Proc. 15th International Conference on Pattern Recognition, Volume 4, pp.102–105, Sep. 2000.

Production of Video Images by Computer Controlled Camera Operation Based on Distribution of Spatiotemporal Mutual Information

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

This paper define a spatiotemporal mutual information on the pixels of a given video image on the basis of information theory (Shannon's communication theory), which can be interpreted as the theoretical estimation of interested spots for human being. As an application of this spatiotemporal mutual information, we propose a method of producing a vivid video image of the distance learning by using the computer controlled camera operation and switching of plural camera images on the basis of the video image processing. The results of questionnaire survey for the produced video image confirm the effectiveness of our approach.

1. Introduction

Recent years, high-speed digital data transmission becomes possible accompanied with development of network technique, and a distant learning system, which uses communications satellite or ISDN network, is developed not only at universities but other kind of school and companies [1] [2]. One of the representative methods takes video image of distant learning by a fixed camera. On the other hand, there is a method taking video image using plural cameras and switch the images by a director. Problems encountered in the former method are easy to be a monotone video image and lack of real atmosphere of classroom. Many operators such as cameramen and directors are necessary in the latter method.

In this paper, we propose a method of producing the most suitable video image for members at a virtual classroom by controlling camera direction and zooming, and switching of plural camera images. In the first, we define a spatiotemporal information for each pixel of the video image taken by fixed camera using the Shannon's communication theory. This information reflects a degree of interest in the pixels of the video image for the members of the virtual classroom. In the next, we examine the distribution of spatiotemporal information on the video image, then control the camera point of view to focus on a spot with large distribution. And, by choosing the most suitable video image among plural camera images under the switching rule, which was decided beforehand, the automated camera controlled system and switching system produces a vivid video image, which includes necessary information for the members of the distant learning class.

2. Distribution of spatiotemporal information

2.1. Probability density function of color parameter

We define an appearance probability of a color parameter of a given pixel $\boldsymbol{P}_i = [\, x_i \; y_i \,]^T$ in the form of a probability density function based on distribution of color parameters of the neighborhood pixels, and suppose that the probability function of color parameter $\boldsymbol{X}_i = [\, R_i \; G_i \; B_i \,]^T$ of the pixel $\boldsymbol{P}_i$ follows a 3-dimensional normal distribution function.

$$p(\boldsymbol{X}_i) = \frac{1}{(2\pi)^{\frac{3}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{X}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})\} \tag{1}$$

Here, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ denote a mean value and a covariance matrix of the color parameter $\boldsymbol{X}_i$ at the neighborhood pixels and are given by following equations.

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_k \tag{2}$$

$$\boldsymbol{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} (\boldsymbol{X}_k - \boldsymbol{\mu})(\boldsymbol{X}_k - \boldsymbol{\mu})^T \tag{3}$$

As the probability $p(\boldsymbol{X}_i)$ is defined by the distribution of the color parameter of the pixel at the neighbor region, this value can be interpreted to be a degree of appearance of the color parameter $\boldsymbol{X}_i$ at the pixel $\boldsymbol{P}_i$.

2.2. Spatial information and temporal information

Here let us define generating information of the pixel of an image. The information $H(\boldsymbol{X}_i)$ of pixel $\boldsymbol{X}_i$ is given by the following equation (1). This definition of the information of the pixel is based on the definition of information, which is given by the Shannon's communication theory.

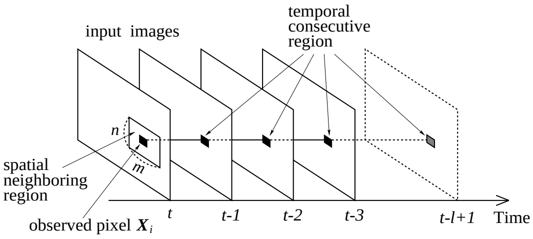
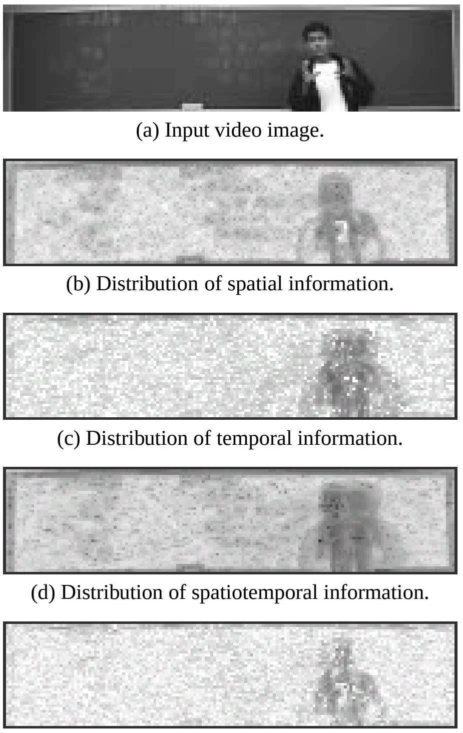

input images
temporal consecutive region
n
m
spatial neighboring region
observed pixel $\boldsymbol{X}_i$
t
t-1
t-2
t-3
t-l+1
Time

Figure 1. Neighboring and consecutive region.



(a) Input video image.

(b) Distribution of spatial information.

(c) Distribution of temporal information.

(d) Distribution of spatiotemporal information.

(e) Distribution of mutual information.

Figure 2. Distribution of mutual information.

$$H(\boldsymbol{X}_i) = -\log_2 p(\boldsymbol{X}_i) \tag{4}$$

This information $H(\boldsymbol{X}_i)$ is getting large as the value of $p(\boldsymbol{X}_i)$ decreases. We can examine the distribution of the generating information of an image by calculating the information of each pixel of the image.

In this paper, we discuss two kind of information, one is originated in a spatial variation of the color parameter of the pixels of an image and the other is temporal variation of the color parameter at a fixed pixel. The spatial information is given by the variation of color parameters on an image at the time t, and is getting large in general in the case when the pattern of an image becomes complicated. On the other hand, the temporal information is defined by the variation of the color parameters along the time axis at the fixed pixel $\boldsymbol{X}_i$.

To show the relationship of the spatial and temporal information, let us examine Figure 1. The figure shows $m \times n$ pixels along the spatial direction on one frame image and l pixels along the temporal direction, namely last l frame images. Taking the $m \times n$ neighbor pixels on a frame image into account, the information $H_s(\boldsymbol{X}_i)$ represents the spatial information. In the case when the l pixels along the time axis at $\boldsymbol{X}_i$, the temporal information is given by $H_t(\boldsymbol{X}_i)$. Figure 2 (a) shows an input gray image at time t, (b) shows the distribution of spatial information and (c) gives the temporal information respectively.

2.3. Mutual information between spatial and temporal information

It is necessary that the video image of the lecture include the spot, which is focused by members at the virtual classroom as much as possible when taking a video image. The necessary spot of the scene for the persons at the virtual classroom is a region composed by pixels with both spatial and temporal information, such as the region of a lecturing person (lecturer) with a motion or letters just written by the lecturer.

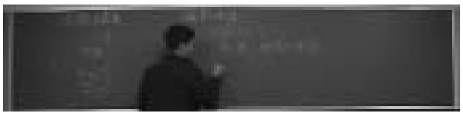
On the other hand, the degree of interest in a region of the letters written on the blackboard at a long time ago has been reduced even if the spatial information is still large. There is also no interest in the erased region on the blackboard even if enough temporal information is large. By applying a way of thinking of mutual information between spatial and temporal information, we define mutual information $I(\boldsymbol{X}_i)$ of the region to which the members at virtual classroom pay attention by the following equation.

$$\begin{aligned} I(\boldsymbol{X}_i) &= \log_2 \frac{p_{st}(\boldsymbol{X}_i)}{p_s(\boldsymbol{X}_i) \cdot p_t(\boldsymbol{X}_i)} \\ &= H_s(\boldsymbol{X}_i) + H_t(\boldsymbol{X}_i) - H_{st}(\boldsymbol{X}_i) \end{aligned} \tag{5}$$

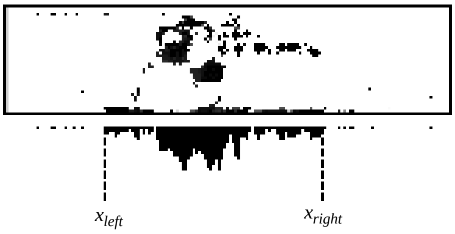
Here, $H_{st}(\boldsymbol{X}_i)$ denotes the spatiotemporal information of the pixel $\boldsymbol{X}_i$ on the domain of the neighbor $l \times m \times n$ pixels where $m \times n$ pixels to the spatial direction and l frame images to the time direction as shown in Figure 1. Figure 2 (d) shows a distribution of the spatiotemporal information for the input image shown in (a), and (e) shows the mutual information.

3. Computer controlled operation based on the distribution of mutual information

As a natural consequence, the person at the virtual classroom concentrates the region with large information. A video image including necessary information for the person present produces so as to control the camera to focus the region with the distribution of large degree of information which is calculated for the video image taken by the fixed monitor camera.



(a) Input video image.



(b) Distribution of mutual information and its histogram.

Figure 3. Distribution of mutual information and its histogram.

In the first, we examine the histogram of the binary mutual information to the direction of horizontal axis as shown in Figure 3 (b). The camera controller sets the line of camera sight to be $(x_{right} + x_{left})/2$ and the zooming range (angleview) to be in proportion to $(x_{right} - x_{left})^{-1}$ so as to be able to catch the interested region for the person present the virtual classroom.

4. Switching of plural video images

Section 3 discusses how to control one camera to take a spot with the necessary information at the video image for the person at the virtual classroom. Suppose that we take a video image by only one camera, there sometimes occurs occlusions of the written letters on the blackboard covered by the lecturer. As this kind of problem could not be solved by one camera, it is necessary to consider switching of the plural camera images so as to obtain the most effective video image. There are several kinds of standards for choosing the best camera angle image. The representative conditions are as follows.

1. The letters written on the blackboard can be always seen without occlusion.
2. The camera line of sight coincides the lecturer's eyes.

Considering the above conditions and experiences through the video image of real distance learning lecture, we prepare lists of switching conditions. In our experiment, the placement of the two shooting cameras and the fixed monitoring camera is shown in Figure 4.

Concrete control conditions are as follows.

1. In the case of zooming out, the blackboard in the video picture may have a distortion when the camera placed in the left hand side of the classroom takes the scene. In this case, the image is taken by the center camera (*CameraI*).
2. In the case when the information is distributed at the right hand side of the blackboard, the image of the center camera (*CameraI*) should be chosen. Conversely, the image of the left camera (*CameraII*) should be chosen on the condition that the information is distributed at the left hand side of blackboard.

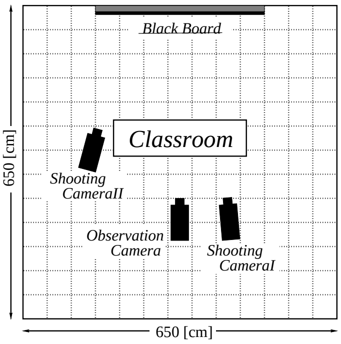


Figure 4. Camera layout in the classroom.

These two conditions intend to avoid the distortion of the image, especially the image of the blackboard.

The lecturer is standing at the left hand side of a block of the distribution of information in the case when the point (x_{right}) does not move and the point (x_{left}) moves to the left. Under this condition, the scene is taken by the *CameraI* to avoid the increase of the dead angle caused by the occlusion of the lecturer. Conversely, the scene is taken by the *CameraII* under the condition when the point x_{left} does not move and the point x_{right} moves to the right. Table 1 shows these rules of camera switching. Under the other conditions except above, there is no necessary to switch the video image. Initial camera is set to be *CameraI*.

5. Experiments and Discussions

5.1. Experiments

Controlling the operations of pan, tilt and zooming of the two shooting cameras placed as shown in Figure 4 on the

Table 1. Rules of camera switching.

Taking the video image by *CameraI*
$(x_{right} - x_{left})^{-1}$: sufficient small
$(x_{right} + x_{left})/2$: around the right edge of the blackboard
x_{right} : no change $\cap$ x_{left} : move to left on a large scale
Taking the video image by *CameraII*
$(x_{right} + x_{left})/2$: around the left edge of the blackboard
x_{left} : no change $\cap$ x_{right} : move to right on a large scale

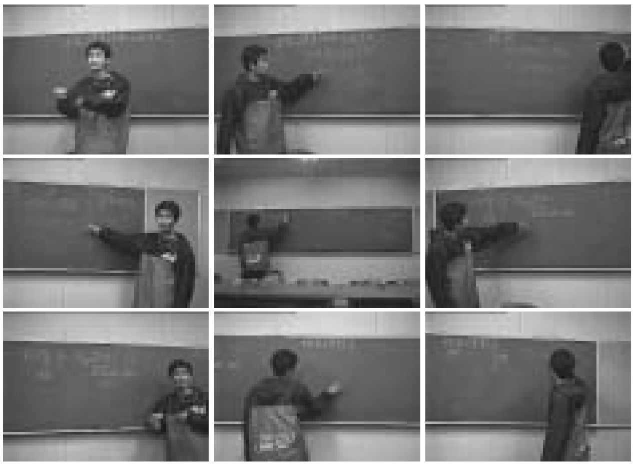


Figure 5. Some examples of video images of the experiment.

basis of the distribution of information using the fixed monitor camera, the lecture scene was taken by the two shooting cameras in real time. Image size of the fixed observation camera is set to be 640×160 pixels on the observation camera and the camera are placed so as to keep the both edges of the blackboard within the field of camera vision. Real image processing was performed with a reduced resolution (160×40 pixels) of the original image considering the processing time.

In this experiment, calculation of the distribution of mutual information on the observation camera image, the control of the shooting cameras and switching of the camera images are done by one personal computer (Pentium 400MHz, memory 128 Mbytes). Every 3 frames of the video image of the observation camera can be processed on this computer.

The duration time of the generating information on the pixels on the observation camera image was set to be 15 seconds, and the parameters of spatiotemporal information are set to be $l = m = n = 3$ taking the processing time into account in the experiments. Figure 5 shows some examples of our video images [1].

5.2. Evaluation of video images by a questionnaire survey

We have examined the effectiveness of the proposed method of the computer controlled camera operations and switching by the visual test. A lecture on explanation of our proposed method is chosen as the sample video image of a lecture (video image was about 20 minute long), and we conduct a questionnaire on 35 persons on the points of camera operation and switching of the images. The questionnaire was set out on the following three case of camera operations.

Table 2. Results of questionnaires.

Question item	Man	Com	F
Easy to recognize the lecturer's expression	4.36	4.23	0.58
Easy to read written letters on the blackboard	4.47	4.23	2.85
Easy to understand the lecturer's gestures	3.86	3.80	0.06
Easy to look at interested points	4.23	3.91	3.23
Atmosphere of the lecture	3.87	3.72	0.51

In this table, "Man" shows manual operation image, and "Com" shows computer operation image.

1. Video image taken by fixed observation camera
2. Video image by experts (Manual operation image)
3. Video image taken by our propose method (Computer operation image)

The video image by experts was taken by 2 cameramen and 1 director who switch the two video images. The placement of video cameras used by the above three cases is shown in Figure 4. The suitability of the camera shooting was evaluated with 5-step estimation method. The level 3 estimation (middle level) was set to be the case of the video image by the fixed observation camera. The result of the questionnaire survey is shown in Table 2 where the value F was defined by a ratio of the mean square and the mean square of error of the shooting methods. The value F is getting large as the difference among the methods increases.

6. Conclusions

In this paper, we have defined the spatiotemporal mutual information on the pixels of a given video image on the basis of information theory (Shannon's communication theory). This can be interpreted as the theoretical estimation of interested spots for human being. As an application of this spatiotemporal mutual information, this paper has proposed a method of producing a vivid video image of the distance learning by using the computer controlled camera operation and switching of plural camera images on the basis of the video image processing. The results of questionnaire survey confirm the effectiveness of our approach.

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] M. Minoh and Y. Kameda, "Distance Learning Environment based on the Interpretation of Dynamic Situation of Lecture Room," *Proc. of 3rd Int. Workshop on Cooperative Distributed Vision*, pp.283–301, Nov. 1999.

[1] See: http://www.com.cs.osakafu-u.ac.jp/˜onishi/research-e.html