shown in Fig.8, where automatic AOI was shown as red rectangle and Participant 1 to 5 were shown as blue, green, purple, cyan, and yellow rectangles, respectively. The results in fig. 8 indicate that the AOI setting results for both modes are similar and appropriate. This further indicates the reliability of the present methods for tracking dynamic human objects and setting them as AOIs.

C. SYSTEM USABILITY SCALE (SUS) QUESTIONNAIRE

To obtain quantitative feedback on the system, we invited 15 participants to use our system and then asked them to complete a standard usability questionnaire system usability scale (SUS) [28]. All the participants had no experience with the system, and they were instructed to use the system, namely to set the dynamic AOIs with both automatic module

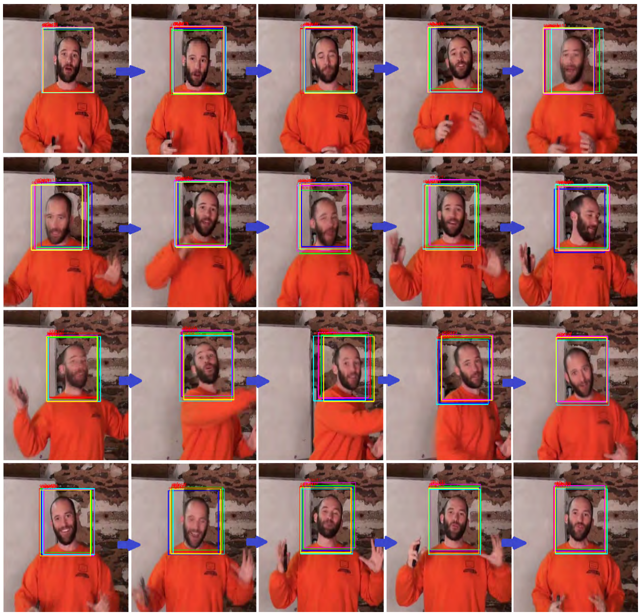


FIGURE 8. AOI setting results of video1 by using "Manual Setting mode" and "Automatic Setting mode."

and manual module. After the instruction, they freely use the system to set the dynamic AOIs for about 10-15 minutes. After the system usage, participants are invited to fill out the SUS questionnaire. The SUS is a ten-item questionnaire giving a global assessment of usability. Each question is a statement and a rating on a five-point scale from 1 (Strongly Disagree) to 5 (Strongly Agree). It has had a great success among usability practitioners since it is a quick and easy-to-use measure of usability [29].

TABLE 2. SUS questionnaire: results of each question.

	Question	Mean	SD
1	I think that I would like to use this system frequently.	4.2	0.89
2	I found the system unnecessarily complex.	1.53	0.69
3	I thought the system was easy to use.	4.6	0.4
4	I think that I would need the support of a technical person to be able to use this system.	1.53	0.41
5	I found the various functions in this system were well integrated.	4	0.57
6	I thought there was too much inconsistency in this system.	1.47	0.41
7	I would imagine that most people would learn to use this system very quickly.	4.27	0.92
8	I found the system very cumbersome to use.	1.33	0.38
9	I felt very confident using the system.	4.4	0.83
10	I needed to learn a lot of things before I could get going with this system.	1.33	0.24
	SUS score	**85.67**	

Results of the SUS study are shown in Table 2, which indicate that the usability results for the system are encouraging. The best-graded question was question 3 and 9, suggesting that the system is very easy to use, and that they are confident using the system themselves. Nevertheless, some users state that the system may still be cumbersome, and that some concepts were necessary to acquire before using the system. Regarding the average SUS score is around 85.67, which corresponds to the adjective rating is about Excellent [30], indicating that the participations a are very satisfied with the proposed system.

D. CASE STUDY

As an illustrative example, we will provide the data of an case study from usability research in video lecture. In this study, we are interested in how the fixation time rate and fixation number represents to each object? Does the fixation duration beyond a certain value indicate an influence on learners' impression formation process to the dynamic learning objects? Furthermore, to demonstrate the proposed system is effective for low-cost eye trackers, the Eye-Tribe is chosen as the recoding device in this experiment. The price of Eye-Tribe is $300, which has a sampling rate of 60 Hz and a tracking accuracy of about 0.5 to 1 degrees of visual angle.

A total of 12 subjects (6 female, mean age = 24) participated in this study. The stimulate material was an 11-minute video picked from cloud computing courses (i.e. video3). Suggested by a cloud computing teacher, 20 objects are defined in stimulate video. The Dynamic AOI Module is used to set this series of selected AOIs. For dynamic human objects, an interval of 10 frames was set for automatic tracking of their AOIs. For other objects, manual module was used to set their AOIs. The AOIs includes human objects, different titles, detail contents and so on. As an example, Fig. 9 shows the screen shot of some defined AOIs.

At the beginning, participants were asked to watch the video lecture. Their eye movement data to the stimulate video were recorded by Eye Movement Recorder connected to an Eye Tribe Tracker. After watching the video, participants were asked to rate their impression for each AOI on a five-point scale ranging from 1 (no impression) to 5 (strong impression). After the experiment, Video Analyzer was used to generate analysis results combining the eye movement data with dynamic AOIs. Finally, visualization methods or statistical data provided by the system will then benefit the eye movement analysis for video lectures. The final analysis shows the calculation results of the three indicators of eye movement for each AOI: fixation time rate, fixation number and the fixation state. The fixation time rate for the whole video is 81.82%. Fig. 10 presents the average fixation time rate computed for all the objects.

The detail fixation information for each AOI is shown in Table 3, including fixation time rate and fixation number. The time rate of AOI "Delivery&Licensing" is the highest (76.90%), which means the participants pay more attention on this area, comparing to other objects. The fixation time rate of brands areas "SAP" (0.42%) and "SAPUA" (1.14%) are the lowest, which may suggest the

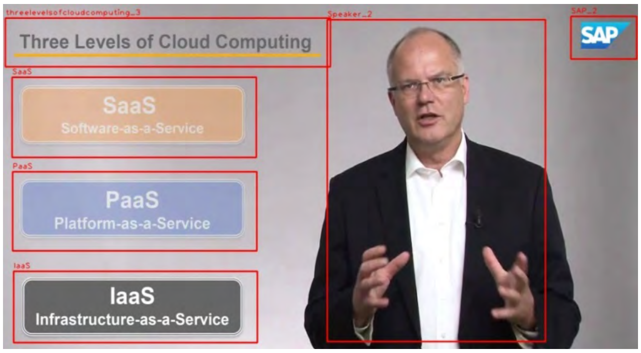


FIGURE 9. The screen shot of some defined AOIs in video3.

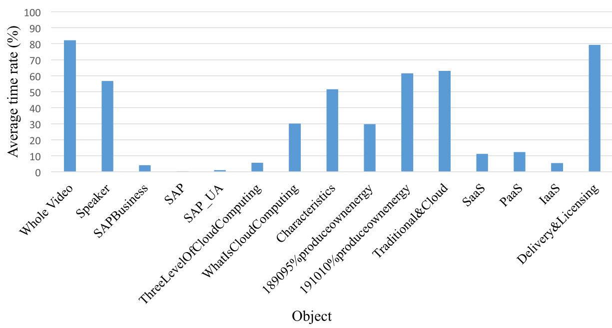


FIGURE 10. The average fixation time rate of each object in video3.

TABLE 3. The average fixation time rate and fixation number of each object in video3.

Object name	Fixation time rate		Fixation	
Speaker_1	63.28	58.22	3.83	21.58
Speaker_2	53.16		17.75	
SAPBusiness_1	4.47	4.4	3.33	9.5
SAPBusiness_2	4.32		6.17	
SAP_1	0.64	0.42	0.17	0.5
SAP_2	0.33		0.33	
SAPUA_1	1.29	1.14	2.33	3.5
SAPUA_2	0.98		1.42	
ThreeLevelsOfCloudComputing_1	11.2	6.13	9	15.5
ThreeLevelsOfCloudComputing_2	5.41		5	
ThreeLevelsOfCloudComputing_3	1.78		1.5	
WhatIsCloudComputing	26.03		10.92	
Characteristics	51.26		14.5	
1890-95%ProduceOwnEnergy	31.11		11.5	
1910-10%ProduceOwnEnergy	60.92		12.17	
Traditional&Cloud	65.16		24.5	
SaaS	13.05		12.08	
PaaS	10.83		10.83	
IaaS	4.45		4	
Delivery&Licensing	76.9		33.75	

concentration of the learning content. The fixation time rate of "1910-10%ProduceOwnEnergy" (60.92%) is much higher than "1890-95%ProduceOwnEnergy" (31.11%). The fixation time rate of "PaaS" is the highest (13.05%), while the time rate of "IaaS" is the lowest (4.45%). Consistency with fixation time rate, the average fixation number of "Delivery&Licensing" is the highest (33.75) among all the AOIs and the average fixation number of "SAP" and "SAPUA" are the lowest (0.5 and 3.5).

TABLE 4. The true positive rate of fixation state to predict impression state for each AOI.

Object name	True positive rate(%)
Speaker	100
SAPbusiness	91.96
SAP	41.67
SAPUA	83.33
ThreeLevelsOfCloudComputing_1	75
ThreeLevelsOfCloudComputing_2	66.67
ThreeLevelsOfCloudComputing_3	58.33
WhatIsCloudComputing	83.33
Characteristics	100
1890-95%ProduceOwnEnergy	100
1910-10%ProduceOwnEnergy	100
Traditional&Cloud	100
SaaS	100
PaaS	100
IaaS	91.67
Delivery&Licensing	100
AVERAGE	87.00

To verify whether the fixation state indicates the impression state for each object, the impression scores were first coded to 0 and 1. The person who evaluating the impression "Strong" and "Very strong" were regarded as "with impression" (coded as 1). The remainder evaluations were regarded as "without impression" (coded as 0). The correct outcomes (true-positive rate) [31] were then calculated for each object, as shown in Table 4. To predict subjective impression state, fixation state achieves the rate of true positive predictions of 87% on average. For most of the AOIs, the correct prediction rates are 100%. For AOI "SAP," fixation state has a lowest correct prediction rate of 41.67% for subjective impression state. The prediction errors all happened to belong to this case: fixation state is 0, while the impression state is 1. One possible reason is that the area of AOI is too small to grab the gaze state. The other reason maybe the less text presented within the area, which means shorter fixation duration is required to make the impression. The prediction rates of AOIs "Three Levels Of Cloud Computing_1," "Three Levels Of Cloud Computing_2" and "Three Levels Of Cloud Computing_3" were all below the average correct rate. The errors of these AOIs were all presented to this case: fixation state is 1, while the impression state is 0. This may due to the similarity of the three AOIs and they were appeared many times in the video, the participations would pay an unconscious attention to these while watching. Overall, the prediction results (87% on average) indicate that the analysis performance of the proposed system for the low cost Eye-Tribe eye tracker is outstanding, and the fixation state could well predict the impression state according to the subjective impression to each AOI rated by participants.

V. CONCLUSIONS

In this paper, the analysis system (VLEYE) for video lectures based on the eye movement of dynamic AOIs is proposed. The proposed system supports both manual and automatic methods for dynamic AOI definition offline. Manual mode allows users to define various objects as AOIs by drawing the rectangles. Automatic mode enables the detection and tracking of human objects as AOIs by system algorithm. Two kinds of operating modes are proposed to satisfy AOI setting requirements for static objects and dynamic objects within the video. In the representation template, the representations for multiple eye movement data combining with dynamic AOIs are shown, as well as heat map visualization representation. Furthermore, the system supports the off-line eye movement data mining or analysis for dynamic objects. The system is verified with a set of lecture video clips. Tracking video with appropriate interval instead of tracking at each frame is confirmed to speed up human object tracking. The out-performance of automatic setting mode of dynamic AOI are verified by comparing with manual setting mode. The SUS results indicate the proposed system rather easily accepted by the users in generally. Furthermore, to support the analysis of dynamic objects in the video lectures, a case study was conducted to verify the effective of eye movement data generated from the system. Especially, among these eye movement data, fixation state is shown to well predict the impression state according to the subjective impression to each AOI rated by participants after watching the lecture video.

VLEYE is being actively maintained and develop, which can be found on GitHub as the following link: https: // github.com/WWChen2310/eye_track_based_video_analyze_system. However, it cannot support the function of automatic detection of text objects so far, user should defined this AIO with manual drawing function in the proposed system. The future work is to implement new functionality, such as the automatic detection of text objects as AOIs, and to extend analysis feathers, such as eye movement measures that can directly reveal the influence of dynamic objects within the video lecture on learners' cognitive and emotion.

REFERENCES

[1] J. Bergmann and A. Sams, *Flip Your Classroom: Reaching Every Student in Every Class Every Day*. Washington, DC, USA: International Society for Technology in Education, 2012.

[2] H. D. Brecht and S. M. Ogilby, "Enabling a comprehensive teaching strategy: Video lectures," *J. Inf. Technol. Educ., Innov. Pract.*, vol. 7, no. 1, pp. 71–86, 2008.

[3] M. N. Giannakos, K. Chorianopoulos, and C. Chrisochoides, "Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course," *Int. Rev. Res. Open Distrib. Learn.*, vol. 16, no. 1, pp. 260–283, 2015.

[4] P. Viviani, "Eye movements in visual search: Cognitive, perceptual and motor control aspects," *Rev. Oculomotor Res.*, vol. 4, no. 4, pp. 353–393, 1990.

[5] F. Schmidt-Weigand, A. Kohnert, and U. Glowalla, "A closer look at split visual attention in system- and self-paced instruction in multimedia learning," *Learn. Instruct.*, vol. 20, no. 2, pp. 100–110, 2010.

[6] H.-C. Liu, "Eye-tracking viewers' processing of Web-based multimedia information," in *Proc. Joint Conf. Pervasive Comput. (JCPC)*, Dec. 2009, pp. 699–704.

[7] J. H. Goldberg and J. I. Helfman, "Eye tracking on visualizations: Progressive extraction of scanning strategies," in *Handbook of Human Centric Visualization*. New York, NY, USA: Springer, 2014, pp. 337–372.

[8] D. S. Wooding, "Fixation maps: Quantifying eye-movement traces," in *Proc. Symp. Eye Tracking Res. Appl.*, 2002, pp. 31–36.

[9] A. Bojko, "Informative or misleading? Heatmaps deconstructed," in *Proc. Int. Conf. Hum.-Comput. Interact.*, 2009, pp. 30–39.

[10] K. Holmqvist, M. Nystrom, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, U.K.: Oxford Univ. Press, 2011.

[11] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension," *Psychol. Rev.*, vol. 87, no. 4, p. 329, 1980.

[12] X. Zhang and S.-M. Yuan, "An eye tracking analysis for video advertising: Relationship between advertisement elements and effectiveness," *IEEE Access*, vol. 6, pp. 10699–10707, 2018.

[13] G. Tien, M. S. Atkins, and B. Zheng, "Measuring gaze overlap on videos between multiple observers," in *Proc. Symp. Eye Tracking Res. Appl.*, 2012, pp. 309–312.

[14] P. Marchant, D. Raybould, T. Renshaw, and R. Stevens, "Are you seeing what I'm seeing? An eye-tracking evaluation of dynamic scenes," *Digit. Creativity*, vol. 20, no. 3, pp. 153–163, 2009.

[15] S. S. Alam and R. Jianu, "Analyzing eye-tracking information in visualization and data space: From where on the screen to what on the screen," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 5, pp. 1492–1505, May 2017.

[16] C. Sennersten. J. Alfredson, M. Castor, J. Hedstrom, and B. Lindahl, "Verification of an experimental platform integrating a tobii eyetracking system with the HiFi game engine," Command Control Syst., Tech. Rep., 2007. [Online]. Available: https://www.researchgate.net/publication/281640602_Verification_of_an_Experimental_Platform_Integrating_a_Tobii_Eyetracking_System_with_the_HiFi_Game_Engine

[17] F. Papenmeier and M. Huff, "DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies," *Behav. Res. Methods*, vol. 42, no. 1, pp. 179–187, 2010.

[18] Y. Zhang and A. J. Hornof, "Mode-of-disparities error correction of eye-tracking data," *Behav. Res. Methods*, vol. 43, no. 3, pp. 834–842, 2011.

[19] R. Jianu and S. S. Alam, "A data model and task space for data of interest (DOI) eye-tracking analyses," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 3, pp. 1232–1245, 2017.

[20] G. J. Zelinsky and M. B. Neider, "An eye movement analysis of multiple object tracking in a realistic environment," *Vis. Cogn.*, vol. 16, no. 5, pp. 553–566, 2008.

[21] M. Friedrich, N. Rußwinkel, and C. Möhlenbrink, "A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: Looking at moving aircraft," *Behav. Res. Methods*, vol. 49, no. 3, pp. 822–834, 2017.

[22] L. Jin, S. Li, X. Luo, Y. Li, and B. Qin, "Neural dynamics for cooperative control of redundant robot manipulators," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2018.2789438.

[23] S. Li, Y. Zhang, and L. Jin, "Kinematic control of redundant manipulators using neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2243–2254, Oct. 2017.

[24] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, vol. 1, Sep. 2002, pp. 900–903.

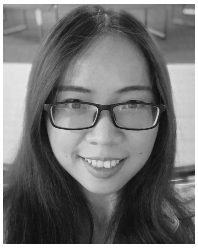
[25] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[26] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL, USA: CRC Press, 1986.

[27] X.-B. Zhang, C.-T. Fan, S.-M. Yuan, and Z.-Y. Peng, "An advertisement video analysis system based on eye-tracking," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom*, Dec. 2015, pp. 494–499.

[28] A. Bangor, P. Kortum, and J. A. Miller, "The system usability scale (SUS): An empirical evaluation," *Int. J. Hum.-Comput. Interact.*, vol. 24, no. 6, pp. 574–594, 2008.

[29] S. Borsci, S. Federici, and M. Lauriola, "On the dimensionality of the system usability scale: A test of alternative measurement models," *Cogn. Process.*, vol. 10, no. 3, pp. 193–197, 2009.
[30] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.
[31] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.



XUEBAI ZHANG received the B.S. degree from Xiamen University, China, in 2011, and the M.S. degree in computer science and information management from Providence University, Taiwan, in 2013. She is currently pursuing the Ph.D. degree with the Department of Computer Science, National Chiao Tung University, Taiwan. She has authored over 10 peer-reviewed international journals and conferences. Her research interests include Internet technologies, digital learning, and human–computer interaction.

SHYAN-MING YUAN received the Ph.D. degree in computer science from the University of Maryland at College Park, College Park, in 1989. In 1990, he was an Associate Professor with the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan.

Since 1995, he has been a Professor with the Department of Computer Science, National Chiao Tung University, where he is also the Director of Library. His current research interests include distance learning, Internet technologies, and distributed computing.

MING-DAO CHEN received the B.S. and M.S. degrees from the Department of Computer Science, National Chiao Tung University, Taiwan, in 2015 and 2017, respectively. His research interests include Internet technologies and digital learning.

XIAOLONG LIU received the B.S. degree from Xiamen University, China, in 2011, the M.S. degree in computer science and information management from Providence University, Taiwan, in 2013, and the Ph.D. degree from the Institute of Computer Science and Engineering, National Chiao Tung University, Taiwan, in 2016.

Since 2016, he has been an Assistant Professor with the College of Computer and Information Sciences, Fujian Agriculture and Forestry University. He has authored over 20 peer-reviewed international journal and conference papers. His research interests include Internet technologies, digital learning, and multimedia security.

Dr. Liu was a recipient of the Taiwan Association for Web Intelligence Consortium Best Dissertation Award in 2016.