Landmark based localization: LBA refinement using MCMC-optimized projections of RJMCMC-extracted road marks

Bahman Soheilian, Xiaozhi Qu, Mathieu Brédif

HAL Id: hal-01882538

https://hal.science/hal-01882538

Submitted on 27 Sep 2018

Landmark based localization : LBA refinement using MCMC-optimized projections of RJMCMC-extracted road marks

Bahman Soheilian, Xiaozhi Qu, Mathieu Brédif

Abstract— Precise localization in dense urban areas is a challenging task for both mobile mapping and driver assistance systems. This paper proposes a strategy to use road markings as localization landmarks for vision based systems. First step consists in reconstructing a map of road marks. A mobile mapping system equipped with precise georeferencing devices is applied to scan the scene in 3D and to generate an ortho-image of the road surface. A RJMCMC sampler that is coupled with a simulated annealing method is applied to detect occurrences of road marking templates instanced from an extensible database of road mark patterns. The detected objects are reconstructed in 3D using the height information obtained from 3D points. A calibrated camera and a low cost GPS receiver are embedded on a vehicle and used as localization devices. Local bundle adjustment (LBA) is applied to estimate the trajectory of the vehicle. In order to reduce the drift of the trajectory, images are matched with the reconstructed road marks frequently. The matching is initialized by the initial poses that are estimated by LBA and optimized by a MCMC algorithm. The matching provides ground control points that are integrated in the LBA in order to refine the pose parameters. The method is evaluated on a set of images acquired in a real urban area and is compared with a precise ground-truth.

I. INTRODUCTION

Precise localization of mobile vehicles in dense urban areas constitutes an important component of several systems such as Mobile Mapping (MMS), Advanced Driver Assistance (ADAS) and Autonomous Navigation (ANS) Systems. The most popular localization system is GNSS (Global Navigation Satellite System). In dense urban areas GNSS localization however suffers from signal outage and multi-paths, thus failing to provide sufficient accuracy. More advanced systems rely on Inertial Navigation Systems (INS) and/or odometers in order to overcome the lack of reliable GNSS signals by dead-reckoning [1]. An innate issue of these systems is drift. The quantity of the drift depends on the quality of the INS and high precision systems are not affordable for applications such as low cost map updating systems and car navigators.

More affordable relative localization systems using vision sensors were proposed by the computer vision and robotic communities. Visual odometry approach proposed by Nistér *et al.* [2] enables successive computation of relative poses in a sequence of mono or stereo images in real time. More optimal pose parameters can be obtained by a local bundle adjustment technique proposed by [3] in real-time. Like in any dead-reckoning method, in vision based methods, errors

The authors are with Universite Paris-Est, IGN, SRIG, MATIS, 73 avenue de Paris, 94160 Saint Mande, France. {firstname.lastname}@ign.fr

are accumulated through the sequence and lead to drifts. Even though loop closure [4] enables the system to estimate the drift, on the one hand many trajectories contain no loops at all and on the other hand adjustment of large loops is computationally expensive for both SLAM (Simultaneous Localization and Mapping) and bundle adjustment methods [5].

An efficient solution for adjusting the drift consists in integrating external absolute information into the pose estimation at short intervals. GNSS observations were integrated with vision based approaches such as visual odometry [6], bundle adjustment [7], SLAM [8] and bundle adjustment SLAM [9].

Other approaches used pre-built maps and applied data association techniques in order to adjust the drift of vision based localization methods. Various types of maps were used in different systems.

Some authors pre-build a set of 3D points by structure from motion or 3D laser scanners in an offline mapping mode. In order to ensure a sufficient precision for these 3D points, precise georeferencing techniques such as differential GPS, precise INS, loop closure and even integration of ground control points (GCP) were used. Then, online localization is performed by matching features that are acquired in real time to the corresponding features in the georeferenced map. Royer *et al.* applied this strategy for localizing a monocular system [10]. A very similar system used planar patches instead of points [11]. In the work presented by Bodensteiner *et al.* a laser scanner was used for mapping. Then, optical images were aligned to the 3D mapped points in the navigation mode [12]. Other authors used laser scanning for both mapping and navigation modes [13]. Point based maps take large storage capacities on-board of the vehicle and make partial updates laborious.

More classical maps such as building models [14], digital elevation models [15], [16], ortho-photos [17], textured facades and road surfaces [18], [19] were also used as landmarks for localization. These GIS objects are easily available since they are produced and updated by national cartographic agencies. In addition, their compactness simplifies their embedment on the vehicle. However their geometric precision is generally rather limited.

Other techniques applied a map containing compact se-

mantic objects such as pole-like objects [20], road markings [21] and traffic signs [22]. The simple geometric shapes of these objects make their detection and matching precise and easy. In addition they take low storage capacity and are manageable to update.

This paper presents an offline localization technique that is useful for low cost street-level mapping systems. The system reaches sub-decimetric absolute accuracies using only affordable sensors (a low cost GNSS receiver and a calibrated camera) and taking benefit from a pre-built 3D road marking map.

II. RELATED WORK

Lane marking detection was investigated in many research works in intelligent transportation community since the 90s [23]. They were usually used for lane keeping [24]. More recently road marking objects were also used as landmarks for localization.

Tournaire *et al.* used 3D zebra-crossings that were reconstructed in street-level stereo pairs as matching objects between ground-level images and multiple georeferenced aerial images. The method applied GPS localization for initialization and enabled to compute 6D pose parameters of the stereo pairs with sub-decimetric precision of the aerial images' poses [25].

Pink proposed a two step mapping-localization schema [26] . Road marking objects are first detected in aerial images with a semi-automatic approach. They assume that all objects lie on a plane referred to as the ground plane. This provides a 2D map of markings. During the localization step, road mark contours are detected with a set of image filters in street-level images. The detected objects are then projected on the ground plane estimated using the V-disparity method. Initial position of objects are known using GPS. The matching is performed by ICP (Iterative Closest Point) algorithm using centroid of markings. A 2D transformation composed of a 2D shift and a heading angle was computed in order to refine the pose of the image.

In another method [21], road markings were captured manually as polylines in an aerial image. The polylines were then projected on the ground surface (supposed to be a plane). Poses of street level images are initialized by a GPS. Given an initial pose, an image of map features, like the camera would see it is simulated. The simulated image is then compared to the image captured by the camera. Pose refinement parameters (a 2D shift and a heading angle) were estimated by maximizing the similarity of the two images. One of the interests of this technique is in comparing map features directly with optical image without intermediate object detection step in ground-level images.

In a system called LaneLoc, road markings were mapped in an offline mode using georeferenced bird-eye views of ground-level 3D point clouds and optical images by a semi-automatic method. In the localization mode, road markings were detected using a filtering method. GNSS is used for initializing the pose of the camera. The features of the map are projected in the current image. A 2D shift and a heading angle were estimated by matching the map features (lines) and the detected markings (points) [27]. Another system following a similar strategy integrated road marking observations obtained from two lateral cameras with GPS and IMU for pose estimation [28].

In the aforementioned papers, the pose parameters were initialized by GNSS and final parameters were obtained by matching road marking objects between the real-time views and the map. Following this strategy, there is no merge of two sources of localization. The localization switches frequently between the GNSS solution (if no map matching available) and map matching solution. Ranganathan *et al.* used a windowed bundle adjustment as initial pose estimator [29]. Then every time a pose is corrected by matching road marks, the refined pose is integrated as a measurement in the bundle adjustment. The benefit of this method is that the refined pose contributes to improve the accuracy of its neighboring images in the sequence as well.

Two main approximations were used by all of the afore-mentioned methods.

a) Flat road surface: In mapping and localization phases, the road surface was approximated by a plane. Moreover the position and orientation of the plane was supposed to be known in relation to the camera. These two approximations introduce errors in the coordinates of the road marking.

b) 2D pose correction: In the pose correction phase, only three degrees of freedom are taken into account instead of six. The height of the camera, roll and pitch are supposed to be exactly known. The corrections were injected to horizontal shifts and heading angle only.

In order to reach higher localization accuracies that are required for our mapping system, we propose a method that is free of these two approximations.

III. OUR STRATEGY

Our system is based on two separate mapping and localization phases. In the mapping phase we use precise georeferencing devices (GNSS/INS/odometer) and a 3D laser scanner embedded on a mobile mapping system such as the STEREOPOLIS MMS [30] to acquire a precise and dense 3D description of the environment. We applied an extension of our previous work on road marking object reconstruction from a 3D point cloud [31]. It will be briefly explained in section IV. In the localization phase we use only one GNSS observation at the start point and one calibrated camera. The localization is performed by our previous work on extension of local bundle adjustment to integrate ground control points (GCP) [32]. This part is summarized in section V. The main contribution of this paper is in associating 3D georeferenced

road marks in the map to the images during the localization phase without using any road mark detector. Uncertainty propagation through the bundle adjustment provides error ellipsoids that will drive the search space for road mark matching (section VI). Once a 3D marking is associated to its corresponding 2D position in an image, it provides a GCP. The obtained GCP together with its uncertainty adds a weighted constraint equation to the bundle adjustment system. The resolution of the system provides refined pose parameters for the entire image sequence. Results on a real dataset are presented in section VII.

IV. ROAD MARKING MAP GENERATION

We propose to generate a road marking map using an extension of the method proposed by Hervieu *et al.* [31]. We first summarize here this approach, then detail the proposed extensions, and finally lift the 2D extractions as a 3D road marking database.

A. Original approach

First, the acquired point cloud by a Mobile Mapping System is projected vertically onto a horizontal plane in order to generate an orthophoto-like Lidar image (figure 2) with two channels (intensity and height) which undergo a hole-filling filter to cope with the irregular Lidar sampling.

Within the Lidar intensity orthophoto, road markings are then searched for as occurrences of a translated/rotated/scaled rectangular road marking template instanced from a library of road markings (figure 3). This search space is modeled as a set of road marking types and for each type a fixed aspect ratio, an interval of scale and a template vector pattern delineating the white road marking area against a dark background. Thus the extraction of road markings boils down to finding a set of road markings $X = (\ell_i, x_i, y_i, \theta_i, \lambda_i)$ parameterized by a type ℓ, a translation (x, y), a rotation by θ and a scaling λ (figure 1). The marking type defines a pattern I_ℓ that may be rasterized into the intensity orthophoto geometry using the affine transform $T_{x,y,\theta,\lambda}$ (denoted T_{X_i} for short).

Hervieu *et al.* [31] formulate the road marking extraction as an energy minimization problem over the varying-dimension search space defined above, with an energy defined over a set of road markings $X = (X_i)_{i=1...n}$ as :

$$U(X) \;=\; \sum_{i=1}^{n} u_1(X_i) + \sum_{i<j} u_2(X_i, X_j) \qquad (1)$$

$$u_1(X_i) \;=\; f^0 - \max(0, ZMNC(I_{\ell_i}, T_{X_i}^{-1}(I))) \quad (2)$$

$$u_2(X_i, X_j) \;=\; \beta \frac{\mid S(X_i) \cap S(X_j) \mid}{min(\mid S(X_i) \mid, \mid S(X_j) \mid)} \qquad (3)$$

where f^0 is the cost of adding an object. A low value of f^0 (0.35 as used in [31]) enables the optimization to add objects with lower correlation scores at lower costs. A high value, in contrast penalizes the objects with low correlation scores. It should provide a trade-off between the number of over-detections and under-detections. In the present work we chose a higher value $f^0 = 0.55$ in order to reduce the

number of over-detections which comes at a cost of higher number of under-detection. $ZMNC(I, I')$ denotes the zero-mean normalized correlation between images I and I' and $S(X) = T_X(I_{\ell_X})$ is the resampled image of the pattern and $|\cdot|$ and $\cap$ denote respectively the area and intersection of white pixels. The coefficient β tunes the tradeoff between the energy terms u_1 and u_2 ($\beta = 100$ in [31]).
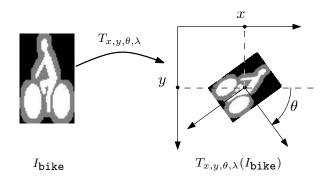

Fig. 1.   The object i with parameters ($\ell_i =$ bike, $x_i, y_i, \theta_i, \lambda_i$).

This energy is minimized using a Reversible-Jump Markov Chain Monte Carlo (RJMCMC) sampler coupled with a simulated annealing, which may cope with search spaces of varying dimensions (the number of road markings to extract itself being unknown) and arbitrary energy functions (cf. Fig. 4). Hervieu *et al.* [31] further discusses both standard and more advanced RJMCMC kernels which may be used to bias the random sampling toward good solutions, thereby improving the convergence rate.

B. Proposed extensions

- New patterns have been introduced, leveraging the extensibility of the original paper (figure 3).
- The data energy term has been scaled by the road-marking perimeter in order to reduce over-detections.

$$u_1'(X_i) = u_1(X_i) \operatorname{perimeter}(X_i) \qquad (4)$$

  It enables to favour larger objects that could be replaced by many smaller objects using previous data energy.
- A new binary orientation energy u_{orient}' has been introduced in order to penalize incompatible orientations of neighboring road markings. Road markings follow usually a same direction and are nearly parallel except in the intersections where perpendicular markings are observed. This energy term is computed for neighboring objects that are situated at a distance lower than $5m$.
- The raster-based intersection energy proved to be very time consuming as it required the resampling of the template pattern and pixel-by-pixel raster comparisons to get the raster area of intersection. This energy has been replaced by a simplified version u_{inter}', penalizing the intersection of the road marking oriented bounding boxes (OBB) instead. This drastically reduced computing times while the approximation is very reasonable as road markings are very rarely sufficiently close that

their oriented bounding boxes intersect.

$$
\begin{aligned}
u_2'(X_i, X_j) &= u_{orient}'(X_i, X_j) + u_{inter}'(X_i, X_j) \\
u_{orient}'(X_i, X_j) &= \max(0, -cos4\Delta_\theta) \qquad (5) \\
u_{inter}'(X_i, X_j) &= u_{inter}(\mathbf{OBB}(X_i), \mathbf{OBB}(X_j))
\end{aligned}
$$

- Road markings tend to follow a regular layout, thus we added a birth/death in a neighborhood kernel which gives the sampler the opportunity to explore more efficiently the possibility that some road marking might exist next to an already detected one. The inclusion of this kernel also resulted in a significant performance boost [33].
- Finally, another kernel was added to enable a uniform type switch, which proved to be necessary in order to help the sampler find the right road marking type.
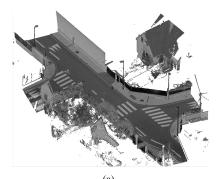
C. 3D road marking database

Once the 2D rectangles labelled with a road marking type have been extracted, they are lifted in 3D using the digital terrain model (DTM) encoded in the height channel of the Lidar orthophoto. A simple height lookup enables the lifting of these 2D rectangles as a 4-sided 3D polygon.

Dictated by the targeted application, and due to the abundance of road markings in street view images, the detection tradeoff has been tuned to minimize false detections at the cost of under-detecting some road markings. This results in an extraction with some under-detection but very limited over-detection. In order to ensure the accuracy of this database the extracted road markings may be validated interactively in order to remove the remaining few false positives. Note that this manual intervention is optional and very limited as the extracted road markings may be sorted using their data attachment term u_1' such that the operator only has to review the few extracted road markings that have the worst data evidence.

V. LOCALIZATION SYSTEM

Similar to other works, we use a monocular calibrated camera embedded on a vehicle as a localization device. The localization is initialized using a low cost GPS at the beginning of the trajectory. Except at the starting point, any GPS observation is used in the localization system. Local bundle adjustment (LBA) [3] and uncertainty propagation [34] are applied in order to estimate the pose parameters and their covariance. Like in any relative positioning algorithm, accumulation of errors in LBA leads to considerable drifts for long trajectories. In order to reduce the drift, we proposed to integrate ground control points (GCP) in the equation system of LBA [32]. GCPs are points whose coordinates are known in both world and image coordinate systems. In order to explain how road markings are integrated as GCPs in the bundle adjustment, we explain briefly the principal concepts of a constrained LBA. For more details about the algorithm the reader is invited to refer to [32].

The principal concept of LBA (cf. Fig. 5) is to compute the system on a low number of images N since the complexity
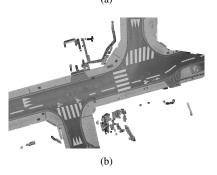

(a)

(b)

Fig. 2. (a) 3D point cloud, (b) corresponding orthophoto I ($GSD = 2cm$) generated from points' intensities.

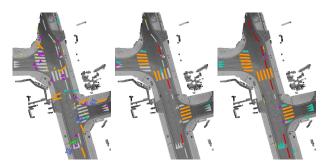Fig. 3. Library of road marking template patterns ($GSD = 2cm$).

Fig. 4. Simulated annealing-coupled RJMCMC optimization.

of the system is $\mathcal{O}(N^3)$. At the very beginning a classical bundle adjustment is computed on the first N images (step 1 in Fig. 5). Then the bundle adjustment window is slid by a step $n < N$ (step 2 in Fig. 5). $N - n$ poses of the new window are already computed by the previous step (C_p) and n new poses are unknown (C_n). The vector X_t contains the 3D coordinates of all tie points. The re-projection errors are estimated by equation 6.

$$
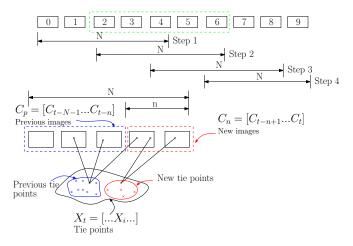v_t = F(C_p, C_n, X_t) - m_t \qquad (6)
$$

Fig. 5. Schematic flow of local bundle adjustment.

F is the function projecting X_t in poses C_p and C_n and m_t is a vector containing observed coordinates in the images. As mentioned before, previous poses C_p were already computed during the previous step. The previously computed values C_p^0 are used to constrain new estimation of C_p, denoting this error as v_p:

$$v_p = C_p - C_p^0 \tag{7}$$

The system is then resolved by minimizing the weighted quadratic error:

$$[\hat{C}_p, \hat{C}_n, \hat{X}_t] = \underset{[C_p, C_n, X_t]}{\arg\min} \frac{1}{2}(v_t^T Q_t^{-1} v_t + v_p^T Q_{C_p}^{-1} v_p) \tag{8}$$

Where, Q_t is the covariance matrix of the 2D tie point positioning in image space and Q_{C_p} is the covariance of poses that are computed in the previous step.

In order to integrate GCPs in the equation system, let us suppose that their world coordinates are coded in vector X_G and their 2D coordinates in image spaces in vector m_g. The re-projection error of these points (v_g) is expressed as:

$$v_g = F_g(C_p, C_n, X_G) - m_g \tag{9}$$

Since the 3D world coordinates of these points are known (X_G^0) the following constraint may be added to the equation system:

$$v_G = X_G - X_G^0 \tag{10}$$

After adding GCPs the system is resolved by equation 11.

$$
\begin{aligned}
[\hat{C}_p, \hat{C}_n, \hat{X}_t, \hat{X}_G] &= \\
\underset{[C_p, C_n, X_t, X_G]}{\arg\min} \quad \frac{1}{2} \quad &(v_t^T Q_t^{-1} v_t + v_p^T Q_{C_p}^{-1} v_p \\
&+ v_g^T Q_g^{-1} v_g + v_G^T Q_G^{-1} v_G)
\end{aligned} \tag{11}
$$

Q_G is the covariance matrix of 3D GCPs and is related to the precision of 3D reconstruction of road marks (section IV). Q_g is the covariance of position of road marks in image space. It is related to the precision of road mark matching (section VI). After resolution of the system the covariance of the parameters can be estimated by partial inversion of normal matrix [32]. In this paper we used SIFT points for

feature detecting and matching. Features of every image are matched to the previous three images. The size of the sliding window is seven ($N = 7$) and sliding step is one ($n = 1$). The output of this step is 6D pose parameters of image sequence together with their covariance.

VI. Road mark matching

This section tackles the problem of providing new GCPs to the LBA by leveraging the LBA-estimated pose with uncertainty by matching the projection of a 3D road marking database into the current image.

A. Initialization using the LBA estimates

Given a LBA-estimated pose with uncertainties for the current pose (C, Q_C), the nearest 3D road markings (X_G) may easily be selected from the database and projected into the current image. Due to the perspective projection (F), this yields for each marking in the database a 4-sided polygon which may be used to resample the road marking template pattern through the homography defined by these 4 points. Furthermore, the pose and road mark uncertainties (Q_C and Q_G) translate directly into a 2D Gaussian re-projection uncertainty ($Q_{\bar{x}_G}$) for each of the 4 points following equation 12.

$$Q_{\bar{x}_G} = \begin{bmatrix} \frac{\partial F}{\partial C} & \frac{\partial F}{\partial X_G} \end{bmatrix} \begin{bmatrix} Q_C & 0 \\ 0 & Q_G \end{bmatrix} \begin{bmatrix} \frac{\partial F}{\partial C} \\ \frac{\partial F}{\partial X_G} \end{bmatrix} \tag{12}$$

B. Search space definition

These four 2D points together with their 2D Gaussian uncertainties allow us to define a sufficiently tight search space: We consider for each 2D point a search space for the refined position as the 2D bounding box of the 99%-confidence region of the Gaussian uncertainty centered around its estimated 2D position (figure 6). This conservative search space definition is both able to cope with small error underestimation, and is convenient to deal with.
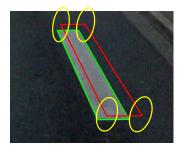

Fig. 6. LBA-based search space definition, initialization and MCMC optimization.

C. Objective function

Given four image points defining the homographic projection of a road marking template into the current view, we can assess the quality of this projection by computing the ZMNC of this homographic projection with the image content. We can then formulate our problem as finding the four road

marking corner projections within their uncertainty-based bounding boxes such that they maximize this ZMNC score. In order to rule out degenerate set ups, we further impose that the four points define a convex polygon.

D. MCMC Optimisation

Contrary to the offline road marking extraction step, this optimization is defined in a fixed dimension setup: the 8 coordinates of the four 2D points. Given the nature of the objective function, a more specific optimizer is not trivially available, thus we propose to perform a (regular) MCMC optimization.

Given the strong correlation between the errors of the 4 projected points, we propose the following transformation kernels for the MCMC modification proposal step:

- An overall rigid translation of the 4 points
- A translation of one point leaving the three other points fixed with a lower amplitude

Similar to the database generation optimization, the MCMC sampler is coupled with a simulated annealing in order to optimize the ZMNC objective function, rejecting all modifications that produce a concave polygon. The initialization is provided by the road marking projection using the estimated pose.

VII. RESULTS AND EVALUATION

In order to evaluate the proposed mapping/localization algorithm, it is applied to a $500m$ trajectory in an urban area (cf. Fig. 7(a)). STEREOPOLIS mobile mapping system [30] was applied to scan the area. It provides point clouds of $5cm$ resolution on the road surface. Thanks to an integrated georeferencing system (GPS/INS/odometer), the acquired data are precisely georeferenced. A full HD camera is embedded on the system. The focal length is quite high and the horizontal angle of view is about $70°$. Intrinsic parameters of the camera and its position and orientation in relation to the system are known. It provides our ground-truth for vision based localization. Fig. 7(c) depicts an example of image acquired by the system. The road mark mapping algorithm presented in section IV was applied to detect the road marks in the orthophoto. The detected objects were projected in 3D using the height value of Laser points at the corners of the road marking. The 3D map was edited manually for removing false objects. Fig. 7(b) depicts the provided 3D map on the test area.

The first image of the sequence was initialized using GPS localization and LBA (cf. section V) was launched using a sliding window containing seven images ($N = 7$) and progressing step of one image ($n = 1$). From the very beginning of the trajectory road marks were matched to the images and provided GCPs. Fig. 8 shows the interest of road marking at the beginning of the trajectory. The green polyline shows the ground-truth trajectory. The red trajectory in Fig. 8(a) was obtained by LBA without using road marks and the blue one in Fig. 8(b) was computed by LBA integrating road marks as GCPs. The error ellipsoids were exaggerated

ten times. The larger size of error ellipsoids as well as larger difference with ground-truth is noticeable when road marking were not integrated in the LBA.
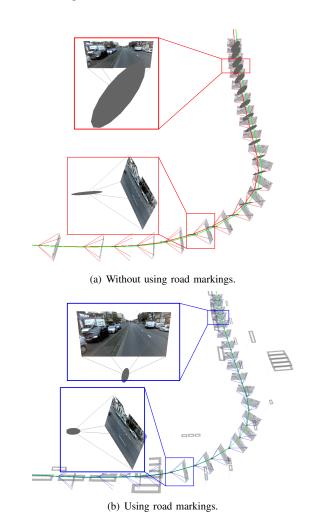

(a) Without using road markings.

(b) Using road markings.

Fig. 8. Vision based pose estimation at the first $20m$ of the trajectory. Ground-truth trajectory is drawn in green.

The results of localization with and without using road marks on the entire trajectory are shown in Fig. 9. The maximum of error if no road mark is integrated in the system is $4m$. This error is reduced to $0.4m$ if road marks are integrated in the system. The accuracy along the trajectory is correlated to the density of road marks. For a large part of the sequence the error is around $0.1m$. This error increased to $0.4m$ at the end of the trajectory since the density decreased. We noticed that in the case where no road marking is used (Fig. 9(a)) the positioning error decreased near image number 130. We would expect the error to increase continuously. The decrease of error starts to happen at a very sharp turning. At this position only about 20 tie points are available between successive images whereas this number is about 100 for normal conditions. We guess that relative pose estimation at this area is erroneous. It seems that this error happens coincidently in the opposite direction to the accumulated error and leads to a decrease of final error to $0.5m$ at the

(a)   (b)   (c)
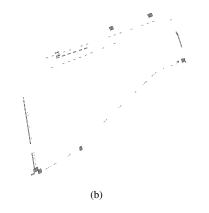
Fig. 7.   (a) Our $500m$ test trajectory. (b) Approximately 200 road mark objects are reconstructed by the mapping step and 150 were kept after manually editing. (c) An example of image acquired by the embedded camera on the system.

end of the trajectory. This problem was not happened for the localization mode using road marks. Only a few numbers of observed GCPs can resolve such degenerate pose estimation cases.

Most of the computation time goes into road mark matching. The number of needed iterations for convergence in the MCMC algorithm depends on the initial estimation and the size of the uncertainty region. Due to some non Gaussian errors, our estimated uncertainty is sometimes underestimated. This is why in practice we enlarge the search area to guarantee the convergence and it slows down the algorithm. In addition the computation time for each iteration is proportional to the number of the pixels of the pattern (50×200 pixels). It takes $10 - 40s$ for each object.

VIII. CONCLUSIONS AND TRENDS

This paper presents a solution to the problem of precise localization for low cost mobile mapping systems equipped only with low cost sensors (camera and GPS). The method is particularly careful about the uncertainties and does few approximations. It enables a real 6D pose estimation using precise landmarks. Evaluation on a test trajectory revealed an accuracy of $10\ cm$ in areas with high density of road marks and $40\ cm$ in areas with low densities. More tests are required in order to compare the presented method using road marks with our previous work using traffic signs as landmarks. Fusion of both methods should obviously provide better result. Actually, the system assumes that the landmark map contains no false object and that the uncertainty of initial pose is sufficiently low to avoid ambiguities in landmark matching step. The robustness of landmark association phase would be improved if all the visible landmarks in a view undergo the matching at the same time (instead of matching every object separately). The high computation time of the matching part constitutes the main barrier for using the proposed localization method in real-time applications. One interesting idea for accelerating this part would be to use image gradient to match the contours and/or corners of objects instead of using all the pixels inside the road marking for costly correlation score computation. Moreover, a smarter
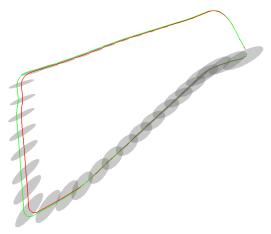
adaptation of MCMC parameters (number of the iterations, starting temperature and temperature decrease rate) for each object can also help to avoid useless iterations and save computation time. Finally, we believe that the real-time Jurie-Dhome [35] tracker can be adapted to the problem of road mark matching.
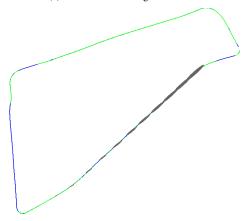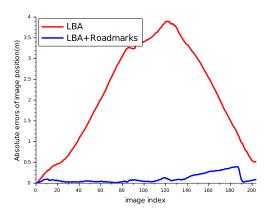
IX. ACKNOWLEDGEMENTS

REFERENCES

[1] D. Lapucha, K. Schwarz, M. Cannon, and H. Martell, "The use of ins/gps in a highway survey system," in *Position Location and Navigation Symposium, 1990. Record. The 1990's - A Decade of Excellence in the Navigation Sciences. IEEE PLANS '90.*, Mar 1990, pp. 413–420.

[2] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of the CVPR*, vol. 1.   IEEE, 2004, pp. 652–659.

[3] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3d reconstruction," in *Proceedings of the CVPR*, vol. 1.   IEEE, 2006, pp. 363–370.

[4] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proceedings of the ICCV*, Oct 2003, pp. 1403–1410 vol.2.

[5] J. Aulinas, Y. R. Petillot, J. Salvi, and X. Lladó, "The slam problem: a survey." in *CCIA*, 2008, pp. 363–371.

[6] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive gps," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3.   IEEE, 2006, pp. 1063–1068.

[7] M. Lhuillier, "Incremental fusion of structure-from-motion and gps using constrained bundle adjustments," *TPAMI*, vol. 34, no. 12, pp. 2489–2495, 2012.

[8] D. Schleicher, L. M. Bergasa, M. Ocaña, R. Barea, and E. López, "Real-time hierarchical gps aided visual slam on urban environments," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*.   IEEE, 2009, pp. 4381–4386.

[9] Y. Shi, S. Ji, Z. Shi, Y. Duan, and R. Shibasaki, "Gps-supported visual slam with a rigorous sensor model for a panoramic camera in outdoor environments," *Sensors*, vol. 13, no. 1, pp. 119–136, 2012.

[10] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest, "Monocular vision for mobile robot localization and autonomous navigation," *IJCV*, vol. 74, no. 3, pp. 237–260, 2007.

[11] B. Charmette, E. Royer, and F. Chausse, "Efficient planar features matching for robot localization using gpu," in *CVPRW*, June 2010, pp. 16–23.

(a) LBA without using road marks.

(b) LBA using road marks.

(c) Difference of localization methods with ground-truth.

Fig. 9. Comparison of localization without and with road marks. Error ellipsoids are exaggerated 10 times. Ground-truth trajectory is drawn in green.

[12] C. Bodensteiner, W. Hübner, K. Jüngling, P. Solbrig, and M. Arens, "Monocular camera trajectory optimization using lidar data," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2018–2025.

[13] K. Yoneda, H. Tehrani, T. Ogawa, N. Hukuyama, and S. Mita, "Lidar scan feature for localization with highly precise 3-d map," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014, pp. 1345–1350.

[14] C. Arth, C. Pirchheim, J. Ventura, D. Schmalstieg, and V. Lepetit, "Instant outdoor localization and slam initialization from 2.5d maps," in *Proceedings of the ISMAR*, 2015.

[15] D. Larnaout, S. Bourgeois, V. Gay-Bellile, and M. Dhome, "Towards bundle adjustment with gis constraints for online geo-localization of a vehicle in urban center," in *Proceedings of the 3DIMPVT*. IEEE, 2012, pp. 348–355.

[16] P. Lothe, S. Bourgeois, F. Dekeyser, E. Royer, and M. Dhome, "Towards geographical referencing of monocular slam reconstruction using 3d city models: Application to real-time accurate vision-based localization," in *Proceedings of the CVPR*. IEEE, 2009, pp. 2882–2889.

[17] M. Jaud, R. Rouveure, P. Faure, and M.-O. Monod, "Methods for fmcw radar map georeferencing," *IJPRS*, vol. 84, pp. 33–42, 2013.

[18] C. Cappelle, M. El Badaoui El Najjar, F. Charpillet, and D. Pomorski, "Outdoor obstacle detection and localisation with monovision and 3D geographical database," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, Sept 2007, pp. 1102–1107.

[19] J. Levinson, M. Montemerlo, and S. Thrun, "Map-based precision vehicle localization in urban environments." in *Robotics: Science and Systems*, vol. 4. Citeseer, 2007, p. 1.

[20] A. Schlichting and C. Brenner, "Localization using automotive laser scanners and local pattern matching," in *Proceedings of the IVS*, June 2014, pp. 414–419.

[21] N. Mattern, R. Schubert, and G. Wanielik, "High-accurate vehicle localization using digital maps and coherency images," in *Proceedings of the IVS*. IEEE, 2010, pp. 462–469.

[22] L. Wei, B. Soheilian, and V. Gouet-Brunet, "Augmenting vehicle localization accuracy with cameras and 3d road infrastructure database," in *Computer Vision-ECCV 2014 Workshops*. Springer, 2014, pp. 194–208.

[23] M. Bertozzi and a. Broggi, "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection." *IEEE TIP*, vol. 7, no. 1, pp. 62–81, jan 1998.

[24] J.-F. Liu, T.-H. Hu, and T.-H. Hsu, "Design of an Automotive Lane Keeping System Based on the Structure of Electric Power Steering," *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, vol. 36, no. 4-5, pp. 391–411, 2001.

[25] O. Tournaire, B. Soheilian, and N. Paparoditis, "Towards a sub-decimetric georeferencing of ground-based mobile mapping systems in urban areas: matching ground-based and aerial-based imagery using roadmarks," in *Proc. of the ISPRS Commission I Symposium*, vol. Part A, Marne-la-Vall ee, France, jul 2006, interne.

[26] O. Pink, "Visual map matching and localization using a global feature map," in *CVPRW*. IEEE, June 2008, pp. 1–7.

[27] M. Schreiber, C. Knoppel, and U. Franke, "Laneloc: Lane marking based localization using highly accurate maps," in *Proceedings of the IVS*. IEEE, 2013, pp. 449–454.

[28] D. Gruyer, R. Belaroussi, and M. Revilloud, "Map-aided localization with lateral perception," *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 674–680, 2014.

[29] A. Ranganathan, D. Ilstrup, and T. Wu, "Light-weight localization for vehicles using road markings," in *IROS*. IEEE, Nov 2013, pp. 921–927.

[30] N. Paparoditis, J.-P. Papelard, B. Cannelle, A. Devaux, B. Soheilian, N. David, and E. HOUZAY, "Stereopolis ii: A multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology," *RFPT*, no. 200, pp. 69–79, 2012.

[31] A. Hervieu, B. Soheilian, and M. Brédif, "Road marking extraction using a model&data-driven rj-mcmc," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W4, pp. 47–54, 2015. [Online]. Available: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3-W4/47/2015/

[32] X. Qu, B. Soheilian, and N. Paparoditis, "Vehicle localization using mono-camera and geo-referenced traffic signs," in *Proceedings of the IVS*, June 2015, pp. 605–610.

[33] O. Tournaire and N. Paparoditis, "A geometric stochastic approach based on marked point processes for road mark detection from high resolution aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 6, pp. 621 – 631, 2009.

[34] A. Eudes and M. Lhuillier, "Error propagations for local bundle adjustment," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2411–2418.

[35] F. Jurie and M. Dhome, "Hyperplane approximation for template matching," *TPAMI*, vol. 24, no. 7, pp. 996–1000, Jul 2002.