Vehicle Ego-localization by Matching In-vehicle Camera Images to an Aerial Image

Masafumi NODA[1,*], Tomokazu TAKAHASHI[1,2], Daisuke DEGUCHI[1], Ichiro IDE[1], Hiroshi MURASE[1], Yoshiko KOJIMA[3] and Takashi NAITO[3]
[1] *Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464–8601, Japan*
[2] *Gifu Shotoku Gakuen University, Nakauzura 1–38, Gifu, 500–8288, Japan*
[3] *Toyota Central Research & Development Laboratories, Inc., 41–1 Aza Yokomichi, Oaza Nagakute, Nagakute, Aichi, 480–1192, Japan*
[*] *mnoda@murase.m.is.nagoya-u.ac.jp*

Abstract. Obtaining an accurate vehicle position is important for intelligent vehicles in supporting driver safety and comfort. This paper proposes an accurate ego-localization method by matching in-vehicle camera images to an aerial image. There are two major problems in performing an accurate matching: (1) image difference between the aerial image and the in-vehicle camera image due to view-point and illumination conditions, and (2) occlusions in the in-vehicle camera image. To solve the first problem, we use the SURF image descriptor, which achieves robust feature-point matching for the various image differences. Additionally, we extract appropriate feature-points from each road-marking region on the road plane in both images. For the second problem, we utilize sequential multiple in-vehicle camera frames in the matching. The experimental results demonstrate that the proposed method improves both ego-localization accuracy and stability.

1 Introduction

The vehicle ego-localization task is one of the most important technologies for Intelligent Transport Systems (ITS). Obtaining an accurate vehicle position is the first-step to supporting driver safety and comfort. In particular, ego-localization near intersections is important for avoiding traffic accidents. Recently, in-vehicle cameras for the ego-localization have been put to practical use. Meanwhile, aerial images have become readily available, for example from Google Maps [1]. In light of the above, we propose a method for accurate ego-localization by matching the shared region taken in in-vehicle camera images to an aerial image.

A global positioning system (GPS) is generally used to estimate a global vehicle position. However, standard GPSs for a vehicle navigation system have an estimation error within about 30–100 meters in an urban area. Therefore, a relatively accurate position is estimated by matching information, such as a geo-location and an image taken from a vehicle, to a map. Among them, map-matching [2] is one of the most prevalent methods. This method estimates

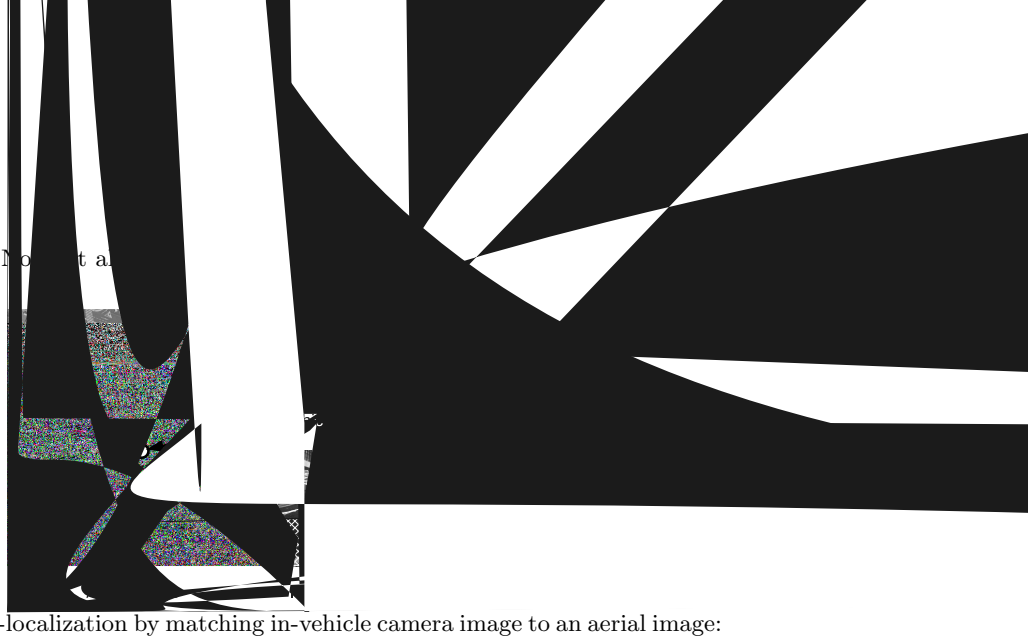


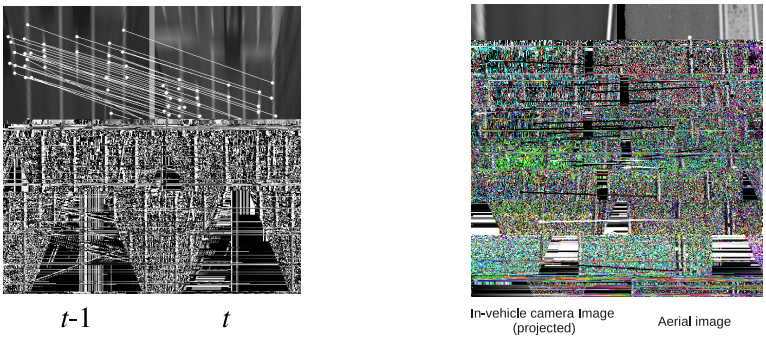
Fig. 1. Vehicle ego-localization by matching in-vehicle camera image to an aerial image: Shaded regions in both images correspond.

a vehicle position by matching a vehicle's driving trajectory calculated from rough estimations using GPS to a topological road map. Recently, in-vehicle cameras have been widely used; therefore, vehicle ego-localization using cameras has been proposed [3–5]. This camera-based vehicle ego-localization matches in-vehicle camera images to a map, which is also constructed from in-vehicle camera images. In many cases, the map is constructed by averaging in-vehicle camera images with less-accurate geo-locations. Therefore, it is difficult to construct a globally consistent map.

In contrast, aerial images that covers a wide region and with a highly accurate geo-location have also become easily available, and we can collect them at low-cost. There are some methods that ego-localize an aircraft by matching aerial images [6, 7]. However, the proposed method estimates a vehicle position. The proposed method matching the shared road-region of in-vehicle camera images and an aerial image is shown in Figure 1. Pink et al. [8] have also proposed an ego-localization method based on this idea. They estimate a vehicle position by matching feature-points extracted from an aerial image and an in-vehicle camera image. An Iterative Closest Point (ICP) method is used for this matching. As feature-points, the centroids of road markings, which are traffic symbols printed on roads, are used. This method, however, has a weakness in that a matching error occurs in the case where the images differ due to illumination conditions and/or occlusion. This decreases ego-localization accuracy.

There are two main problems to be solved to achieve accurate ego-localization using in-vehicle camera images and an aerial image. We describe these problems and our approaches to solve them.

1) **Image difference between the aerial image and the in-vehicle camera image**: The aerial image and the in-vehicle camera image have large difference due to viewpoints, illumination conditions and so on. This causes difficulty in feature-point matching. Therefore, we use the Speed Up Robust Feature (SURF) image descriptor [9]. The SURF image descriptor is robust for such differences of view and illumination. Additionally, since the road-plane region in the images has a simple texture, the feature-points extracted by a general method tend to be too few and inappropriate for the matching.



(a) Sequential matching between projected images.

(b) Matching the projected image to the feature-points map.

Fig. 5. Two step matching (Corresponding feature-point pairs in the projected images: The dots represent the feature-point in each image and the lines show their correspondence).

to obtain the accurate corresponding region $\mathcal{R}_t$. In this matching, in order to improve the accuracy and stability in a situation where occlusions occur in the in-vehicle camera image, multiple in-vehicle camera frames are used. We first explain a matching method the only uses a single frame, and then how to extend it to that uses multiple frames.

Matching using a Single Frame We extract the feature-points from the projected images in the same manner as described in Section 2. The position of a feature-point extracted from I_t is represented as $\mathbf{y}_{t,l_t}(l_t = \{1, \ldots, L_t\})$, where L_t is the number of feature-points. The SURF descriptor of $\mathbf{y}_{t,l_t}$ is represented as $\mathbf{g}_{t,l_t}$. Thus, the feature-points could be represented as $\{(\mathbf{y}_{t,1}, \mathbf{g}_{t,1}), \ldots, (\mathbf{y}_{t,L_t}, \mathbf{g}_{t,L_t})\}$.

For the matching, each feature-point position $\mathbf{y}_{t,l_t}$ is transformed to $\mathbf{y}'_{t,l_t}$ in the map as

$$\mathbf{y}'_{t,l_t} = \mathbf{A}_{t-1}\mathbf{M}_t\mathbf{y}_{t,l_t}. \tag{3}$$

Feature-point pairs are chosen so that they meet the following conditions:

$$\begin{cases} ||\mathbf{y}'_{t,l_t} - \mathbf{x}_n|| < r \\ \min\limits_{l_t} ||\mathbf{g}_{t,l_t} - \mathbf{f}_n|| \end{cases}, \tag{4}$$

where r is the detection radius. Figure 5(b) shows the feature-point pairs. Then, $\boldsymbol{\Sigma}_t$ is obtained by minimizing the LMedS criterion by selecting the correspondences.

Matching using Multiple Frames To achieve accurate matching in a situation where occlusions occur in some in-vehicle camera images, we integrate the feature-points in the multiple in-vehicle camera frames. The feature-points at t' are represented as $\mathcal{Y}_{t'} = \{\mathbf{y}_{t',1}, \ldots, \mathbf{y}_{t',L_{t'}}\}$. They are transformed to $\mathcal{Y}'_{t'} =$

Table 1. Dataset

		Aerial image	In-vehicle camera image	
Set No.	Length (m)	Occlusion	Occlusion	Time
1	85	small	small	day
2	100	small	small	night
3	100	small	large	day
4	75	large	large	day

$\{\mathbf{y}'_{t',1}, \ldots, \mathbf{y}'_{t',L_{t'}}\}$ in the map coordinate. $\mathbf{y}'_{t',1}$ is transformed as

$$\mathbf{y}'_{t',l_{t'}} = \begin{cases} \mathbf{A}_{t'-1}\mathbf{M}_{t'}\mathbf{y}_{t',l_{t'}} & t' \text{ is current frame} \\ \mathbf{A}_{t'}\mathbf{y}_{t',l_{t'}} & \text{otherwise} \end{cases}. \tag{5}$$

Then, the feature-points in the F multiple frames including the current frame are used for the matching. Then, we obtain $\boldsymbol{\Sigma}_t$ in the same manner as in the case of a single frame.

3.5 Estimation of the Vehicle Position

Finally, $\mathbf{A}_t$ is calculated by Equation 2, and the vehicle position $\mathbf{p}_t$ is estimated by Equation 1. As for the matrix $\mathbf{A}_0$ at the initial frame, it is obtained by a global matching method in the map without the estimation of $\hat{\mathcal{R}}_0$

4 Experiment

4.1 Setup

We mounted a camera, a standard GPS and a high accurate positioning system (Applanix, POSLV) [10] on a vehicle. The standard GPS contains an error of about 5–30 meters, which was used for the initial frame matching. The high-accuracy positioning system was used to obtain the reference values of vehicle positions. We used four sets of an aerial image and an in-vehicle camera image sequence with different capturing conditions. Table 1 shows the specification of the datasets and Figure 6 shows examples. The resolution of the aerial image was 0.15 meters per pixel. The resolution of the in-vehicle camera image was 640×480 pixels, and its frame-rate was 10 fps. Occlusions in the aerial image occurred due to vehicles, trees and so on. Occlusions in the road regions in an aerial image occurred due to vehicles, trees and so on. We defined a road segment in an aerial image which was occluded less than 10% as a small occlusion, and that occluded more than 50% as a large occlusion by visual judgment. Occlusions in the in-vehicle camera images were due to forward vehicles.

4.2 Evaluation

We evaluated the ego-localization accuracy by the Estimation Error and the Possible Ratio defined by the following equations:

$$\text{Estimation error} = \frac{\text{The sum of estimation errors in available frames}}{\text{The number of available frames}}, \tag{6}$$