VIRTUAL CAMERAWORK FOR GENERATING LECTURE VIDEO FROM HIGH RESOLUTION IMAGES

Takao Yokoi, Hironobu Fujiyoshi

Department of Computer Science, Chubu University
Email: {taka, hf}@vision.cs.chubu.ac.jp

ABSTRACT

We propose a method for generating a dynamic lecture video from the high resolution images recorded by a HDV camcorder. The lecture images are cropped to track the region of the instructor. Our approach uses bilateral filtering to avoid jittery motion caused by temporal differencing, and pseudo camera motion (panning) based on shooting technique of broadcast cameraman is generated. We evaluated our method, and showed that the effectiveness of our algorithm was verified through subjective experiments

1. INTRODUCTION

Recently, E-learning such as Web Based Training (WBT) has become a popular method used in higher education. One way of making content on WBT is to record the usual lectures by multiple cameras and to broadcast the archived lecture video to students through the Internet. However, video recording by multiple cameras and video editing of the archived lectures take a long time and cost a great deal.

There are many previous works about automated lecture archive systems using multiple cameras[1, 2, 3]. These methods extract events by motion detection and speech recognition and decide camerawork to control camera from the detected events. However, [1, 2] are not effective at generating dynamic video like shooting by broadcasting cameraman. [3] contains broadcasting cameraman's expression technology about camera position and target object in the screen, etc. However, the component does not contain the expression technology for the speed transition of the panning.

In this paper, we present a method for generating lecture video by cropping the high resolution image recorded by a HDV camcorder. Our approach uses bilateral filtering to avoid jittery motion caused by temporal differencing, and detects a period of pseudo camera panning to track the region of interest(ROI) such as the instructor. Finally, virtual camerawork (panning) based on shooting techniques of broadcast cameraman is generated.

This paper is organized as follows. Overview of our proposed system is described in Section 2. Detection of timing for camera panning and generation of virtual camera motion are explained in Section 3 and 4, respectively. Experiments are shown in Section 5 and the conclusion is presented in Section 6.

2. GENERATING LECTURE VIDEO FROM HIGH RESOLUTION IMAGES

Fig 1 shows our camera setting in the classroom. A HDV (1080*i*) camcorder is located at the back of the classroom to videotape images with high resolution (1,400 $\times$ 810), which contain the whole area of the chalkboard, so that students can read the handwritten characters on the chalkboard. However, it is impossible to display the high resolution image on the small screen of a general notebook PC (XGA).

To solve this problem, our approach is to reproduce the lecture video by cropping from the high resolution image to track the ROI such as the instructor (Fig. 2). Cropped image has 720 $\times$ 480 resolution which is the same as DVD format. So, it looks like a real video shot by broadcasting cameraman controlling the camera like a panning. To generate the dynamic lecture video, a period of pseudo camera panning should be detected and then a virtual camerawork should be calculated.

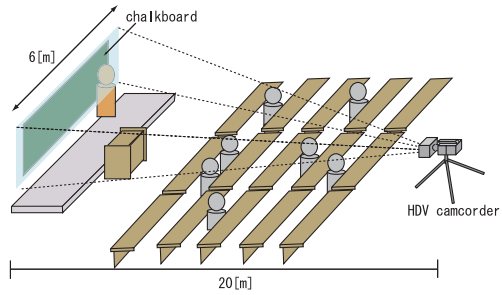

Fig. 1. Camera location in the classroom.

3. TIMING DETECTION FOR CAMERA PANNING

3.1. Detection of the ROI by temporal differencing

To determine the position to crop from the high resolution image, the ROI such as an instructor should be automatically detected. Foreground regions in the area of chalkboard

0-7803-9332-5/05/$20.00 ©2005 IEEE

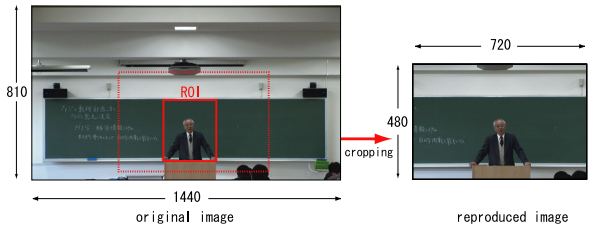

Fig. 2. Cropped image from high resolution image.

are initially detected by temporal differencing [6]. If I_n is the intensity of the n^{th} frame, the pixelwise difference function Δ_n is

$$\Delta_n = \max\{|\ I_n - I_{n-k}\ |\} \qquad 0 < k < 5 \tag{1}$$

and a motion image M_n can be extracted by thresholding.

$$M_n(u,v) = \begin{cases} I_n(u,v), & \Delta_n(u,v) \geq Th \\ 0 \quad , & \Delta_n(u,v) < Th \end{cases} \tag{2}$$

After the motion image is determined, moving sections are clustered into motion regions $R_n(i)$. The region R_n which has the largest foreground pixels in $R_n(i)$, is selected as the region of the instructor. At each frame, center position (x_c,y_c) of the R_n is obtained by temporal differencing. These center positions are expressed as an one dimensional discrete function,

$$R(n) = (x_c, y_c) \tag{3}$$

Fig. 3 shows x coordinate points of the $R(n)$ at each frame. It is clear that the positions of the ROI fluctuate because of temporal differencing. This causes jitter motion in the video generated by cropping based on the position of the ROI.

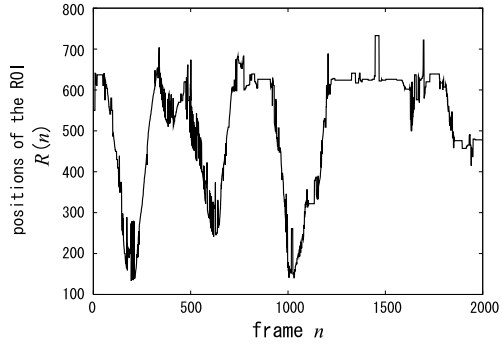

Fig. 3. Positions of the ROI detected by temporal differencing.

3.2. Bilateral filtering

Our approach uses the bilateral filtering method to avoid jitter motion. Bilateral Filtering smoothes images while preserving edges, by means of a nonlinear combination of nearby image values [5]. We applied bilateral filtering to smooth the $R(n)$, which is deformed as one dimensional filtering as follows:

$$\hat{R}(n) = \frac{\sum_{k=-w}^{w} W(n,k) \cdot R(n+k)}{\sum_{k=-w}^{w} W(n,k)} \quad , \tag{4}$$

$$W(n,k) = \exp\{-\frac{k^2}{2\sigma_S^2}\} \cdot \exp\{-\frac{R(n) - R(n+k)}{2\sigma_R^2}\}$$

where ω means [2W+1] samples around the n^{th} sample. Fig. 4 shows results of bilateral filtering with 10 iterations. We can see that the noise, which causes the jitter motion in the cropped images, is suppressed, and the edges are preserved.

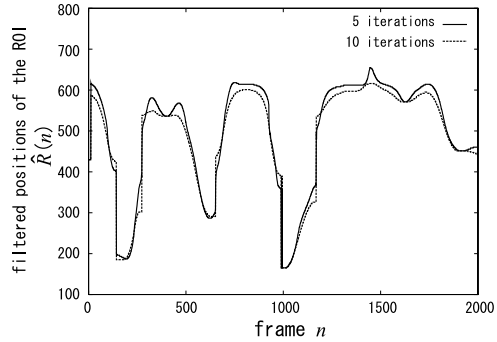

Fig. 4. Results of bilateral filter.

3.3. Detection of a period for camera panning

Extremal points of the $R(n)$ are detected by finding zero-crossings of the difference function.

$$\delta(n) = \hat{R}(n) - \hat{R}(n-1) \tag{5}$$

After calculating the distance of the detected next points, the moving period of the ROI is determined by threshold. Fig. 5 shows detected points and these are used to make pseudo camera panning described in Section 4.

4. VIRTUAL CAMERA PANNING

4.1. Camerawork by broadcasting cameraman

In order to reproduce dynamic images automatically, virtual camerawork based on cameraman's shooting technique should be implemented. Analysis of camerawork of broadcasting cameraman has been reported in [4], and they found the following points regarding the shooting technique to track moving subjects.

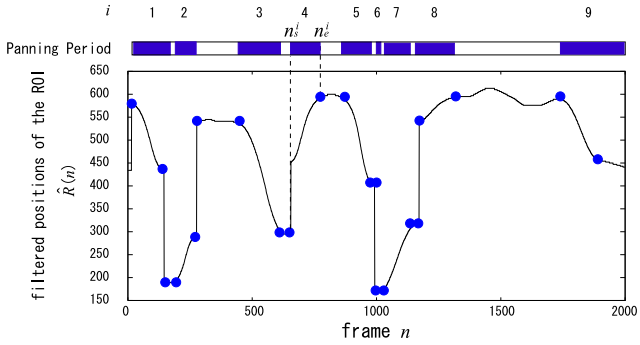

Fig. 5. Detection of panning period.

- Panning-speed curve is asymmetrical: quick acceleration is followed by slow deceleration. The deceleration time is 60% longer than the acceleration time.
- In the case of camera panning within 20[deg], the maximum panning speed ranges from 5 to 10 [deg/s]. Otherwise the maximum panning speed ranges from 15 to 20 [deg/s].
- The maximum panning speed is observed around the transition point from acceleration to deceleration.

4.2. Virtual camerawork algorithm

To implement the virtual camerawork based on shooting technique described above, the positions to crop from the high resolution images are calculated according to the following steps:

step1 Given the start frame n_s^i and the end frame n_e^i of the period i for camera panning, a transition point n_t from acceleration to deceleration is calculated based on the fact that the deceleration time is 60% longer than acceleration time (See Fig. 6).

$$n_t = \frac{0.6n_s + 0.4n_e}{0.4 + 0.6} \quad (6)$$

$$\hat{R}(n_t) = \frac{0.6\hat{R}(n_s) + 0.4\hat{R}(n_e)}{0.4 + 0.6} \quad (7)$$

step2 Acceleration α of camera panning velocity is calculated by the following equation.

$$\alpha = \begin{cases} \dfrac{2 \cdot (\hat{R}(n_t) - \hat{R}(n_s))}{(n_t - n_s)^2} & , n_s < n \leq n_t \\ \dfrac{2 \cdot (\hat{R}(n_e) - \hat{R}(n_t))}{(n_e - n_t)^2} & , n_t < n \leq n_e \end{cases} \quad (8)$$

step3 Positions to crop from high resolution image $R(n)$ for virtual camerawork is calculated by the following equation.

$$R(n) = \begin{cases} \frac{1}{2}\alpha \cdot (n - n_s) + \hat{R}(n_s) & , ns < n \leq n_t \\ \alpha \cdot (n_t - n_s) + \frac{1}{2}\alpha \cdot (n_e - n_t) + \hat{R}(n_t) & , n_t < n \leq n_e \end{cases} \quad (9)$$

Step1-3 are repeated for each period i of camera panning. Fig. 7 shows cropped images by virtual camerawork based on the shooting technique of cameraman described in 4.1.

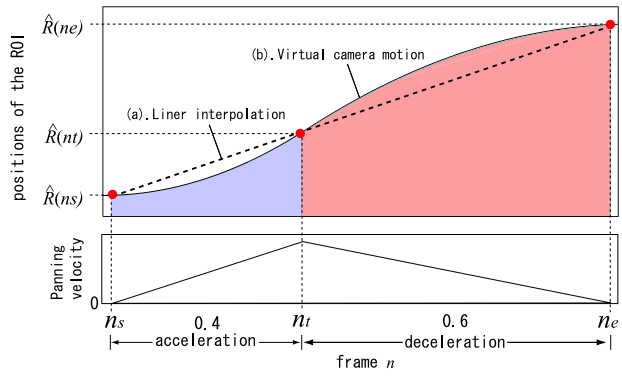

Fig. 6. Virtual camerawork (panning).

5. EXPERIMENTAL RESULTS

We evaluated the lecture video generated by the following four methods:

A. Temporal Differencing
Video is generated by cropping from the detected position by temporal differencing (Fig. 3).

B. Bilateral Filter
Video is generated by cropping from the filtered position by bilateral filtering with 10 iterations (Fig. 4).

C. Linear Interpolation
First, a period of camera panning is detected as described in Section 3.3 and then pseudo camera motion is determined by linear interpolation (Fig. 6(a)).

D. Virtual Camerawork
Video is generated by virtual camerawork based on cameraman's shooting technique described in Section 4.2 (Fig. 6(b)).

5.1. Subjective experiment

We carried out experiments examining the naturalness of the generated lecture video. The subjects were 20 students. After watching the 10 videos, the subjects rate naturalness on a 5-degree-scale (very good, good, neither good nor bad, bad, very bad) by answering the following questions.
Q1. Is the ROI such as instructor easy to see?

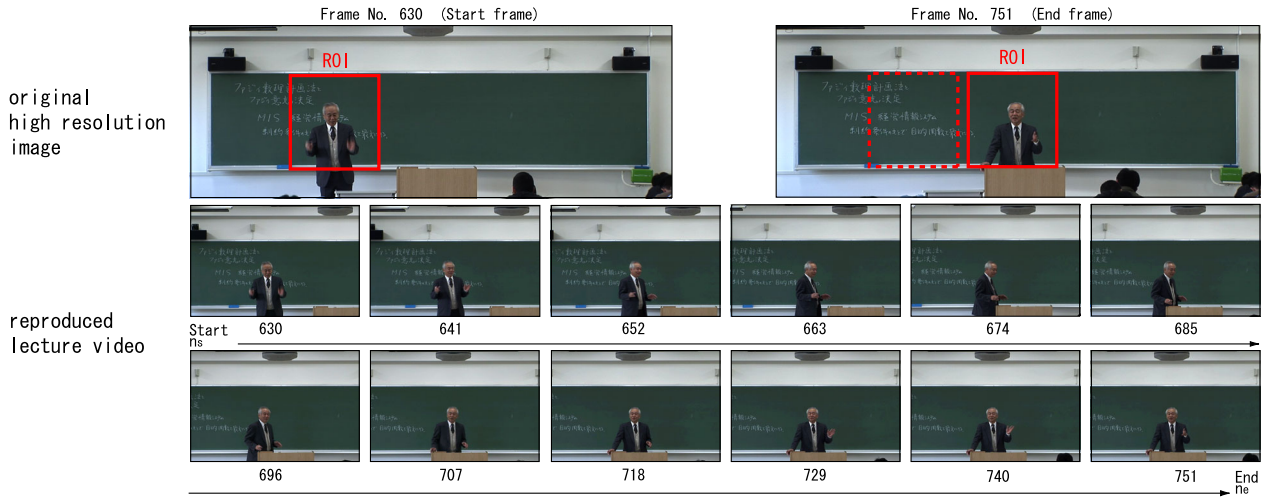

Fig. 7. Cropped images by virtual camerawork.

Q2. Is the motion by the instructor visible clearly?
Q3. Are the areas of the chalkboard that you are interested in easy to see?
Q4. Is the camera motion (panning) natural?
Q5. Is the lecture video comfortable to watch?

5.2. Results

After translating the subjective evaluation to a scale running from -2 to +2 (-2 = very bad, +2 = very nice), statistical analyses were performed. Fig. 8 shows results by subjective evaluation. The worst evaluations were given by method

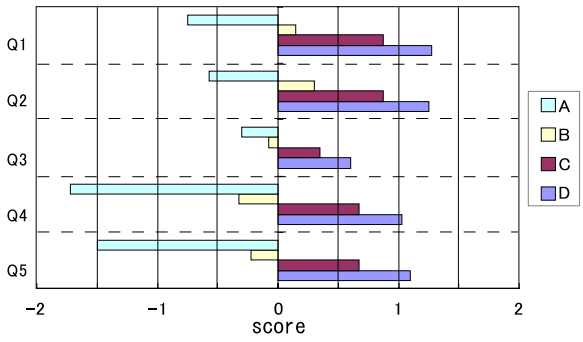

Fig. 8. Results by subjectvie evaluation.

A, because the positions of temporal differencing are not stable. On the other hand, method C and D have positive scores for all the above questions. The difference between method C and D for question Q.5 was statistically significant (Student's t-test; p less than or equal to 5%). Therefore, method D is better than method C, because the virtual camera panning is calculated based on broadcasting cameramen's shooting technique.

6. CONCLUSION

This paper has presented a novel method for generating a lecture video by cropping the high resolution image recorded by a HDV camcorder. We described a method for the detection of timing for camera panning using bilateral filtering and we showed virtual camerawork based on shooting technique of broadcasting cameraman. We showed the effectiveness of the proposed method using the results of subjective evaluation.

Virtual camera zooming for focusing the instructor's face would be our future work.

7. REFERENCES

[1] Y. Kameda, K. Ishizuka, and M. Minoh, "A Live Video Imaging Method for Capturing Presentation Information In Distance Learning," Proc. ICMCS'99, Vol. 2, pp. 897-902, 1999.
[2] M. Ozeki, Y. Nakamura, Y.Ohta, "Camerawork For Intelligent Video Production – Capturing Desktop Manipulations," Proc. ICME'01, pp. 41-44, 2001.
[3] Y. Rui, A. Gupta, J. Grudin, "Videography for Telepresentations", Proc. of the ACM Conference on Human Factors in Computing Systems, pp. 457 - 464, 2003.
[4] D. Kato, M. Yamada, and K. Abe, "Analysis of the Work and Eye Movement of Broadcasting-Studio Cameramen," The Journal of Institute of Television Engineers of Japan, Vol. 49, No. 8, pp. 1023-1031, 1995.
[5] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," ICCV, pp. 839-846, 1998.
[6] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving target classification and tracking from real-time video," Proc. of the WACV, IEEE, pp. 8-14, 1998.