2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Real-Time Hierarchical GPS Aided Visual SLAM on Urban Environments

David Schleicher, Luis M. Bergasa, Manuel Ocaña, Rafael Barea and Elena López

Abstract— **In this paper we present a new real-time hierarchical (topological/metric) Visual SLAM system focusing on the localization of a vehicle in large-scale outdoor urban environments. It is exclusively based on the visual information provided by both a low-cost wide-angle stereo camera and a low-cost GPS. Our approach divides the whole map into local sub-maps identified by the so-called fingerprint (reference poses). At the sub-map level (low level SLAM), 3D sequential mapping of natural landmarks and the vehicle location/orientation are obtained using a top-down Bayesian method to model the dynamic behavior. A higher topological level (high level SLAM) based on references poses has been added to reduce the global accumulated drift, keeping real-time constraints. Using this hierarchical strategy, we keep local consistency of the metric sub-maps, by mean of the EKF, and global consistency by using the topological map and the MultiLevel Relaxation (MLR) algorithm. GPS measurements are integrated at both levels, improving global estimation. Some experimental results for different large-scale urban environments are presented, showing an almost constant processing time.**

I. Introduction

THE interest in visual SLAM has grown tremendously in recent years as cameras have become much more inexpensive than lasers, and also 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 standalone information is not always as accurate as needed, specially on urban environments, mainly due to satellites occlusion because of high buildings, tunnels, etc.

One of the most popular methods to solve the SLAM problem is the Extended Kalman Filter (EKF) and more recently FastSLAM [1]. The first one has the covariance matrix growing problem while the second one discretizes the problem by using particle filters. Both of them are limited, in terms of computing time, when the environment becomes larger. To cope with that issue, two different approaches have been developed that try to divide the whole map into smaller ones in a hierarchical way. The original idea of having a set of sub-maps with uncertain relations dates back to [2] and [3]. The first approach introduces a high metric level over pieces of the metric map in the so-called *Metric-Metric* approach [4] [5] [6]. The second one is referred as the *Toplogical-Metric* one, which adds a high topological level over the metric sub-maps [7] [8] [9] [10] [11]. A third alternative to face the large scale SLAM problem is to use only *topological* maps without sub-maps associated to their vertex [12] [13]. These maps lack the details of the environments but they can achieve good results for certain applications.

Some of the last contributions to large scale path estimation using visual sensors have focused on recovering only 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. One example of this method is presented in [14], where they present an implementation that estimates ego-motion over a large path using a simple laptop webcam. Global consistency is achieved by using the RatSLAM method [15]. Another approach is presented in [16], based on the quadrifocal relations between stereo image pairs, using dense greyscale information. The method, taking a reference image pair, predicts the visual appearance of the scene by warping the reference using the quadrifocal geometry.

Our final goal is the autonomous outdoor navigation of a vehicle in large-scale environments where GPS signal does not exist or it is not reliable (tunnels, urban areas with tall buildings, mountainous forested environments, etc). Focusing on the SLAM approach, in one hand Metric-Metric methods do not keep a topological structure that helps on a global optimization in large scale environments as well as path planning techniques for navigation purposes. Topological approaches do not provide accurate information of vehicle state estimations instead. Therefore, our proposal to solve the large-scale problem is based on the hierarchical topological-metric approach, resembling on the NCFM algorithm [8]. Our metric level works in a similar way than visual odometry because the main goal, at this level, is the local positioning of the vehicle. Nevertheless, our proposal keeps local consistency, by

Manuscript received September 14, 2008. This work was supported in part by the Spanish Ministry of Education and Science (MEC) under grant TRA2005-08529-C02 (MOVICON Project) and grant PSE-370100-2007-2 (CABINTEC Project) as well as by the Community of Madrid under grant CM: S-0505/DPI/000176 (RoboCity2030 Project).

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

mean of the EKF, and global consistency by mean of the topological map and the MultiLevel Relaxation (MLR) algorithm [17]. The work presented in [11] proposes a hierarchical SLAM approach similar to ours, which defines a set of independent local maps linked to a global topological map, where each node (fingerprint) defines the reference frame for each of the local map. The global map is optimized using a recursive non-linear constrained least squares optimization algorithm instead of the MLR method. MLR improves the computation time of the least squares formula thanks to the Multi Level Relaxation technique. This leads to a slower running than the method presented on this work. Moreover, this system is based on a laser range finder sensor instead of a visual sensor as the used on this work. Our proposal uses visual information combined with the robust SIFT descriptors providing a more reliable way for loop detection.

This paper is organized as follows: the general structure of the system is described in section II. Section III presents the low level SLAM implementation, section IV studies the high level SLAM. In section V a large set of results is given to test the behavior of our system. Section VI contains our conclusions and future work. A demonstration video is published together with the paper.

II. IMPLEMENTATION

Our approach defines a *Low Level SLAM*, where the system uses stereo vision to feed an EKF to create local sub-maps which are expressed in local coordinates relative to some reference frames (*fingerprints*). Local poses are periodically fused with GPS measurements by using (1) (2). The only output used from the low level is the relation of the final vehicle frame (current fingerprint) relative to the reference vehicle frame (previous fingerprint). Over this low level a *High Level SLAM* is defined, where fingerprints uncertain relations are stored in a graph of relations defining stochastic constraints on the reference vehicle frames (fingerprints), as shown on Fig. 1. GPS is also added there as an absolute constraint on such a frame. This graph of relations is fed into MLR, that computes the least square estimate for the graph. Unfortunately, current implementation of MLR does not provide covariance information for this estimate. So, to derive uncertainty information, our approach implements, in parallel, another procedure. The algorithm exploits, that uncertain metrical relations can be compounded by (3). So to obtain uncertainty information about a reference vehicle pose (fingerprint), the shortest path in the above mentioned graph is taken, where the different relations from local maps are compounded. To detect loop closing, some of the fingerprints add visual information to the pose that helps to identify previously visited places. These kind of fingerprints are called SIFT fingerprints because they are based on SIFT features (*Scale Invariant Feature Transform).* They are taken under significant vehicle turns because in urban environments loop closing will appear only if the vehicle turns in perpendicular or quasi-perpendicular streets. In case of *long-term* GPS signal lost, at the time of signal recovering, vehicle pose is corrected and the global map is optimized by mean of the MLR as well.

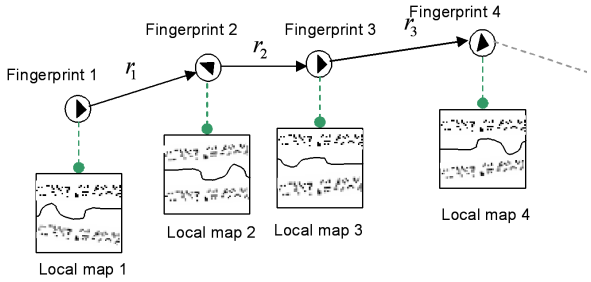

Fig. 1. General architecture of our two hierarchical levels SLAM. Each sub-map has an associated fingerprint.

III. LOW LEVEL SLAM

A. EKF implementation

This level is inspired on A. Davison monocular approach [18], however it has been modified for a stereo implementation as detailed in [19]. The low level state vector for the EKF is defined as $X_l = (X_v \quad Y_1 \quad Y_2 \quad \cdots)^T$, which is composed by the vehicle state vector $X_v = (X_{rob} \quad q_{rob} \quad v_{rob} \quad \omega)^T$ plus all local landmarks on the sub-map Y_i. Landmarks are identified by their corresponding features, which on this implementation are defined by the whole set of pixels of the patch. On this equation, X_{rob} is the 3D position of the vehicle relative to the local frame, $q_{rob} = (q_0 \quad q_x \quad q_y \quad q_z)^T$ is the orientation quaternion, v_{rob} is the linear speed and ω is the angular speed. For clarity reasons the sub-map notation is omitted.

B. Low Level GPS Fusion

Each time a new GPS reading $X_{GPS} = (x_{GPS} \quad y_{GPS})^T$ is available, which under normal conditions occur at 1s 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. Although GPS provides height data, as the high level SLAM manages only 2D information only 2D GPS data will be used. Then, to calculate the final position estimation we merge both estimates making use of their respective 2D uncertainty covariances, as we depict in equation (1). This is obtained by applying a two-dimensional statistical approach based on Bayes Rule and Kalman filters. Here, $X_{\mathrm{Pr}ob}$ and $P^0_{\mathrm{Pr}ob}$ are the 2D vehicle global position and global covariance respectively, obtained as the 3D vehicle position projection on the ground plane. This global 3D uncertainty P^0_{rob} is calculated using the procedure explained in the next

section. Resultant estimation improves uncertainty distribution because it is calculated as the product of the two original ones.

$$X^{fusion} = X_{\Pr ob} + P^0_{\Pr ob}\left(P^0_{\Pr ob} + P_{GPS}\right)^{-1}\left(X_{GPS} - X_{\Pr ob}\right) \quad (1)$$

GPS uncertainty P_{GPS} is obtained as a function of the HDOP (Horizontal Dilution Of Probability), containing the variable error provided by the GPS, and the UERE (User Equivalent Range Error), covering the estimated constant errors along time. In the same way, the following estimated covariance is calculated by mean of equation (2):

$$P^{fusion} = P^0_{\Pr ob} - P^0_{\Pr ob}\left(P^0_{\Pr ob} + P_{GPS}\right)^{-1} P^0_{\Pr ob} \quad (2)$$

Once the vehicle uncertainty has been updated, subsequent landmarks measurements will be based on this covariance. Therefore, these landmarks covariances will be calculated based on the already updated vehicle covariance. Vehicle orientation is estimated from the vector that joints the two last GPS position updates. In order to obtain the best estimation for the MLR nodes we generate them in a synchronized way with the GPS updates. Therefore, when conditions for a new node generation are ready, we wait until a new GPS update is available.

IV. High Level SLAM

Our SLAM implementation adds an additional topological level, called high level SLAM, to the explained low level SLAM in order to keep global map consistency with almost constant processing time. This goal is achieved by using the MLR algorithm over the reference poses. Therefore, the global map is divided into local sub-maps referenced by the mentioned fingerprints, one by one. There are two different classes of fingerprints: *Ordinary Fingerprints* and *SIFT fingerprints*. The first ones are denoted as $FP = \{fp_l | l \in 0...L\}$. Their only purpose is to store the vehicle reference pose $X^{fp_l}_{rob}$ and local covariance $P^{fp_l}_{rob}$ relative to the previous one, i.e., the reference frame of the current sub-map. The sub-map size is defined to keep the EKF linearization error low enough using the criterion explained in [10] in one hand. The other constraint is to keep the processing time, highly dependent on the low level landmarks per sub-map, below the real time constrain. Therefore the resultant sub-map size, after experimental testing, is limited to 10 m of covered path.

SIFT fingerprints are a sub-set of the first ones, denoted as $SF = \{sf_q \in FP | q \in 0...Q, Q < L\}$. 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. These fingerprints are taken only under the condition of having a significant change on the vehicle trajectory, defined by a maximum angular speed increase $\gamma_{\max}$ followed by a minimum decreasing $\gamma_{\min}$, both experimentally obtained. 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^0_{rob} because it keeps the global uncertainty information of the vehicle. If the matching is positive, it means that the vehicle is in a previously visited place and a *loop closing* is identified. Then, the MLR algorithm is applied in order to determine the maximum likelihood estimate of all nodes poses. Finally, nodes corrections are transmitted to theirs associated sub-maps.

A. Local Sub-maps

Each time a new fingerprint is created, an associated sub-map is created as well. Each of the old sub-maps defines the pose $X^{fp_{l-1}}_{fp_l}$ and covariance $P^{fp_{l-1}}_{fp_l}$ of a fingerprint relative to the previous node. The current sub-map defines the vehicle pose $X^{fp_l}_{rob}$ and covariance $P^{fp_l}_{rob}$ relative to the previous node. So, the global pose of the vehicle is computed by compounding these relations with uncertainty using the equation $X^0_{rob} = X^0_{fp_1} \oplus X^{fp_1}_{rob}$, where X^0_{rob} and $X^0_{fp_l}$ define the vehicle and previous reference absolute poses respectively. Due to the need of being aware about the current global uncertainty at any time, we need to maintain P^0_{rob} updated (see Fig. 2). We calculate it by using the *coupling summation formula* (see [8]), obtained from the *compounding* operation, in a recursive way. The process can be summarized as follows: first, to obtain P^0_{rob} we need to evaluate (3).

$$P^0_{rob} = \frac{\partial X^0_{rob}}{\partial X^0_{fp_l}} \cdot P^0_{fp_l} \cdot \left(\frac{\partial X^0_{rob}}{\partial X^0_{fp_l}}\right)^T + \frac{\partial X^0_{rob}}{\partial X^{fp_l}_{rob}} \cdot P^{fp_l}_{rob} \cdot \left(\frac{\partial X^0_{rob}}{\partial X^{fp_l}_{rob}}\right)^T \quad (3)$$

Second, to obtain the global covariance of the current fingerprint $P^0_{fp_l}$, we must apply (3) again, but this time to the previous fingerprint. We apply the same incremental procedure until we reach the first fingerprint, where $P^0_{fp_l} = P^{fp_0}_{fp_l}$ can be directly solved. At the time of sub-map creation, the current vehicle local uncertainty $P^{fp_{l+1}}_{rob}$, is assumed to be 0. The visible landmarks at that time are removed from the previous sub-map and incorporated to the new one, with local coordinates $Y_i^{fp_{l+1}}$. To calculate them from their expression on the previous sub-map $Y_i^{fp_l}$, we apply the *common root coupling* formula proposed on [8]. It allows changing the base reference from fp_l to fp_{l+1} by expressing landmarks on the vehicle reference frame $X^{fp_{l+1}}_{rob}$, which at that time is the actual new sub-map reference frame. To obtain the landmarks covariances expressed on fp_{l+1} base frame we make use of the *common root coupling* as well (4). As the local initial vehicle covariance is assumed 0, at the initial step, landmarks covariances depend only on the uncorrelated measurement noise contribution (see [18]). Therefore new landmarks are initially uncorrelated among them, so the final total sub-

map covariance at sub-map creation is composed as a block-diagonal matrix of vehicle covariance plus landmarks covariances.

$$P_{Y_i Y_i}^{fp_{l+1}} = \frac{\partial Y_i^{fp_{l+1}}}{\partial Y_i^{rob}} \cdot P_{Y_i Y_i}^{rob} \cdot \left(\frac{\partial Y_i^{fp_{l+1}}}{\partial Y_i^{rob}} \right)^T + \frac{\partial Y_i^{fp_{l+1}}}{\partial X_{fp_{l+1}}^{rob}} \cdot P_{fp_{l+1}}^{rob} \cdot \left(\frac{\partial Y_i^{fp_{l+1}}}{\partial X_{fp_{l+1}}^{rob}} \right)^T \quad (4)$$

B. SIFT Fingerprints

Our system identifies a specific place using the SIFT fingerprints. These fingerprints, apart of the vehicle pose, are composed by a number of SIFT [20] [21] [22] landmarks distributed across the reference image and characterize the visual appearance of the image. The identification is achieved by the association of a 128 length descriptor to each of the features, which will identify *uniquely* all of them. These SIFT feature descriptors are loaded in each SIFT fingerprint joint to the left image coordinates and the 3D vehicle position for the fingerprints matching process. To evaluate the distinctiveness of SIFT fingerprints a matching study has been carried out, as shown on Table 2.

C. Loop closing detection

To detect a loop closure, once a new SIFT fingerprint is generated, it is matched with all stored SIFT fingerprints within the uncertainty area defined by P_{rob}^0. The matching process is based on the Euclidean distance between the descriptors of the SIFT features on both fingerprints, as well as the geometrical relations between matched features on both images. A total matching probability is calculated taking into account both the number of matched features, as well as the fitting of the lines connecting each pair of matched SIFT features, into a model [23]. RANSAC method is used for that purpose.

D. Map Correction

Once a loop-closing has been detected, the whole map must be corrected according to the old place recognized. To do that, we use the MLR algorithm [17], which has proved to show a high efficiency in terms of computation cost and map complexity. 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 MLR inputs are the relative poses and covariances of the fingerprints. As outputs MLR returns the most *likely* set of reference poses, i.e., the set already corrected $X_M = \left(Xc_{fp_1}^0 \quad Xc_{fp_2}^0 \quad \cdots \quad Xc_{fp_L}^0 \right)^T$. The MLR algorithm manages only 2D information, therefore we need to obtain the 2D related fingerprint pose $X_{2D}^{fp_l} = \left(x_{2D} \quad y_{2D} \quad \theta_{2D} \right)^T$ and covariance $P_{2D}^{fp_l}$ from $X_{fp_l}^{fp_{l-1}}$ and $P_{fp_l}^{fp_{l-1}}$. Then the corresponding corrected fingerprints X_M are obtained, assuming flat terrain. To calculate the global vehicle uncertainty P_{rob}^0 after closing a loop, there is a situation where one fingerprint has relations with more than one additional fingerprint, as occurs, for example, to sf_3 (see Fig. 2). To calculate the current P_{rob}^0 we apply the recursive coupling summation formula (3) to the shortest possible path from the first fingerprint to the current position, which leads to the lowest P_{rob}^0. Being aware of the current global uncertainty is important in order to increase the fingerprints search process efficiency because the number of matched SIFT fingerprints will be lower.

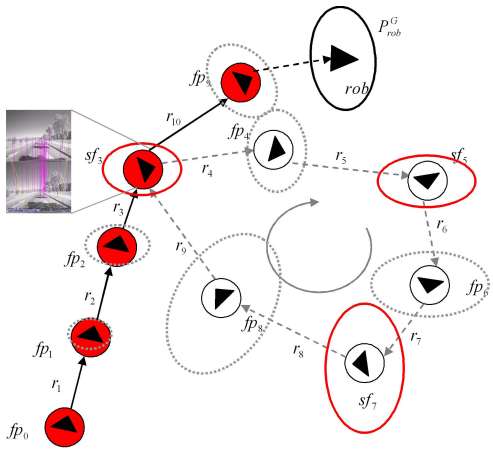

Fig. 2. Representation of the vehicle global uncertainties P_{rob}^0, increasing along the vehicle path at each of the reference poses. Solid red lines represent vehicle global uncertainties at SIFT fingerprints places. Numbers represent each fingerprint. Graph also shows an example of shorter path selection for global uncertainty calculation after a loop-closing situation.

The last step is to transfer the correction performed on the high SLAM level into the Low SLAM level. This is implicitly done by the transformation of each sub-map reference frame, i.e. all the landmarks within each sub-map will be moved according to their corresponding reference frame. By doing this, we keep the relative positions and covariances of the landmarks unchanged respect to their corresponding local sub-map reference frame. Therefore, sub-maps independency is kept as well.

E. High Level GPS Data Fusion

At this level, GPS data fusion is taken into account only at *long term* signal loose, which is usually the case within tunnels or in urban areas with high buildings. In that case, the state correction implies a global map correction that concerns mainly the section where the GPS signal was unavailable (see Fig. 4). Because GPS uncertainty is global, when GPS signal is available, fusion is carried out on global coordinates and nodes are introduced within the graph as global relations, i.e. MLR algorithm is fed with the global pose $X_{fp_l}^0$ and covariance $P_{fp_l}^0$ of each fingerprint.

V. Results

In order to test the behaviour of our system several video

sequences were collected from a commercial car manually driven in large urban areas, covering more than 20 km. 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 100º horizontal and vertical with a resolution of 320x240. The baseline of the stereo camera was 40 cm. Both cameras were synchronized at the time of commanding the start of transmission. The cameras were mounted inside the car on the top of the windscreen and near the rear-view mirror. We used a low-cost standard GPS was GlobalSat BU-353 USB.

Part of the path covered by the vehicle is shown on Fig. 3. The average speed of the car was around 30 km/h. The complete covered path was 3.17 km long. It contained 5 loops inside, taking 8520 low level landmarks and 281 nodes. More landmarks are located on high buildings areas, while GPS signal has more strength in open-spaced areas providing better location estimation. This shows that both sensors complement each other, providing good estimations for different situations. To evaluate the performance of our system we compared our results with a ground truth reference, obtained with an RTK-GPS Maxor GGDT, with an estimated accuracy of 2 cm. On Fig. 4 we show the high level map before and after a GPS signal recovering, as well as the optimization performed by the MLR. The Euclidean error relative to the ground truth of both the standard GPS and our combined SLAM implementation is depicted in Fig. 5. We obtain an average error of around 4 m and a reasonably low error at the moments of total GPS loose. This error is compared to the global uncertainty covariances for each node using the Euclidean formula applied to the X and Z components as well, showing reasonably consistent error estimates. As expected, uncertainty monotonically grows on GPS unavailable sections due to the relative measurements provided by the visual sensor. Fig. 3 depicts the estimation of our combined SLAM system and the standard GPS alone compared to the ground truth. The GPS signal was lost at different moments at the beginning of the path. The longest signal neglect period corresponds to the one shown on Fig. 4. The period begins on frame 453 and finishes at frame 1436, as shown on Fig. 5. The increased estimation error can be easily observed on that segment. However, we still have a relatively accurate estimation to be able to locate the vehicle. Respect to the processing time, the real-time implementation imposes a time constraint, which shall not exceed 33 ms for a 30 frames per second capturing rate. All results were taken using an AMD Turion 2.0 GHz CPU.

On Table 1 we show the average processing times for some of the most important tasks in the process. Low level SLAM tasks are limited in regards of time consuming due to the limited sub-map size. High level SLAM tasks slightly increase over time, but as they do not belong to the continuous self-locating process carried out by the low level SLAM they can be calculated apart in a parallel process.

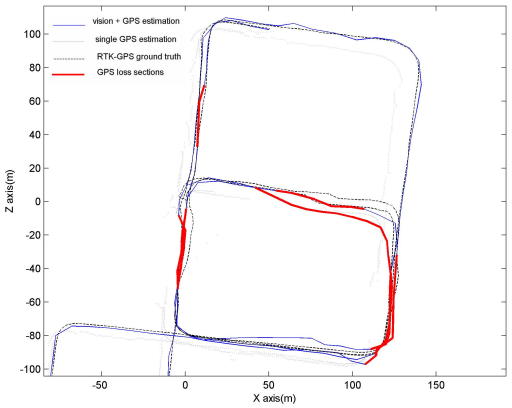

Fig. 3. Path estimation using only a standard low-cost GPS (dotted line), our SLAM method by means of vision and GPS (solid line), and the ground truth (dashed line). Thick red lines indicate path sections where GPS was unavailable.

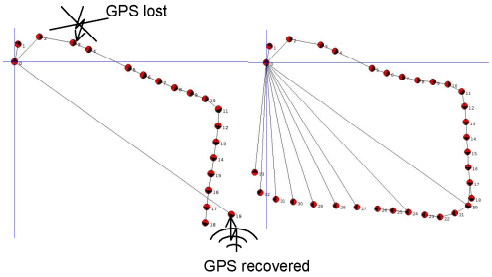

Fig. 4. MLR diagram before (left) and after (right) GPS recovering.

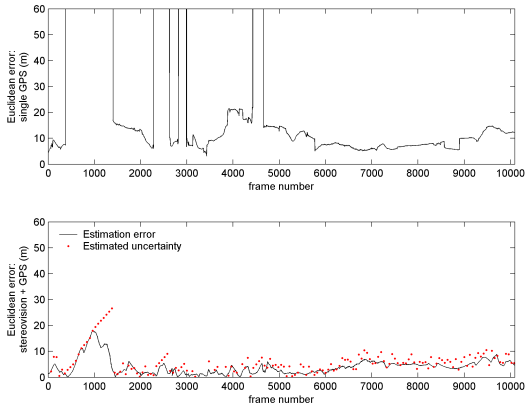

Fig. 5. Euclidean distance error ($\varepsilon = \sqrt{X^2 + Z^2}$) using standard single GPS (up) and our combined SLAM system (down). Global covariances uncertainties for each node are shown as well.

Therefore, the total processing time is proved to remain below the real time constraint within all our testing environments. On Table 2 we show a comparative study of the robustness to illumination changes. We focused on the

SIFT fingerprints matching process. We took 40 images database of the same place at different times along the day. We registered the number of erroneous matching (false positives) as well as missing ones (false negatives). From the results we can conclude that the probability of a false positive is extremely low, keeping reasonable values for false negatives at daylight. During the night results get worse on false negatives, mainly due to the decreasing of illuminated areas.

TABLE I
PROCESSING TIMES

Low level SLAM processing times		High level SLAM processing times (parallelized).	
Number of features / frame	5	Number of features	8520
		Number of nodes	281
Filter step	Time		Time
Measurements	3 ms	Fingerprint matches	3 s
Filter update	5 ms	Loop closing + graphic representation time	1 s + 10s
Feature initializations	7 ms		
GPS processing (1s sampling period)	4ms		

TABLE II
ROBUSTNESS TO ILLUMINATION CHANGES

% False positives / % False negatives	Daylight morning	Daylight afternoon	Sunset	Night
Daylight morning	0 / 7.5	1 / 10	0 / 10	0 / 40
Daylight afternoon		1 / 7.5	0 / 12.5	0 / 32.5
At sunset			0 / 10	0 / 35
At night				0 / 20

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a two levels (topological/metric) hierarchical SLAM that allows self-locating a vehicle in a large-scale urban environment using a low-cost wide-angle stereo camera and a standard low-cost GPS as sensors. We have shown the positioning improvements of our system regarding to use a simple standard GPS, opening the possibility to improve current vehicle navigation systems. Our system use a compounding and fusing algorithm to derive uncertainty information on a MLR process. This procedure may appear strange, but it is correct because MLR computes a least-square estimate and the compounding and fusion operations provide actual upper bounds on the uncertainty of this estimate. One limitation of our system is that flat terrain is assumed for matching the 2D map of the topological level with the 3D maps of the metric one.

As future work, we plan to use an estimator on the high level SLAM that can provide covariance information in order to get the estimate and its uncertainty from the same procedure. Then, we plan to generalize the MLR algorithm in order to manage 3D characteristics, as well as to replace the low level SLAM by Visual Odometry. 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 as ours can be very useful.

REFERENCES

[1] M. Montemerlo, "FastSLAM: A factored solution to the simultaneous localization and mapping problem with unknown data association." Ph.D. thesis, Carnegie Mellon Univeristy, 2003.

[2] H. F. Durrant-Whyte, "Uncertain geometry in robotics." IEEE Trans. Robotics, vol. 4, no. 1, pp. 23-31, 1988.

[3] R. Smith and M. Self and P. Cheeseman, "Estimating Uncertain Spatial Relationships in Robotics." Autonomous Robot Vehicles, pp. 167-193, 1988.

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

[5] P. Piníes, and J. D. Tardós, "Scalable SLAM building conditionally independent local maps." IROS 2007.

[6] U. Frese, "Treemap: An O(log n) algorithm for indoor simultaneous localization and mapping". Autonomous Robots, pp. 103-122. 2006.

[7] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós, "Mapping large loops with a single hand-held camera." Robotics: Science and Systems, RSS, 2007.

[8] T. Bailey, "Mobile robot localisation and mapping in extensive outdoor environments." PhD Thesis, University of Sydney, 2002.

[9] M. Bosse, P. Newman, J. Leonard, and S. Teller, "An Atlas Framework for Scalable Mapping". ICRA, pp. 1899-1906. 2003.

[10] E. Eade, T. Drummond, "Monocular SLAM as a Graph of Coalesced Observations," ICCV, pp.1-8. 2007.

[11] C. Estrada, J. Neira, and J. D. Tardós, "Hierarchical SLAM: real-time accurate mapping of large environments," IEEE Trans. Robotics, vol. 21, no. 4, pp. 588-596, August 2005.

[12] H. Andreasson., T. Duckett and A. Lilienthal, "Mini-SLAM: minimalistic visual SLAM in large-scale environments based on a new interpretation of image similarity," ICRA 2007.

[13] M. Cummins, and P. Newman, "Probabilistic Appearance Based Navigation and Loop Closing," IEEE International Conference on Robotics and Automation, pp. 2042-2048, 2007.

[14] M. J. Milford, and G. Wyeth, "Single camera vision-only SLAM on a suburban road network." ICRA 2008.

[15] M. J. Milford, G. Wyeth, and D. Prasser, "Simultaneous Localization and Mapping from Natural Landmarks using RatSLAM," Australasian Conference on Robotics and Automation, 2004.

[16] A. I. Comport, E. Malis, and P. Rives, "Accurate quadrifocal tracking for robust 3D visual odometry." ICRA, pp. 40-45, 2007.

[17] U. Frese, P. Larsson, and T. Duckett. "A multilevel relaxation algorithm for simultaneous localization and mapping." IEEE Transactions on Robotics, 21(2): pp.196–207, 2005.

[18] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera." ICCV 2003.

[19] 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." IROS2006.

[20] David G. Lowe, "Object Recognition from Local Scale-invariant Features". International Conference on Computer Vision, pp. 1150-1157. 1999.

[21] David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 2004.

[22] D.G. Lowe, and J. Little, "Vision-based mobile robot localization and mapping using scale-invariant features." In International Conference on Robotics and Automation, pp. 2051-58. 2001.

[23] 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." IROS 2007.