

Late Breaking Results: Adding Sound to Silent Video by Identifying Corresponding Motion Vectors

Tomoya Hirose¹⁾, Miwa Nishimura¹⁾, Tsuyoshi Kobayashi²⁾, Jun Murayama³⁾,
Yukihiro Hirata⁴⁾, Tetsuya Harada³⁾ and Makoto Sato²⁾

- 1) Graduate School of Industrial Science and Technology, Tokyo University of Science, 2641 Yamasaki, Noda-shi, Chiba, 278-8510 Japan
2) Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda-cho, Yokohama-shi, Kanagawa, 226-8503 Japan
3) Faculty of Industrial Science and Technology, Tokyo University of Science, 2641 Yamasaki, Noda-shi, Chiba, 278-8510 Japan
4) Department of Electronics Engineering, Tokyo University of Science, Suwa, 5001-1 Toyohira, Chino-shi, Nagano, 391-0292 Japan

ABSTRACT

This paper proposes a system for adding appropriate sound to silent videos. The proposed system gives more reality to scenes by adding sounds – it generates appropriate sounds by using the motion vectors of objects in the videos. The motion vectors are obtained by applying the watershed algorithm. Appropriate sounds are added to the silent video according to the timing of the motion vectors.

KEYWORDS: Auditory sense, Sound generation, Optical Flow

INDEX TERMS: H5.1 [Multimedia Information Systems]: Audio input/output—Artificial, augmented, and virtual realities; I.2.20 [Vision and Scene Understanding]: Video analysis.

1 INTRODUCTION

Recently, it has become possible to use old videos easily through video sharing services. However, video based on old media such as 8mm film often lacked sound. The audience can better imagine a scene in a video by hearing appropriate sounds with the video. Silent video make it more difficult for audiences to imagine the scene.

Moving objects make certain sounds when the video was recorded. Those sounds become part of the environmental sound in the scene. We can ascertain the movement of objects reflected in the video by processing digitized old video.

This study aimed to restore sounds to silent video similar to the original sounds. We prepared a database of sounds. A suitable sound corresponds to the kind of movement portrayed. That is, a suitable sound is chosen from a database on the basis of the movement of the object.

This paper describes how to make sounds correspond realistically to the movement of an object in an image by using sounds prepared beforehand.

Using the watershed algorithm, the system detects the two-dimensional movement vector of the object, selects a sound suitable for it from the database, and adds it to the video at the timing based on the amount of movement.

2 METHOD OF ADDING SOUNDS

The method of the processing system is shown in Figure 1. First, the system loads the silent video. Next, it extracts two-dimensional motion vectors from the video by using the watershed algorithm [1]. Then, it determines the frames for which sounds should be added with the displacement of the motion vectors.

20th International Conference on Artificial Reality and Telexistence (ICAT2010), 1-3 December 2010, Adelaide, Australia
ISBN: 978-4-904490-03-7 C3450 ©2010 VRSJ

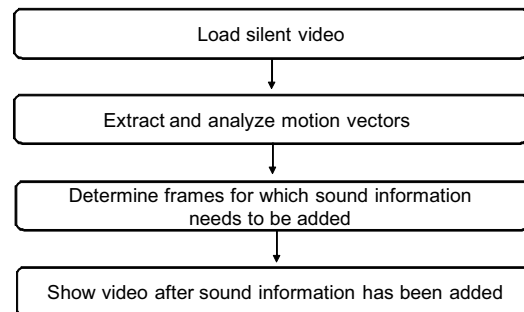


Figure 1. Processing procedure

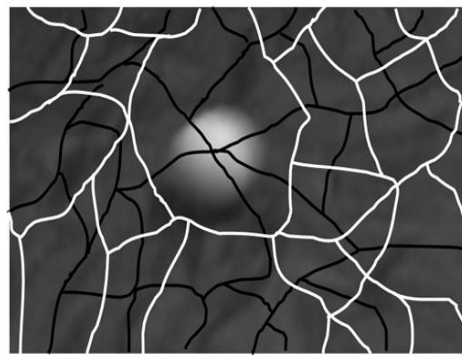


Figure 2. Watershed region division (Black: dale [Black outlined regions are dales.], White: hill [White outlined regions are hills.])

2.1 Extraction and Analysis of Motion Vectors

The watershed algorithm [1] was used to extract motion vectors from the silent video. It divides a picture into hill and dale regions. Hill regions are groups of pixels separated by dark edge lines in a video frame, and dale regions are groups of pixels separated by bright edge lines. Figure 2 shows a picture divided into hill and dale regions - the regions surrounded by white edges are hill regions, and those surrounded by black edges are dale regions.

The watershed algorithm also obtains the central point of each region for every frame of the video. A two-dimensional motion vector of an object in the video is obtained by searching for the displacement of the central points in two adjacent frames. And the object's position information is obtained with respect to the coordinate data of the central point.

Every frame of the video was smoothed with a Gaussian filter as a pretreatment to provide more precise watershed divisions [2].

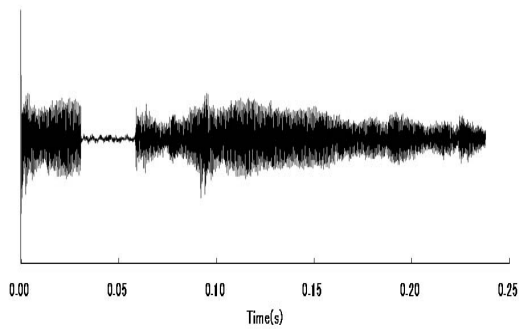


Figure 3. Waveform of sampling audio

2.2 Determining the Object and the Frames Causing Sounds

The object that caused sounds was selected by selecting a region in which the amount of movement exceeded a specified reference value. The timing of adding sounds was based on the moving distance. The system used a short sampled sound prepared beforehand.

3 EXPERIMENT AND RESULTS

3.1 Experiment

In the experiment, we added sounds to previously recorded videos. The audio track was removed from the video and compared the sound generated by the proposed method with the original sound of the video.

We prepared a sample video that featured rolling balls. Each ball had a bell inside, which rang as the ball rolled. The sample showed two rolling balls colliding. The ball moved over about 65 pixels in the sample video as it executed a 360-degree roll.

We also prepared a short sampling sound to record as a bell ring. Figure 3 shows the waveform of the prepared sound.

The sounds were added to every frame in which the movement distance of the ball exceeded 65 pixels.

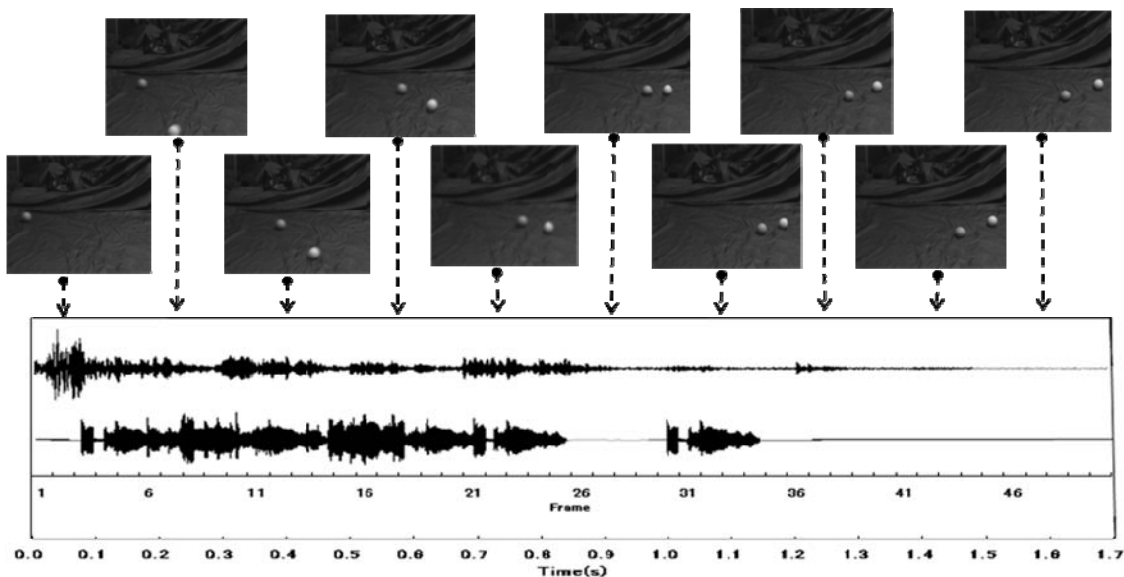


Figure 4. Comparison of waveforms (Top: sound waveform obtained at the time of animation photography, Bottom: sound waveform with added sampling sound)

3.2 Results and Observations

Figure 4 shows the sound waveforms added by the proposed method and the original. In the waveforms, the termination time of the waveform was around 1.2 seconds.

The original sound of video started earlier than the added sound. This was caused by the experimental conditions for adding sounds. It took time to determine whether a frame had a certain displacement from the beginning of the video. Moreover, the volume of the original sound became smaller as the velocity got smaller. Therefore, the relationship between the velocity and the sound volume should be considered.

4 CONCLUSION AND FUTURE SUBJECTS

In this study, we added sounds to silent videos by using two-dimensional motion vectors of objects in the video and the watershed algorithm. The sounds were added to the silent video by referring to the distance through which the objects moved in the videos.

This study focused on motion vectors of only one kind of object and used a single prepared sound for it. However, it seems important to distinguish between two or more different kinds of objects and to add the different sounds to them. To deal with this problem, we will employ the size and the shape of the regions created by the watershed algorithm. In addition, the depth of the objects in the video will be estimated by using their relative size.

REFERENCES

- [1] A. Bieniek and A. Moga, "A connected component approach to the watershed segmentation", In *Mathematical Morphology and its Applications to Image and Signal Processing*, Vol. 12 of *Computational Imaging and Vision*, pp. 215-222 (1998).
- [2] T. Kobayashi, X. Zhang, M. Sato, "Watershed Region Tracking for Video Motion Description", In *Proceedings of the IET International Conference on Visual Information Engineering (VIE2006)*, pp. 488-493 (2006).