# Real–Time Hierarchical GPS Aided Visual SLAM on Urban Environments

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

Department of Electronics, University of Alcala, Alcalá de Henares, 28805 Madrid, Spain
dsg68818@telefonica.net,
{bergasa,barea,elena,mocana}@depeca.uah.es

**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.

**Keywords:** SLAM, Intelligent Vehicles, Computer Vision, Real-Time.

## 1 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, especially 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

R. Moreno-Díaz et al. (Eds.): EUROCAST 2009, LNCS 5717, pp. 326–333, 2009.

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]. The second one is referred as the *Toplogical-Metric* one, which adds a high topological level over the metric sub-maps [6] [7] [8]. A third alternative to face the large scale SLAM problem is to use only *topological* maps without sub-maps associated to their vertex [9] [10]. These maps lack the details of the environments but they can achieve good results for certain applications.

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). 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 the MultiLevel Relaxation (MLR) algorithm [11], which 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 another procedure in parallel. 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).* 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.

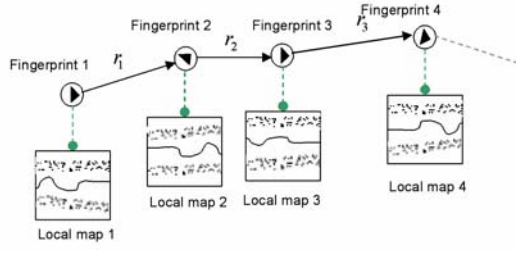


**Fig. 1.** General architecture of our two hierarchical levels SLAM. Each sub-map has an associated fingerprint.

## 2 Low Level SLAM

This level is inspired on A. Davison monocular approach [12], however it has been modified for a stereo implementation as detailed in [13]. 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 $\omega$ is the angular speed. For clarity reasons the sub-map notation is omitted.

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 by applying a two-dimensional statistical approach based on Bayes Rule and Kalman filters, as shown in (1). Here, $X_{\Pr ob}$ and $P^0_{\Pr ob}$ are the 2D vehicle global position and global covariance respectively.

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

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).

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

## 3 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 submaps 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, 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. When 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. When 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. Then, 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* (described in [6]), obtained from the *compounding* operation, in a recursive way: first, to obtain $P^0_{rob}$ we need to evaluate (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, repeating it until we reach the first fingerprint, where $P^0_{fp_l} = P^{fp_0}_{fp_l}$ can be directly solved.

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

### 3.1 Loop Closing and Map Correction

Our system identifies a specific place using the SIFT fingerprints. These fingerprints, in addition to the vehicle pose, are composed by a number of SIFT [14] landmarks distributed across the reference image and characterize the visual appearance of the image, allowing loop closing detection as explained in [15]. 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 [11], 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 $X_M = \begin{pmatrix} Xc^0_{fp_1} & Xc^0_{fp_2} & \cdots & Xc^0_{fp_L} \end{pmatrix}^T$. The MLR algorithm manages only 2D information, therefore we need to obtain the 2D related fingerprint pose $X^{fp_l}_{2D} = \begin{pmatrix} x_{2D} & y_{2D} & \theta_{2D} \end{pmatrix}^T$ and covariance $P^{fp_l}_{2D}$ from $X^{fp_{l-1}}_{fp_l}$ and $P^{fp_{l-1}}_{fp_l}$. Then the corresponding corrected fingerprints $X_M$ are obtained, assuming flat terrain. To calculate the global vehicle uncertainty $P^0_{rob}$ 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 $P^0_{rob}$ 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 $P^0_{rob}$. 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.

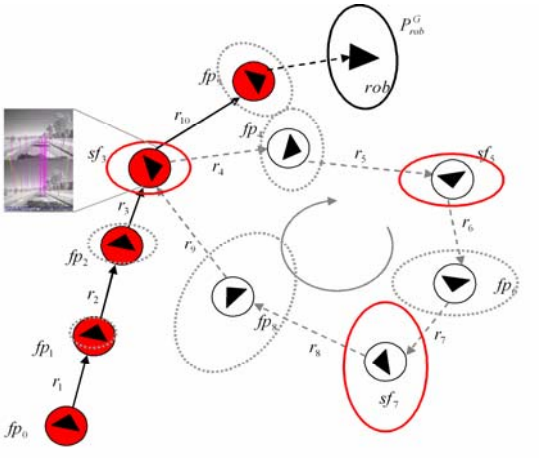


**Fig. 2.** Representation of the vehicle global uncertainties $P^0_{rob}$, increasing along the vehicle path at each of the reference poses. Solid red lines represent vehicle global uncertainties at SIFT fingerprints places. Numbers represent each fingerprint. Graph also shows an example of shorter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

The last step is to transfer the correction performed on the High level SLAM into the Low level SLAM. 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. 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. 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 $X^0_{fp_l}$ and covariance $P^0_{fp_l}$ of each fingerprint.

## 4 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, the GlobalSat BU-353 USB. 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. 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. The Euclidean error relative to the ground truth of both the standard GPS and our combined SLAM implementation is depicted in Fig. 4. 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 nearly 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. In spite of the increased estimation error on some segments at the beginning of the path, as shown on Fig. 4, 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.

**Table 1.** 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 | | |

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. Therefore, the total processing time is proved to remain below the real time constraint within all our testing environments.

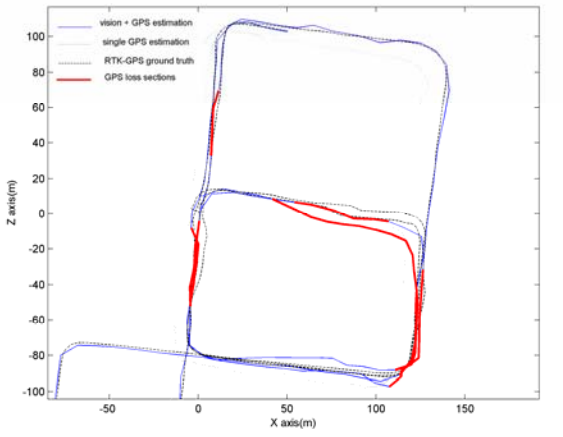


**Fig. 3.** Path estimation using only a standard low-cost GPS (dotted line), our SLAM method by means of vision and GPS (solid line), and the ground truth (dashed line). Thick red lines indicate path sections where GPS was unavailable.

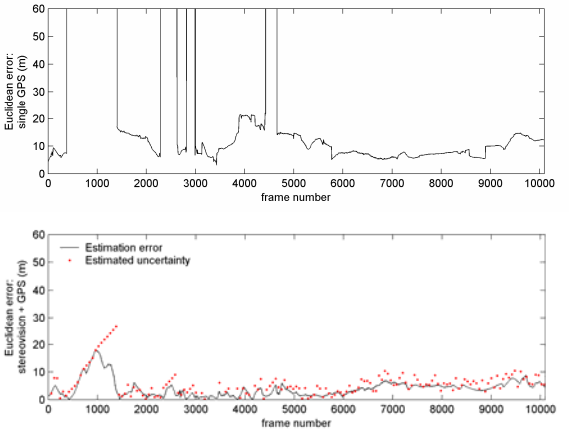


**Fig. 4.** Euclidean distance error ( $\varepsilon = \sqrt{X^2 + Z^2}$ ) using standard single GPS (up) and our combined SLAM system (down). Global covariances uncertainties for each node are shown as well.

## 5 Conclusions

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. 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 generalize the MLR algorithm in order to manage 3D characteristics, as well as to replace the Low Level SLAM by Visual Odometry.

## References


1. Montemerlo, M.: FastSLAM: A factored solution to the simultaneous localization and mapping problem with unknown data association. Ph.D. thesis, Carnegie Mellon Univeristy (2003)
2. Durrant-Whyte, H.F.: Uncertain geometry in robotics. IEEE Trans. Robotics 4(1), 23–31 (1988)
3. Smith, R., Self, M., Cheeseman, P.: Estimating Uncertain Spatial Relationships in Robotics. Autonomous Robot Vehicles, 167–193 (1988)
4. Piníes, P., Tardós, J.D.: Scalable SLAM building conditionally independent local maps. IROS (2007)
5. Frese, U.: Treemap: An O(log n) algorithm for indoor simultaneous localization and mapping. Autonomous Robots, 103–122 (2006)
6. Bailey, T.: Mobile robot localisation and mapping in extensive outdoor environments. PhD Thesis, University of Sydney (2002)
7. Bosse, M., Newman, P., Leonard, J., Teller, S.: An Atlas Framework for Scalable Mapping. In: ICRA, pp. 1899–1906 (2003)
8. Eade, E., Drummond, T.: Monocular SLAM as a Graph of Coalesced Observations. In: ICCV, pp. 1–8 (2007)
9. Andreasson, H., Duckett, T., Lilienthal, A.: Mini-SLAM: minimalistic visual SLAM in large-scale environments based on a new interpretation of image similarity. In: ICRA (2007)
10. Cummins, M., Newman, P.: Probabilistic Appearance Based Navigation and Loop Closing. In: IEEE International Conference on Robotics and Automation, pp. 2042–2048 (2007)
11. Frese, U., Larsson, P., Duckett, T.: A multilevel relaxation algorithm for simultaneous localization and mapping. IEEE Transactions on Robotics 21(2), 196–207 (2005)
12. Davison, A.J.: Real-time simultaneous localisation and mapping with a single camera. In: ICCV (2003)
13. Schleicher, D., Bergasa, L.M., Lopez, E., Ocaña, M.: Real-Time simultaneous localization and mapping using a wide-angle stereo camera and adaptive patches. In: IROS (2006)
14. Lowe, D.G.: Object Recognition from Local Scale-invariant Features. In: International Conference on Computer Vision, pp. 1150–1157 (1999)
15. Schleicher, D., Bergasa, L.M., Barea, R., Lopez, E., Ocaña, M., Nuevo, J.: Real-Time wide-angle stereo visual SLAM on large environments using SIFT features correction. In: IROS (2007)