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

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 [3,4,5]. 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.