Vehicle localization using mono-camera and geo-referenced traffic signs

Xiaozhi Qu[1], Bahman Soheilian[1] and Nicolas Paparoditis[1]

Abstract— **Vision based localization is a cost effective method for indoor and outdoor application. However, it has drift problem if none global optimization is used. We proposed a geo-referenced traffic sign based localization method, which integrated the constraints of 3D traffic signs with local bundle adjustment to reduce the drift. Comparing to global bundle adjustment, Local Bundle Adjustment(LBA) has low computational cost but suffers the drift problem for large scale localization because of the random error accumulation. We reduced the drift by means of the constraints from geo-referenced traffic signs for bundle adjustment process. The original LBA model was extended for the constraints and the traffic signs were detected in images and matched with 3D landmark database automatically. From the experiments of simulated and real images, our approach can reduce the drift and have better locating results than none-constraint LBA based localization method.**

I. INTRODUCTION

Precise localization is the basic issue for autonomous navigation and driving assistant systems. GPS (Global Positioning System) is widely applied for positioning in such systems, however, in urban areas due to high buildings and narrow streets it suffers often from localization errors caused by multi-paths and signal masks. INS (Inertial Navigation System)S and/or odometer can help filtering errors and interpolating between GPS interruptions [1]. Intrinsic drifts of theses sensors can accumulate during long term GPS interruptions and lead to localization errors as well. High cost of precise INSs motivated many systems to use inexpensive vision based sensors for localization.

Simultaneous Localization and Mapping (SLAM) enables to create a set of 3D points and to estimate the motion progressively[2]. This method is based on Kalman filter but suffer from computational inefficiencies due to the need for inversion of large matrices [3].

Visual Odometry (VO) using mono image or stereo pairs was proposed for estimating mobile poses in relation to a fixed point [4]. Some authors proposed to integrate proprioceptive information coming from the sensors in the classical VO, in order to accelerate the computation time [5]. As VO is based on relative measurements, error accumulation leads to drift over time [6].

An alternative technique is Bundle Adjustment (BA) [7] integrating matched points within a sequence of images and finding a solution simultaneously optimal with respect to both camera parameters and 3D points.The strength of bundle adjustment is in providing a globally optimal solution. However, the size of equation system increases with the growing number of images and matched points and becomes computationally expensive for large set of images. To cope with this problem Local Bundle Adjustment (LBA) was proposed by Mouragnon et al. [8]. The method is incremental and runs a bundle adjustment when new key frames are added. An extension of the same method that takes into account the uncertainties was proposed by Eudes and Lhuillier [9]. The method enables propagating the errors through the bundle adjustment blocks and obtains uncertainties of pose parameters.

All the aforementioned vision-based methods are based on relative measurements. So the random errors are accumulated trough the trajectory and cause drifts. In order to reduce the drift external georeferenced data should be added to the processes.

a) GPS measurements:

An incremental structure from motion fusing GPS signal in a constrained bundle adjustment was proposed by Lhuillier[10]. The method enables to decrease regularly the drift effect by associating constraints from GPS position on camera centers. The GPS measurements were also integrated in SLAM based structure from motion technique[11]. The quality of these methods is directly related to the precision of GPS measurements. The GPS mask or multi-path could lead to drifts.

b) Sparse point cloud:

The geo-referenced low level landmarks such as interest points [12] and planar patches [13] are also investigated . In order to align the trajectory, geo-referenced point clouds acquired by laser scanners were also used as maps for vehicle localization. Yoneda et al [14] used an embedded Velodyne scanner on the vehicle and proposed a matching approach to align the real time scanning data to the geo-referenced map . Maddern et al [15] proposed an illumination invariant color space allowing to register color images to the 3D map . This family of strategy enables to reduce the drift significantly by aligning the trajectory. The first drawback is the requirement for high storage volume due to the high number of points or patches. The second inconvenience is in the difficulty of resolving the matching ambiguities. The last inconvenience is the difficulty of updating the 3D point cloud.

c) GIS data:

Lothe et al [16] proposed to register a sparse 3D point cloud obtained with SLAM to geo-referenced 3D building models. In a similar approach, the Digital Elevation Model (DEM) was applied to reduce the number of camera parameters from six to five supposing that the height of the embedded camera in relation to the ground surface is fixed [17]. Geo-

[1]Xiaozhi Qu, Bahman Soheilian and Nicolas Paparoditis are with Universite Paris-Est, IGN, SRIG, MATIS, 73 avenue de Paris, 94160 Saint Mande, France. {firstname.lastname}@ign.fr

referenced textured objects such as facades [18] and road surfaces [19] were also applied as landmarks for localization through image registering. Comparing to 3D point clouds the advantages of using geographic databases are in lower storage volume requirement and easier updating process. However The quality of registration is related to the precision of GIS data. Nowadays the precisions of large scale geographic databases are far from being sub-decimeter.

d) Punctual semantic features:

Schlichting and Brenner [20] applied pole-like objects as georeferenced landmark database. An embedded laser scanner on the vehicle allows to detect automatically vertical objects. Then the detected objects are matched to the landmark database and provide corrections to the vehicle position. Road markings were also applied as matching objects for image registration. Tournaire et al. proposed an algorithm for registering street-level images to multiple-view georeferenced aerial images using zebra-crossings [21]. Road marking maps were also applied for image-based localization [22][23]. Wei et al [24] applied a 3D database of geo-referenced road markings and traffic signs in order to reduce the drift of INS based localization. A calibrated stereo base allows to reconstruct the signs and markings. The predicted pose by INS was corrected by matching the reconstructed features to the geo-referenced objects in the landmark database.

The main benefits of using punctual features as landmarks are:

- Precision: Their simple geometric forms make their detection and reconstruction precise and robust.
- Low storage volume: Their standard shapes enable compact representations.
- Low matching ambiguity: Their shape and texture are stable and unambiguous.

These are the reasons why we adopted traffic signs as landmark database. In this paper we present a localization approach using a sequence of mono images and a set of geo-referenced 3D traffic signs. We suppose that the vehicle starts from a known point (using a valid GPS positioning for example). The local bundle adjustment algorithm presented by Mouragnon et al. [8] together with its uncertainty propagation [9] were used to estimate the poses of cameras in relation to the starting point. Every time the estimated pose is near enough to at least one traffic sign in the landmark database, a traffic sign detection algorithm is run and the detected signs will be matched to 3D signs in the landmark database. This will provide Ground Control Points (GCP) that will be integrated in the bundle adjustment. The main contribution of our paper is in generating GCP automatically and integrating it in a bundle adjustment process that should decrease the drift of structure from motion algorithm.

The papers is organized as follows: Section II describes our geo-referenced database. Section III introduces the matching and initial parameters estimation methods. The new approach of local bundle adjustment is proposed in section IV and section VI presents the experiment results. Finally, conclusions are shown in section VII.

II. Landmark database generation

The geo-referenced traffic sign database was generated from the images acquired by Stereopolis Mobile Mapping System [25] in a dense urban area. [26] proposed the method for traffic sign detection, recognition and reconstruction. Not only the 3D model of the traffic signs was reconstructed, but the categories and semantic meaning of traffic sings were also recognized. Thanks to the precise detection (sub-pixel level) and integration of multi-view algorithm and prior knowledge, the precision of the 3D traffic signs reaches $3.5cm$ for planimetric position and $1cm$ for elevation. Figure 1 depicts an example of reconstructed signs embedded in a virtual model of the city.

Fig. 1: Real and virtual view of the reconstructed scene [27].

III. Matching and resection

The initial values of parameters for bundle adjustment, which are 6 parameters (3 for position and 3 for orientation) for each image pose and 3 parameters for each 3D points are needed. Image matching, pose estimation and multi-view resection are the key techniques for initial values calculation.

A. Feature extraction and matching

In this paper, SIFT(Scale Invariant Feature Transform) algorithm is chosen to detect and describe the key points [28] and FLANN (Fast Library for Approximate Nearest Neighbors) [29] is used to speed up corresponding point searching. We take one image pair[I,J] for example and every feature points in image I will be searched the corresponding point in image J. 1) construct K-D(K-Dimensional) tree for image J feature descriptors; 2) search the matching point in the constructed K-D tree for each feature point in image I according to the euclidean distance between two descriptor vectors. The rule for good matching is $d_1/d_2<0.6$, where d_1 is minimal distance and d_2 is second minimal distance. We reject outliers using the epipolar constraint. The distance between an image point and the corresponding epipolar line should be less than 2 pixel. The fundamental matrix which is estimated with an AC-RANSAC (A Contrario RANSAC) [30], is employed to compute the equation of epipolar line.

The aforementioned content is about the matching method for one image pair. In this paper, we just match the current image with latest 3 images. In this case, every new image has three time pair-wise matching.

B. Pose estimation and triangulation

We assume that the geo-coordinate system is known for us and the first image pose is fixed as(R_0, C_0); where R_0 is rotation matrix from image coordinate system to geo-coordinate system, C_0 is the position of camera center. The second key frame is chosen which has sufficient motion

from first image to second key frame. The relative pose(R,t) from first image to second key frame is estimated with the 5-point algorithm proposed in [31]. The actual distance from first image to second key frame is set as D_{01} and the scale is defined as $\lambda = D_{01}/\|t\|$. Therefore, the geo-referenced pose of second key frame can be calculated as: $R_1 = RR_0^{-1}, \quad C_1 = -\lambda R_1 t + C_0$.

After initialization of the first pair, a resection-triangulation process is used to estimate the initial values for other images. For instance, the pose of third key frame is estimated with the 3D points that are reconstructed by the first pair. We apply P3P (Perspective three Point) algorithm [32] to estimate the image pose according the reconstructed 3D points. With these known image poses, the new 3D points that are from new key frame can be calculated with DLT (Direct Linear Transformation) [33].

IV. LOCAL BUNDLE ADJUSTMENT(LBA)

In order to speed up the procedure of optimization, we apply local bundle adjustment [9]. It means that only latest N images are taken into account for bundle adjustment. We take an example with ten images and set $N = 5$, $n = 2$. N is the size of sliding window, n is the number of new images of current LBA step. $N - n$ is the number of images that have already been optimized in previous step. The procedure of LBA is shown in figure 2.

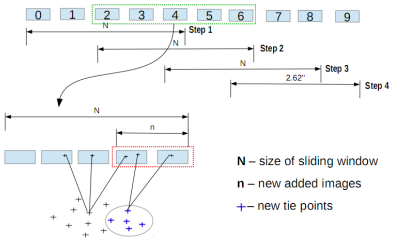


Fig. 2: Schematic flow of local bundle adjustment.

We define the parameters for LBA at time t as following notation:

$$\begin{aligned} C_p &= \begin{bmatrix} C_{t-N-1} ... C_{t-n} \end{bmatrix} \\ C_c &= \begin{bmatrix} C_{t-n+1} ... C_t \end{bmatrix} \\ X_t &= \begin{bmatrix} ... X_i ... \end{bmatrix} \end{aligned} \tag{1}$$

where, C_p is the vector of camera poses estimated by previous step, C_c corresponds to image poses of latest n frames and X_t is a vector containing 3D points.

Bundle adjustment aims at minimizing the re-projection error between the observed position of points and the predicted positions. Supposing that image errors are Gaussian distribution, the bundle adjustment is formulated as a maximum likelihood estimator. It looks for x that make $f(x) = V^T PV$ reach minimum. Where, $f(x)$ is called cost function and V is the vector of residuals of image points. The P is the wight matrix. We define $F(x)$ as the function of pinhole projection and $x = [C, X]^T$ as unknowns including camera poses and 3D points. So the V can be estimated with equation $V = F(x) - l$, where l is the 2D observations. Only the first order Taylor expansion of $F(x)$ is considered, which is $F(x) \approx F(x_0) + \frac{\partial F}{\partial x}\delta_x$, so the error equation is:

$$V = \frac{\partial F}{\partial x}\delta_x - r \tag{2}$$

where, $r = l - F(x_0)$ is the vector of differences between the observations and the predicted values.

In current step, we regard C_p^0 as observation of C_p and its covariance matrix is defined as Q_{C_p}. The error equation of C_p is defined as $V_p = C_p - C_p^0$. If we apply the Taylor expansion at C_p^0 for V_p, the original error equations of LBA is:

$$\begin{bmatrix} V_p \\ V_t \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ \frac{\partial F_t}{\partial C_p} & \frac{\partial F_t}{\partial C_c} & \frac{\partial F_t}{\partial X_t} \end{bmatrix} \begin{bmatrix} \delta_{C_p} \\ \delta_{C_c} \\ \delta_{X_t} \end{bmatrix} - \begin{bmatrix} 0 \\ r \end{bmatrix} \tag{3}$$

where σ is the standard error of image noise and the weight matrix is $diag(Q_{C_p}^{-1}, \frac{1}{\sigma^2} I)$.

The new approach of LBA will take into account geo-referenced traffic signs as GCPs (Ground Control points). We aim at integrating the 3D coordinates of GCPs as weighted constraints in the bundle adjustment equation system. We define X_s^0 as measured 3D coordinates GCPs with covariance Q_s. The X_s is noted as a vector containing the adjusted coordinates of the GCPs. The error equation of X_s is defined as $V_s = X_s - X_s^0$. We also apply the first order Taylor expansion at X_s^0 for V_s. Finally, back-projection error of the GCPs in image space is noted by V_g and is obtained similarly to equation 2. Our new LBA error equations are deduced as equation 4.

$$\begin{bmatrix} V_p \\ V_t \\ V_s \\ V_g \end{bmatrix} = \begin{bmatrix} I & 0 & 0 & 0 \\ \frac{\partial F_t}{\partial C_p} & \frac{\partial F_t}{\partial C_c} & 0 & \frac{\partial F_t}{\partial X_t} \\ 0 & 0 & I & 0 \\ \frac{\partial F_g}{\partial C_p} & \frac{\partial F_g}{\partial C_c} & \frac{\partial F_g}{\partial X_s} & 0 \end{bmatrix} \begin{bmatrix} \delta_{C_p} \\ \delta_{C_c} \\ \delta_{X_s} \\ \delta_{X_t} \end{bmatrix} - \begin{bmatrix} 0 \\ r_t \\ 0 \\ r_g \end{bmatrix} \tag{4}$$

Where:

- δ_{C_p} is the correction to the previous camera parameters
- δ_{C_c} is the correction to the current camera parameters
- δ_{X_s} is the correction to the 3D control points (traffic signs) coordinates
- δ_{X_t} is the correction to the 3D tie points coordinates
- σ_g is the standard deviation of 2D measurement for traffic signs
- σ_t is the standard deviation of 2D measurement for tie points(SIFT points)
- r_g is the residual of 2D measurement of traffic sign
- r_t is the residual of the 2D measurement of tie points

The weight matrix for GCP based error equations is $diag(Q_{C_p}^{-1}, \frac{1}{\sigma_t^2} I, Q_s^{-1}, \frac{1}{\sigma_g^2} I)$. The covariance matrix of parameters is estimated as $[J^T PJ]^{-1}$, where J is the jacobian matrix of error equation. The covariance matrix of camera poses is the top-left sub-matrix of $[J^T PJ]^{-1}$ and a block-wise inversion method is employed in this paper to reduce the computation.

V. LANDMARK DETECTION AND MATCHING

Traffic signs are automatically detected with the method proposed in [26]. The category (indication, obligation, prohibition, warning) of traffic signs and their types (semantic meaning of the sign) are determined after automatic detection. The full work flow of traffic sign detection and matching is shown in figure 3.

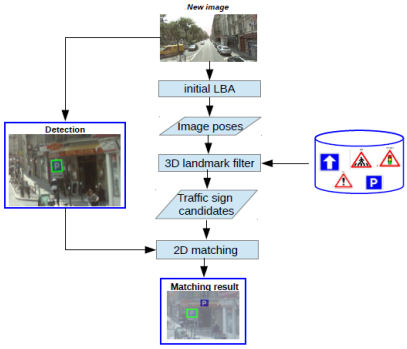


Fig. 3: Work flow of traffic sign matching.

There are two main steps for the matching between detected 2D traffic signs and 3D landmark database. First step is to select the possible 3D landmarks from database which are called as traffic sign candidates. There are two principles for these candidates. 1) the distance between 3D landmarks and new image position is less then the threshold $T1 = 30m$. 2) the 3D landmark should be in the field view of new image, which means the re-projection position of the 3D landmark should in the frame of image. After first step, we have some candidates for each detected traffic sign in new image. The second step is to reject the false matches and keep the right matched traffic signs as the 2D matching in figure3. 1) we classify the candidates according to the category of each new detected traffic sign. If there is no corresponding 3D landmarks for the detected traffic sign, it means there is no valid matching for this detected traffic sign in new image. Otherwise, we need to more principles to check. 2) We reject the false matches based on the re-projection position. We only keep the 3D landmark that the re-projected position is nearest to the detected 2D traffic sign and the nearest distance should be less than $T2 = 50pixel$.

VI. EXPERIMENT

Both simulated and real images are used for experiment. Real images was captured by the optical camera equipped on STEREOPLIS, and the referenced poses were recorded by GPS/IMU system. Our simulated images were generated based on an integrated 3D city model [27] and referenced poses and intrinsic parameters of camera. Compared with real images, the simulated images have two advantages. Firstly, there is no image distortion, we can avoid the impact of camera calibrating error for our experiment. Secondly, the image poses are exact which can be used as actual reference to evaluate the quality of our localization method. A subsample of experiment images are shown in figure 4.

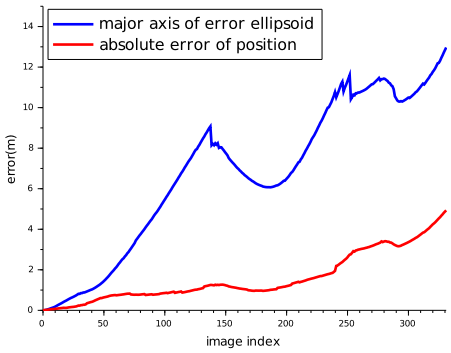
(a) simulated images (b) real images

Fig. 4: The left images are generated according to intrinsic and extrinsic parameters of right real images

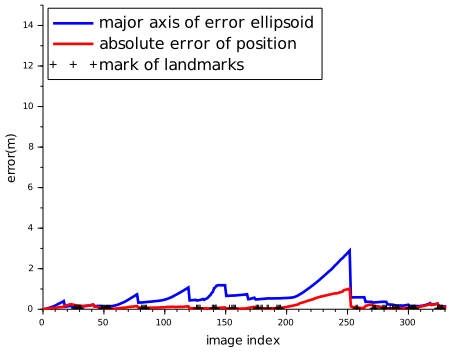
A. Experiment of simulated data

We generate 331 images (f = 10.33mm, image size 1920x1080) in dense urban environment and the whole trajectory is about 1.1km. The initial values for LBA are estimated with the method mentioned in section III and there are 18 3D traffic signs that are successfully matched for the simulated images.

We estimate the error ellipsoid of image position with 99% confidence level for each image and define the absolute error as $\sqrt{(X-X_0)^2+(Y-Y_0)^2+(Z-Z_0)^2}$. (X, Y, Z) are the estimated values while (X_0, Y_0, Z_0) are referenced values. The results of our approach and original none-constraint method are shown in figure 5.



(a) Original LBA



(b) Our method

Fig. 5: Major axis of error ellipsoid of image position and absolute error.

If we integrate the geo-referenced traffic signs with LBA process, both predicated and actual errors are smaller than

none-constraint method. In 5b, represent that the images have constraints from 3D landmarks. When the curves of major axis of error ellipsoid and absolute error meet the road signs (shown by dark crosses on the horizontal axis), both of them decrease immediately. It means that the drift of trajectory reduces and the precision of localization improves.

In this paper, we also discuss the impact of the number of 3D traffic signs for LBA result. Figure 6 depicts the impact of number of the applied road signs in the process. Even if two traffic signs(around image 80 and image 250) are added for LBA, the absolute errors are reduced. However, the absolute errors start to increase between the two traffic signs. So proper density of 3D traffic sign is very important to compensate error accumulation. In this experiment, if we can detect and match one traffic sign every 50-100 meters, we can get satisfying localization result.

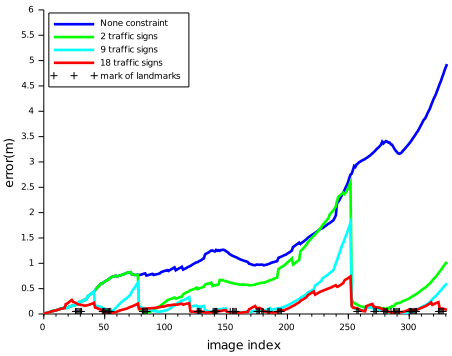


Fig. 6: The absolute error for different number of traffic sign situation.

Figure 7 shows the comparison of our localization result with the reference data. The process order obeys anti-clockwise from starting point in 7. The error ellipsoids of position are also shown in this figure and the center of ellipsoid is the position of image. The size of error ellipsoid reduces immediately when the image contains the constraint of geo-referenced 3D traffic sings.

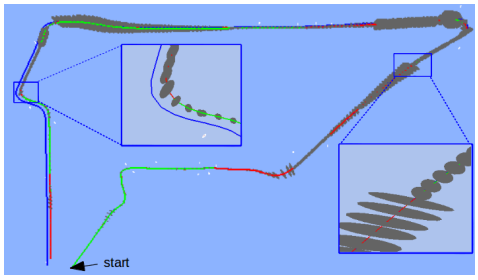


Fig. 7: The green line is the referenced trajectory, the blue shows the none-constraint results, the red line is the results of our method, the highlight points are the 3D traffic signs. The error ellipsoids are exaggerated seven times.

B. Experiment on real data

Real images are captured with the calibrated camera and they have the same size and the focal length with simulated images. The comparison between none-constraint and geo-referenced constrained localization results is shown in figure 8. The absolute error is calculated with the same method mentioned in simulated image experiments. We notice that none-constraint result has about 40m error at the end of trajectory. However, if we take into account the constraints from traffic signs, the result becomes much better. The red line in figure 8 is always under the blue line and the maximal absolute error drops to 5 meters.

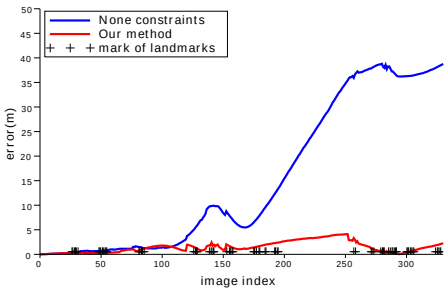


Fig. 8: Absolute error for none-constraint and our method.

We also notice that the absolute errors of real images don't always decrease when geo-referenced constraints are added, such as the red line in figure 8 around image 120. This phenomenon is caused by the noise of referenced poses. Even if precise navigation system are used to capture the referenced image poses, we can't avoid the noise of localization in dense urban environment. The trajectories of reference, none-constraint and our approach are demonstrated in figure 9. It is obvious that our traffic sign based localization method has smaller drift than original LBA based localization method. The geo-referenced traffic signs can improve the precision of localization, which can be seen in figure 9.

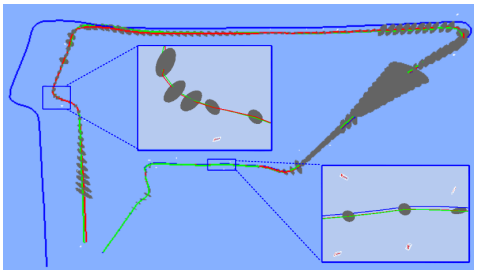


Fig. 9: The red line is our localization result and the blue is the none-constraint result. The green line is trajectory of GPS/INS. The highlight points are geo-referenced traffic signs. The size of error ellipsoids were exaggerated seven times.

VII. Conclusion

A 3D geo-referenced traffic sign based vision localization method was proposed in this paper. We extended the original LBA with the constraints of 3D traffic signs and integrated the traffic sign detection and matching with image based localization process that made it possible to generate the GCPs automatically for bundle adjustment. Both simulated and real images were used for experiments and the localization results

with geo-referenced constraints were more precise than non-constraint method. We reduced the drift of original LBA and the trajectory estimated by our approach was much closer to the reference data than the original method. 331 images were used in our simulated and real experiments, the processing time on LBA was less than 100 seconds and most of the time was spent on feature dection and image matching. This should be improved in the future work to adapt real-time application. We aim at integrating other visual landmarks such as building facades and road markings in the bundle adjustment in our future work.

VIII. Acknowledge

we gratefully thank Emmanuel Habets and Bertrand Cannelle for the discussion and advices about bundle adjustment.

References

[1] J. D.A.Tazartes, "Integration of gps receivers into existing inertial navigation system," *Navigation: Journal of The Institute of Navigation*, vol. 35, no. 1, pp. 105–119, 1988.

[2] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ser. ICCV '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 1403–1410.

[3] J. Castellanos, J. Neira, and J. Tardós, "Limits to the consistency of EKF-based SLAM," in *5th IFAC Symp. on Intelligent Autonomous Vehicles, IAV'04*, Lisbon, Portugal, 2004.

[4] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004.*, vol. 1. IEEE, 2004, pp. I–652–I–659 Vol.1.

[5] S. Rodríguez F, V. Fremont, P. Bonnifait, and V. Cherfaoui, "An embedded multi-modal system for object localization and tracking," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, June 2010, pp. 211–216.

[6] R. Jiang, R. Klette, and S. Wang, "Modeling of unbounded long-range drift in visual odometry," in *Proceedings - 4th Pacific-Rim Symposium on Image and Video Technology, PSIVT 2010*, 2010, pp. 121–126.

[7] A. society for photogrammetry, remote sensing, C. C. Slama, C. Theurer, and S. W. Henriksen, Eds., *Manual of photogrammetry*. Falls Church, Va. American Society of Photogrammetry, 1980.

[8] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3D reconstruction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, vol. 1. IEEE, 2006, pp. 363–370.

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

[10] M. Lhuillier, "Incremental fusion of structure-from-motion and gps using constrained bundle adjustments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2489–2495, 2012.

[11] G. Bresson, R. Aufrere, and R. Chapuis, "Making visual SLAM consistent with geo-referenced landmarks," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 553–558.

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

[13] B. Charmette, E. Royer, and F. Chausse, "Efficient planar features matching for robot localization using gpu," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010*, June 2010, pp. 16–23.

[14] 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*, June 2014, pp. 1345–1350.

[15] W. Maddern, A. Stewart, and P. Newman, "Laps-ii: 6-dof day and night visual localisation with prior 3D structure for autonomous road vehicles," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 330–337.

[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 *IEEE Conference on Computer Vision and Pattern Recognition, 2009.* IEEE, 2009, pp. 2882–2889.

[17] 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 *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012*. IEEE, 2012, pp. 348–355.

[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*, W. Burgard, O. Brock, and C. Stachniss, Eds. The MIT Press, 2007.

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

[21] O. Tournaire, B. Soheilian, and N. Paparoditis, "Towards a sub-decimeter georeferencing of ground-based mobile mapping systems in urban areas : Matching ground-based and aerial-based imagery using roadmarks," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 36 (Part 1)*, Paris, France, 2006.

[22] N. Mattern, R. Schubert, and G. Wanielik, "High-accurate vehicle localization using digital maps and coherency images," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*. IEEE, 2010, pp. 462–469.

[23] M. Schreiber, C. Knoppel, and U. Franke, "Laneloc: Lane marking based localization using highly accurate maps," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 449–454.

[24] L. Wei, B. Soheilian, and V. Gouet-Brunet, "Augmenting vehicle localization accuracy with cameras and 3D road infrastructure database," in *ECCV workshop on Computer Vision in Vehicle Technology, 2014*, Sept 2014.

[25] 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," *Revue française de photogrammétrie et de télédétection*, no. 200, pp. 69–79, 2012.

[26] B. Soheilian, N. Paparoditis, and B. Vallet, "Detection and 3D reconstruction of traffic signs from multiple view color images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 77, pp. 1–20, 2013.

[27] B. Soheilian, O. Tournaire, N. Paparoditis, B. Vallet, and J.-P. Papelard, "Generation of an integrated 3D city model with visual landmarks for autonomous navigation in dense urban areas," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 304–309.

[28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[29] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP'09)*. INSTICC Press, 2009, pp. 331–340.

[30] P. Moulon, P. Monasse, and R. Marlet, "Adaptive structure from motion with a contrario model estimation," in *Computer Vision–ACCV 2012*. Springer, 2013, pp. 257–270.

[31] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 26, no. 6, pp. 756–770, 2004.

[32] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2969–2976.

[33] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.