METHOD FOR GENERATING VIDEO WITH VIRTUAL CAMERAWORK USING BI-DIRECTIONAL OBJECT TRACKING BETWEEN KEYFRAMES

Yudai Shinoki and Hironobu Fujiyoshi

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

ABSTRACT

Personal video sharing services such as YouTube have become popular because videos can easily be recorded in high-definition (HD) using a personal camcorder. However, it is difficult to broadcast an HD video via the Internet because of the large amount of data involved. We present a novel method for generating videos with virtual camerawork that is based on object tracking technology. Once a user specifies the positions of the object on keyframes, our method can be used to generate virtual camerawork between two keyframes in a row on the basis of the results of bi-directional tracking. We evaluated our method with subjective experiments and demonstrated its effectiveness.

Index Terms— Keyframe, virtual camerawork, cropping, bi-directional tracking, video generation

1. INTRODUCTION

Users have recently been able to record videos in high-definition (HD) because personal camcorders have become more affordable. This has allowed personal video sharing services such as YouTube to increase. However, it is difficult to broadcast HD videos via the Internet because of the large amounts of data needed. Although image compression is the most widely used way to solve this problem, the objects in the videos made with this method cannot be seen clearly because they are highly compressed.

An effective method for generating video is by using digital camerawork by cropping the region of interest (ROI). Doing this enable us to handle a large amount of data without causing rough images. Digital camerawork methods [1, 2, 3, 4] are foremost among the conventional methods for automatically generating video. However, videos generated by Onishi et al's methods [1, 2] are not natural because not enough research has been done on panning and zooming, and one's by Zhang et al's method [3] is also not natural because it makes use of a linear camerawork model. Shinoki et al [4] proposed a method that uses virtual camerawork modeled on shooting technique of broadcast cameraman to successfully generate natural camerawork. However, it is impossible to apply the method for personal videos, which are typically recorded under various conditions and with a moving camera, because the input videos are recordings of a lecture using a fixed camera only.

In this paper, we propose a method for generating videos with virtual camerawork by object tracking based on several keyframes that the user specifies. Once a user specifies the positions of the object on keyframes, our method can be used to generate virtual camerawork between two keyframes in a row based on the results of bi-directional tracking. Thus, the proposed method is a semi-automated system and it can be used to generate a video with natural camerawork.

2. PROPOSED METHOD

A framework of the proposed method is shown in Fig.1. Our method can generate video automatically with virtual camerawork by using the following three steps:

1. Specifying the object on two keyframes
2. Bi-directional object tracking between two keyframes
3. Generation of camerawork by using results of tracking

Our method can be used to generate natural camerawork on the basis of the results of bi-directional object tracking by controlling functions with a bilateral filter, detecting features by zero-crossing, and using virtual camerawork.

2.1. Specifying the object

The user specifies an object as the bounding box in the ROI on two keyframes. These keyframes are the first and the last frame that user wants to extract from input video. The color histogram (RGB color space of 512 dimensions ($8 \times 8 \times 8$)) is used for bi-directional tracking.

2.2. Bi-directional tracking

Many methods of tracking objects have been proposed. Our approach uses bi-directional tracking [5] because it is robust to sudden motion, ambiguity, and occlusion. First, the user specifies the positions of the same object on two keyframes.

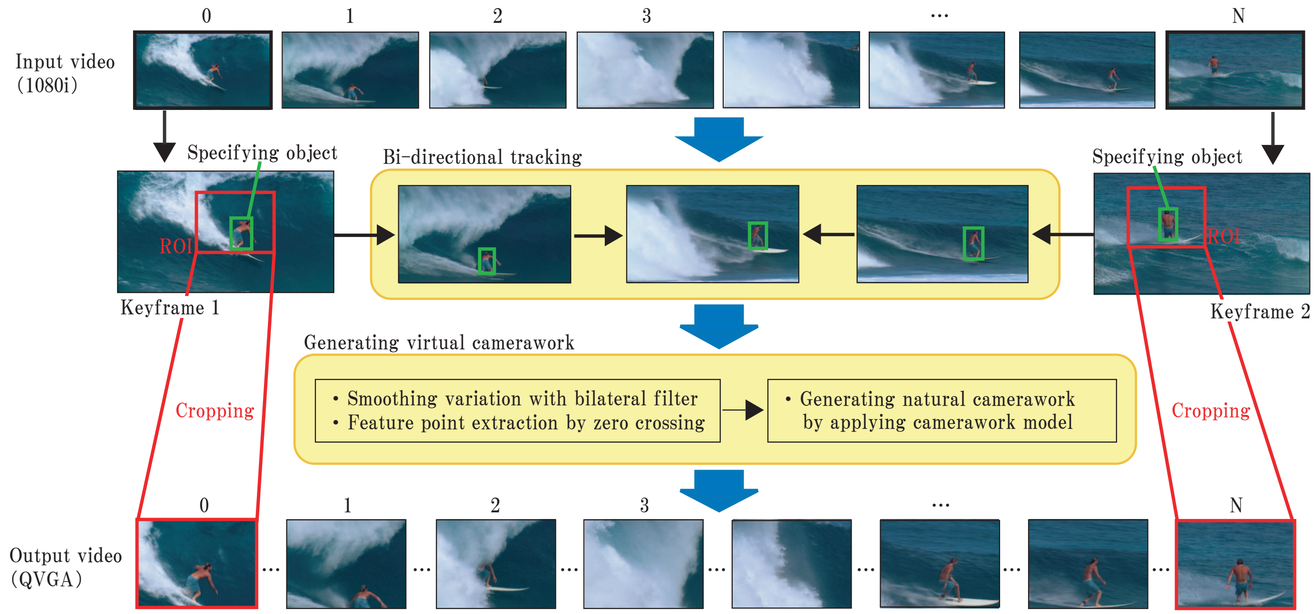

Fig. 1. Framework of generating camerawork based on keyframes.

Next, the positions of objects in the frames are estimated respectively by using the Bhattacharyya distance of the color histogram and mean-shift seeking. The local trajectories are then calculated from the obtained positions of the object by using spectral clustering and joined by using a tree-growing algorithm. Thus, this joined trajectory becomes a long trajectory of the object.

2.3. Generation of camerawork

Our method uses virtual camerawork [4] that is based on the trajectory of the object obtained by the bi-directional tracking.

2.3.1. *Virtual camerawork*

Virtual camerawork [4] is modeled based on shooting techniques of broadcast cameraman [6] and has the following features:

- The panning speed curve is asymmetric, and the deceleration time is about 60 percent longer than the acceleration time.
- The maximum panning acceleration starts immediately before the movement of the camera reaches maximum speed and ends immediately after deceleration from maximum speed.

2.3.2. *Correction of trajectory*

To apply virtual camerawork model to tracking results, the obtained trajectory is filtered for a suitable camerawork. We apply the bilateral filtering to avoid jitter motion of the tracking results. Bilateral filtering smoothes trajectory while preserving edges by using nonlinear combination of nearby trajectory values.

2.3.3. *Applying camerawork model*

The timing for camerawork is determined from feature points obtained by the zero-crossing method. Virtual camerawork is calculated using these feature points as follows:

Step 1 The starting position in the panning section is assumed to be t_f and the end position is assumed to be t_e. The acceleration and deceleration time ratio is 60 and 40 percent of camerawork time respectively. The deceleration time determines the end position of acceleration t_c and the point value $\hat{\mathbf{m}}_c$is computed as follows:

$$t_c = 0.6t_f + 0.4t_e, \quad (1)$$
$$\hat{\mathbf{m}}_c = 0.6\hat{\mathbf{m}}_f + 0.4\hat{\mathbf{m}}_e. \quad (2)$$

Step 2 Acceleration α of panning is computed according to the acceleration area ($t_f < t \leq t_c$) and deceleration area ($t_c < t \leq t_e$) using the following equation.

$$\alpha = \begin{cases} \dfrac{2\cdot(\hat{\mathbf{m}}_c - \hat{\mathbf{m}}_f)}{(t_c - t_f)^2} & , t_f < t \leq t_c, \\ \dfrac{2\cdot(\hat{\mathbf{m}}_e - \hat{\mathbf{m}}_c)}{(t_e - t_c)^2} & , t_c < t \leq t_e. \end{cases} \quad (3)$$

Step 3 Virtual panning is archived from cropped position $\mathbf{m}'_i$. Cropped position $\mathbf{m}'_i$ is computed by using acceleration α in

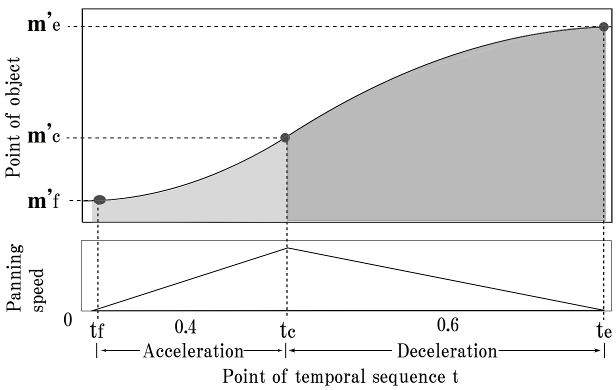

Fig. 2. Virtual panning model.

Step2 and the following equation:

$$\mathbf{m}'_i = \begin{cases} \frac{1}{2}\alpha \cdot (t - t_f)^2 + \hat{\mathbf{m}}_f & , t_f < t \leq t_c, \\ \frac{1}{2}\alpha \cdot (t_e - t_c)^2 - \frac{1}{2}\alpha \cdot (t_e - t)^2 + \hat{\mathbf{m}}_c & , t_c < t \leq t_e. \end{cases} \quad (4)$$

The virtual panning model is outlined in Fig.2.

3. EXPERIMENTAL RESULTS

We evaluated camerawork videos generated by the following five methods:

A.Tracking Result + Virtual Camerawork (proposed method)

Video is generated by using the tracking result and virtual camerawork. An example of the camerawork video by the proposed method is shown in Fig.3.

B.Tracking Result + Linear Camerawork

Video is generated by using the tracking result and Zhang et al's liner camerawork [3].

C.Tracking Result Only

Video is generated by using the tracking result only.

D.Two Joined ROIs

Video is generated by two joined ROIs without the tracking result.

E.Manually Joined ROIs

Video is generated by manually joined ROIs every 50 frames without the tracking result.

3.1. Subjective experiment

We carried out experiments to investigate the naturalness of generated video. The subjects were 20 adults. After watching the 20 videos, which contain various scenes, the subjects rated the naturalness on a 5-degree scale (very bad, bad, neither bad nor good, good, very good) by answering the following questions.

Q1. Was the camerawork natural?

Q2. Was the video easy to watch?

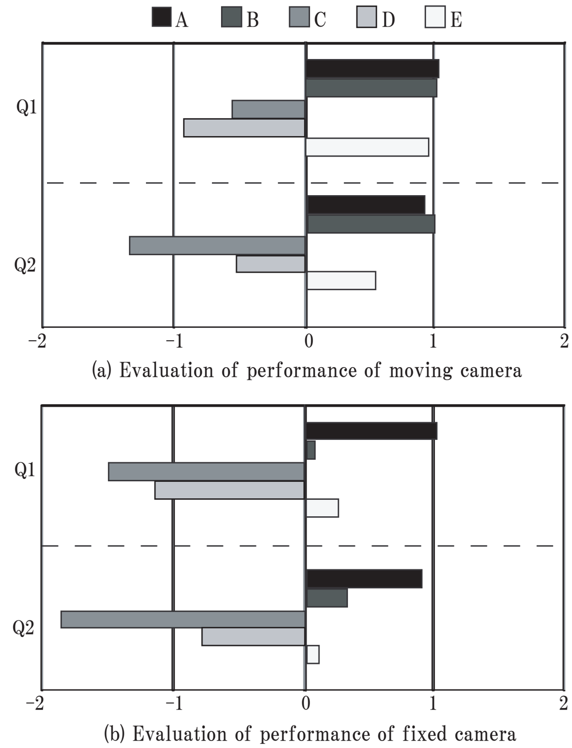

Fig. 4. Result of subjective experiment.

3.2. Results

After translating the subjective evaluation to a scale running from -2 to +2 (-2 = very bad, +2 = very good), statistical analyses were performed. Figure 4 shows the results of the subjective evaluation.

The worst evaluation for both the moving camera and fixed camera was received by methods C and D because the object was sometimes out of frame in the camerawork video. This was because there was no correction of trajectory for method C and no tracking results for method D. On the other hand, both method A and B had positive scores for the above two questions. As seen in Fig.4 (a), there are no large differences between both methods A and B for the moving camera. This was because the movement of these two cameraworks after cropping ROIs is similar because of the original camerawork in the input video. As seen in Fig.4 (b), the difference between method A and B for the fixed camera was statistically significant. Therefore, our method using the virtual camerawork based on tracking result is effective for both a moving and a fixed camera. Moreover, using our method obtained an equivalent evaluation to when method E, with manual camerawork was used. In this way, we confirmed that our method can be used to generate a video with natural camerawork.

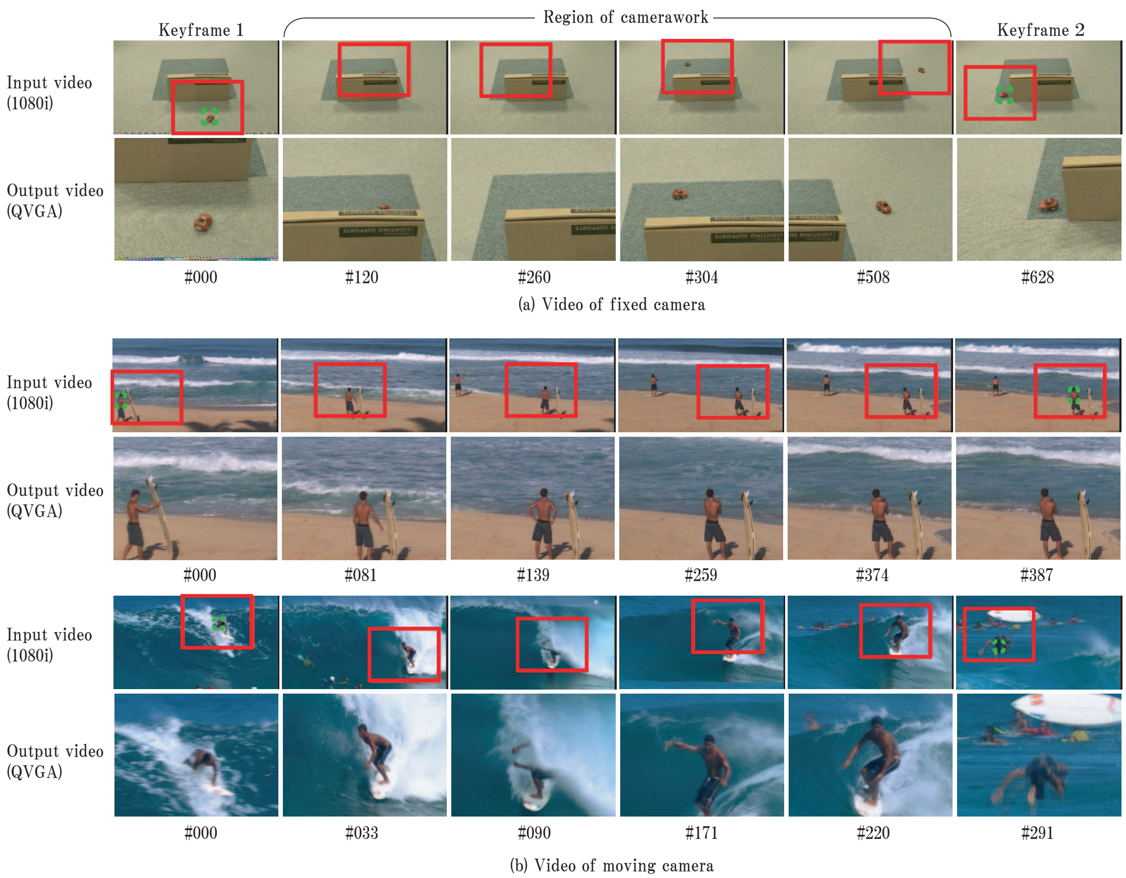

Fig. 3. Virtual camerawork videos by using tracking result (top: input videos; bottom: output videos).

4. CONCLUSION

We proposed a method for generating video with virtual camerawork using bi-directional object tracking between keyframes. Natural camerawork was automatically generated by applying a virtual camerawork model, and its effectiveness for both a moving and a fixed camera was confirmed from the subjective experiments. Since the proposed method generates a small video by cropping the ROI, the amount of data in the generated video is about 10MB where the original HD video is about 200 MB (about 5 percent of the original video). We intend to generate a virtual camerawork video that includes zooming in future research.

5. REFERENCES

[1] M. Onishi, I. Masao, and K. Fukunaga, "Automatic Image Production Using Digital Camera Work", MIRU2000, pp. I-331-I-336, Jul. 2000.

[2] M. Onishi, M. Murakami, and K. Fukunaga, "Computer-Controlled Camera Work at Lecture Scene Based on Situation Understanding and Evaluation of Video Images" The transactions of the Institute of Electronics, Information and Communication Engineers. Vol.J85-D-II, No. 4, pp. 594-603, Nov. 2002.

[3] C. Zhang, Y. Rui, L. He, and M. Wallick, "Hybrid Speaker Tracking in an Automated Lecture Room", Proc. of IEEE ICME 2005, Jul. 2005.

[4] Y. Shinoki and H. Fujiyoshi, "A Method for Generating Lecture Videos Relevant to Viewers from High Resolution Images", The Journal of The Institute of Image Information and Television Engineers Vol. 62, No. 2, pp. 240-246, 2008.

[5] J. Sun, W. Zhang, X. Tang, and H.-Y. Shun, "Bi-directional Tracking using Trajectory Segment Analysis", IEEE ICCV 2005, Vol. 1, pp. 717-724, Oct. 2005.

[6] D. Kato, M. Yamada, and K. Abe, "Analysis of the Work and Eye Movement of Broadcasting-Studio Cameramen", The Journal of The Institute of Television Engineers of Japan Vol. 49, No. 8, pp. 1023-1031, 1995.