Change detection in streetscapes from GPS coordinated omni-directional image sequences

Junji Sato, Tomokazu Takahashi, Ichiro Ide and Hiroshi Murase
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, Japan
{jsato, ttakahashi}@murase.m.is.nagoya-u.ac.jp, {ide, murase}@is.nagoya-u.ac.jp

Abstract

As part of ITS technology, to achieve quick map updates, we propose a method for automatically detecting changes in streetscapes from images captured by car-mounted omni-directional cameras. It comprises two stages; accurate alignment of a map and street images taken at various times, and detection of changes in streetscapes from the aligned data. The system will collect data via many free-running cars fitted with low-cost equipment to obtain images at various times and along routes. In the first stage, we process the alignment of the image frames taken at same locations and determine the accurate position information of each frame by a method composed of dimension reduction and DP matching. Then in the second stage, we detect changes in streetscapes from images taken at various times. Experiments with 44 data items which were collected over about a year, demonstrate the effectiveness of our method.

1. Introduction

Recent years have seen great advances in the development of ITS technology. One part of this is the enhancement of the car navigation systems in terms of offering better information to drivers [1, 2]. In these systems, navigation map is important. However, it needs frequent update because of streetscape changes; redecorations and constructions of buildings. Updating a map, however, costs much, since many people have to walk through the city and collect a lot of relevant information.

For quick and efficient updating of maps, we propose a method of automatically detecting changes in streetscapes from omni-directional images taken from cars. It is composed of two stages. The first stage accurately aligns a map and street images taken at various times, while the second stage detects changes in streetscapes from aligned data. In collecting street image data, we assume that many cars with general GPS and an omni-directional camera run freely without purposing to collect the data. Consequently, we can obtain a large amount of GPS coordinated images taken at various times and routes with a low-cost system. We use an omni-directional camera to efficiently capture images in all directions from the cars. To collect position information, high-accuracy GPS such as RTK GPS may be available, but it does not suit our application because of its cost. Therefore, we assume the use of a general GPS or D-GPS fitted to conventional car navigation systems. Such GPS, however, has about a 10-meter margin of error, meaning that we cannot correctly take images at same locations by simply collecting images that have the same coordinates.

For the reasons mentioned above, we need to solve the following three problems.

1. Accurate alignment of images of the same location from images collected by freely running cars at various times.

2. Improvement of the position information attached to image frames.

3. Detection of changes in streetscapes from images taken at various times.

In this paper, to deal with these issues we propose a novel method comprising alignment of images and calculation of the difference between aligned image frames. In the first stage, we solve the first two problems. We align image frames at the same location by matching image sequences taken along a roughly identified GPS coordinate route. For image matching, we integrate dimension reduction by PCA and DP matching, then accurately determine the position information of each frame by calculating average coordinates for the aligned images. We call this aligned image data a *Street Image Map*. Then, in the second stage, we calculate the difference between aligned images taken at various times and detect any change in streetscapes.

2. Detection of changes in streetscapes

We collect a large amount of images with their GPS coordinates, construct a Street Image Map, and detect changes

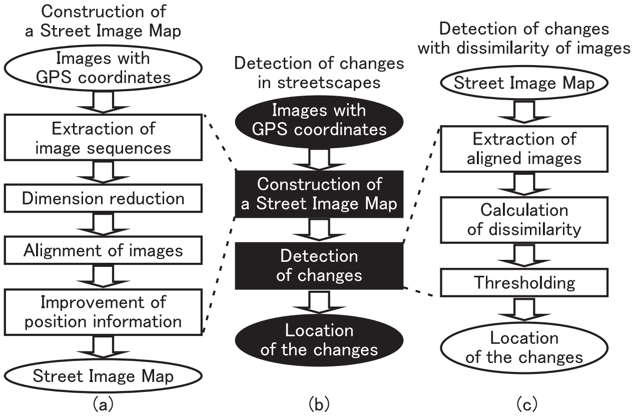


Figure 1. Detection of changes in streetscapes.

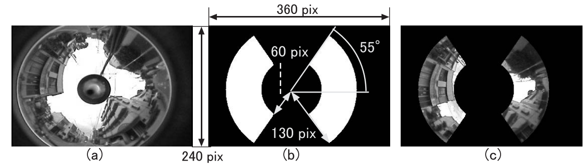


Figure 2. (a) Omni-directional image, (b) Mask to extract feature vectors (N=25,538), (c) Masked image.

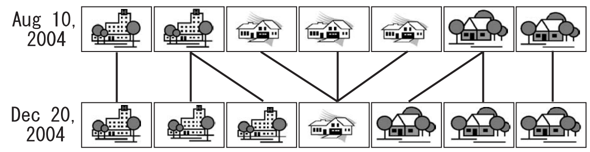


Figure 3. Alignment of two sequences of video streams by DP matching.

in streetscapes according to the process shown in Fig. 1.

2.1. Collection of images with GPS coordinates

A full implementation of this system involves many cars running freely with an omni-directional camera and GPS. The omni-directional camera is attached to the top of the car to take images in all directions. We take coordinates from GPS synchronized with the images. Generally, however, the frequency of GPS updates does not correspond to video frequency, so we provide coordinates to the frames without GPS coordinates by linear interpolation.

2.2. Stage 1: Construction of a Street Image Map

We accurately align image frames at the same location from large amounts of cityscape data. We call this a Street Image Map. Fig. 1(a) shows the process.

2.2.1 Alignment of frames

Since the data we collect is taken by a lot of cars runnning freely, we must extract image sequences taken on the same route. We can distinguish the route with GPS coordinates because GPS has an accuracy of about 10 m, thus we extract image sequences taken along the same route using such GPS coordinates. As a result, we can obtain images along the same route on various dates.

Next, we apply PCA[3] to reduce the dimensions of the feature vectors of each image frame. This makes it possible to some extent reduce the amount of calculation, the required storage space and the influence of illumination changes accompanied by weather changes. The feature vector is an $N \times 3$ dimension vector that has R, G and B values for pixels in the masked area shown in Fig. 2, where N is the number of pixels in the area. This vector is normalized so that the average of its components should be zero and its norm one. We limit the area by the mask because pedestrians or other cars may be on the edges, top, or bottom of omni-directional images, which can be detrimental for matching. Before dimension reduction, with PCA we create a lower-dimension eigenspace than the feature space using various city images, then project each frame feature vector to the eigenspace and obtain the sequence of points $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p\}$ on the eigenspace. Here, p is the number of frames.

Next, we align dimension-reduced images frame by frame. This is the first point of our method. By this process, it is possible to align frames that reflect the same location at various times. We use DP matching to absorb temporal expansion and contraction caused by differences of car speed and to achieve alignment through all the images (Fig. 3). We apply Eq. 1 below recursively to the two sequences of points on the eigenspace : $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p\}$ and $\{\mathbf{y}'_1, \mathbf{y}'_2, \cdots, \mathbf{y}'_q\}$, and employ Euclidean distance on the eigenspace $d(i,j) = ||\mathbf{y}_i - \mathbf{y}'_j||^2$ as the dissimilarity.

$$D(i,j) = \min \begin{cases} D(i-1, \quad j \ \ ) + \omega_1 \cdot d(i,j) \\ D(i-1, j-1) + \omega_2 \cdot d(i,j) \\ D(\quad i \ \ , j-1) + \omega_3 \cdot d(i,j) \end{cases} \qquad (1)$$

Here, $D(1,1) = d(1,1)$. In an experiment, we set certain values for $\omega_1, \omega_2, \omega_3$. Furthermore, the sequence of frame number pairs (i,j) chosen up to $D(p,q)$ have been calculated to show matches of two image frames. If the matching breaks down and we cannot obtain any correct matching results at all, $D(p,q)$ will be a very large value. Consequently, we set a threshold d_D to discard the results that exceed the threshold as outliers.

2.2.2 Improvement of position information accuracy

The second point of our method is to average GPS coordinates attached to the aligned images and aim to determine the accurate position information. Generally, it is assumed that the average of the coordinates measured at a particular location for a long time converges at the true coodinates. Based on this assumption, it is assumed that the average coordinates are more accurate than the collected data.

2.3 Stage 2: Detection of changes

It is the third point of our method to detect changes in streetscapes from images of a specified location aligned in the first stage. Figure 1(c) shows this process.

We sort the aligned frames of the specified location in order of time, namely, $P_0, P_1, P_2, \ldots$, and calculate the dissimilarity between P_0 and the other frames P_i $(i = 1, 2, \ldots)$ by approximately adjusting their positions to reduce the influence of the lane position of the car. We transform the omni-directional images P_0 and P_i to panoramic images [4], divide them into the right side and the left side of the car, and label them P_{0l}, P_{0r}, P_{il}, and P_{ir}. In a similar fashion, we transform the mask image to M_l and M_r. The process of calculating the dissimilarity of P_0 and P_i is as follows.

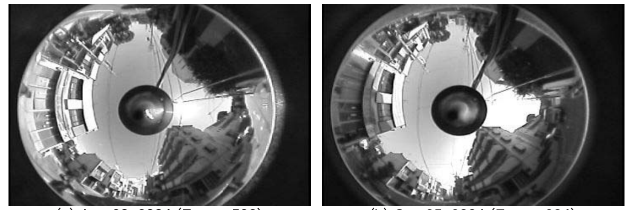
1. Enlarge P_{0l} and P_{ir} to match them with P_{il} and P_{0r} respectively, or vice versa.

2. To equalize the number of pixels, move M_l onto P_{0l} and P_{il} and M_r onto P_{0r} and P_{ir}. Extract the feature vectors of the R, G, and B values of the masked area and normalize them. Then calculate the dissimilarity $d_i(P_0, P_i) = ||\mathbf{x}_0 - \mathbf{x}_i||$, where x_0 and x_i are the feature vectors.

3. Repeat steps 2. while moving the mask within a certain area, and calculate the minimum dissimilarity $d_k(P_0, P_i)$.

4. Repeat steps 1. to 3. while changing the enlargement factor within a experimentally given range. The dissimilarity of the two frames is defined as the minimum value $d_{min}(P_0, P_i)$ of $d_k(P_0, P_i)$.

This process enables us to obtain the dissimilarities between P_0 and the other frames of the same location. If the variance of them exceeds the threshold d_T, we determine that there is a change, but if the camera catches a reflection off a large car or if image alignment fails at certain frames, the dissimilarity will rise temporarily. Therefore we restrain it from temporarily rising by median filter and gaussian filter smoothing over the time sequence of the dissimilarities.



(a) Aug 20, 2004 (Frame 533) (b) Oct 25, 2004 (Frame 284)

Figure 4. Example of aligned frames.

3. Experiments

3.1 Stage 1: Construction of a Street Image Map

First, we experimented on the construction of a Street Image Map, confirming the accuracy of image alignment. We used 44 data items collected for about a year, and aligned the oldest data with the other data. For the alignment, we extracted a route (route B, about 170 m) from the data, and reduced the 64,458 dimensions of the feature vectors to 20 dimensions. In terms of DP matching, the weight factor $(\omega_1, \omega_2, \omega_3)$ is $(2, 1, 2)$ from Eq. 1. Though this was the best weight in the pilot study, we assume there would be better values than that. We chose threshold d_D of the image sequence dissimilarity $D(p, q)$ as $\mu + \sigma$, where μ is the average of the dissimilarities of all results and σ is their standard deviation.

We judged the results by manual checking. If a frame was aligned to its most similar frame, we judged it to be correct, while if there were some frames more similar than it, we judged it as false. We evaluated the rate of correctly aligned frames in all frames, which resulted in an average of 94.1%. Figure 4 shows a part of frames aligned at a certaion location on route B, and Fig. 5 shows the GPS coordinates attached to the correctly aligned frames at that location (Fig. 4) and their average coordinates. Their standard deviation was 6.86 m, which is assumed to be the accuracy of the GPS coordinates for measurement over a long time. The average of these standard deviations over all frames was 7.98 m. The data are collected at speeds under 40 km/h, and the frame rate was 30 fps, so the distance between the locations of two consecutive frames was less than 0.4 m. Thus the average coordinate has a 0.4 m margin of error in the car's direction of travel even if the alignment is correct and the position information is accurate. Therefore, it is assumed that if the number of samples increases, we can converge the error of position information to about 0.4 m in the direction of travel.

To examine the alignment performance of DP matching, we conducted an experiment on matching by aligning frames with the most similar frame in all frames on the eigenspace. As a result of the experiment applied to all data, while we achieved a success rate of 94.1% by DP matching, we obtained only 35.4% by full-search frame match-

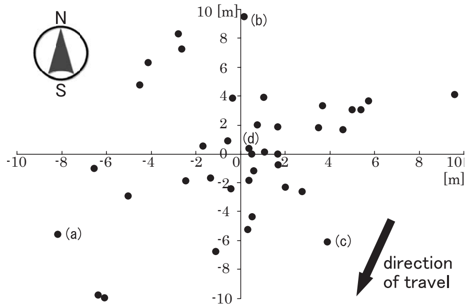


Figure 5. GPS coordinates attached to the aligned frames (Fig. 4): The origin is the average.

ing. When checking the result manually, the most similar frames on the eigenspace did not correspond to the correct frames: some had shifted and some were out of sequence. Thus it is assumed that even though the similarity on the eigenspace is not necessarily high, we could still achieve high performance thanks to the effective alignment by DP matching over all images.

3.2 Stage 2: Detection of changes

We alse experimented on detection of changes using the same data used in Section 3.1. This time, there were four extracted routes. They included four major changes in the streetscapes. Threshold d_T was set experimentally.

According to the results, we correctly detected three changes out of four. Figure 6 shows images of detected locations. Though we used omni-directional images in this experiment, images taken with a digital camera at the same location on the same dates are shown here for easy viewing. These images reveal that there were changes at the detected locations.

We detected three changes out of four by our method. For the remaining case, the change was very little even from human eyes, our method could not distinguish a sufficiently large dissimilarity.

4. Conclusion

In this paper, we proposed a new method that detects changes in streetscapes from many street images taken at various times. In the first stage we used data comprising images and GPS coordinates taken together over a long period. We aligned images taken at the same location at various times and then averaged their position information. The alignment of images was achieved by DP matching on eigenspace. In the second stage, we detected changes using dissimilarities between aligned images. Experiments produced the following three results.

- We aligned images with a high accuracy of 94.1%. It is therefore possible to align images taken at the same location at various times.

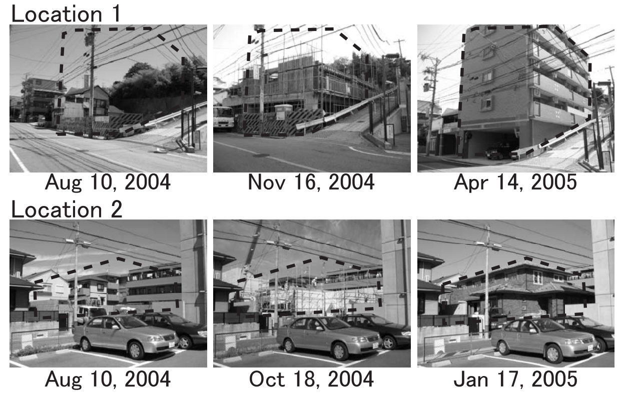


Figure 6. Example of images at change-detected locations: The area where the streetscape changed significantly is marked by dashed lines.

- GPS coordinates attached to the aligned images had a 7 to 8 m accuracy. We confirmed that if the alignment is accurate and we collect more data, we could converge the error to less than 0.4 m in the direction of travel.
- We detected three changes of streetscapes out of four from real-world data.

Future work will include applying a larger amount of data. We plan to create a Steet Image Map, featuring street images taken at various times attached to accurate position information in the first stage of our method. We assume this will have various applications in addition to detection of changes, for car navigation systems or driving simulation systems [5] that use real-world images.

Acknowledgement

Parts of this research were supported by the Grant-In-Aid for Scientific Research (16300054) and the 21st century COE program from the Ministry of Education, Culture, Sports, Science and Technology. This work is developed based on MIST library (*http://mist.suenaga.m.is.nagoya-u.ac.jp/*).

References

[1] J. Y. Zheng and X. Wang: "Pervasive Views: Area exploration and guidance using extended image media", Proc. 13th ACM Intl. Conf. on Multimedia, pp.986-995, Nov. 2005

[2] E. Verbree and A. V. Anrooij: "Interactive navigation through distance added valued panoramic images", ISPRS, Panoramic Photogrammetry Workshop, Feb. 2004

[3] H. Murase and S. Nayar: "Visual learning and recognition of 3-D object from appearance", Int. J. Comput. Vision, vol. 14, pp. 5-24, Jan. 1995

[4] Y. Yagi and M. Yachida: "Real-time omnidirectional image sensors", Int. J. Comput. Vision, vol. 58, no. 3, pp. 173-207, Jul. 2004

[5] S. Ono *et al*: "A photo-realistic driving simulation system for mixed-reality traffic experiment space", IEEE IV 2005, pp.747-752, Jun. 2005