MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Image Rectification for Robust Matching of Car-mounted Camera Images

Naoko Enami, Norimichi Ukita and Masatsugu Kidode
Nara Institute of Science and Technology
{naoko-e, ukita, kidode}@is.naist.jp

Abstract

We propose a matching method for images captured at different times and under different capturing conditions. Our method is designed for change detection in streetscapes using normal automobiles that has an off-the-shelf car mounted camera and a GPS. Therefore, we should analyze low resolution and frame-rate images captured asynchronously. To cope with this difficulty, previous and current panoramic images are created from sequential images which are rectified based on the view direction of a camera, and then compared. In addition, several image deformations are rectified for robust panoramic image matching.

1 Introduction

A car navigation system is important for a smooth car-oriented society. For accurate navigation, the map should be up-to-date. However, most map information (locations of buildings/signs facing the road that can be landmarks and destinations for drivers) is updated on the basis of surveillance fieldwork by a small number of probe cars.

Our future goal is to update map information by automatically analyzing streetscape's images captured by cameras. In this analysis, the images captured in the same location at different times are selected from a number of captured images. The selected images are then compared in order to find the change in the streetscape. In this paper, we tackle the former problem, namely a matching method for images captured under different capturing conditions. Methods for detecting a change in streetscapes, Sato et al.[1] achieved change detection only with image matching and comparison by using a car-mounted omnidirectional camera and an off-the-shelf GPS. In this method, images observed at the same location are found from a number of images observed at different times and locations by GPS position data. Matching using only position data is impossible because the off-the-shelf GPS has a margin of error of about 15m. This method employ DP matching [2] for making the correspondences of time-series images observed in close locations. However, this method can work well under the assumption that a probe car with expensive equipments captures images under convenient capturing conditions (i.e., high frame-rate omnidirectional capturing in a low car speed). Therefore, each location is rarely observed because it is impossible to prepare a number of these probe cars. To solve the problems of the previous methods, we aim at a map-update system with the features below:(1) wide areas can be observed simultaneously at low cost by a number of normal automobiles, and (2) each automobile has an ordinary low-resolution car-mounted camera,which is widely used in an event data recorder(EDR), and an off-the-shelf GPS, which is used widely in a car navigation system. As with the previous method [1], our system first selects previous and current images captured in the same location from image sequences. Even if a part of a streetscape has been changed (e.g., change of a sign), matching for selecting the previous and current images must be successful. In addition, imaging conditions under which the images are observed in our system (i.e., asynchronous low resolution and frame-rate capturing by a conventional camera) are more severe than those in the previous system [1].

In this paper, therefore, we have to cope with the following difficulties:

- Images at different times look different due to changes in illumination.
- Even if a streetscape is changed and/or moving objects are observed, images observed in the same location must be matched.
- Since images are captured from a number of automobiles, the images are asynchronized and the frame-rate of the images is low due to wireless communication. These result in the large difference in the image-capturing positions.
- Since the camera angle is directed towards the front, they are not suitable for image matching and change detection of the streetscapes.
- Since the camera angles and the driving lanes are different among automobiles, the appearance of the streetscape in the images changes significantly even in the same location.

While the first two issues are dealt with by existent methods, the last three are our own issues. We propose an image matching method that solves all the issues.

2 Image Matching Robust to Changes in the Capturing Conditions

2.1 Panoramic Image Matching

Images with the position information are collected from automobiles. It is impossible to make a correspondence between the images observed in the same location only with a GPS position due to its errors of

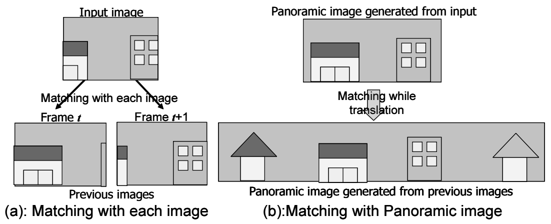

Figure 1: Image matching.

about 15m. Therefore, a currently observed image at GPS position P must be compared with previous images observed at GPS positions $P\pm$ 15m. Even if the current image is compared with one of the previous images captured in the location nearest to where the current image is captured ("Previous images" in Fig. 1 (a)), searching for an overlapping region between these two images is difficult because the overlapping region is small due to the long capturing intervals. Therefore, our method compares current and previous images using their panoramic images. Each panoramic image is created by connecting time-series images as shown in Fig. 1(b). The width of the panoramic image is determined empirically so that a correct matching result is obtained stably as follows. In this paper, we assume that an automobile moves at 40km/h and the images are sent at 3 fps in a standard situation. Under the assumption, the current/previous panoramic image consists of 12/32 a part of image [1] (an image of interest captured at time t + next 11/31 frames); the width of the previous panoramic image is that of the current panoramic image with $\pm$ 15m.

For matching, the current panoramic image is translated on the previous panoramic image and compared with the overlapping region. Since the overlapping is large in the panoramic images, matching can be stable.

The following panoramic image matching methods similar to our method have been proposed: (1) [4] creates a panoramic image by concatenating a part of omnidirectional and temporally dense images captured while a camera moves along a straight line parallel to a streetscape and stop at regular intervals, and (2) [5, 6] create a panoramic image by concatenating sequential isometric line-scanned images. However, all of these methods cannot deal with (1) a conventional camera directed towards a different direction from a target streetscape and (2) the different capturing conditions such as different moving speeds and different camera angles. As mentioned before, our system must deal with these difficulties.

Our method copes with these difficulties by (1) rectifying each captured image based on the camera angle and (2) dynamically adjusting the width of the image region, which becomes a part of a panoramic image, extracted from each captured image based on the moving speed of the camera. With the images rectified and adjusted, panoramic images acquired at different times can be similar. The similar panoramic images facilitate matching an overlapping region between them.

The features for the matching are extracted from less-changed streetscape area in each observed image. This is because the streetscape area has the characteristic features for identifying the capturing position. Its detail is described in Sec. 2.4).

[1]hereafter, it is defined as the rectangle image

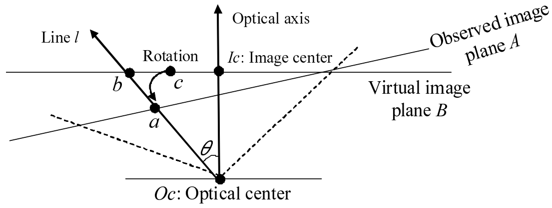

Figure 2: Rotation rectification.

Figure 3: Pitch rectification result.

2.2 Rotation Rectification

First, we describe how to estimate the FOE(f_x,f_y) for camera angle estimation. When an automobile with the camera moves forward, all optical flows in observed images caused by this car's motion pass a point. The intersection of all the optical flows is called FOE that indicates the automobile's moving direction. For optical flow estimation, in our system, feature points are tracked by Lucas-Kanade tracker [7].

If the optical axis of the camera coincides with the 3D line from the optical center of the camera to the FOE in the image plane, the image plane is considered to be perpendicular to the moving direction of the automobile. The rotation angle of the camera is represented by pan (θ) and tilt (ϕ) angles between the optical axis, and the 3D line is obtained by the following formulas: $\theta = \tan^{-1}\frac{f_x}{fl}$ and $\phi = \tan^{-1}\frac{f_y}{\sqrt{(fl^2+f_x^2)}}$, where fl denotes the focal length of the camera, which is known because all camera parameters required are sent with observed images from the automobile.

Note that the processes mentioned above in this section (i.e., FOE estimation and camera angle estimation) are executed once as long as the camera angle is not changed in each automobile.

Each observed image is then rectified with the rotation angle of the camera. Fig. 2 illustrates the plane consisting of the optical axis and the x axis of the image in a 3D coordinate system. Let virtual image plane B be perpendicular to the optical axis, while observed image plane A is rotated by θ and ϕ. To obtain a pixel value for each point b in B for generating a rectified image, the intersection, a, of A and the line, l, that is determined by b and the optical center O_c is computed. The 2D coordinates of a in A can be calculated by rotating a by $-\theta$ and $-\phi$ and computing its 2D coordinates (i.e., 2D coordinates in c in Fig. 2) in B. a is rotated towards c. This process is performed for all pixels in B to generate the rotation-rectified image.

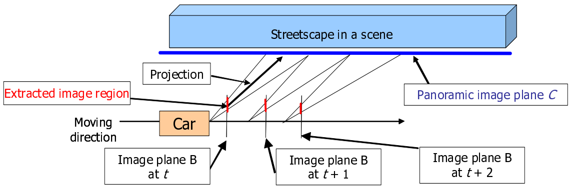

Figure 4: Panoramic image generation.

2.3 Pitch rectification

If a panoramic image is generated (will be described in Sec. 2.4) without taking into account up and down shakes of cars, it might have zig-zag effects as shown in a left-hand image in Fig. 3. To suppress this effect, up and down motions are rectified as follows.

Feature points are extracted and tracked in rotation-rectified images t and $t+1$. These points are extracted from a region which is used for generating a panoramic image. If an automobile goes straight, corresponding points should be along a line determined by the FOE and one of the points. Therefore, the image observed at $t+1$ is translated upward/downward so that three points (i.e., the FOE and the corresponding points) are on a line. The translated displacement is determined by the median of displacements of all the feature points. With this rectification, the zig-zag effects can be suppressed as shown in a right-hand image in Fig. 3.

2.4 Panoramic Image Generation

A rectangle image is created by extracting a part of left side of an observed image, in which a streetscape is captured. Fig. 4 illustrates how to generate a panoramic image by concatenating the rectangle images extracted from the observed images. We assume that a panoramic image plane C (as shown in Fig. 4) is parallel both to the frontal surface of a streetscape and the automobile's moving direction. Time-series images ("Image plane B at t, $t+1$, and $t+2$" in Fig 4) are projected onto the panoramic image plane. The panoramic image is generated as follows.

(1) Region extraction: The rectangle image is extracted from the mid point between the image center and the leftmost side of the image. The width of the rectangle image should be determined by a horizontal displacement between two consecutive frames for smoothly concatenating them in a panoramic image plane. The displacement is determined by the horizontal components of optical flows.

(2) Image projection: The rectangle images extracted from sequential images are projected onto a panoramic image plane and concatenated. All the projections are defined by the perspective projection. Panoramic image plane C is perpendicular to rotation-rectified image plane B. The projection between B to C at time t is, therefore, defined only by the position of the optical center at t (the position is denoted by Oc_t). Oc_t is determined by (1) the distance between Oc_t and C (denoted by d) and (2) the position of Oc_t along the optical axis (denoted by Oa).

It is obvious that d is just a scaling factor in the perspective projection between C and each B. d can be, therefore, determined arbitrarily. In our experiments,

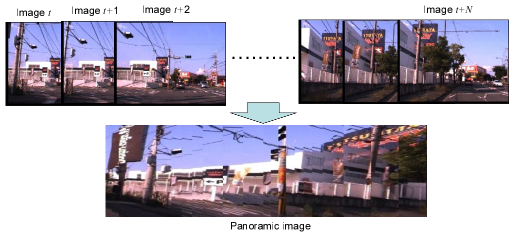

Figure 5: Generated panoramic image.

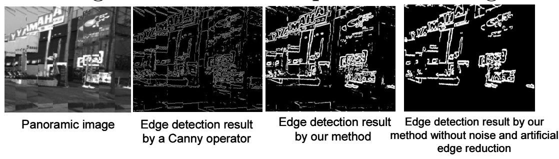

Figure 6: Edge detection and noise extracted result.

d was determined empirically so that horizontal lines that face a road (e.g, boundary lines of buildings) in a 3D scene were lined horizontally in C and concatenated as smooth as possible.

As described above, extracted images are projected onto C and concatenated side by side. Given d, therefore, the plane displacement $dis_{t,t+1}$ is not required for panoramic image generation. That is, all the projected images in C are concatenated side by side in order of their capturing times after the extracted image in B_t is projected on C from an arbitrarily located Oc_t at t.

An example of the generated panoramic image is shown in the bottom in Fig. 5. The panoramic image has zig-zag effect on the boundaries of concatenated images. This zig-zag effect is caused because the images captured from different optical centers are simply concatenated. Our objective is, however, not producing a smooth panoramic image but generating a panoramic image that enables stable image matching. All processes in our system should be as fast as possible because a number of images are analyzed. This is why panoramic image generation based on simple perspective projection, which is described in this section, is employed in our system.

2.5 Features for Robust Matching

In our method, edge features are employed for image matching robust to a change in illumination.First of all, the edges are extracted using Sobel operator, and thresholding. The threshold was determined manually based on extensive experiments using various images.

This edge detection is not applied to the boundaries between connected rectangle images. This is because the boundaries in our panoramic image is not smooth geometrically and optically;the non-smooth boundaries result in artificial edge lines.

Only using Sobel operator, if the hairline lines are congested like characters in signs and roadside trees, they might be detected undesired. Next, therefore, these edges are removed based on the size of connected edge pixels: (1) all detected edge pixels are segmented based on connection of 8-neighboring pixels and (2) the segmented edge pixels are removed if its size (the number of the connecting pixels) is smaller than a predefined threshold(as shown in Fig. 6). Furthermore, a previous panoramic image is blurred by a Gaussian filter. This blurring is required for matching robust to minute differences between images.

2.6 Image Matching

The image matching is performed using the previous and current edge panoramic edge images. In order to cope with the difference between these images due to observation from different lanes, correlation between the images is calculated as follows:(1) the current image is reshaped to various sizes, and (2) each size of the current image is translated on the previous image for calculating the normalized correlation between them at each position.

The current image matches the part of the previous image in which the normalized correlation becomes maximum. In our method, the image size is changed between 0.5 and 1.5 times at intervals of 0.1 times.

3 Experiments

Experimental data has been collected for half a year in Nara, Japan. All the data was obtained by a monocular camera and an off-the-shelf GPS. Two cars were used for data collection. The cars ran at 30-60 km/h. Although the camera was always set under a rearview mirror so that its view direction matched roughly the automobile's moving direction, the view direction changed slightly. The JPEG-compressed image size was 640x480 pixels. The frame-rate was 3-4fps. All internal camera parameters were estimated using [8]. The GPS receiver got its location once per second.

First of all, the ground truth of the matching region is found manually. Let ld/rd be the distances between the leftmost/rightmost line of the ground truth and the leftmost/rightmost line of the matching result. If both of ld and rd are less than 20 pixels, the matching result is regarded as "Success".

In order to confirm the robustness to the change under various capturing conditions (i.e., illumination, weather, road, change in a signs, change in a driving lane), image sequences captured in 20 locations were selected and compared for image matching. Two sequences captured in the same location were selected and compare with each other.

Our matching method could establish successful results in 17/20 locations. Unsuccessful results were obtained in locations H, I, and Q. This means that matching was successful without being disturbed by the changes in illumination, observed streetscapes, and driving lanes between the current and previous images, which are the target problems in this paper. Note that the imaging conditions in location H was more severe than the change in driving lanes between the images; in location H, an automobile changed its driving lane while the images were being captured.

Unsuccessful matching results in locations H, I, and Q are shown in Fig. 7. In these image, red and white rectangles in each previous image denote the region matched with the current image and the groundtruth region, respectively. The unsuccessful results might be obtained due to the following reasons:

Location H: The previous panoramic image was deformed significantly due to lane changing.

Figure 7: Current and previous panoramic images in locations H, J, N, and Q.

Location I: An automobile was going round a curve. This resulted in the failure in rotation and pitch rectifications.

Location Q: Optical flow estimation was failed due to large and wide roadside trees. Then the rectangle images extracted from sequential observed images were discontinuous.

The changes in location H was temporary. Therefore, matching is successful in these locations after the changes disappear and then the unsuccessful results are ignored. It is essentially difficult to get a successful matching result in a scene, in which good flows and/or matching features are not available, such as location I and Q. The images observed in such a scene should be matched based on successful matching results obtained before and after this scene. This is included in our future work.

4 Concluding Remarks

We proposed a method for matching streetscape images captured at different times by a conventional low-resolution camera and an off-the-shelf GPS.

While our method can get plausible results in many cases, it should be improved more in terms of improving panoramic images and robustness to extreme changes in weather conditions.

References

[1] J. Sato, T. Takahashi, I. Ide, and H. Murase, "Change detection in streetscapes from GPS coordinated omni-directional image sequences," *ICPR*,pp.I–531–534,2006.

[2] R. Bellman and R. Kalaba, "Dynamic Programming and Modern Control Theory," Academic Press,1965.

[3] K.Kato, H.Ishiguro and M.Barth, "recording of street views by using omnidirectionalvision sensors," *IECON*, Vol.4, pp.2571–2576,2000.

[4] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, R. Szeliski, "Photographing long scenes with multiview-point panoramas,"ACM *TOGr*,Vol.25, No.3, pp.853–861, 2006.

[5] J. Yu Zheng and S. Tsuji, "Panoramic Representation for Route Reconstruction by a Mobile Robot,"*IJCV*, Vol.9, No.1, pp.55–76, 1992.

[6] Min Shi and J. Yu Zheng, "Spatial resolution analysis of route panorama,"*ICIP*, Vol.25, pp.311–314,2003.

[7] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *IJCAI*, pp.674–679, 1981.

[8] Z. Zhang, "A Flexible New Technique for Camera Calibration," PAMI, Vol.22, pp.1330–1334, 2000.