Detection of Dozing Students from an Ordinary Classroom Scene with CV Technique

Taichi Kuriyama^a, Akira Suganuma^{b,*}

^a Advanced Production and Information Systems Engineering Course

National Institute of Technology, Ariake College, 150 Higashi-Hagio Omuta, Fukuoka, 836-8585, Japan

^b Division of Human and Welfare Engineering

National Institute of Technology, Ariake College, 150 Higashi-Hagio Omuta, Fukuoka, 836-8585, Japan

*Corresponding Author: suga@ariake-nct.ac.jp

Abstract

Abstract—In the recent years, a development of the Internet and information technique is enable to use IT devices in an ordinary classroom. One of them is videoing with a camera. We decided to take a video of a lecture with information about students who is taking part in it. The proposed system takes a video so that student's face comes to the front. The system extracts a rectangle around each student from each frame in the video. It estimates state of student by using the face detection technique and the flesh color detection one. The system assumes the flesh color of Japanese people. In this paper, we describes a method detecting dozing students. A simple experiment is performed with 90-minute lecture. In the experimental result, state of student is effectively estimated. The system takes a class with a single camera and detects dozing of multiple people.

Keywords: image processing, lecture support, detecting dozing students, state of student.

1. Introduction

In the recent years, a development of the Internet and information technics is enable us to use IT devices in an ordinary classroom. One of them is to take a video with a camera. Another is to get useful information of students taking a class. Many systems, however, require IT device, such as a clicker or a tablet PC, assigned to each student for an input device. Moreover we still hold a lecture in the form of a traditional group class. In this form, it isn't easy to prepare the input device assigned to each student.

We developed a system to take a video of lecture scene

including information of state of the teacher during the lecture⁽¹⁾. This information is useful for searching particular scene and analyzing student attendance situation. We have developed the system on condition that we use a small, inexpensive camera. We used the open license library such as OpenCV, which facilitates the making of the image processing program.

We have developed a system to detect students who are dozing during a lecture. Our system takes a video of students taking a class by the camera placed in front of them. The system extracts each student from each frame in the video. It detects student's behavior from information obtained by the face detection method and the flesh color detection. There are two functions in our final system. The first is to detect students in real time and warn them awake based from the number of students in a dozing state. The second is to use as a tool for a teacher's reflection after lecture. For example, when many students are dozing, the lecture at that time may be boring. It is, therefore, preferable to change the way of explaining in the next lecture.

2. System Overview

Our system consists of a PC and a steady camera. We developed the system with Visual Studio so that it runs on Windows running on the PC. In this research, we coded the program of our system with the language C++. Our system uses some functions of OpenCV⁽²⁾, which is a library for image processing developed by Intel Corporation.

Our system has a steady camera in a place where it can take pictures including approximately one-third of the

Fig. 1. Example image of lecture scene.

Fig. 2. An image extracted with mouse. classroom area. A sample shot is shown in Fig. 1. This camera is connected to the PC with an USB cable.

3. Detect Students

3.1 Overview of the extracting method

Our system extracts the students attending lecture by combining three techniques. One is the method to extract each student. Our camera takes a video of the approximately 10 students. The rectangular domain that surrounded only one student is advantageous to the following two methods. The second technique is the face detection. For the face detection method, a Haar feature based cascade type classifier was used. The last one is the detecting a flesh color region. Many students attend a lecture while exposing a face and a hand. Therefore, several pixels of the students in the image would be flesh color. We combine these three methods to detect student's face movement. For example, a dozing student faces down for a long time. A chatting student has face sideways for a long time.

3.2 Extract a Student from the Lecture Scene

Our system detects multiple students with one camera. It is, however, difficult to detect more than one person at a time and distinguish individuals. In the first step, we decided to extract one student by user's mouse operation as

Fig. 3. An image succeeds face detect.

Fig. 4. Result of extracting flesh-color-pixels.

an experimental operation to move the system. Fig. 2 shows a rectangular image cropped by the above operation from the photo of Fig. 1.

There are caveats in this operation. Our system performs detection using human flesh-color area. Therefore, it is preferable to operate so that the student's hand information is not reflected as much as possible.

3.3 Face Detection

In the second step, face detection is performed on the rectangular image extracted in the first step. The face detection method uses the Haar feature based cascade type classifier. This detection method was published by Paul Viola and Michael Jones in 2001⁽³⁾. This method is based on machine learning. This learning used the Haar features based on a large number of positive example images and negative ones.

Our system uses face detectors provided by OpenCV. Fig. 3 shows the image successfully detected for Fig. 2. Our system does not perform face detection afterward if it succeeds in one frame. In the system, the student's face is necessary to get size of his face in the frame. Once the size was gotten, our system estimates state of student by using it after that frame.

3.4 Flesh-color detection

In the third step, the flesh colored pixels are detected from the rectangular image. Our system generates the image which extracted only pixels whose color is the student's flesh-color. The result of applying the flesh-colored region detection to the image of Fig. 2 is shown in Fig. 4. Our system, firstly, converts the color system of the image in order to be easy to detect

Fig. 5. Binary image after the noise reduction.

Fig. 6. Face extraction image.

flesh-colored pixels. The image captured by the steady camera is expressed in RGB color system. On the other hand, HSV color system is superior to RGB one about the color search⁽⁴⁾. Therefore, our system converts the image into HSV color system.

We previously decided an upper limit level and a lower limit level of the flesh color that we wanted to detect. Our system decides that the pixels whose value is between the lower limit and the upper limit are flesh colored ones. They are output the white pixel value into the result image. Other pixels are output the value of the black pixel.

Our system detects flesh color region assuming Japanese. But even Japanese have different flesh-color. Our system performs flesh color detection in the next frame based on the average value of the detected flesh-color. Therefore, our system generates the image which indicates the flesh-colored part of the student.

3.5 Face Extraction

Our system performs student detection using flesh-color region. Therefore, even hands are detected. They adversely effect in our system. Our system removes them using two methods. One is noise reduction processing for opening processing and closing processing. An isolated small white lump can be erased by closing processing. Also, in the opening process, you can fill the detected lump. The result of applying is shown in Fig. 5.

The second method performs labeling processing and detects the size of each lump. Our system outputs the largest of detected lump. We think that the flesh color region of the face is the largest. The result of applying the above process to the image of Fig. 5 is shown in Fig. 6.

3.6 Detection of Dozing Students

Our system detects dozing students. Most dozing students face down for a long time. Our system preserves the face size of the frame where face detection succeeded. We set the size of the face of the frame for which face detection succeeded as the reference value. Our system sets the size of the dozing state from the reference value. Our system judges that students are dozing when the size of dozing continues for a certain period of time. Also, the time during which the dozing state is judged is set as the dozing check time. Our system sets it as dozing when it is less than half of the reference value at present. Also, the doze check time is set to 1 minute. The above two values can be changed by the user.

3.7 Feedback to Teacher

Our system is also used for reviewing the lecture. Our system saves detection data in relation to the time axis. At that time, when judging that the system is dozing, the dozing check time is also stored as a dozing state and the data is saved. Our system has set it to dozing when it is less than half of the current reference value.

4. Experimental Evaluation

4.1 Detection in Real Time

We took a video of a lecture which is held for 90 minutes. We decided that the video of the lecture is stored as a sequence of the still image without compression. We stored an image every 0.5 seconds. We sorted the saved images into 3 states: staying up state, dozing state, and dozing check state. Table 1 shows a state in which the dozing state and the state of not dozing judged by the human and the system are separated for each frame. The total number of frames is 28,500 frames, which is the sum of 9,500 frames photographed in the lecture for three people. The not doze state the system has determined also includes the dozing check state.

The recall, which is ratio of the frames judged as dozing by our system among the frames judged as dozing by human, is calculated by equation (1). On the other hand, the precision, which is ratio of the frames judged as dozing by human among the frames judged as dozing by the system, is expressed by equation (2).

$$(2,706 / 4,852) * 100 = 55.8 \% \quad (1)$$

$$(2,706 / 2,859) * 100 = 94.6 \% \quad (2)$$

Table 1. Determination of student status by system and human (real time).

		human		total
		dozing	not doze	
system	dozing	2,706	153	2,859
	not doze	2,146	23,495	25,641
total		4,852	23,648	28,500

Fig. 7. Detection of personnel and system (real time).

Next, we examined the status of the students from all frame. We synchronized the state examined and the state detected by the system with the time axis. Fig. 7 is a graph of data obtained by the above work performed on one student. The blue vertical bars mean that a human presumed the student is dozing. The orange ones mean that the system estimated he is dozing. The graph shows that oversight happens. The recall obtained by equation (1) was 55.8%. The value was not so good. Our system enters a dozing check time of 1 minute when the student begins to doze. Therefore, our method miss the frames of the time when the student have begun to doze. However, if the system can determine the dozing state in a frame somewhere the student is dozing, the teacher can wake the student.

4.2 Detection After Lecture

A teacher can use our system to review his lecture. Our system enters the dozing check time for 1 minute when detecting dozing. As a result, if the system judges that the student is dozing, the dozing check time should also be estimated dozing. Table 2 summarizes the frames of the dozing state and the dozing state judged by human and the system in consideration of the dozing check time.

The recall, which is ratio of the frames judged as dozing by our system among the frames judged as dozing by human, is calculated by equation (3). On the other hand, the precision, which is ratio of the frames judged as dozing

Table 2. Determination of student status by system and human (after lecture).

		human		total
		dozing	not doze	
system	dozing	4,006	647	4,653
	not doze	846	23,001	23,847
total		4,852	23,648	28,500

Fig. 8. Detection of personnel and system (after lecture).

by human among the frames judged as dozing by the system, is expressed by equation (4).

$$(4,006 / 4,852) * 100 = 82.6 \% \quad (3)$$

$$(4,006 / 4,653) * 100 = 86.1 \% \quad (4)$$

The recall obtained by equation (3) was 82.6%. Therefore, in the after time, the recall when considering the dozing check time is greatly improved compared with the case of the detection in the real time.

Fig. 8 is a graph showing the doze detection accuracy when the dozing check time is considered. It can be confirmed that there are more orange regions in blue than in Fig. 7.

5. Conclusions

In this paper, we proposed a system that detects student's dozing states only from lecture images by using two methods: face detection by machine learning and flesh color region detection. It is possible to grasp the states of the student in the image captured by the steady camera. Consequently, our extracting method is useful for state of student estimation. We were able to grasp the state of the students in the lecture. However, it cannot be said that it can be accurately detected due to the influence of the student sitting behind the student to be detected. Also, because of

detection using color, hands may be erroneously detected. Therefore, it is necessary to devise a method that can accurately detect the state of the students without being influenced by the students behind. We have a future work that we will design a system to improve the recall of dozing students. Also, we have a future work to detect students who are operating smartphones, and detect students who are talking about. Then, detection data of many people is acquired.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP16K00502.

References

- (1) T. Nakayama, T. Noguchi, and A. Suganuma, "Extracting a Teacher on the Platform for Estimating His Motion from a Lecture Scene," 4th International Conference on Applied Electrical and Mechanical Engineering, 2017, pp.59–62.
- (2) G. Bradski and A. Kaehler, (translation: K. Matsuda), "Learning OpenCV: Computer Vision with the OpenCV Library," OReilly Japan, Inc., Aug. 2009 (in Japanese).
- (3) P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Conference on Computer Vision and Pattern Recognition, 2001, pp.1–8.
- (4) Y. Adachi, A. Imai, M. Ozaki and N. Ishii, "Study on Extraction of Flesh Colored Region," Transactions on Electrical and Electronic Engineering C, Vol.120-C, No.12, Dec. 2000, pp.1927–1932 (in Japanese).