Real-Time Hierarchical Outdoor SLAM Based on Stereovision and GPS Fusion

David Schleicher, Luis M. Bergasa, *Member, IEEE*, Manuel Ocaña, Rafael Barea, and María Elena López

Abstract—**This paper presents a new real-time hierarchical (topological/metric) simultaneous localization and mapping (SLAM) system. It can be applied to the robust localization of a vehicle in large-scale outdoor urban environments, improving the current vehicle navigation systems, most of which are only based on Global Positioning System (GPS). Then, it can be used on autonomous vehicle guidance with recurrent trajectories (bus journeys, theme park internal journeys, etc.). It is exclusively based on the information provided by both a low-cost, wide-angle stereo camera and a low-cost GPS. Our approach divides the whole map into local submaps identified by the so-called fingerprints (vehicle poses). In this submap level (low-level SLAM), a metric approach is carried out. There, a 3-D sequential mapping of visual natural landmarks and the vehicle location/orientation are obtained using a top-down Bayesian method to model the dynamic behavior. GPS measurements are integrated within this low-level improving vehicle positioning. A higher topological level (high-level SLAM) based on fingerprints and the MultiLevel Relaxation (MLR) algorithm has been added to reduce the global error within the map, keeping real-time constraints. This level provides nearly consistent estimation, keeping a small degradation with GPS unavailability. Some experimental results for large-scale outdoor urban environments are presented, showing an almost constant processing time.**

Index Terms—**Global Positioning System (GPS), outdoor simultaneous localization and mapping (SLAM), stereovision, vehicle navigation system.**

I. Introduction

Autonomous vehicle guidance interest has increased in recent years thanks to events like the Defense Advanced Research Projects Agency Grand Challenge and, more recently, the Urban Challenge. Most of the research has focused on different location and navigation techniques using a previously known map [1], [2]. More recently, simultaneous localization and mapping (SLAM) has become a key component in vehicle navigation [3]–[5], following the trend of the robotics area, which has seen significant progress in the last decade. The interest in SLAM based in cameras has tremendously grown in recent years.

Cameras have become much more inexpensive than lasers and provide texture-rich information about scene elements at practically any distance from the camera. Currently, the main goal in SLAM research is to apply consistent, robust, and efficient methods for large-scale environments in real time. On the other hand, one of the most popular sensors in outdoor navigation is the GPS. However, their information is not always as accurate as needed, mainly due to the satellite's occlusion because of high buildings, tunnels, etc. To improve the usual vehicle navigation systems mainly in very populated urban areas, where the GPS information is not reliable, is an interesting goal. Onboard navigation systems solve this problem by coupling dead reckoning and GPS positions. Dead reckoning is based on the use of an onboard low-cost inertial measurement unit (IMU) and the measurement of the covered distance (available through ABS wheel speed sensors). Microelectromechanical system (MEMS) accuracy is a well-known problem [6]. Despite the many recent improvements in the characterization of acceleration and rotation rate, measurement errors due to the thermal stability of MEMS components cause drifts in pure integration cycle even with the aid of odometry. Fiber-optic gyrometers offer more accuracy than MEMS, but their cost does not comply with the automotive cost requirements. As a consequence, the behavior of a low-cost GPS and IMU fusion for urban scenarios with recurrent trajectories is not highly reliable. In addition, there is no standard access to the onboard sensors, being this information own and confidential for each manufacturer. On the other hand, most of the portable navigation systems are only based on GPS. With this background, our general approach consists of adapting SLAM strategies from the robotic area to the vehicle-localization problem.

One of the most popular methods to solve the SLAM problem is the extended Kalman filter (EKF). As it is well known, the EKF implementation is limited by the complexity of the covariance matrix calculation, which quadratically increases in large-scale maps as a function of the landmarks introduced into the filter. To deal with that problem, several approaches, like the so-called FastSLAM [7]–[9] or some others that try to reduce the complexity of EKF, either by modifying its intrinsic principles [10], [11] or by dividing the map into smaller ones using a metric [12]–[14] or topological approach [15], [16], were presented.

This paper relies on the topological/metric philosophy using local maps to represent the world and locate the vehicle within. Our approach basically generates a series of local submaps taken on an equally spaced basis (low-level SLAM). Each of them consists of a number of visual landmarks precisely taken and is handled by using a standard EKF. A topological

Manuscript received February 14, 2008; revised October 4, 2008 and February 25, 2009. First published July 21, 2009; current version published September 1, 2009. This work was supported in part by the Spanish Ministry of Education and Science under Grant TRA2005-08529-C02 (MOVICON Project) and Grant PSE-370100-2007-2 (CABINTEC Project) and in part by the Community of Madrid under Grant CM: S-0505/DPI/000176 (RoboCity2030 Project). The Associate Editor for this paper was M. Trivedi.

The authors are with the Department of Electronics, University of Alcala, 28805 Madrid, Spain (e-mail: dsg68818@telefonica.net; bergasa@depeca.uah.es; mocana@depeca.uah.es; barea@depeca.uah.es; elena@depeca.uah.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2009.2026317

map, along with local metric submaps, is built (high-level SLAM). The topological map is a graphlike map consisting of vertices and edges. Each vertex represents a topological place, a vehicle pose that we call *fingerprint*, and includes a local metric submap. When a vehicle is traveling between two vertices, an edge is inserted to connect these two vertices, which represents a link between two poses. Meanwhile, the edges store transformation matrices and uncertainties to describe the relationship between connected vertices. Using this hierarchical strategy of two levels, on one hand, we keep the local consistency of the submaps by means of the EKF, and on the other hand, we keep the global consistency by using the topological level and the MultiLevel Relaxation (MLR) method of Frese *et al.* [17]. The MLR algorithm determines the maximum-likelihood estimate of all vehicle vertices along the whole path. Vertex corrections are transmitted to the landmarks of their corresponding submaps.

Our final goal is the autonomous vehicle outdoor navigation in large-scale environments and the improvement of the current vehicle navigation systems based only on standard GPS. This includes the ability to dynamically improve the vehicle navigation maps (building new streets where nothing was previously mapped, correcting their paths, etc.). Our system is particularly efficient in areas where the GPS signal is not reliable or even not fully available (tunnels, urban areas with tall buildings, mountainous forested environments, etc.). Our research objective is to develop a robust localization system based on SLAM using only a low-cost stereo camera and a standard GPS sensor for vehicle navigation assistance. Then, this paper is focused on real-time localization as the main output of interest. A map is certainly built, but it is a sparse map of landmarks optimized toward enabling localization. However, this map is enough for updating obsolete navigator maps in real time as the vehicle covers new paths.

To obtain vehicle dead reckoning, our system uses visual information instead of an IMU because our goal is to develop a low-cost standard system that is independent of the manufacturer protocol confidentiality. Moreover, as a difference of IMU systems, our proposal generates a map and is able to detect loop closings using visual appearance information. This way, the accumulated drifts, which are typical of odometry sensors, are removed from time to time, even with GPS unavailability.

Finally, our hierarchical proposal of two levels (topologic and metric) works well in large-scale environments, producing topologically correct and geometrically accurate submaps at minimal computational cost. On the other hand, the topological level facilitates the path-planning strategies, the fusion with the GPS information, and the future generalization of the system to a multivehicle SLAM.

II. Related Work

In [18], Davison presented an impressive work of real-time 3-D visual SLAM carried out by using a handheld single camera. It was the main basis of our research. In his recent paper [19], Davison presented a revision of his method called MonoSLAM. MonoSLAM is an EKF SLAM system and cannot be used to map large environments. To solve the covariance complexity problem, several strategies have been developed in recent years. We will focus our study in the submapping strategies.

One possible solution to the large-scale problem is the *Metric–Metric* approach, which divides the whole map into smaller parts using a high-metric-level approach over the metric submaps. One of the first methods that applied techniques for map splitting was presented by Tardós *et al.* [13] and more recently in [20], where a conditionally independent divide-and-conquer SLAM is proposed. To extend the MonoSLAM method to larger environments, a hierarchical visual SLAM is presented in [12]. One of the last contributions is the work presented in [14]. A 6-degree-of-freedom (DOF) stereo-in-hand system based on the commercial Bumblebee stereo system is used to capture visual landmarks. An EKF submap strategy is also applied here.

Another alternative to solve the large-scale problem is to use a high topological-level approach over the metric submaps, which leads to the *topological metric* methods. In [21], they present the decoupled stochastic mapping, where a global map is divided into smaller *cells* containing parts of the global map. The hierarchical local map (HLM) method is presented in [22]. It consists of a hierarchical set of submaps locally referenced in this case. The constrained relative submap filter (CRSF) presented in [23] is essentially equal to HLM but introduces improvements on the way coupling estimates are stored. The network-coupled feature map (NCFM) presented in [24] is based on CRSF as well. The *Atlas framework* [25] is also based on the graphs of local frames; however, it lacks cycle optimization. In [26], Frese presented the *TreeMap* algorithm. The idea is to build a hierarchical map consisting of several levels. The measurements are based on landmarks. The approach of Eade and Drummond [27] is based on the NCFM method. It consists of a set of interconnected nodes containing Kalman filter map estimates. A third alternative to face the large-scale SLAM problem is to only use *topological* maps without submaps associated with their vertex. These maps lack the details of the environments, but they can achieve good results for certain applications. In [28], a minimalist visual SLAM for large-scale environments is presented. The approach is based on a graphical representation of robot poses and links between poses based on odometry and omnidirectional image similarity. Another approach is presented in [29], where a topological map that captures and stores images frame by frame and compares them with the previous images is built.

Some of the last contributions to large-scale path estimation using visual sensors have focused on only recovering the estimated vehicle local path using *visual odometry* and adding a topological level for a globally consistent solution. These methods avoid the estimation of external features because they use other strategies for loop closing and global positioning correction. Some examples are [30] and [31].

Related to sensor fusion for navigation tasks, in [32], a multisensor SLAM and navigation system is presented. It is applied to a mobile robot to be able to navigate outdoors. It is based on wheel odometry together with periodic real time kinematic (RTK)-GPS and laser range finder (LRF) measurements. A drawback is the fact that the vehicle must periodically stop to obtain GPS and LRF measurements. This makes it unsuitable to perform automatic SLAM and, therefore, to navigate within unknown environments from the beginning. Another example of sensor fusion is presented in [33]. In this case, they use a pair of Bumblebee stereo cameras, an IMU, and a wheel encoder odometry as relative measurement sensors. On the other hand, a low-cost GPS is also used as an absolute measurement sensor. The system was tested outdoors on an open-spaced nonurban area. Therefore, the GPS accuracy increases, and the availability is almost always guaranteed. The fact that there is no large-scale SLAM management method prevents using the system for much larger environments, even more if we do not have either wheel encoder odometry or IMU available. In [34], a vehicle location estimation method for navigation applications is presented. It is based on a high-performance GPS sensor and an INS inertial unit, as well as the odometry information from the vehicle, fused using an EKF. High-accuracy results are shown. As a drawback, the high cost of the system can be highlighted.

In [35], an onboard vehicle pose estimation exclusively based on a stereo camera is presented. The approach is focused on estimating the pose relative to the environment dominant surface area. In this paper, however, no absolute position coordinates are estimated; therefore, there is no path estimation performed. This is the main objective of this paper.

To choose one of the three main alternative approaches regarding map management, we take into account that, in one hand, although Metric–Metric methods provide accurate estimations, they do not keep a topological structure that helps global optimization in large-scale environments and path-planning techniques for navigation purposes. Topological approaches do not provide accurate information of vehicle state estimations. Therefore, our proposal to solve the large-scale problem is based on the hierarchical topological-metric approach. The behavior of our metric level is similar to visual odometry because a local map is built, but it is a sparse map of landmarks optimized toward enabling localization. The main contributions of our method compared with more relevant proposals presented in this section can be summarized in a more robust data-association strategy for large loop closing based on *scale invariant feature transform* (SIFT) fingerprints, a simpler node-relation management that is well suited for large outdoor urban environments, and the fusion of a cheap stereovision and low-cost GPS sensors to build a precise and real-time vehicle global-localization system.

This paper is organized as follows: The general structure of the system is described in Section III. Section IV presents the low-level SLAM implementation focused on the visual system, Section V studies the high-level SLAM, and Section VI describes the fusion process with the GPS data. In Section VII, a large set of results is given to test the behavior of our system. Section VIII contains our conclusions and future work. This paper relies on previous papers presented by the authors at two conferences [36], [37]. The first paper is focused on the low-level SLAM development using stereovision. The second paper introduces a preliminary version of the high-level SLAM, without the use of the MLR algorithm, for indoor applications.

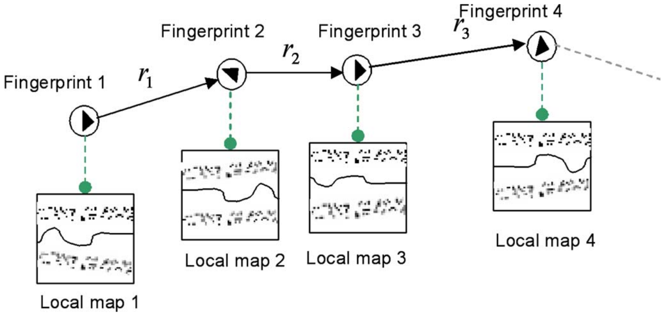

Fig. 1. General architecture of our two-hierarchical-level SLAM. Each submap has an associated fingerprint.

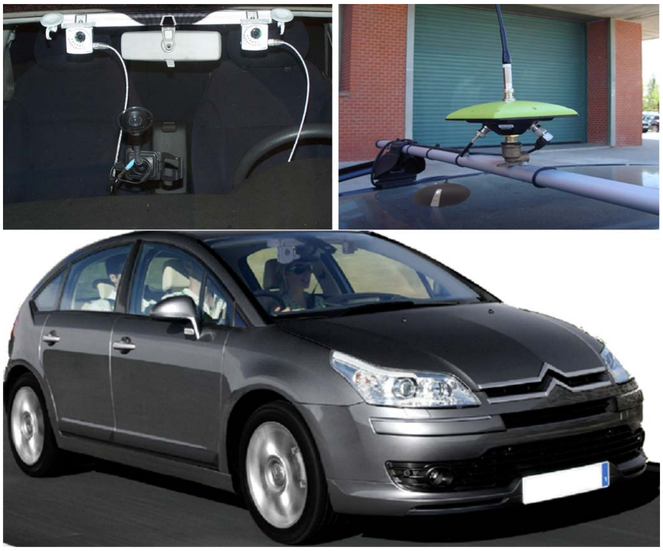

Fig. 2. (Bottom) System architecture mounted on a commercial car. (Top left) Stereo-vision system and low-cost GPS. (Top right) Ground truth RTK-GPS.

III. Implementation

In this paper, we present a real-time SLAM method for large-scale outdoor environments based on the fusion of stereovision and GPS. To deal with the covariance-matrix-growing problem intrinsic to the visual SLAM, we divide the global map into local submaps. Each of these submaps has its own metric SLAM process, independent of the other submaps. Over these local submaps, we define a higher topologic SLAM level that relates them to keeping the global map consistency (see Fig. 1). On this task, GPS provides a valuable contribution, because the positioning error increases over time in visual SLAM systems unless loop-closing situations take place. On the other hand, the GPS (when available) errors on estimations are limited but, at the same time, can locally grow much more quickly than visual estimation does. Therefore, at the end, both sensors are complementary.

The visual system is based on a stereo wide-angle camera mounted on a vehicle in the windshield area and looking forward of the vehicle (see Fig. 2). For each local submap, several visual landmarks are sequentially captured using the Shi and Tomasi operator [36] and introduced on an EKF filter to model the probabilistic behavior of the system. A measurement model is used for landmark perception, and a motion model is implemented for the dynamic behavior of the vehicle. GPS measurements contribute to improve both the vehicle and the map estimation.

We present a hierarchical SLAM implementation that adds an additional processing level called "high-level SLAM" to the explained metric SLAM that we will call "low-level SLAM." The whole map is divided into local submaps identified by *fingerprints*. These fingerprints store the vehicle pose at the moment of submap creation and define its local reference frame. The submap generation is periodically performed in space so that after a certain covered section of the path, a new submap is created, and a fingerprint is associated with it. When the vehicle is traveling between two fingerprints, an edge is inserted to connect these two vertices, which represents a link between two poses. Meanwhile, the edges store transformation matrices and uncertainties to describe the relationship between the connected fingerprints. To optimize the loop-closing detection, when a significant vehicle turn is detected, an additional fingerprint called SIFT *fingerprint* is taken. This adds to the vehicle pose some visual information to identify the place where it was taken. Matching between the previously captured SIFT fingerprints, within an uncertainty area, and the current fingerprints is carried out to detect previsited zones. In the case of positive matching, a loop closing is detected, and the topological map is corrected by using the MLR algorithm [17] over the whole set of fingerprints.

The MLR determines the maximum-likelihood estimate of all fingerprint poses. After that, the landmarks of each submap are corrected as a function of the correction applied to its associated fingerprint.

Each time a new GPS measurement is available, it is introduced into the system. This is carried out by fusing visual and GPS 2-D position coordinates and taking into account the uncertainty covariances from both of them. Orientation is obtained through an interpolation of the two last updates. The confidence level of the measurement is taken into account by estimating its uncertainty area, which is obtained by fusing both visual and GPS estimation uncertainties as well. This GPS uncertainty is obtained using the information provided by the GPS (satellite visibility, geometrical distribution, etc.) and other error sources assumed constant along time.

IV. Low-Level SLAM

This level implements all the algorithms and tasks needed to locate and map the vehicle on its local submap using the visual information. It is based on the monocular approach by Davison [18] and its adaptation to stereo developed by the authors [36]. The GPS sensor contribution will be explained in Section VI. For clarity reasons, the submap notation is omitted, so a unique submap for the low-level SLAM implementation is assumed.

A. EKF Application

To apply an EKF, a state vector X and its covariance matrix P need to be defined. The purpose of the algorithm is to continuously estimate the position and orientation of the vehicle via the linearization of the *next state function* $f(X)$ at each time step. The vehicle coordinate system has been set in camera frame one. Due to the *motion model* used for vehicle movement, which will be explained later, linear and angular speeds are added to the vehicle state vector $X_v = (\, X_{vh} \quad q_{vh} \quad v_{vh} \quad \omega \,)^T$. In this equation, $X_{vh} = (\, x_{vh} \quad y_{vh} \quad z_{vh} \,)^T$ is the 3-D position of the camera relative to the global frame, $q_{vh} = (\, q_0 \quad q_x \quad q_y \quad q_z \,)^T$ is the orientation quaternion, v_{vh} is the linear speed, and ω is the angular speed. On the other hand, as the whole submap has to be included into the filter, all feature global positions Y_i are added to the state vector $X = (\, X_v \quad Y_1 \quad Y_2 \quad \cdots)^T$.

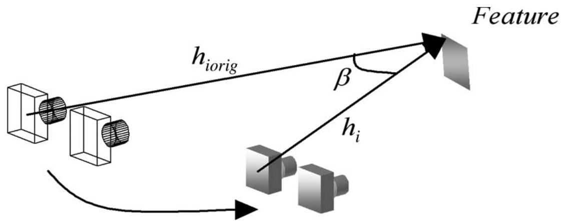

Fig. 3. Original and current feature measurement vectors.

B. Motion Model

To construct a motion model for a camera mounted on a mobile vehicle only using visual information, a practical solution is to apply the so-called *impulse model.* This assumes a constant speed (both linear and angular) during each time step and random speed changes between steps in three directions. Some restrictions have been applied to adapt the 6DOF generic model to the vehicle's movement dynamics. According to this model, to predict the next state of the camera the function, $f_v = (\, X_{vh} + v_{vh} \cdot \Delta t \quad q_{vh} \times q[\omega \cdot \Delta t] \quad v_{vh} \quad \omega \,)^T$ is applied. The term $q[\omega \cdot \Delta t]$ represents the transformation of a three-component vector into a *quaternion*. Assuming that the map does not change during the whole process, the absolute feature positions Y_i should be the same from one step to the next step. This model is subtly effective and gives the whole system important robustness even when visual measurements are sparse.

C. Measurement Model

Visual measurements are obtained from the "visible" feature positions. In our system, we define each individual *measurement prediction* vector $h_i = (\, h_{ix} \quad h_{iy} \quad h_{iz} \,)^T$ as the corresponding 3-D feature position relative to the camera frame. To choose the features to measure, some selection criteria have to be defined. These criteria will be based on the feature "visibility," that is, whether its appearance is close enough to the original (when the feature was initialized). This is based on the relative distance and point of view angle with respect to that at the feature initialization phase (see Fig. 3), as explained in [36].

The first step is to predict the measurement vector h_i. To look for the actual measurement vector z_i, we have to define a search area on the projection images. This area will be around the projection points of the predicted measurement h_i on both *left* and *right* images: $U_L : (u_L, v_L)$, $U_R : (u_R, v_R)$. To obtain the image projection coordinates, first, we apply the simple "pin-hole" model, and then, it is distorted using the radial and tangential distortion models, which are detailed in [36]. To obtain z_i, we need to solve the inverse geometry problem, applying the distortion models as well (see [36]).

Regarding the search areas, they will be calculated based on the uncertainty of the feature 3-D position, which is the called *innovation covariance* S_i (see [38]). As we have two different image projections, S_i needs to be transformed into the projection covariance P_{U_L} and P_{U_R} using

$$P_{U_L} = \frac{\partial U_L}{\partial h_i} \cdot S_i \cdot \left(\frac{\partial U_L}{\partial h_i}\right)^T$$
$$P_{U_R} = \frac{\partial U_R}{\partial h_i} \cdot S_i \cdot \left(\frac{\partial U_R}{\partial h_i}\right)^T. \quad (1)$$

These two covariances define both elliptical search regions, which are obtained by taking a certain number of standard deviations (usually three) from the 3-D Gaussians. Once the areas where the current projected feature should lie are defined, we can look for them. At the initialization phase, the left and right images representing the feature *patches* are stored. Then, to look for a feature patch, we perform normalized *sum-of-squared-difference correlations* across the whole search region (see [38]). The path appearance is modified depending on the vehicle point of view using the *Patch Adaptation* method described in [36]. This helps on the search correlation phase in the sense of extending the tracking of the patch.

In our application, the camera provides a baseline of $T_{\text{int}} = 400$ mm. We do not make any explicit differentiation between near and far landmarks, as done in [14]. However, our method implicitly does that. Far landmarks provide more useful information when the vehicle turns, and near landmarks provide more useful information when the vehicle goes straight ahead. The reason is due to the innovation covariance S_i, which at the end provides the weight of each landmark within the filter. In straight movements, distant landmarks appear to be almost static, i.e., their innovation from frame to frame is relatively low. However, on the vehicle turns, the innovation on distant landmarks is higher, increasing their weights on that situation. As long as landmarks become more distant, their location errors increase. Nevertheless, the distance information from far landmarks is almost useless. Therefore, to handle very far landmarks, we limit the maximum-distance estimation to a fixed value so that only the angle information is relevant.

D. Feature Initialization

The selected criteria to initialize new landmarks are to always maintain at least five visible features and four successfully measured features.

Then, when a new feature initialization needs to take place, its corresponding patch will be searched within a rectangular area randomly located on the left camera image. To obtain the right image feature correspondence, we search over the *epipolar line*, which is restricted to a certain segment around the estimated right projection coordinates (see [36]).

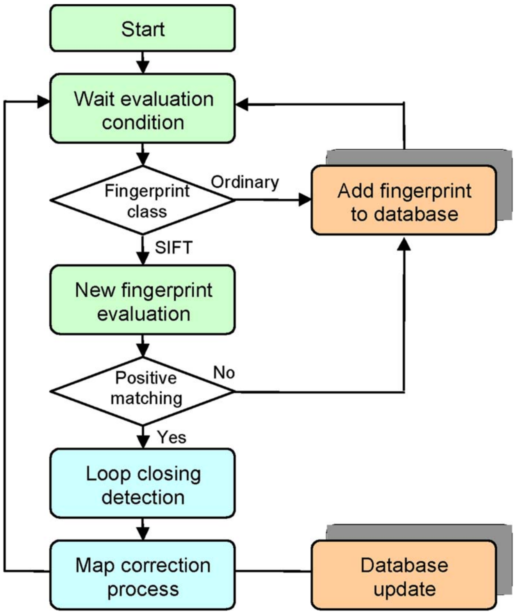

Fig. 4. High-level map management.

V. High-Level SLAM

Our SLAM implementation adds an additional topological level called high-level SLAM to keep global map consistency with almost constant processing time. This goal is achieved by using the MLR algorithm over the so-called *Fingerprints*. Therefore, the global map is divided into local submaps identified by the mentioned fingerprints. There are two different classes of fingerprints: 1) *ordinary fingerprints* and 2) *SIFT fingerprints*.

The first classes are denoted as $FP = \{fp_l | l \in 0, \ldots, L\}$. Their purpose is to store the vehicle local pose $X_{vh}^{fp_l}$ and local covariance $P_{vh}^{fp_l}$ relative to the previous fingerprint, i.e., the reference frame of the current submap. These fingerprints are periodically obtained approximately each 10 m of the covered path, synchronized with the GPS measurements.

The second classes are a subset of the first classes, which are denoted as $SF = \{sf_q \in FP | q \in 0, \ldots, Q, Q < L\}$. Their additional functionality is to store the visual appearance of the environment at the moment of being obtained. That is covered by the definition of a set of *SIFT features* associated to the fingerprint, which identifies the place at that time $YF^q = \{Yf_m^q | m \in 0, \ldots, M\}$. These fingerprints are only taken under the condition of having a significant change on vehicle trajectory (see Fig. 4). Each time a new SIFT fingerprint is taken, it is matched with the previously acquired SIFT fingerprints within an uncertainty search region. This region is obtained from the vehicle global covariance P_{vh}^G because it keeps the global uncertainty information of the vehicle. If the matching is positive, then it means that the vehicle is in a previously visited place, and a *loop closing* is identified. Then, the MLR algorithm is launched to determine the maximum-likelihood estimate of all fingerprint poses. Finally,

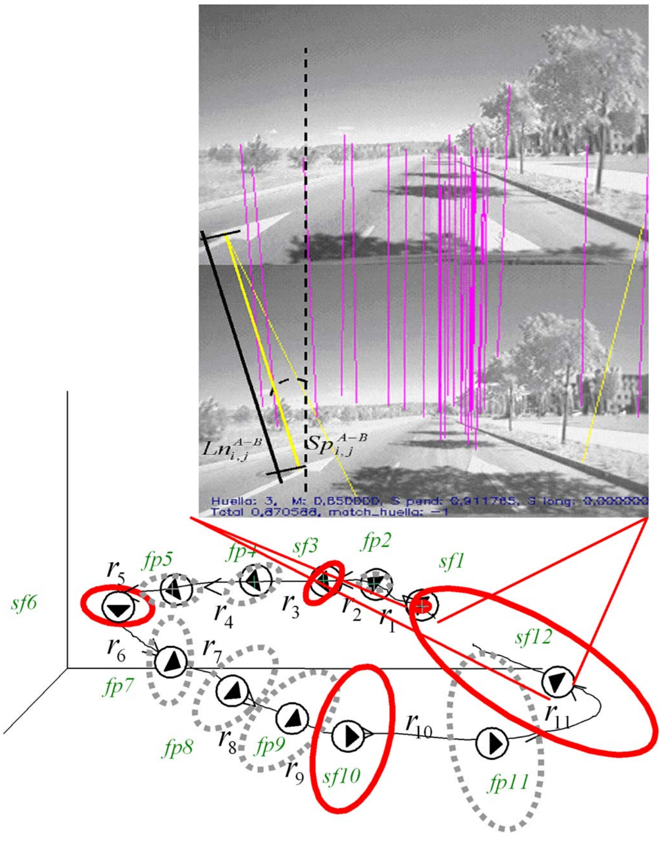

Fig. 5. Representation of the fingerprint global uncertainties $Pc^0_{2fp_l}$ increasing along the vehicle path. Solid line ellipses represent SIFT fingerprints uncertainties. Numbers represent each fingerprint. Fingerprint SIFT feature matching is shown on the corner image. Outliers are marked as light color lines.

the fingerprint corrections are transmitted to their associated submaps.

A. Local Submaps

Each time a new fingerprint is taken, an associated submap is created. The vehicle relative local pose $X_{vh}^{fp_l}$ and its covariance $P_{vh}^{fp_l}$ are stored in the fingerprint at that moment. Due to the need of being aware about the global uncertainty at any time, we need to maintain P_{vh}^{G} updated (see Fig. 5). We calculate it by expressing the current local uncertainty $P_{vh}^{fp_l}$ in the global reference frame as

$$P_{vh}^{G} = \frac{\partial X_{vh}^{0}}{\partial X_{vh}^{fp_l}} \cdot P_{vh}^{fp_l} \cdot \left(\frac{\partial X_{vh}^{0}}{\partial X_{vh}^{fp_l}}\right)^{T} \tag{2}$$

where X_{vh}^{0} is the current global vehicle position.

Two different high-level approaches can be done either using metric submaps [12]–[14] or topologic submaps [15], [16]. Both of them use detailed local maps, but the submaps employed in the metric approach do not maintain a topological structure of an environment. On the other hand, in the metric approach, a global map is built by joining all the local maps. Then, a process is carried out to identify all the duplicated landmarks, closing all possible loops inside. Instead, we continuously keep the accumulated global uncertainty, allowing the global map correction at any time as soon as a closed-loop situation is detected.

B. SIFT Fingerprints

Our system identifies a specific place using the SIFT fingerprints. These fingerprints, apart from the vehicle pose, consist a number of SIFT landmarks distributed across the reference image and characterize the visual appearance of the image. The SIFT features introduced by Lowe and Little [39] are invariant to image scaling and rotation and partially invariant to change in illumination and 3-D camera viewpoint. In addition, the features are highly distinctive, which allows a single feature to correctly be matched with high probability. This is achieved by the association of a 128-length descriptor to each of the features, which will uniquely identify all of them. These SIFT feature descriptors $\vec{\delta}$ are loaded in each SIFT feature joint to the left image coordinates and the 3-D vehicle position $Yf_m^q = (\,u_L \quad v_L \quad X \quad Y \quad Z \quad \vec{\delta}\,)$ for the fingerprint matching process.

C. Loop Closing Detection

One of the main issues on SLAM in large environments is the *loop-closing* problem. The first issue to solve is the recognition of previously visited places. Once a new SIFT fingerprint is generated, it is matched with all stored SIFT fingerprints within the uncertainty area defined by P_{vh}^{G}. This matching is carried out for each pair of SIFT fingerprints (sf_A, sf_B), taking into account both the number of recognized SIFT features and their relative positions within the images to be compared. The overall process is as follows:

1) The Euclidean distance between the descriptors $\vec{\delta}_i^A$, $\vec{\delta}_j^B$ of all detected SIFT features on both fingerprints (sf_A, sf_B), which is shown as

$$\left\{\left\|\vec{\delta}_1^A - \vec{\delta}_1^B\right\|, \ldots, \left\|\vec{\delta}_1^A - \vec{\delta}_{mB}^B\right\|,\right.$$
$$\left.\left\|\vec{\delta}_2^A - \vec{\delta}_1^B\right\|, \ldots, \left\|\vec{\delta}_{mA}^A - \vec{\delta}_{mB}^B\right\|\right\} \tag{3}$$

are computed. Then, we select those close enough as correctly matched. The trigger value is empirically selected.
2) Lines connecting each pair of matched features are calculated. The corresponding lengths $Ln_{i,j}^{A-B}$ and slopes $Sp_{i,j}^{A-B}$ are computed as well, as depicted in Fig. 5.
3) Outlier features are excluded from the computation by using the *RANSAC* method. The model to fit is defined as the vector $(\text{avg}(Ln_{i,j}^{A-B}), \text{avg}(Sp_{i,j}^{A-B}))$ containing the average lengths and slopes of the connecting lines. RANSAC is applied to the whole set of lines, calculating the Euclidean distance of all the individual length/slope pairs to the average. Features whose connecting line pairs are close enough to the model are considered as inliers; otherwise, they are declared as outliers.
4) The global *fingerprint matching probability* is computed as a weighted function of two parameters: 1) *number*

of matched feature probability $P(num\_matches) = num\_matches/m_3$ and 2) *inliers/num_matches relation*, where (m_1, m_2, m_3) were experimentally obtained as

$$P_{fp\_match} = m_1 \cdot P(num\_matches) + m_2(n_I/num\_matches). \quad (4)$$

Obviously, $P(num\_matches)$ can eventually be higher than 1 so that we limited the function to avoid this situation. The typical values for our experiments are $m_1 = 2/3$, $m_2 = 1/3$, and $m_3 = 40$.

D. Map Correction

Once a loop closing has been detected, the whole map is corrected according to the old place recognized. To do that, we use the MLR algorithm [17]. The purpose of this algorithm is to assign a globally consistent set of Cartesian coordinates to the fingerprints of the graph based on local inconsistent measurements by trying to maximize the total likelihood of all measurements.

The reasons for using it are its highly efficient implementation in terms of computational cost and the extremely high complexity allowed for the relations between new and previously visited places. This provides the ability of closing multiple loops even in a hierarchical way. On the other hand, as we will explain later, we can correct the map, not only when closing loops, but also when GPS has been unavailable for a long time and it recovers again.

The MLR inputs are the relative poses and the covariances of the fingerprints. As outputs, the MLR returns the most "likely" set of fingerprint poses, i.e., the set already corrected. Because the standard MLR does not provide corrected covariances, we have modified the MLR to calculate them.

The MLR algorithm only manages 2-D information; therefore, we need to obtain the 2-D relative fingerprint pose $X_{2fp_l}^{fp_{l-1}}$ and covariance $P_{2fp_l}^{fp_{l-1}}$ from the corresponding 3-D relative fingerprint pose $X_{fp_l}^{fp_{l-1}}$ and covariance $P_{fp_l}^{fp_{l-1}}$. First, the 2-D pose is defined as $X_{2fp_l}^{fp_{l-1}} = (x_{2fp_l}^{fp_{l-1}} \;\; y_{2fp_l}^{fp_{l-1}} \;\; \theta_{2fp_l}^{fp_{l-1}})^T$, i.e., the two planar coordinates and the orientation angle. Therefore, we can relate both 2-D and 3-D poses as

$$X_{2fp_l}^{fp_{l-1}} = \left(x_{fp_l}^{fp_{l-1}} \quad z_{fp_l}^{fp_{l-1}} \quad 2\arccos\left(q_{0_{fp_l}}^{fp_{l-1}}\right)\right)^T \quad (5)$$

where $x_{fp_l}^{fp_{l-1}}$, $z_{fp_l}^{fp_{l-1}}$, and $q_{0_{fp_l}}^{fp_{l-1}}$ are coordinates of $X_{fp_l}^{fp_{l-1}}$. In addition, we compute the 2-D covariance by using the corresponding Jacobians.

The MLR algorithm, as explained in [17], is based on the quadratic error function computation of the fingerprints, and we then try to minimize it. The expression $X_M = (Xc_{fp_1}^0 \quad Xc_{fp_2}^0 \quad \cdots \quad Xc_{fp_L}^0)^T$ represents the total vector of the whole set of 2-D-corrected fingerprint poses, which are denoted as states in [17].

Once the 2-D-corrected vector has been calculated, we obtain the corresponding 3-D-corrected fingerprints. At the step of obtaining 2-D from 3-D poses [see (5)], we lost the $y_{fp_l}^{fp_{l-1}}$ coordinate information (altitude). Therefore, when going back from 2-D to 3-D again, we have to set this value. We assume a flat terrain because our system is mounted on a commercial car driving in a flat urban area; therefore, this value will be taken as 0. Then, we form the corrected absolute pose vector for each fingerprint as

$$X_{fp_l}^0 = \left(xc_{fp_l}^0 \quad 0 \quad yc_{fp_l}^0 \quad \cos\left(\theta c_{fp_l}^0/2\right) \quad 0 \quad \sin\left(\theta c_{fp_l}^0/2\right) \quad 0\right)^T \quad (6)$$

where $xc_{fp_l}^0$, $yc_{fp_l}^0$, and $\theta c_{fp_l}^0$ are the coordinates of the corrected absolute fingerprint poses. As explained before, the standard MLR method does not provide means to obtain the corrected global covariances of the fingerprints. The reason is because this method is uniquely based on the relative covariances between poses. However, our system needs to obtain them to keep the global uncertainty of the vehicle updated. Then, the first step is to express the initially estimated relative covariances in the global frame. This is done, as in (5), by means of the corresponding Jacobians. Taking the expression of the measurement function shown in [17], we can find the relative pose as

$$X_{2fp_i}^{fp_{i-1}} = \begin{pmatrix} x_{2fp_i}^0 \cos\theta_{2fp_{i-1}}^{fp_{i-2}} - y_{2fp_i}^0 \sin\theta_{2fp_{i-1}}^{fp_{i-2}} + x_{2fp_{i-1}}^{fp_{i-2}} \\ x_{2fp_i}^0 \sin\theta_{2fp_{i-1}}^{fp_{i-2}} + y_{2fp_i}^0 \cos\theta_{2fp_{i-1}}^{fp_{i-2}} + y_{2fp_{i-1}}^{fp_{i-2}} \\ \theta_{2fp_i}^0 + \theta_{2fp_{i-1}}^{fp_{i-2}} \end{pmatrix} \quad (7)$$

where $x_{2fp_i}^0$, $y_{2fp_i}^0$, and $\theta_{2fp_i}^0$ are the 2-D global fingerprint coordinates, whereas $x_{2fp_{i-1}}^{fp_{i-2}}$, $y_{2fp_{i-1}}^{fp_{i-2}}$, and $\theta_{2fp_{i-1}}^{fp_{i-2}}$ are the 2-D relative previous fingerprint coordinates. The Jacobian $\partial X_{2fp_l}^0/\partial X_{2fp_l}^{fp_{l-1}}$ is easily calculated from (7). The next step is to calculate the corrected covariance as a function of the uncorrected covariance through the corresponding Jacobians. As we do not have any equation to calculate the corrected 2-D absolute fingerprint coordinates $Xc_{2fp_l}^0$ as a function of the uncorrected coordinates $X_{2fp_l}^0$, we cannot obtain the Jacobian earlier shown. We already know the whole set of both estimated and corrected fingerprints. Assuming a dense grid of them, we can approximate the Jacobian as a discrete differentiation

$$\frac{\partial Xc_{2fp_i}^0}{\partial X_{2fp_i}^0} = \begin{pmatrix} \frac{\delta xc_{fp_i}^0}{\delta x_{fp_i}^0} & \frac{\delta xc_{fp_i}^0}{\delta y_{fp_i}^0} & \frac{\delta xc_{fp_i}^0}{\delta \theta_{fp_i}^0} \\ \frac{\delta yc_{fp_i}^0}{\delta x_{fp_i}^0} & \frac{\delta yc_{fp_i}^0}{\delta y_{fp_i}^0} & \frac{\delta yc_{fp_i}^0}{\delta \theta_{fp_i}^0} \\ \frac{\delta \theta c_{fp_i}^0}{\delta x_{fp_i}^0} & \frac{\delta \theta c_{fp_i}^0}{\delta y_{fp_i}^0} & \frac{\delta \theta c_{fp_i}^0}{\delta \theta_{fp_i}^0} \end{pmatrix}$$

$$\cong \begin{pmatrix} \frac{xc_{fp_i}^0 - xc_{fp_{i-1}}^0}{x_{fp_i}^0 - x_{fp_{i-1}}^0} & \frac{xc_{fp_i}^0 - xc_{fp_{i-1}}^0}{y_{fp_i}^0 - y_{fp_{i-1}}^0} & \frac{xc_{fp_i}^0 - xc_{fp_{i-1}}^0}{\theta_{fp_i}^0 - \theta_{fp_{i-1}}^0} \\ \frac{yc_{fp_i}^0 - yc_{fp_{i-1}}^0}{x_{fp_i}^0 - x_{fp_{i-1}}^0} & \frac{yc_{fp_i}^0 - yc_{fp_{i-1}}^0}{y_{fp_i}^0 - y_{fp_{i-1}}^0} & \frac{yc_{fp_i}^0 - yc_{fp_{i-1}}^0}{\theta_{fp_i}^0 - \theta_{fp_{i-1}}^0} \\ \frac{\theta c_{fp_i}^0 - \theta c_{fp_{i-1}}^0}{x_{fp_i}^0 - x_{fp_{i-1}}^0} & \frac{\theta c_{fp_i}^0 - \theta c_{fp_{i-1}}^0}{y_{fp_i}^0 - y_{fp_{i-1}}^0} & \frac{\theta c_{fp_i}^0 - \theta c_{fp_{i-1}}^0}{\theta_{fp_i}^0 - \theta_{fp_{i-1}}^0} \end{pmatrix}.$$

Finally, the global corrected covariance, which is expressed in 3-D coordinates, is again obtained by means of the corresponding Jacobians.

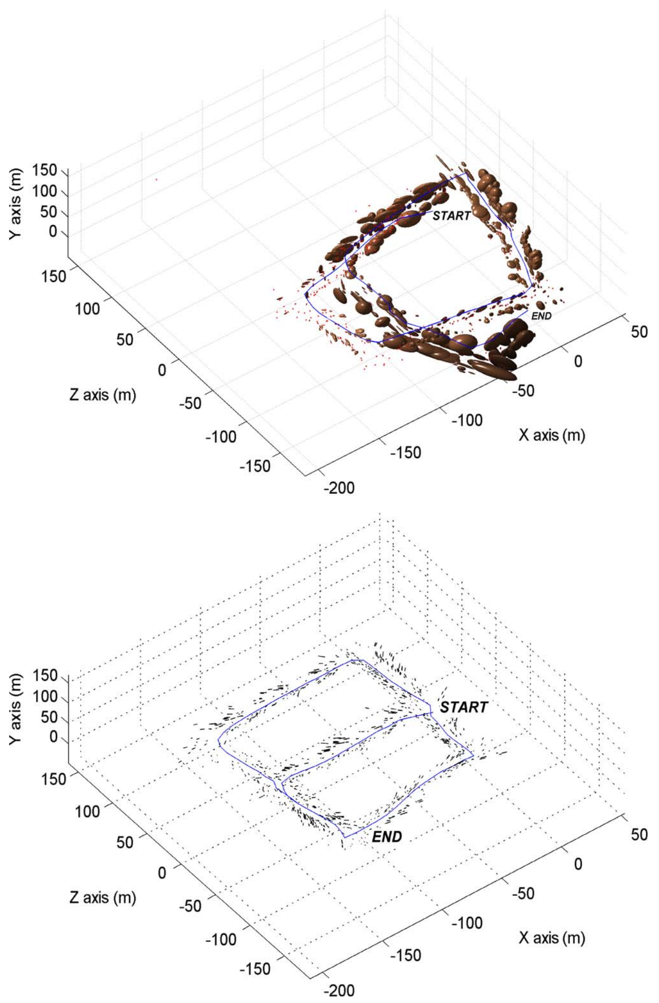

Fig. 6. Detail on the first test path showing landmark global uncertainties (top) as a result of only the low-level SLAM estimation and (bottom) after applying high-level SLAM optimization.

After the topological map has been corrected, the associated global uncertainties of all the fingerprints are expected to be reduced. This increases the fingerprint search process efficiency because the number of SIFT fingerprints matched will be lower.

The last step is to transfer the correction performed on the high SLAM level into the low SLAM level. This is done by applying the same transformation of the fingerprint pose to all the landmarks within the submap. By doing this, we keep the relative positions of the landmarks unchanged with respect to their corresponding local submap reference frame. Therefore, the landmark covariances remain unchanged in the frame of each submap. However, to represent their global uncertainties, we show in Fig. 6 a portion of one of the paths used for testing purposes. We represent the global feature covariances using just the EKF on the local maps (see the top of Fig. 6) and after applying the MLR optimization (see the bottom of Fig. 6).

VI. GPS Sensor Fusion

In this section, we explain the way we introduce the information provided by the low-cost GPS into the system. This is carried out on two hierarchical levels: first, at the low-level SLAM by updating the local state and covariance estimations, and, second, by updating the fingerprints and correcting the global map in case of long-term GPS unavailability. Due to the 2-D implementation of the MLR algorithm, we lose the vertical estimation of the vehicle path. Because of that, it does not make sense to use the altitude information provided by the GPS.

A. GPS Uncertainty Estimation

To perform sensor fusion, we need to quantify the confidence level on the GPS measurements. The error sources on GPS data are multiple (receiver noise, satellite clock, ionosferic model, etc.), and most of them are difficult or impossible to quantify. To summarize all of these errors, the statistical user equivalent range error (UERE) is defined in [40]. We assume that this UERE is 4 m, which seems reasonable if we look at the different studies carried out [41]. On the other hand, the final GPS uncertainty estimation will also depend on the number of visible satellites and their spatial distribution. This is quantified by the so-called dilution of probability (DOP), which is provided in real time by the GPS. This value is used as a ratio of the positioning accuracy σ_R, which is defined as the UERE at the two-sigma level (95% UERE). As we only pay attention to the horizontal error, we make use of the horizontal DOP (HDOP). To obtain the x and y standard deviations, we make use of the following expression defined in [40]:

$$\mathrm{HDOP} = \sqrt{\sigma_x^2 + \sigma_y^2}/(95\%\ \mathrm{UERE}). \tag{8}$$

As we are not able to know each of the individual σ_x and σ_y, we assume that both of them are equal, and therefore, the final uncertainty region defined by

$$P_{\mathrm{GPS}} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \tag{9}$$

will have a planar circular shape.

B. Low-Level Data Fusion

Each time a new GPS reading $X_{\mathrm{GPS}} = (\,x_{\mathrm{GPS}} \quad y_{\mathrm{GPS}}\,)^T$ is available, which under normal conditions occur at 1-s period, we proceed to fuse it with our visual estimation. As GPS does not provide orientation information, initially, we only calculate the position, and then, the orientation is estimated, as explained later. To fuse the two positions, first, we express the initial visual estimation in a two-component vector $X_{Pvh} = (\,x_{vh} \quad z_{vh}\,)^T$. Then, to calculate the final position estimation, we merge both estimates by making use of their respective 2-D uncertainty covariances, as we depict in (10). This is obtained by applying a 2-D statistical approach based on Bayes' rule and Kalman filters.

The resultant estimation improves the uncertainty distribution because it is calculated as the product of the two original estimations. In [42], an improved way, in terms of computing time, to fuse these data by using the corresponding covariance matrices is presented as

$$X^{\mathrm{fusion}} = X_{Pvh} + P_{Pvh}^{G}\left(P_{Pvh}^{G} + P_{\mathrm{GPS}}\right)^{-1}\left(X_{\mathrm{GPS}} - X_{Pvh}\right) \tag{10}$$

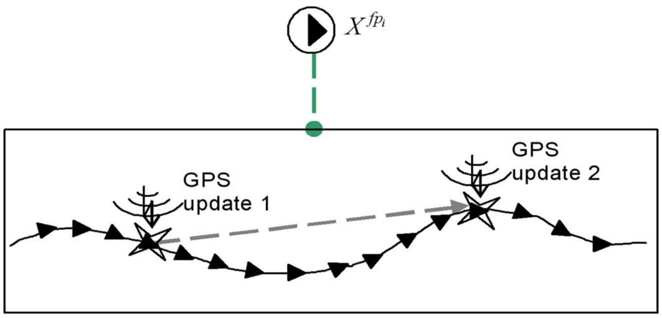

Fig. 7. Submap detail showing the vehicle orientation estimation from GPS data fusion. The local map estimation is represented through local pose estimations (black arrows). The grey arrow indicates the estimated orientation of the vehicle at the second GPS update point.

where P_{Pvh}^{G} is the adapted 2-D position estimation covariance

$$P_{Pvh}^{G} = \begin{pmatrix} P_{xx}^{G} & P_{xz}^{G} \\ P_{zx}^{G} & P_{zz}^{G} \end{pmatrix}. \qquad (11)$$

In the same way, the following estimated covariance is calculated by means of the following equation:

$$P^{\text{fusion}} = P_{Pvh}^{G} - P_{Pvh}^{G} \left(P_{Pvh}^{G} + P_{\text{GPS}}\right)^{-1} P_{Pvh}^{G}. \qquad (12)$$

Then, we update the global state and covariance as follows:

$$P_{vh}^{G} = \begin{pmatrix} P_{xx}^{\text{fusion}} & 0 & P_{xy}^{\text{fusion}} & \cdots \\ 0 & 0 & 0 & \cdots \\ P_{yx}^{\text{fusion}} & 0 & P_{yy}^{\text{fusion}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix};$$
$$X = \left(X_{x}^{\text{fusion}} \quad 0 \quad X_{y}^{\text{fusion}} \quad \cdots \right)^{T}. \qquad (13)$$

To estimate the orientation, each time a new GPS measurement is taken, we obtain the relative position of the current GPS update related to the previous update. Then, we calculate the absolute angle of the vector that joints the two positions and obtain the corresponding estimated quaternion q_{vh} (see Fig. 7). We also update the linear speed v_{vh} according to the new estimated orientation.

To obtain the best estimation for the MLR fingerprints, we generate them in a synchronized way with the GPS updates. Therefore, when conditions for a new fingerprint generation are ready, we wait until a new GPS update is available.

C. High-Level Data Fusion

One of the most common problems when using GPS in very populated urban areas is the complete unavailability. Usually, this is because of a low number of visible satellites, which is caused by different circumstances like tunnels, bridges, or even high buildings. We consider a "long term" GPS loss when we do not have GPS measurements available for more than two consecutive fingerprints. In that case, the state correction implies a global map correction that mainly concerns the section where the GPS signal was unavailable, as can be seen in Fig. 8. The way we introduce new fingerprints into the database when GPS is available is slightly different than when is not. In the first case, the GPS measurement uncertainty is directly referred to the global reference frame; therefore, there is no cumulative error. Therefore, the fingerprints will be defined relative to the global frame. However, when GPS is not available, we only rely on visual information, which, at the end, only provides local relative information. Then, fingerprint poses are introduced relative to the previous one (see Fig. 8). At the time of map correction, these different procedures make that the relative fingerprint poses are modified in a higher degree than absolute poses. This makes sense, because in these fingerprints, the cumulative uncertainty continuously increases over time, unlike absolute uncertainties, which have a limited absolute error. Therefore, when losing the GPS signal for a long time and recovering it again, the new estimated pose will have a lower uncertainty than the accumulated up to this time.

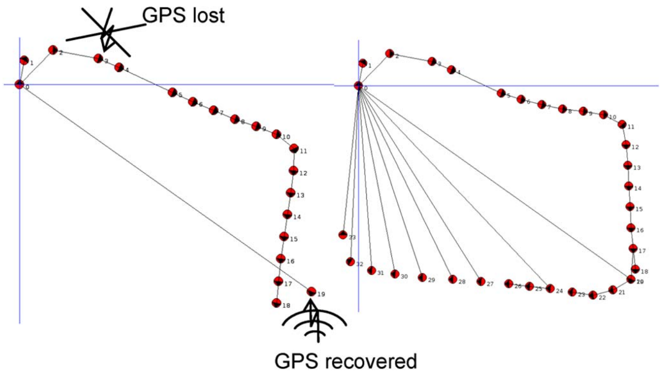

Fig. 8. Fingerprints MLR diagram (left) before and (right) after GPS recovering. Fingerprints with GPS available are expressed in global coordinates (19...33). Fingerprints without GPS are expressed relative to the previous coordinates.

Therefore, we can exploit this fact to reduce the uncertainty of the visual-estimated section of the path. Then, to perform map correction, as soon as the GPS signal is back, we create a new fingerprint relative to the global frame and add a relation between the last fingerprint and the current fingerprint (see Fig. 8). To be able to retrieve the vehicle orientation, because we do not have a previous GPS update, we must wait for two consecutive updates and calculate the new orientation from them.

VII. Results

To test the behavior of our system, several video sequences were collected from a commercial car manually driven in large urban areas. The employed cameras for the stereo pair were the Unibrain Fire-i IEEE1394 with additional wide-angle lens, which provide a field of view of around 100° horizontal and vertical with a resolution of 320 × 240. The baseline of the stereo camera was 40 cm. Both cameras were synchronized at the time the start of the transmission is commanded. The cameras were mounted inside the car on top of the windscreen. The calibration was performed offline using a chessboard panel using the method referenced in [43]. The employed low-cost standard GPS was GlobalSat BU-353 USB. The metric coordinates were obtained from the geographic coordinates taking into account the WGS84 ellipsoid. To refer the global positions provided by the GPS to our local reference frame, we subtract the first measurement at the origin to the rest of the measurements.

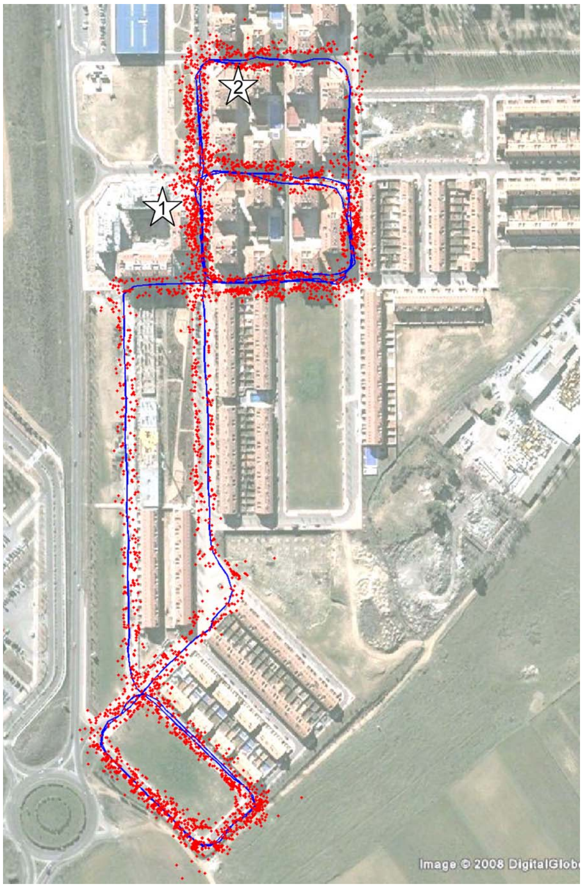

Fig. 9. Aerial view of the path covered by the vehicle. The starting point is marked with number 1, and the endpoint is marked with number 2. Landmarks are showed as dots. The loop shown on the lower part of the picture contains several buildings that still do not appear at that time.

One of the video sequences was taken by covering with a car the urban path shown in Fig. 9. The average speed of the car was around 30 km/h. The complete covered path was 3.17 km long. It contained five loops inside, taking 8520 low-level landmarks and 281 fingerprints. In Fig. 9, we show the estimation of the path covered by the vehicle. It can be appreciated that the areas where high buildings are located contain a higher number of landmarks (marked as dots). This helps on a more precise location provided by the vision system. On the other hand, open-spaced areas without high buildings do not provide accurate visual information, whereas the GPS signal has more strength and provides better location estimation. This shows that both sensors complement each other, providing good estimations in different situations. Therefore, combining them in a proper way, we can obtain better estimations. A perspective view of the same estimation is shown in Fig. 10, where we appreciate landmarks distributed on the whole volume. To evaluate the performance of our system, we compared our results with a ground truth reference.

This ground truth was obtained with an RTK-GPS Maxor GGDT, which provides an estimated accuracy of 2 cm. On the other hand, we collected together car positions obtained by only using the low-cost GPS to analyze them and compare with our system. Fig. 11 depicts the estimation of our combined SLAM system, the standard GPS alone, and the visual SLAM only compared with the ground truth.

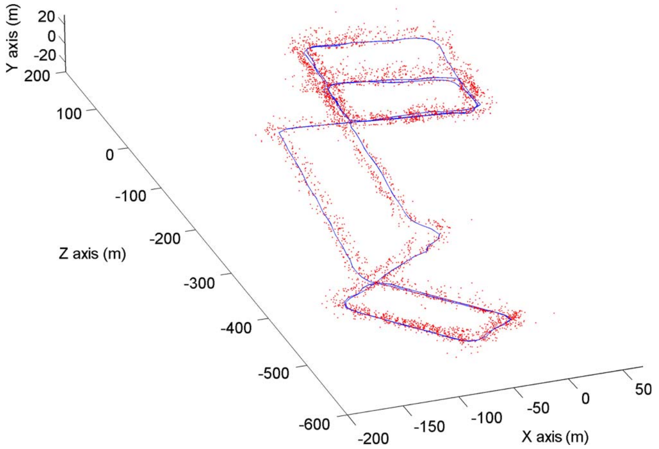

Fig. 10. (Solid line) Perspective view of the path covered by the vehicle. Low-level landmarks are shown as dots.

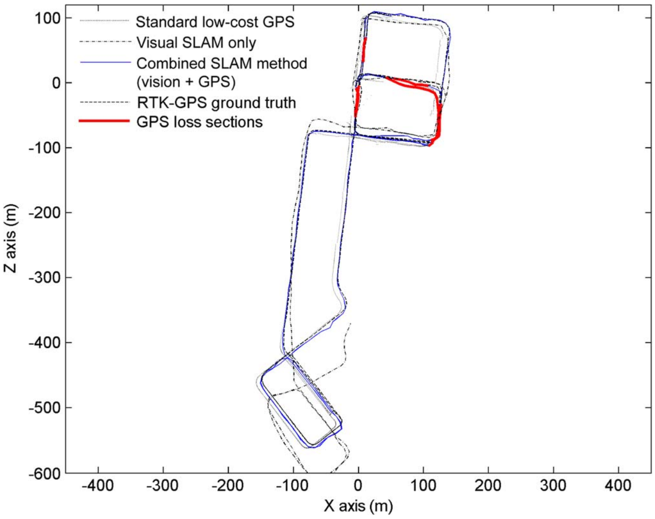

Fig. 11. Path estimation using (dotted line) only a standard low-cost GPS, (dash-dot line) visual SLAM only, (solid line) our combined SLAM method by means of vision and GPS, and (dashed line) the ground truth. GPS loss sections are marked with thick lines.

The GPS signal was lost at different moments at the beginning of the path, as shown in Fig. 11. On the longest signal neglect section, the increased estimation error can easily be observed; however, we still have a relatively accurate estimation to be able to locate the vehicle. As can be seen, using visual SLAM only, the error becomes larger at the end of the long straight segment. This is as a result of the low amount of landmarks captured on this area. Then, we have calculated the X- and Y-axis errors relative to the ground truth using the visual SLAM only and our combined SLAM implementation (see Fig. 12 for the X error).

One observation is that, at the moments of total GPS loss, the error on our system remains quite low. The longest period of GPS neglect is shown at the beginning of the graph. Within this period, the system showed the highest error, which on the Y-axis was around 20 m. However, even at that time, the error on the X-axis remained quite low. In Table I, we show the numerical errors obtained on several tests carried out over more than 20-km urban paths. The mean and standard deviation of errors at both GPS loss sections and GPS available sections

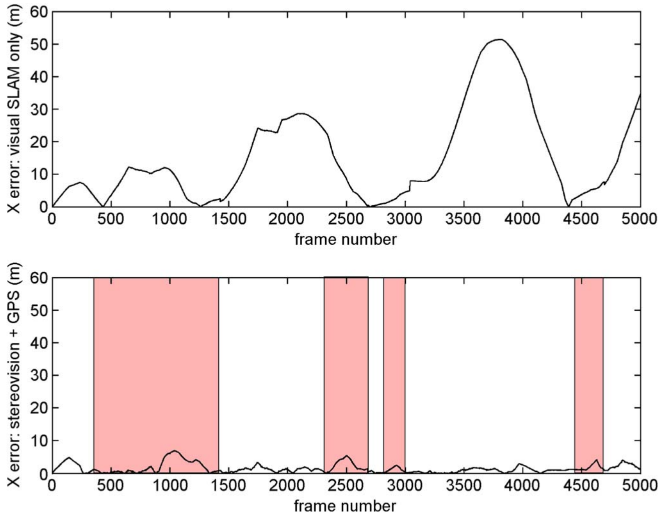

Fig. 12. Distance error on the X-axis using (top) visual SLAM only and (bottom) our combined SLAM system. GPS neglect sections are marked with bars.

TABLE I
ESTIMATION ERRORS. COMBINED SLAM

	Mean x (m)	Std x (m)	Mean z (m)	Std z (m)
Total path	1.75	2.52	4.16	3.02
GPS loss sections	4.11	2.17	9.67	4.93
GPS available sections	1.65	0.41	2.42	1.15

TABLE II
ESTIMATION ERRORS. GPS ONLY/VISUAL SLAM ONLY

	Mean x (m)	Std x (m)	Mean z (m)	Std z (m)
GPS only (available sections)	4.64	1.24	6.46	1.17
Visual SLAM only	16.89	15.58	19.50	17.23

are shown. As expected, the mean errors are higher on GPS loss sections. These errors are compared in Table II with the errors obtained by using either the standard GPS only or the visual SLAM system only. Both of them are higher than the errors obtained by the combined SLAM system. With respect to the processing time, the real-time implementation imposes a time constraint, which shall not exceed 33 ms for a 30 frame/s capturing rate. All of the results were taken using an AMD Turion 2.0-GHz CPU. Fig. 13 depicts the total processing times along the whole vehicle path for the first test.

As we can see, our method is able to work under the real-time constraint, with the average processing time remaining constant along the whole path. In Table III, we show the average processing times for some of the most important tasks in the process. Focusing on the low-level SLAM tasks, we can see that the higher time is used on the landmark initialization phase due to the large search area along the epipolar line, although we restricted its length for the 1 m → ∞ search range.

Regarding the high-level SLAM, the time dedicated to the *SIFT fingerprint-matching* process and the correction of the map at the time of loop closing, because it has 8520 landmarks, is significantly higher than real time. The use of SIFT features implies an increase of the fingerprint matching computing time over other appearance-based methods. However, the use of this kind of feature, besides the high distinctiveness, provides better accuracy on vehicle relocation, as their 3-D positions are used to geometrically estimate the actual vehicle pose.

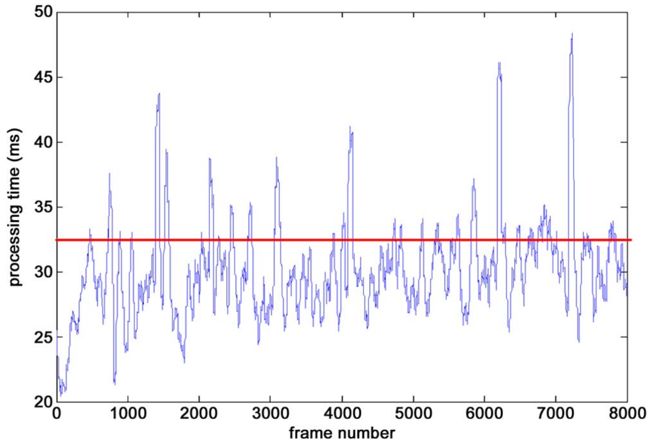

Fig. 13. Processing times for the whole tasks. The real-time limit is represented as a constant in 33 ms.

TABLE III
PROCESSING TIMES

Low level SLAM processing times		High level SLAM processing times (parallelized).	
Number of features / frame	5	Number of features	8520
		Number of fingerprints	281
Filter step	Time		Time
Measurements	3 ms	Fingerprint matches	3 s
Filter update	5 ms	Loop closing	1 s
Feature initializations	7 ms		
GPS processing (1s sampling period)	4ms		

On the other hand, both tasks do not belong to the continuous self-locating process carried out by the low-level SLAM, and therefore, there is no need to complete them within a single frame time slot. Therefore, we can obtain a positive fingerprint matching result of some few frames after it was really detected. Then, we can go back and start the loop-closing task. This implies that both of these tasks can be computed in *parallel*, keeping them outside the real-time computation. Regarding the GPS processing time, it was around 4 ms. Moreover, this task is executed at no more than 1 Hz; therefore, the impact on the processing times is negligible.

VIII. CONCLUSION

In this paper, we have presented a two-level (topological/metric) hierarchical SLAM that allows self-locating a vehicle in a large-scale outdoor urban environment using a cheap wide-angle stereo camera and a standard low-cost GPS as sensors. Using this hierarchical strategy, on one hand, we keep the local consistency of the metric submaps by means of the EKF (low SLAM level) and the global consistency by using a topological map and the MLR algorithm. On the other hand, our method is able to work under the real-time constraint, with the average processing time remaining constant for a very large scale environment. We have shown the positioning improvements of our system compared with using a simple standard GPS, opening the possibility to improve the current vehicle navigation systems. One limitation of our system is that a flat terrain is assumed for matching the 2-D map of the topological level with the 3-D maps of the metric level. This can cause graceful map accuracy degradation in highly rough terrains. On the other hand, loop-closing detection strongly depends on the visual appearance of images taken almost in the same place.

As a future work, we plan to generalize the MLR algorithm to manage 3-D characteristics. In addition, we plan to evaluate the addition of an IMU to improve the estimation from the visual sensor. Then, we have in mind to develop a vehicle navigation assistance prototype based in our system. Our final goal is the autonomous outdoor navigation of a vehicle in large-scale urban environments with recurrent trajectories (bus journeys, Theme Parks internal journeys, etc.), where a SLAM system such as ours can be very useful.

References

[1] A. Simon and J. C. Becker, "Vehicle guidance for an autonomous vehicle," in *Proc. ITSC*, 1999, pp. 429–434.

[2] J. Goldbeck, B. Huertgen, S. Ernst, and L. Kelch, "Lane following combining vision and DGPS," *Image Vis. Comput.*, vol. 18, no. 5, pp. 425–433, Apr. 2000.

[3] P. Newman, D. Cole, and K. Ho, "Outdoor SLAM using visual appearance and laser ranging," in *Proc. ICRA*, 2006, pp. 1180–1187.

[4] L. C. Bento, U. Nunes, F. Moita, and A. Surrecio, "Sensor fusion for precise autonomous vehicle navigation in outdoor semi-structured environments," in *Proc. ITSC*, 2005, pp. 245–250.

[5] J. Guivant and R. Katz, "Global urban localization based on road maps," in *Proc. IROS*, 2007, pp. 1079–1084.

[6] J.-H. Wang and Y. Gao, "The aiding of MEMS INS/GPS integration using artificial intelligence for land vehicle navigation," *IAENG Int. J. Comput. Science*, vol. 33, no. 1, pp. 33-1–33-11, 2007.

[7] M. Montemerlo, "FastSLAM: A factored solution to the simultaneous localization and mapping problem with unknown data association," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 2003.

[8] T. Bailey, J. Nieto, and E. Nebot, "Consistency of the FastSLAM algorithm," in *Proc. ICRA*, 2006, pp. 424–429.

[9] C. Stachniss, G. Grisetti, W. Burgard, and N. Roy, "Analyzing Gaussian proposal distributions for mapping with Rao–Blackwellized particle filters," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3485–3490.

[10] E. D. Nerurkar and S. I. Roumeliotis, "Power-SLAM: A linear-complexity, consistent algorithm for SLAM," in *Proc. IROS*, 2007, pp. 636–637.

[11] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. J. Robot. Res.*, vol. 25, no. 9, pp. 1181–1203, 2006.

[12] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, "Mapping large loops with a single hand-held camera," in *Proc. RSS*, Jun. 2007, pp. 38_1–38_8. [Online]. Available: http://www.roboticsproceedings.org/rss03/p38.html

[13] J. Tardós, J. Neira, P. Newman, and J. Leonard, "Robust mapping and localization in indoor environments using sonar data," *Int. J. Robot. Res.*, vol. 21, no. 4, pp. 311–330, 2002.

[14] L. M. Paz, P. Piniés, J. D. Tardós, and J. Neira, "Large scale 6DOF SLAM with stereo-in-hand," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 946–957, Oct. 2008.

[15] M. Bosse, P. Newman, J. Leonard, and S. Teller, "SLAM in large-scale cyclic environments using the Atlas framework," *Int. J. Robot. Res.*, vol. 23, no. 12, pp. 1113–1139, Dec. 2004.

[16] H. J. Chang, C. S. G. Lee, Y. C. Hu, and Y. Lu, "Multi-robot SLAM with topological/metric maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 1467–1472.

[17] U. Frese, P. Larsson, and T. Duckett, "A multilevel relaxation algorithm for simultaneous localization and mapping," *IEEE Trans. Robot.*, vol. 21, no. 2, pp. 196–207, Apr. 2005.

[18] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. 9th Int. Conf. Comput. Vis.*, Nice, France, 2003, pp. 1403–1410.

[19] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[20] P. Piníes and J. D. Tardós, "Scalable SLAM building conditionally independent local maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3466–3471.

[21] J. J. Leonard and H. J. S. Feder, "Decoupled stochastic mapping," Marine Robot. Lab., Mass. Inst. Technol., Cambridge, MA, 1999. Tech. Rep.

[22] K. S. Chong and L. Kleeman, "Large scale sonarray mapping using multiple connected local maps," in *Proc. Int. Conf. Field Service Robot.*, 1997, pp. 538–545.

[23] S. B. Williams, "Efficient solutions to autonomous mapping and navigation problems," Ph.D dissertation, Australian Centre Field Robot., Univ. Sydney, Sydney, Australia, 2001.

[24] T. Bailey, "Mobile robot localisation and mapping in extensive outdoor environments," Ph.D dissertation, Univ. Sydney, Sydney, Australia, 2002.

[25] M. Bosse, P. Newman, J. Leonard, and S. Teller, "An Atlas framework for scalable mapping," in *Proc. ICRA*, 2003, pp. 1899–1906.

[26] U. Frese, "Treemap: An O(log n) algorithm for indoor simultaneous localization and mapping," *Auton. Robots*, vol. 21, no. 2, pp. 103–122, Sep. 2006.

[27] E. Eade and T. Drummond, "Monocular SLAM as a graph of coalesced observations," in *Proc. ICCV*, 2007, pp. 1–8.

[28] H. Andreasson, T. Duckett, and A. Lilienthal, "Mini-SLAM: Minimalistic visual SLAM in large-scale environments based on a new interpretation of image similarity," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 4096–4101.

[29] F. Fraundorfer, C. Engels, and D. Nister, "Topological mapping, localization and navigation using image collections," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3872–3877.

[30] M. J. Milford and G. Wyeth, "Single camera vision-only SLAM on a suburban road network," in *Proc. ICRA*, 2008, pp. 3684–3689.

[31] A. I. Comport, E. Malis, and P. Rives, "Accurate quadrifocal tracking for robust 3D visual odometry," in *Proc. ICRA*, 2007, pp. 40–45.

[32] K. Ohno, T. Tsubouchi, and S. Yuta, "Outdoor map building based on odometry and RTK-GPS positioning fusion," in *Proc. ICRA*, 2004, pp. 684–690.

[33] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive GPS," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2006, pp. 1063–1068.

[34] R. Toledo-Moreo, M. A. Zamora-Izquierdo, B. Ubeda-Miarro, and A. F. Gomez-Skarmeta, "High-integrity IMM-EKF-based road vehicle navigation with low-cost GPS/SBAS/INS," *IEEE Trans. Intell. Trans. Syst.*, vol. 8, no. 3, pp. 491–511, Sep. 2007.

[35] A. D. Sappa, F. Dornaika, D. Ponsa, D. Geronimo, and A. Lopez, "An efficient approach to onboard stereo vision system pose estimation," *IEEE Trans. Intell. Trans. Syst.*, vol. 9, no. 3, pp. 476–490, Sep. 2008.

[36] D. Schleicher, L. M. Bergasa, E. Lopez, and M. Ocaña, "Real-time simultaneous localization and mapping using a wide-angle stereo camera and adaptive patches," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 2090–2095.

[37] D. Schleicher, L. M. Bergasa, R. Barea, E. Lopez, M. Ocaña, and J. Nuevo, "Real-time wide-angle stereo visual SLAM on large environments using SIFT features correction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3878–3883.

[38] A. J. Davison, "Mobile robot navigation using active vision," Ph.D dissertation, Univ. Oxford, Oxford, U.K., 1998.

[39] D. G. Lowe and J. Little, "Vision-based mobile robot localization and mapping using scale-invariant features," in *Proc. ICRA*, 2001, pp. 2051–2058.

[40] *NAVSTAR Global Positioning System Surveying Engineering Manual*, Dept. Army, U.S. Army Corps Eng., Washington, DC, 2003.

[41] A. Yasuda, "Satellite navigation system, GPS," *Advanced Topics for Marine Technology*, 2005.

[42] A. W. Stroupe, M. C. Martin, and T. Balch, "Distributed sensor fusion for object position estimation by multi-robot systems," in *Proc. ICRA*, 2001, pp. 1092–1098.

[43] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. CVPR*, 1997, p. 1106.

David Schleicher received the M.S. (with First-Class Honors) and Ph.D. degrees in electronic engineering in 2002 and 2009, respectively, from the University of Alcalá, Madrid, Spain, where he is currently working toward the Ph.D. degree.

His current research interests include computer vision, autonomous vehicles, simultaneous localization and mapping in robotics, and machine learning.

Luis M. Bergasa (M'04) received the M.S. degree in electrical engineering from the Technical University of Madrid, Madrid, Spain, in 1995 and the Ph.D. degree in electrical engineering from the University of Alcalá, Madrid, in 1999.

He is currently an Associate Professor with the Department of Electronics, University of Alcalá. His research interests include real-time computer vision and its applications, particularly in the field of robotics, assistance systems for elderly people, and intelligent transportation systems. He is the author of more than 100 publications in international journals, book chapters, and conference proceedings. He is an Associate Editor for the *Physical Agents Journal*. He is a habitual Reviewer of *Image and Vision Computing*, *Robotica*, etc.

Dr. Bergasa received the First Prize in the III contest of ideas for the creation of technology-based companies at the University of Alcala in 2008, the Best Research Award for the 3M Foundation Awards in the category of Industrial in 2004, and the Best Spanish Ph.D. Thesis Award in Robotics from by the Automatic Spanish Committee in 2005, as the director of the work. He is a member of the IEEE Robotics and Automation Society Technical Committee on Autonomous Ground Vehicles and Intelligent Transportation Systems and the IEEE Computer Science Society. He has been a member of the international program committee of several conferences, including the 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems (CIVVS), WIS 2007, and the Workshop de Agentes Físicos (WAF) in 2008. He is a habitual Reviewer of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Manuel Ocaña received the Ing. and Ph.D. degrees from the University of Alcalá, Madrid, Spain, in 2002 and 2005, respectively, both in electrical engineering.

He was a Researcher from 2002 to 2005 and is currently an Associate Professor with the Department of Electronics, University of Alcalá. He is the author of more than 20 refereed publications in international journals, book chapters, and conference proceeding. His research interests include robotics localization and navigation, assistant robotics and computer vision, and control systems for autonomous and assisted intelligent vehicles.

Dr. Ocaña was the recipient of the Best Research Award for the 3M Foundation Awards in the category of eSafety in 2003 and 2004.

Rafael Barea received the B.S. degree (with First-Class Honors) in telecommunications engineering from the University of Alcalá, Madrid, Spain, in 1994, the M.S. degree in telecommunications engineering from the Polytechnic University of Madrid in 1997, and the Ph.D. degree in telecommunications engineering from University of Alcalá in 2001.

He is currently an Associate Professor with the Department of Electronics, University of Alcalá, where has also been a Lecturer since 1994. He is the author of many refereed publications in international journals, book chapters, and conference proceedings. His current research interests include bioengineering, medical instrumentation, personal robotic aids, computer vision, system control, and neural networks.

María Elena López received the B.S. degree in telecommunications engineering, the M.Sc. degree in electronics engineering, and the Ph.D. degree from the University of Alcalá, Madrid, Spain, in 1994, 1999, and 2004, respectively.

Since 1995, she has been a Lecturer with the Department of Electronics, University of Alcalá. She is the author or coauthor of numerous publications in international journals and conference proceedings. Her current research interests include intelligent control and artificial vision for robotic applications.