2010 International Conference on Pattern Recognition

# Removal of Moving Objects from a Street-view Image by Fusing Multiple Image Sequences

Hiroyuki Uchiyama†, Daisuke Deguchi†, Tomokazu Takahashi‡, Ichiro Ide†, Hiroshi Murase†
†*Graduate School of Information Science, Nagoya University, Japan*
‡*Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan*
*uchiyama@murase.m.is.nagoya-u.ac.jp*

## Abstract

*We propose a method to remove moving objects from an in-vehicle camera image sequence by fusing multiple image sequences. Driver assistance systems and services such as Google Street View require images containing no moving object. The proposed scheme consists of three parts: (i) collection of many image sequences along the same route by using vehicles equipped with an omni-directional camera, (ii) temporal and spatial registration of image sequences, and (iii) mosaicing partial images containing no moving object. Experimental results show that 97.3 % of the moving object area could be removed by the proposed method.*

## 1. Introduction

In recent years, street-view images are widely used in many applications [1, 2] such as driver assistance systems, ego-localization and forward obstacle detection. In order to realize these systems, street-view images containing no moving object are required. On the other hand, Google Street View [3] exhibits street-view images on the Internet. However, there is a problem that our privacies may be violated in these images, e.g. faces, running vehicles or bicycles. Although automatic detection and blurring of them are applied, sufficient quality is not achieved in the current system. Therefore, removal of moving objects from these images is one of the solutions (Fig. 1).

To remove obstacles from an image, there are three major approaches: (1) from an image [4], (2) from an image sequence [5, 6], and (3) from multiple images captured independently [7]. Most of them require obstacle areas to be specified manually [4, 5] or detected precisely [6]. However, it is time consuming to specify obstacle areas manually. Meanwhile, it is also very difficult to detect those areas automatically due to a large variety of targets in an urban area. In contrast, the work presented in [7] can remove obstacles without specifying them by using multiple images. However, rough camera positions of these images should be input manually. Therefore, this method cannot be applied for a large number of images.

This paper proposes a method to remove moving objects from an in-vehicle camera image sequence without any manual interaction by using multiple image sequences. To obtain an omni-directional image containing no moving object, the following two problems are solved in this paper.

- Difference of camera positions
- Selection of background-like partial images from multiple image sequences

The difference of camera position occurs due to the different speed and the different lateral position of vehicles. This difference causes an appearance change. To deal with this problem, the proposed method utilizes registration of image sequences in both time and space direction. As for the second problem, we assume that the occurrence of moving objects is relatively few at a same sub-window in images captured at a same place when selecting the most background-like partial image.

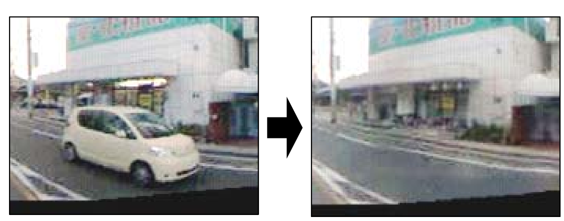


**Figure 1. Removal of moving objects by the proposed method.**

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.844

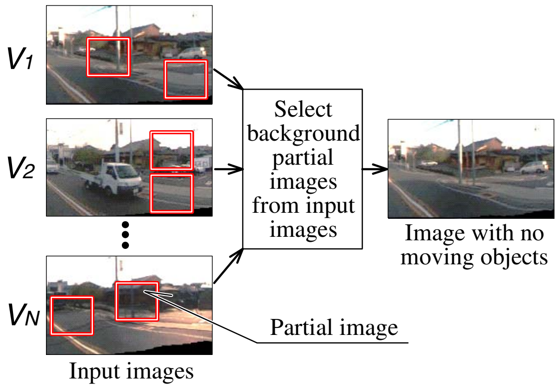



**Figure 2. Omni-directional camera image containing no moving object is obtained from many images captured at the same place in a different timing independently.**

## 2. Removal Method of Moving Objects

### 2.1. Overview

This paper defines a moving object as an unfixed object on the road such as a vehicle, a bicycle, a motorcycle, or a pedestrian. The proposed method consists of three steps: (i) collecting image sequences, (ii) temporal and spatial registration, and (iii) mosaicing partial images containing no moving object. $N$ image sequences are obtained by driving vehicles equipped with an omni-directional camera along the same route many times. Then temporal and spatial registrations are applied for compensating for the difference of the camera position. After these registrations are applied, we can obtain images captured at almost the same position. Finally, by assuming the occurrence of moving objects is relatively few at a same sub-window, the partial images of the background are selected and mosaiced to obtain an omni-directional camera image having no moving object (Fig. 2). The details of steps (ii) and (iii) are explained below.

### 2.2. Temporal and Spatial Registrations between Image Sequences

A total of $N$ image sequences are used in the proposed method: one target image sequence and $N-1$ source image sequences. As shown in Fig. 3, all source image sequences are registered to the target image sequence in the registration step.

**Temporal registration** aligns all image sequences along the time direction. A same frame index does not correspond to the same location since it is difficult to capture images by keeping the same driving speed.

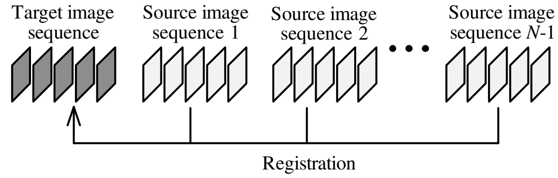



**Figure 3. All source image sequences are registered to the target image sequence.**

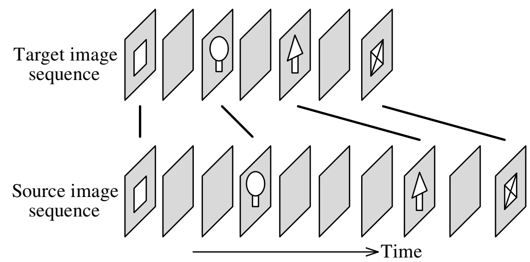



**Figure 4. DTW is applied for aligning image sequences in the time direction.**

Therefore, Dynamic Time Warping (DTW) [8] is applied to solve the non-linear frame alignment problem (Fig. 4).

**Spatial registration** is applied to image sequences aligned by DTW. Since the lateral position of the vehicle may be different and the frame rate of the camera is limited, positional error still exists even if the temporal registration is applied. In order to reduce the small difference of the camera position, registration along space direction should be performed. The different camera positions cause a non-linear distortion to captured images due to complex structures in a scene. In addition, pixel-wise correspondence is required for mosaicing partial images precisely. Therefore, the proposed method approximates appearance variations by using B-spline, and non-rigid registration (NRR) [9] is applied to the image sequences.

### 2.3. Selection and Mosaicing of Partial Images

First, many sub-windows are positioned on the registered images. Here, the size of each sub-window is $W \times W$ pixels, and the sub-windows are slightly overlapped with each other. Then a partial image corresponding a sub-window is treated as a $3W^2$-dimensional vector containing RGB pixel values. Next, the most background-like vector is selected by using the vector median filter [10].

The vector median filter is a median filter extended so that a multiple dimensional vectors could be input.

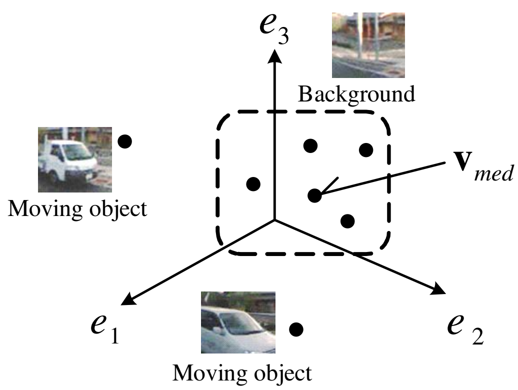



**Figure 5. Vector median filter tends to select a background image instead of a moving object image.**

It tends to exclude outliers and select the most common one. Under the assumption that the occurrence of moving objects is relatively few at a same sub-window in the images captured at a same place, a background image tends to be selected instead of a moving object image (Fig. 5). Using $M$ input vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M$, an output of the vector median filter is calculated as

$$\mathbf{v}_{med} = \underset{\mathbf{v} \in \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}}{\arg\min} \sum_{i=1}^{M} |\mathbf{v} - \mathbf{v}_i|, \quad (1)$$

where $|\cdot|$ is L2 norm of a vector.

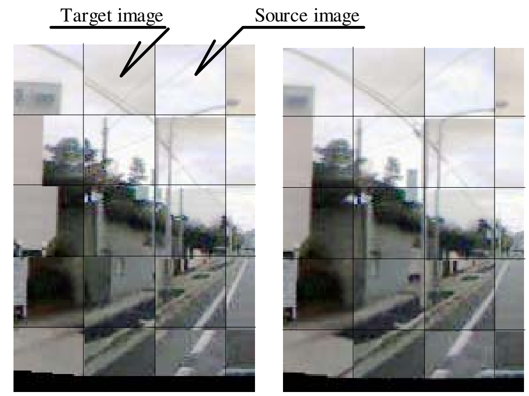
Finally, the selected images of all sub-window positions are mosaiced. Alpha blending is applied at the overlapped area of the partial images.

## 3. Experiment

We performed an experiment to evaluate the effectiveness of the proposed method. Point Grey Research Ladybug2 was used as an omni-directional camera. The frame rate was 15 fps and the original panorama image size was $1{,}024 \times 512$ pixels. One image sequence was used as a target image sequence and 2 to 14 image sequences were used as source image sequences. Each sequence contained about 300 frames. The window size for vector median filter was $30 \times 30$ pixels.

Result of the registration is shown in Fig. 6. The left image shows the result of DTW and the right one shows the result of DTW + NRR. Fig. 7 shows the result of the removal of the moving objects. Although a pedestrian, vehicles and a bicycle are observed in the input image (target image), they were successfully removed in the result.

Effectiveness of the proposed method was evaluated by the removal rate of the moving objects shown in




(a) DTW (b) DTW + NRR

**Figure 6. Example of registration results. The target image and the source image are placed as a checkerboard.**

Fig. 8. The removal rate was calculated by $(1 - B/A)$, where $A$ is the number of pixels corresponding to moving objects in the target image and $B$ is the number of pixels corresponding to moving objects in the output image. This was evaluated by using 11 target image frames selected randomly. In order to avoid the dependency of the selection of the source image sequence, all combinations were examined, and the average of them were used for evaluation (e.g. $3{,}432$ combinations were examined for 7 source image sequences). In the case of using 15 image sequences, 97.3 % of the pixels compositing the moving objects were successfully removed. Fig. 8 shows that use of many image sequences improves the removal of moving objects. This is because partial image selection by the vector median filter is sensitive to illumination change in a small number of image sequences.

## 4. Conclusion

We proposed a method to remove moving objects from an in-vehicle camera image sequence by fusing multiple image sequences at a same location taken in a different timing independently. First, temporal and spatial registrations are applied to compensate for the difference of camera positions. Then image sequence having no moving object is obtained by selection and mosaicing of partial background images obtained from different image sequences. The proposed method removed moving objects accurately with a high rate of 97.3 %. Future work includes the improvement of the removal using a smaller number of sequences.

**Acknowledgement** Parts of this research were supported by JST CREST and MEXT KAKENHI. This


(a) Before removal: input image (target image)

(b) After removal: output image

**Figure 7. Result of the proposed method. Although a pedestrian, vehicles and a bicycle are observed in the input image (a), they were removed in the output image (b).**

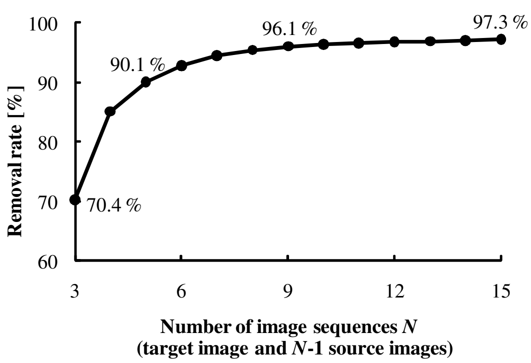



**Figure 8. Removal rate of moving object when changing the number of image sequences.**

work was developed based on the MIST library (*http://mist.murase.m.is.nagoya-u.ac.jp*).

## References

[1] H. Uchiyama, D. Deguchi, T. Takahashi, I. Ide, and H. Murase, “Ego-localization using Streetscape Image Sequences from In-vehicle Cameras,” Proc. 2009 IEEE Intelligent Vehicles Symposium, pp.185–190, Jun. 2009.

[2] M. Jabbour, V. Cherfaoui, and P. Bonnifait, “Management of Landmarks in a GIS for an Enhanced Localisation in Urban Areas,” Proc. 2006 IEEE Intelligent Vehicles Symposium, pp.50–57, Sep. 2006.

[3] “Google Maps,” http://maps.google.com/

[4] A. Levin, A. Zomet, and Y. Weiss, “Learning How to Inpaint from Global Image Statistics,” Proc. IEEE 9th Int. Conf. on Computer Vision, pp.305–312, Oct. 2003.

[5] Y. Wexler, E. Shechtman, and M. Irani, “Space-Time Video Completion,” Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol.1, pp.120–127, Jun. 2004.

[6] A. Yamashita, I. Fukuchi, T. Kaneko, and K. Miura, “Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing, ” Proc. 2008 IEEE Int. Conf. on Robotics and Automation, pp.2386–2391, May 2008.

[7] J. Böhm, “Multi-image Fusion for Occlusion-free Façade Texturing,” Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol.35, Part B5, pp.867–872, Jul. 2004.

[8] J. Sato, T. Takahashi, I. Ide, and H. Murase, “Change Detection in Streetscapes from GPS Coordinated Omni-Directional Image Sequences,” Proc. 18th Int. Conf. on Pattern Recognition, Vol.4, pp.935–938, Aug. 2006.

[9] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, “Nonrigid Registration Using Free-form Deformations: Application to Breast MR Images,” IEEE Trans. Medical Images, Vol.18, No.8, pp.712–721, Aug. 1999.

[10] J. Astola, P. Haavisto, and Y. Neuvo, “Vector Median Filters,” Proc. IEEE, Vol.78, No.4, pp.678–689, Apr. 1990.