HANDS-ON EXPERIENCE IN IMAGE PROCESSING: THE AUTOMATED LECTURE CAMERAMAN

Andrea Cavallaro, Ruchira Chandrasekera, Murtaza Taj

Multimedia and Vision Group, Queen Mary, University of London
Mile End Road, E1 4NS London, United Kingdom

ABSTRACT

We present a system for live recorded lecture-based distance learning delivery that was designed and implemented based on the framework defined in [1]. The system is built by students based on previous students' projects and is deployed in a real distance learning scenario. The video capturing process of a lecture is automated using a robotic camera that tracks the movements of a lecturer during the delivery of a traditional class. The robotic camera is guided by the results of an image processing module based on face detection. The video of the lecturer is synchronized with the presentation slides and with the audio of the lecture. The system was evaluated based on students' feedback.

Index Terms— Image processing, robotic camera, tracking.

1. INTRODUCTION

Laboratory hands-on experiments and projects are a very effective way to learn subjects such as signal processing and computer vision. In [1] we presented the framework we developed and used to enhance the quality of learning in image analysis and computer vision based on the OpenCV library[1]. Providing additional documentation to the library and project examples provides an opportunity for student projects that was not available before. The framework and the documentation are continuously updated with new projects as the number of applications increases. The projects based on the framework produce also a user guide and a functional documentation that complement and enhance the functions available in the OpenCV library. In this paper, we present a student project, the Automated Lecture Cameraman, which was developed to enhance the learning experience to distance learning students. The Automated Lecture Cameraman was selected for its implications in teaching and was evaluated by other students in a real distance learning scenario.

In recent years there has been an increasing interest in effective forms for delivering educational programmes remotely by electronic means. Many of the available distance learning solutions are web-based literature courses. Some of these solutions, such as the MIT OpenCourseWare project[2], are enjoying a considerable success. However, web based solutions have three major drawbacks: they require a lot of self-motivation from the students; they offer a limited student experience; and, if a certain level of interactivity is desired, they require a considerable staff time to prepare the material, which is different from traditional lecture material. Improved e-learning solutions make use of advances in technology, such as broadband access from home and the availability of large storage space on personal computers, which have enabled the delivering of lectures in forms that were not possible few years ago [2] [3]. For example, a number of universities offer to students lecture pod casts or videos displaying the presentation slides synchronised with the voice of the lecturer. The opportunity to capture a video of the lecturer (and not only her or his voice) to be displayed in parallel to the presentation slides is a more effective way to improve students' e-learning experience. However, because these solutions require the presence of one cameraman for each lecture room, they are not economically viable for many universities.

The project presented in this paper aims to overcome this problem with a cost effective solution that automates the video capturing process of a lecture. Our aim is to automate the entire lecture recording production cycle using a robotic camera that automatically tracks the movements of a lecturer during the delivery of a traditional class. The ultimate goal is to provide students with an augmented e-learning experience without the costs that are currently necessary to deliver such an experience. The evaluation of students' feedback includes a comparison with the existing distance learning modality that is offered to the students.

The paper is organised as follows. In Section 2, we discuss related work and define the problem. Section 3 describes the proposed distance learning system and presents sample experimental results. Section 4 introduces the students' evaluation of the project. Finally, Section 5 concludes the paper.

2. PREVIOUS WORK

Existing solutions for speaker/lecturer tracking are based on single camera [2, 4], on multiple cameras [5, 6, 7], or on other types of sensors [5, 8]. Detection techniques include skin color matching [9], motion-based tracking [2], shape-based tracking [10] and also adaptive background modeling [11]. Active camera tracking may use faces to pilot the pan and tilt angles of the camera [4, 9]. A pan-tilt-zoom (PTZ) camera to capture lecturer's movements may also operate by capturing a sub-region (cropped area) from a larger field of view of the camera. In this case, tracking is performed using the horizontal motion histogram of the displaced frame difference [9]. Pan and tilt operations are carried out when motion is detected in the immediate outer region of the cropped area.

Other approaches to video lecture recording combine a static master camera, which handles the detections, with a slave PTZ camera for tracking [6]. Close-ups of the target are obtained using the PTZ camera, based on the information from a static master camera [7]. The use of multiple cameras requires handling the correspondence between object position and cameras using either homography or fixed landmarks, thus making such systems complex in terms of configuration and setup. Non-camera based methods are also used to locate the movements of the lecturer for video recording. Exam-

[1] http://www.intel.com/technology/computing/opencv/
[2] http://ocw.mit.edu/index.html

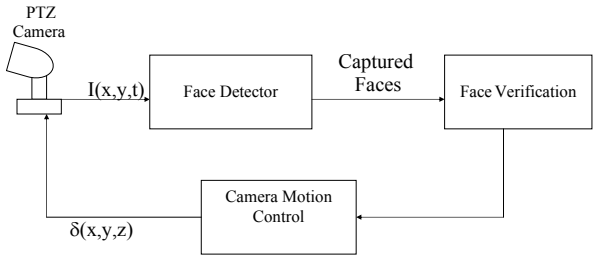

Fig. 1. *The main building blocks of the camera control system based on face detection and face verification*

Fig. 3. *Example of false positive (green) face detection (a) identified (red) and removed by the face verification post-processing (b)*

ples of sensors are microphone arrays [5] and wearable magnetic tags [8]. Microphone arrays may be mislead by background noise in large classes and wearable magnetic tags need to be worn by the lecturer, thus making the acceptance of the system more difficult.

3. THE AUTOMATED LECTURE CAMERAMAN

We aim at reproducing the classroom environment by capturing the presentation and, in a separate window, the lecturer and her/his gestures. We use a single PTZ camera for detection and tracking to make the system more portable and affordable, and let the lecturer move freely in the classrom. The video capturing is based on a robotic IP camera, whose positioning is guided by the results of an image processing module. The module includes a face detector enhanced with a face verification step (Figure 1). The face of the lecturer is detected using a Haar feature-based face detector [12] that uses the integral image, $\mathcal{I}(x, y)$, defined as

$$\mathcal{I}(x, y) = \sum_{i=1}^{x} \sum_{j=1}^{y} I(i, j), \tag{1}$$

where $I(i, j)$ represents the original image intensity. First, features similar to Haar basis functions (Figure 2) are extracted from the integral image. Next, a small number of relevant features are selected using AdaBoost. This learning step selects a small number of relevant features while removing a large number of the available features that do not fit the training samples. This pruning process selects weak classifiers that depend on one feature only. The resulting classifiers are combined in a cascade structure to create a final strong classifier. If there are combinations of structures in the background that fit the Haar features, then false detections are produced that mislead the camera control. For this reason, a post processing is added as to validate the results of the face detector. This post-processing step evaluates the intensity value consistency of the area identified by the face detector.

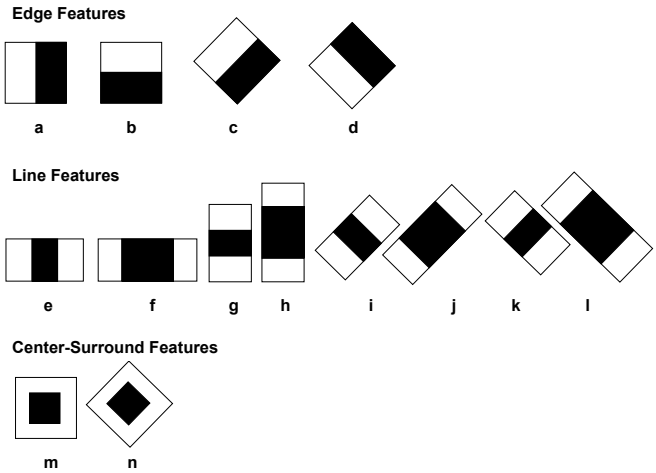

Fig. 2. Haar features for faces. (a-d) Edge features; (e-l) line features; (m-n) are center surround features

Figure 3 shows an example of a false positive detection (Figure 3(a)) that is recognized and rejected by the face verification step. In Figure 4, face detection results in a real lecture scenario captured with the combination of the face detector and the face verification are presented. It is possible to notice how the face is detected under different lighting and pose conditions.

The result of the face detection and post-processing module is a region of interest (ROI) defining the estimated position of the face of the lecturer. The ROI is then used to guide the panning and tilting of the camera in order to align the detected face so that the lecturer remains in the middle of the field of view. We place the upper part of

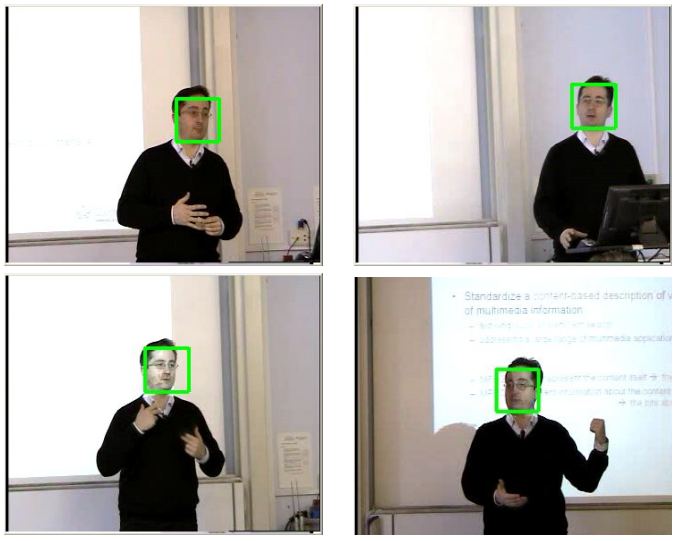

Fig. 4. *Example of face detection results during a lecture. The face detection results are used to control the positioning of the robotic camera*

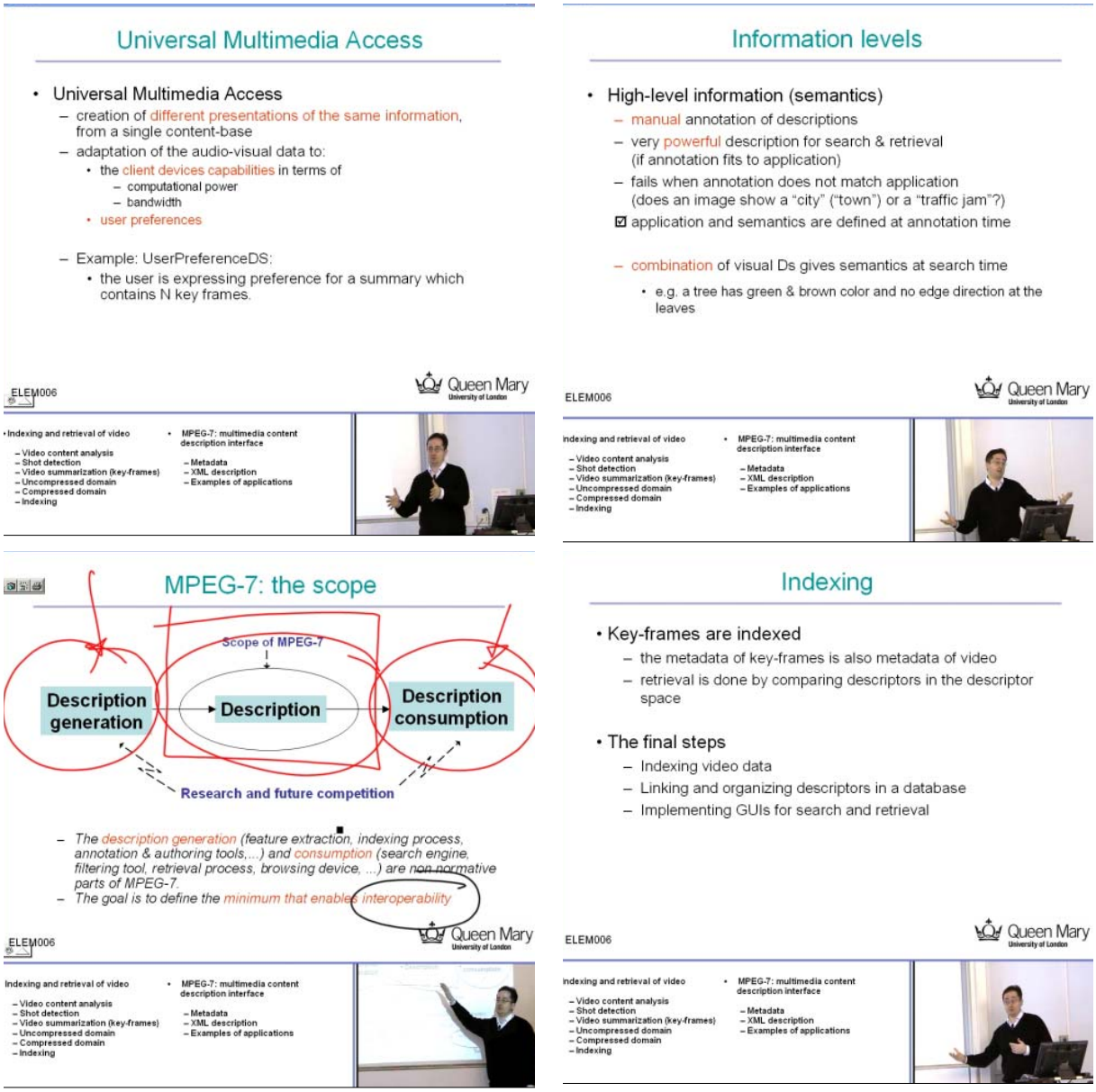

Fig. 5. *Sample snapshots of the distance learning interface*

the body of the presenter (and not only his or her head) in the center of field of view to provide a better coverage of the body language. The control of the camera is performed using a LIFO queue policy, as the image analysis results are generated at a higher rate than the control inputs of the cameras. Let (x_f, y_f) be the detected face center, and (x_c, y_c) the center of the image received as camera input on which detection is performed. If the displacement of the face is larger than a threshold ($T_{ptz} = 80$ for CIF images), then the camera is sent the request to move to the new face position (x_f, y_f).

In the servoing mechanism utilized, two issues needed to be addressed, namely the effects of network data transmission latency (from the IP camera to the computer) and the use of uni-directional camera instruction (i.e., the lack of camera state information). Latency is a significant issue in real-time tracking as it can cause the loss of the target. Because of the lack of camera state information, instructions are sent to the camera sequentially by taking care not to override the previous command, as this causes the camera to perform too rapid movements. Also the temporal gap between two signals is important to enable the computer to rectify any errors in the camera movement. We use a camera controlling instructions frequency of 5Hz. In order to obtain a smooth video output, we have introduced a time delay (250ms) within which the camera may not perform any further movements after the previous one (unless the instruction requests a wide angle movement). Sample results from the resulting automated camera system in real lectures are presented in Figure 5.

4. EVALUATION FROM STUDENTS

The improvement of the students' quality of experience obtained through the students' project presented in this paper was evaluated

Evaluation questionnaire	Agree (%)
I have used RLs of other courses too	62.5
I would benefit if there were more RLs	100
RLs do not offer any advantages to me	16.7
I believe that I can learn more through RLs than through traditional lectures	66.7
RLs without video recording of the lecturer are sufficient for me	37.5
RLs with video recording of the lecturer are useful to me	91.7
A moving camera following the movements of the lecturer is more appropriate than a fixed camera	70.8
Video is of acceptable quality	75.0
I would accept to download a much bigger file if the quality of the video was better	83.3
The electronic message board is useful	95.8
The RLs layout (slides, video of the lecturer, agenda) is good	91.7
I prefer RLs to traditional lectures	50.0
RLs are useful for revision	95.8
RLs save me time	70.8
Rate RLs without video recording of the lecturer	63.6
Rate the RLs with video recording of the lecturer	82.3

Table 1. *Summary of the feedback from students that used the recorded lectures (RLs). The right column shows the percentage of students who agreed with the corresponding sentence on the left column*

with a questionnaire (Table 1) filled by 25 students, including distance learning students from Asia and Europe, and local students. The evaluation includes a comparison with the existing distance learning modality that is offered to the students. Students have been asked their opinion (true or false answer) on the statements presented on the left column of Table 1. The right column shows the percentage of students who considered the corresponding sentence to be true. All students (100%) agreed on the usefulness of recorded lectures (RLs) and an unexpected 50% of the students prefer RLs to traditional ones. The main reason is that RLs allow students (i) to save travel time and (ii) to repeat parts of a lecture that are more difficult. This is confirmed by the 66% of students who believe they can learn more through RLs than through traditional lectures. Also 95% believes that RLs are useful for revision and 83% would accept to download a much bigger file if the quality of the video was better (the reference file was 100MB for a single lecture). Students were also asked to rate RLs with and without video recording of the lecturer. RLs with video scored 20% higher than those without video, despite occasional jerkiness and tracking errors.

To conclude, here are some feedbacks from students about the system: "The quality is pretty good and is enough for the purpose that is serving but it could be better. Very good effort though!"; "I think there is still room for a bit of improvement, and a few image distortions should be taken care of"; "Video quality should be improved"; "It is a good system but the running of the video could be a lot smoother. However it is very useful for revision and I think it would improve exam results overall if all lecturers were recorded".

5. CONCLUSIONS

We presented an example of use of the framework developed in [1] for a real image processing application, the Automated Lecture Cameraman. The outcome of this students' project is used by students themselves in distance learning, which is currently a pilot study used for delivering teaching material to students in remote locations in Asia and in Europe and to local students for revision. The system integrates audio, video and presentation slides from a live lecture and allows the automation of a task that otherwise would require the permanent presence of a staff member (cameraman) and therefore would not be viable for many universities. The distance learning material is playable on a standard personal computer without the need of dedicated software or hardware.

The next students' project based on the proposed framework will focus on the improvement of the camera action movements for smoother pursuit of the lecturer and the reduction of the video frame rate to improve image quality with a different frame rate for the video of the lecturer and for the slide show. A sample Recorded Lectures video together with the tutorials and project documentation are available at http://www.elec.qmul.ac.uk/staffinfo/andrea/edu.html . We hope that this will help both teaching image and signal processing and further enhancing the quality of distance learning delivery at a reasonable cost.

6. REFERENCES

[1] A. Cavallaro, "Image analysis and computer vision for undergraduates," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, 19–23 March 2005.

[2] C. Zhang, Y. Rui, L. He, and M. Wallick, "Hybrid speaker tracking in an automated lecture room," in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, Amsterdam, The Netherlands, 6–8 July 2005.

[3] B. Erol and Y. Li, "An overview of technologies for e-meeting and e-lecture," in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, Amsterdam, The Netherlands, July 2005, vol. 1, p. 6.

[4] D. Comaniciu and V. Ramesh, "Robust detection and tracking of human faces with an active camera," in *Proc. Third IEEE Int. Workshop on Visual Surveillance*, 2000.

[5] Y. Rui, A.p Gupta, and J. Grudin, "Videography for telepresentations," in *Proc. of the SIGCHI Conf. on Human factors in computing systems*, 2003, pp. 457–464.

[6] X. Zhou, R. Collins, T. Kanade, and P. Metes, "A master-slave system to acquire biometric imagery of humans at distance," in *ACM Int. Workshop on Video Surveillance*, November 2003.

[7] C. Micheloni and G.L. Foresti, "Zoom on target while tracking," in *IEEE Int. Conf. on Image Processing*, September 2005.

[8] S. Mukhopadhyay and B. Smith, "Passive capture and structuring of lectures," in *Proc. of the seventh ACM Int Conf. on Multimedia*, 1999, pp. 477–487.

[9] J. Yang and A. Waibel, "A real-time face tracker," in *Proc. of Workshop on Application of Computer Vision*, 1996.

[10] A. M. Baumberg and D. C. Hogg, "An efficient method for contour tracking using active shape models," Tech. Rep., April 1994.

[11] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians," in *Proc. of the Second IEEE Workshop on Visual Surveillance*, 1999, p. 74.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2001.