

Blackboard Segmentation Using Video Image of Lecture and Its Applications

Masaki Onishi, Masao Izumi and Kunio Fukunaga

Department of Computer and Systems Sciences, College of Engineering,
Osaka Prefecture University, 1-1 Gakuen-cho, Sakai, Osaka, 599-8531 Japan
onishi@com.cs.osakafu-u.ac.jp, {izumi, fukunaga}@cs.osakafu-u.ac.jp

Abstract

We propose a method for segmentation of written regions on a blackboard in the lecture room using a video image. We firstly detect static edges of which locations on the image are stationary. And next, we extract several rectangular regions in which these static edges are located densely. Finally by use of fuzzy rules, extracted rectangles are merged as contextual regions, where letters and figures in each contextual region explain each context. And, we apply our method to automatic production of lecture video, and archives system of written regions on the blackboard in lecture rooms.

1. Introduction

Recently, universities and other kinds of schools begin to take an interest in a distance learning system, which can perform lectures at various locations far from lecturers at the same time [1]. State of the art low cost wide band communication system makes such systems realistic. On the other hand, usual lecture videos and lecture notes can be used for similar objective. And also lecture videos can be broadcasted at the learning system mentioned above. For that reason, automatic systems to make these contents have been applied to real lectures and technical talks in recent works [2, 3].

When you make lecture videos, it is important to take a shot as students are able to look at letters and figures on blackboard. Even if students want to take closer look only at the part of the blackboard, the conventional method tends to make video images which include the lecturer only or the whole blackboard. When we make lecture archives, we also record letters and figures on blackboard. Considering making lecture videos and archives automatically, it is necessary to divide blackboard into several written regions.

Here we propose a method of blackboard segmentation. And we propose a procedure for making lecture videos by

use of proposed blackboard segmentation method. We confirm effectiveness of our method by experimental results. And we demonstrate to produce a lecture videos and WWW contents.

2. Separation of lecturer and characters

For understanding a lecture from lecture video images, there are two important objects in lecture video images, a lecturer and characters on blackboard written by the lecturer. (Here, we call letters and figures as 'characters'.) To separate them from video images, we use a method based on edge detection from spatiotemporal images of a video image sequence, because it is difficult to separate edges of lecturer and edges of written characters from an input image only. Since a lecturer is constantly moving around to explain and write characters on a blackboard, we detect edge images from a spatiotemporal image. The spatiotemporal image took advantage of several researches to detect camera operations [4], because edge direction of the spatiotemporal image shows moving flow of the camera. We use edge directions from several cross sections of the spatiotemporal image, in order to segment two kinds of edges, edges with moving objects (We call *dynamic edges*.) and edges with stationary objects (*static edges*).

A cross section perpendicular to the x axis of the spatiotemporal image $F(x, y, t)$ is called y - t cross section. And a cross section perpendicular to the y axis is called x - t cross section. The moving object such as a lecturer tends to make horizontal edges in x - t cross section and vertical edges in y - t cross section as shown in Figure 1. On the contrary, static objects such as characters on blackboard tends to make vertical edges in x - t cross section and horizontal edges in y - t cross section. Here we define intensity of a *dynamic edge* using Sobel operators.

$$|\nabla f^d(x, y)| = \sqrt{f_t(x, t | y)^2 + f_t(y, t | x)^2}, \quad (1)$$

where $f(x, t | y)$ and $f(y, t | x)$ are x - t and y - t cross section respectively, and f_t is a differentiation of f . Similarly,

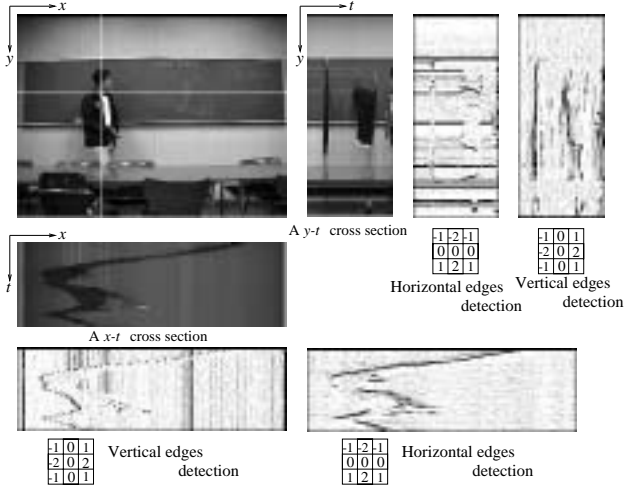


Figure 1. Edge detection of cross section.

intensity of a *static edge* is defined as follows:

$$|\nabla f^s(x, y)| = \sqrt{f_x(x, t | y)^2 + f_y(y, t | x)^2}. \quad (2)$$

However, there is also a case when a lecturer keeps standing without movement for a short period time. Then we assume that *static edges* keep static positions for several periods of time. The static edges are extracted as points which keep certain periods of time as edge points extracted by use of Sobel operators to both $x-t$ and $y-t$ cross sections of the spatiotemporal image.

Figure 2 shows an input image (a) and a *static and a dynamic edge image* (b). We can assure that characters on blackboard are extracted as *static edges* (gray level) and a lecturer is extracted as *dynamic edges* (black level) on usual lecture video images taken by fixed camera. This assumption can be confirmed in Figure 2 (b).

3. Blackboard segmentation

In this section, we extract *written blocks* by using *static edges*. Each *written block* is a unit area and holds letters and figures which, explain the same subject. *Written blocks* are characterized by the clock time when appeared on the video images, and ordered in time sequence. So ordered *written blocks* can be used as a resume of the lecture and so on. We assume that each *written block* consists of several *written rectangles*, which are formed by a following procedure.

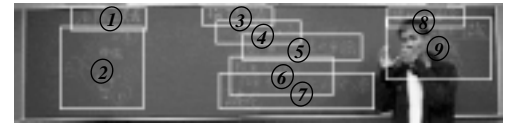
A *written rectangle* is defined as a circumscribed rectangle of *static edges*, which are densely extracted by the method explained in the previous section. While *static edges* are being extracted in time consequently and densely, *written rectangle* is becoming larger. If dense *static edges*



(a) input image.



(b) static and dynamic edge image.



① time 171, no 155	⑥ time 1515, no 212
② time 332, no 283	⑦ time 1653, no 186
③ time 590, no 103	⑧ time 1994, no 241
④ time 630, no 193	⑨ time 2248, no 132
⑤ time 1002, no 220	

(c) Extracted rectangles and their information.

Figure 2. Example of processing result.

are not extracted during certain time period, and if a size and a location of already extracted *written rectangle* are not changed, the *written rectangle* is fixed and stored in terms of the time of being extracted, the location, the size, and the number of *static edges* which exist within the rectangular area on the blackboard. And the number of *static edges* decreases when a lecturer erases some of letters or figures on the blackboard. If the number of *static edges* decreases rapidly and becomes close to zero, the *written rectangle* which holds such erased *static edges* is assumed as erased. Figure 2 (c) shows results of extract *written rectangles* on blackboard (white lines) and their appeared time and number of *static edges*.

In the next step, we try to extract *written blocks* by combining several *written rectangles* in considering time sequential and spatial situations of them. Usually lecturers draw letters and figures on the blackboard as a multi-column article. They tend to write letters from the top of the blackboard to the bottom to make another column on the blackboard. With this consideration, we assume next two rules that the blocking procedure must obey.

- (1) If a presently considering *written rectangle* is located near the top of the blackboard and already extracted *written block* is located to the bottom of the blackboard, a considering *written rectangle* is not combined to the extracted *written block* and becomes a part of new *written block*.

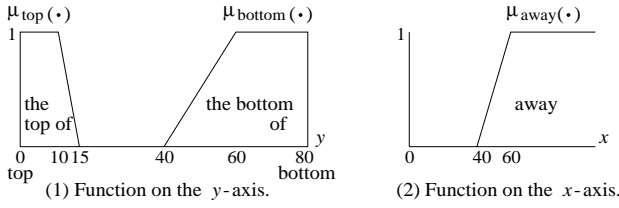


Figure 3. Membership functions.

- (2) If a presently considering *written rectangle* is located horizontally far from the already extracted *written block*, a considering *written block* becomes a part of new *written block*.

Fuzzy rules in Figure 3 are introduced to make above two rules apply a blocking procedure. The argument of function $\mu_{\text{top}}(\cdot)$ is y element of the location of presently considering *written rectangle*, the argument of $\mu_{\text{bottom}}(\cdot)$ should be y element of the location of the bottom of the extracted *written block*. And the argument of $\mu_{\text{away}}(\cdot)$ is the smallest distance between presently considering *written rectangle* and other rectangles. If $\mu_{\text{top}}(\cdot) \times \mu_{\text{bottom}}(\cdot) \geq \text{threshold}$ or $\mu_{\text{away}}(\cdot) \geq \text{threshold}$, presently considering *written rectangle* should not be combined with already extracted *written block* and becomes a part of new *written block*.

4. Its Applications

We have proposed the method of blackboard segmentation. Next we demonstrate an example of applications using this method, such as an automatic production of lecture videos and lecture archives. Our systems use two cameras, the one is fixed and takes images for block extraction, and another one is for taking lecture video and can be panned, tilted and zoomed. We call the former *fixed camera*, and the latter *controlling camera*.

(i) Automatic production of lecture videos

Our system automatically decides the area should be viewed in a lecture video, on the location of extracted blocks and a lecturer. Students tend to look at a lecturer and written regions of blackboard during a lecture. The lecturer can be extracted as a region of *dynamic edges*, and written regions of blackboard are extracted some *written blocks*. So, in our system it is assumed that images in a video sequence should include a lecturer (*dynamic edges*) and some of written regions on the blackboard (*written blocks*). In the case that there are plural blocks on blackboard, we have to decide which the lecturer is using explaining or writing (we call *the current blocks*). The block, which includes a new *written rectangle*, is regarded as a *current block*, because it is supposed as an unfinished block. And the block, which

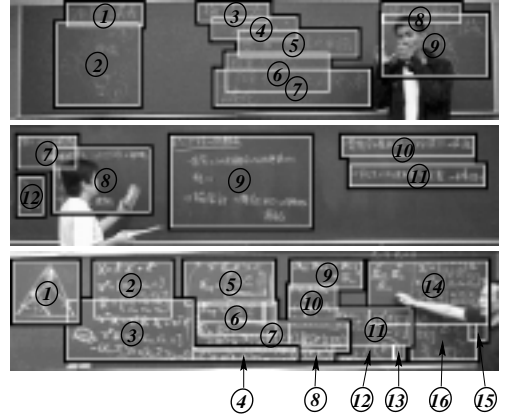


Figure 4. Experimental results of block extraction.

includes *dynamic edges*, is also regarded as a *current block*, because it is supposed to be used for explaining by a lecturer. *The controlling camera* changes a camera angle and zooming parameter to take a shot such that we can see a lecturer and *current blocks*.

Dynamic edges and *written blocks* are extracted from input image sequence taken by *the fixed camera*. Panning, tilting and zooming of *the controlling camera* are manipulated in order to produce lecture videos. In this experiment, we assume only one *controlling camera*, and the positions of *the controlling camera* and *the fixed camera* are known in advance.

(ii) Automatic production of lecture archives

Also our system can produce archives of the lecture as taking a note or a resume of the blackboard. Using results of *written blocks* extraction from fixed camera images, *the controlling camera* can take each *written block* as large as possible. When there is no *dynamic edge* in the *written block*, *the controlling camera* takes a *written block* on

Table 1. Rate of correct block extraction.

	Time (m's'')	Rate (%)	Error of Segmentation			
			OS	OU	OE	NE
Mock	69'29''	94.8%	0.7%	3.0%	0.0%	1.6%
Real	170'53''	83.1%	1.9%	13.8%	1.2%	10.1%
Sum	240'22''	86.5%	1.5%	10.7%	0.8%	7.6%
	No (no)	Rate (%)	Error of Segmentation			
			OS	OU	OE	NE
Mock	87	96.6%	1.1%	2.3%	0.0%	—
Real	90	86.7%	3.3%	8.9%	1.1%	—
Sum	177	91.5%	2.3%	5.6%	0.6%	—

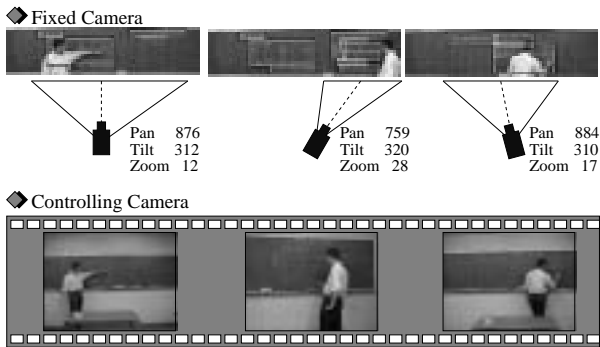


Figure 5. Example of demonstration of lecture video.

blackboard. And in the case that a lecturer adds new letters in a written block, the controlling camera retakes the written block by detecting an increase of the number of static edges. These written block images can be used as wide accessible images for reviewing, such as Web pages demonstrated in a following section.

5. Experimental results

We test our algorithm of blackboard segmentation on 6 mock lectures (3 lecturers, total about 60 min.) and 3 real lectures (3 lecturers, total about 170 min.) image sequence. Figure 4 shows extracted *written rectangles* surrounded by white lines and *written blocks* by black lines¹. The *written rectangles* are numbered in order of finishing to be written on time scale. And Table 1 shows the rate of correct *written block* extraction. From top to bottom are ratios in terms of whole lecture time, and ratios in terms of the number of rectangles. Error of segmentation can be classified into four categories: over segmentation (OS), over merging (OM), over extraction (OE) and no extraction (NE). OS is the case where the region of a single block is erroneously segmented into two or more regions, OM is the case where the regions of multiple blocks are merged into one region. And, OE is the case where others than characters are extracted erroneously, NE is the case where written characters can not be extracted.

Besides, Figure 5 shows a demonstration of lecture videos which is produced automatically. The upper row shows results of *written blocks* extraction, and parameters of 'Pan', 'Tilt' and 'Zoom' are calculated from the coordinates of *dynamic edges* and *current rectangles*. The lower row shows produced lecture video from the controlling camera which is set parameters of 'Pan', 'Tilt' and

¹See: <http://www.com.cs.osakafu-u.ac.jp/~onishi/research-e.html>

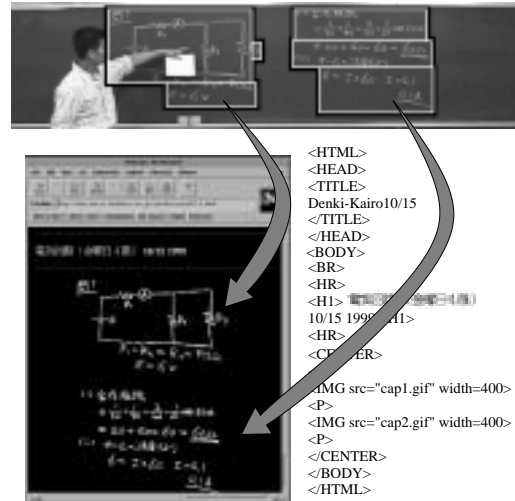


Figure 6. Example of demonstration on WWW.

'Zoom'. And Figure 6 shows a demonstration of lecture archives.

6. Conclusions

In this paper, we proposed a method to separate the written region on blackboard into dense segments using fuzzy rule. The edge detection from spatiotemporal image divides into moving objects and static objects. And the rectangles which enclosed edges of the static objects as written region on blackboard are separated into several blocks with regard to time sequential and spatial situations. As we have shown in experiments, the proposed method yields reliable blackboard segmentation on mock and real lectures. We assume that these results are quite useful to understand the circumstances of lectures. And as examples of its application, we demonstrated automatic production of lecture videos and lecture archives.

References

- [1] S. C. Brofferio, "A University Distance Lesson System: Experiments, Services, and Future Developments," *IEEE Trans. on Education*, vol.41, no.1, pp.17–24, Feb. 1998.
- [2] Y. Kameda, H. Miyazaki and M. Minoh, "A Live Video Imaging for Multiple Users," *Proc. of ICMCS'99*, vol.2, pp.897–902, July 1999.
- [3] S.X.Ju, M.J.Black, S.Minneman, D.Kimber, "Summarization of Videotaped Presentations: Automatic Analysis of Motion and Gesture," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.8, no.5, pp.686–696, Sep. 1998.
- [4] F.M.Idris and S.Panchanathan, "Spatio-Temporal Indexing of Vector Quantized Video Sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.7, no.5, pp.728–740, Oct. 1997.