The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Improved Visual Localization and Navigation Using Proprioceptive Sensors

Nadir Karam, Hicham Hadj-Abdelkader, Clement Deymier and Datta Ramadasan

Abstract— **We address the problem of vehicle (mobile robot) navigation by combining visual-based reconstruction and localization with metrical information given by the proprioceptive sensors such as the odometry sensor. The proposed approach extends the navigation system based on a monocular vision [1] which is able to build a map and localize the vehicle in the real time way using only one camera. An extended kalman filter is used to integrate odometric information to estimate the vehicle position. This position is updated by the localization obtained from the vision system. Experimental result carried out with an urban electric vehicle will show the improvement of the navigation system and its robustness to the temporary loss of images.**

Keywords: robot navigation, urban vehicles, data fusion, visual navigation, odometry, kalman filter.

I. INTRODUCTION

Autonomous vehicle concept has been for few years a topic of great interest in the urban traffic. They appear to be very suitable in specific areas where the public demand is properly structured, as in airport terminals, attraction resorts, university campus, hospitals, or inner-cities pedestrian zones. Such that system should meet the public expectation with a high flexibility to answer to many different individual request. Usually, passengers call a vehicle from any station to reach any other one in an automatic way, and vehicles could park autonomously to the station supply for refilling and reuse. In order to autonomously navigate in an unknown environment, the vehicle must be able to built a representation of the surrounding map and self-localize with respect to it. These information can be provided to the vehicle as a following task of a learned trajectory. Several kinds of sensor are used for outdoor vehicle localization. The most popular of them is the GPS receiver. The Real-Time Kinematic (RTK) GPS allows a consistent accuracy of 1 cm for the navigation applications. However, this accuracy drops considerably in dense urban environment where buildings can mask some satellites. To overcome this limitation, the fusion with other sensors such as odometry can be considered [2]. Vehicle localization can also be supplied by odometry (see [3]) since the odometry sensor is a proprioceptive sensor (often embedded on the vehicle). Nevertheless, this sensor has a major problem of sensitivity to errors due to the integration of velocity measurements over time for position estimates. As a consequence, to efficiently use odometry information it often requires complementary information to enable a correction of the cumulative drift errors. Odometry is usually combined with other exteroceptive sensors using known landmarks as GPS (see [4]), numerical map (see [5]) or visual landmarks (see [6]).
Vision system is well know to provide an accurate localization. Moreover, the use of camera sensor is very attractive since in urban environment such as city centers, there are usually a lot of visual features. We think so that the vision-based localization system could make a good complementary sensor to the odometry.

Two main approaches for visual navigation with respect to a learned trajectories, have been proposed: appearance-based approach (or visual memory-based approach) and map-based approach. The former use only images to represent the reference trajectory. From bounded quantity of images (called key frames) gathered in a database, the vehicle moves from one frame to the next frame using for example visual servoing [7]. The latter needs to build a map of the trajectory and the environment, most often done off-line. The localization process is then fast since the map has been built beforehand. Furthermore, data fusion from different sensors can be considered since the localization is incurred in the global coordinates system. In [8], two cameras and odometry are used to build a map of the environment, which is used to compute a global localization of the robot. In the same approach but opposed to stereo vision, a monocular vision-based map building and localization has been proposed in [1]. Since only one camera is used, the cost and size of the localization system are considerably reduced.

When the vehicle is localized in the map, it will follow the same learned path. To do that, its error deviation to the reference path should be regulated to zero. If we consider the system as proposed in [1], the camera position is obtained in the global camera frame (the frame of the first camera). So as to control the vehicle, its position should be given in the metric coordinates system with respect to a global frame system. Moreover, the vehicle positions provided by the camera are up to a scale factor. It may be set knowing the real length of the path for example. Unfortunately, this scale factor is not all along fix. It can vary along the trajectory.

We show in this paper that the fusion of data given by the odometry and camera sensors overcomes the scale factor problem and improve the navigation system. Indeed, the odometry sensor provide the distance between each key frames to set the local scale factor rather than the global one. Then, the vehicle position provided by the camera will update the estimated position using the odometry information. We show also that this approach will increase the robustness of

This work was supported by the VIPA project

Authors are with the LASMEA laboratory, Blaise Pascal University, 24, Av des Landais 63177 Aubiere Cedex, France {first name.name}@lasmea.univ-bpclermont.fr

the navigation system. Precisely, if for example the camera field of view is masked, the proposed navigation system continue with success since the odometry data is always present.

The remainder of this paper is organized as follows. In section II, the building map process from the reference video sequence and the localization algorithm are recalled. In section III, the localization using data fusion is presented. Section IV is devoted to the control law design. Finally, experimental results are presented in section V.

II. MAP BUILDING AND REAL TIME LOCALIZATION

Real time vision-based localization can be done in two steps. After an off-line building of a 3D map of the learning video sequence, the camera is localized within this map in real-time. In the sequel, we recall briefly these steps. More details can be found in [1].

A. 3D map reconstruction

The Map building process using a single camera can be addressed to the structure from motion problem which has been studied for several years [9]. The reconstruction process is realized from a set of key frames selected from the reference image sequence. The criterion of the selection is based on the motion between key frames which should be the longer while still being able to match the images. This will improve the geometry epipolar computation since it is an ill conditioned problem for a small motion. For each image, interest points are detected with Harris corner detector. The image matching is established by thresholding the correlation score (Zero Normalized Cross Correlation) between the neighborhoods of each couple of detected points within two images. The camera motion computation is done using the triplet of key frames. The first image of the reference sequence is ever selected as the first key frame. The second key frame is selected such as is the farthest image with at least N common interest points with the first key frame. The third key frame is chosen so that is the farthest image with respect to the second key frame. It has at least N common interest points with the second key frame and has at least M common interest points with the first key frame. These conditions are used for the next key frames selection until the end of the image sequence. The camera locations are computed using the hierarchical bundle adjustment as proposed in [1].

B. Real time localization

Before starting the autonomous navigation, the vehicle should be localized over the reconstructed 3D map. To do that, the current image is localized in the set of the key frames by matching interest points between the current image and each key frame. The position of the current camera is then obtained with RANSAC. This step is required only at the start and it takes only few seconds. Note that in order to accelerate this step, one can initialize it by an approximate pose using a low-cost GPS. The current image may be thus localized over a few key frames.

After this step, The vehicle position is updated at each acquired image. This process can be established through the following steps (details can be found in [1]):

- visible landmarks are selected from the closest key frame,
- image points of these landmarks are approximated in the current frame using an estimated pose based on the motion model,
- from these image points, the landmarks are matched to the detected points in the current frame,
- the position of the current frame is computed using a bundle adjustment.

Note that in the second step, the odometry could be incorporated rather than using the motion model to incur a better initial estimate of the camera pose. Even though, this will not change the localization accuracy.

In the next section, we fuse the visual-based localization and the odometry data.

III. LOCALIZATION BASED ON DATA FUSION

Under natural conditions, the vision-based localization could be come up against bad visibility such as obstacle hiding or camera blinding. In order to overcome these situations and keep the vehicle well localized, a fusion-based localization approach is presented. It consists in fusing information provided by proprioceptive and exteroceptive sensors using an Extended Kalman filter ([10], [11]). It can be thought of as operating in two distinct phases: prediction and update. In the prediction phase, the old state is used to generate the current state of the vehicle with the tricycle model. Next, in the update phase, the current observation is used to correct the predicted state for accuracy purpose.

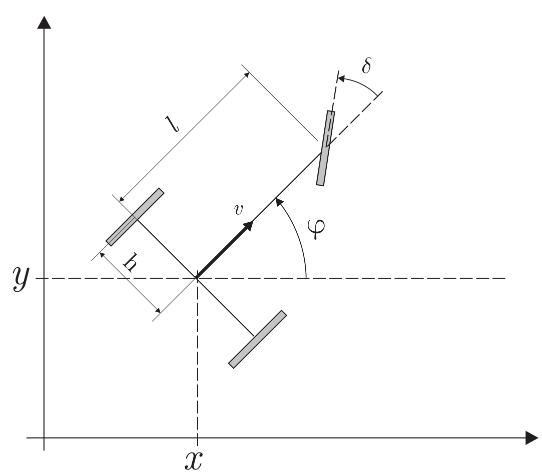

Fig. 1. Tricycle model of the vehicle

A. state prediction and evolution

Let x and y the coordinates of the vehicle, υ its speed and φ its orientation (see Figure 1). The vehicle can be described by the state vector $\mathbf{x} = [x\ y\ \upsilon\ \varphi]^\top$ and the covariance matrix $\mathbf{P}$. The motion of the vehicles can be modeled by the function f using the tricycle model. We assume that the error affecting the encoders data vector $\mathbf{u}$ can be modeled by a normal distribution as:

$$\tilde{\mathbf{u}} \sim \mathcal{N}(\mathbf{u}, \mathbf{Q}) \tag{1}$$

where $\mathbf{u}$ is the measured data and $\mathbf{Q}$ the covariance matrix of the noise affecting it. The evolution equation of the state vector $\mathbf{x}$ and its covariance matrix $\mathbf{P}$ can be written as following:

$$\hat{\mathbf{x}}_{k+1|k} = f(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k) \tag{2}$$

$$\mathbf{P}_{k+1|k} = \mathbf{J_x}\mathbf{P}_{k|k}\mathbf{J_x}^\top + \mathbf{J_u}\mathbf{Q}_k\mathbf{J_u}^\top + \mathbf{B} \tag{3}$$

where k is the time, $\hat{\mathbf{x}}_{k+1}$ is the predicted $\hat{\mathbf{x}}_k$, $\mathbf{J_x}$ and $\mathbf{J_u}$ are respectively the jacobian matrix of the function f with respect to the state $\mathbf{x}$ and $\mathbf{u}$, and $\mathbf{B}$ is the covariance of the noise affecting the motion model of the vehicle. The predicted state $\hat{\mathbf{x}}_{k+1}$ at the time $k+1$ will be updated with the information provided by the vision system.

B. state updating

The vehicle update its state vector $\mathbf{x}$ using information received from the vision system and represented by the vector $\mathbf{z}$. We assume that these information are noised according to the normal distribution model:

$$\tilde{\mathbf{z}} \sim \mathcal{N}(\mathbf{z}, \mathbf{B_z})$$

with $\mathbf{z} = [x\ y\ \varphi]^\top$ is the pose of the vehicle and $\mathbf{B_z}$ is the covariance matrix of noise affecting it. The state vector $\mathbf{x}$ is updated with the observation $\mathbf{z}$ as:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1}\mathbf{H}^\top(\mathbf{H}\mathbf{P}_{k+1}\mathbf{H}^\top + \mathbf{B_z})^{-1} \tag{4}$$

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \mathbf{H}\mathbf{x}_{k+1|k}) \tag{5}$$

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H})\mathbf{P}_{k+1|k} \tag{6}$$

with the observation matrix $\mathbf{H}$ given as:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Typically, the predict and update phases alternate. However, if observations are unavailable for some reason during a short time, the update may be skipped and multiple prediction steps performed. In our approach, we exploit this idea to keep the vehicle localized in lacking of visual observation. Certainly, the vision system may meet some problems to provide the position of the vehicle in some situation such as hiding the camera field of view with obstacles or blinding the camera because of sunrays.

Likewise, the observations may be affected with delays and thus penalize the estimation accuracy. This problem can be solved by dating the observations in an absolute time reference. Indeed, the observations are affected to their true arrival date and put back in chronological order in an observations table. Then, the state is updated with the new observations.

An example is shown in figure 2. The state of the vehicle at t_2 is computed with the observation arrived at that instant and the previous state at t_1. Let us suppose that a camera image arrived at t_3 but the image processing which provides the pose estimation of the vehicle takes time and the result is given at t_5. During this time, other observations can arrive. Those observations can be used to compute the current state using the previous one. When the pose estimation is provided by the image processing (at t_5), the observation is replaced at the right time (t_3). The state at t_5 is recomputed by updating the state at t_2 with the observations at t_3 and t_4.

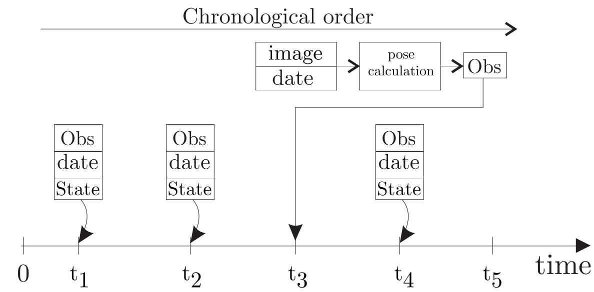

Fig. 2. Observations scheduling

In the next section, a control law is established to follow the reference path using the accurate localization of the vehicle.

IV. CONTROL LAW

A. Vehicle modeling

In order to design a control law, the vehicle must be modeled. We consider a non-holonomic system with classical kinematics since the vehicle is supposed to move on an urban horizontal ground under the conditions of pure rolling and non-slipping at low speed. The well known tricycle model is used. Using the estimated state vector $\hat{\mathbf{x}}$ and the reconstructed reference path, the new state vector of the vehicle may be described with respect to the path by $\begin{bmatrix} s & \tilde{y} & \tilde{\theta} \end{bmatrix}^\top$ where:

- s is the curvilinear coordinates of the vehicle rear axle center along the path,
- $\tilde{y}$ and $\tilde{\theta}$ are respectively the lateral and angular deviations of the vehicle with respect to the path,

The control vector is constituted in the vehicle linear velocity υ and the front wheel steering angle δ. The state and control vectors are related by the following kinematics equations:

$$\left\{\begin{array}{lll}\dot{s} & = & v\,\dfrac{\cos\tilde{\theta}}{1-\tilde{y}\,c(s)} \\ \dot{\tilde{y}} & = & v\,\sin\tilde{\theta} \\ \dot{\tilde{\theta}} & = & v\,\left(\dfrac{\tan\delta}{l}-\dfrac{c(s)\,\cos\tilde{\theta}}{1-\tilde{y}\,c(s)}\right)\end{array}\right. \tag{7}$$

where:

- $c(s)$ is the curvature of the path at s,
- l is the vehicle wheelbase.

Note that this model has two singularities for $v = 0$ and $y = \dfrac{1}{c(s)}$ (this occurs if the vehicle is on the path curvature center). These situations never arise in experiment since the vehicle is assumed to be always running and remains close to reference path.

B. Control law design

The control objective is to bring and maintain the lateral deviation $\tilde{y}$ and angular deviation $\tilde{\theta}$ to 0. The vehicle model (7) is clearly nonlinear. However, it has been established in [12] that mobile robot models can generally be converted in an exact way into almost linear models, named chained forms. This property offers two very attractive features: on one hand, path following control law can be designed and tuned according to celebrated Linear System Theory, while controlling nevertheless the actual non linear vehicle model. Control law convergence and performances are then guaranteed whatever the vehicle initial configuration is. On the other hand, chained form enables to specify, in a very natural way, control law in term of distance covered by the vehicle, rather than in term of time. Vehicle trajectories can then easily be controlled using the expression of the control variable δ given as (details are given in [13]):

$$\delta = \arctan\left(l\left[A\frac{\cos^3\tilde{\theta}}{\alpha^2}+\frac{c(s)\cos\theta}{\alpha}\right]\right) \tag{8}$$

with:

$$\left\{\begin{array}{lll}\alpha & = & 1-c(s)\tilde{y} \\ A & = & \dfrac{d\,c(s)}{d\,s}\tilde{y}\tan\tilde{\theta}-K_d(1-c(s)\tilde{y})\tan\tilde{\theta}-K_p\tilde{y} \\ & & +c(s)(1-c(s)\tilde{y})\tan^2\tilde{\theta}\end{array}\right.$$

V. EXPERIMENTAL RESULTS

The proposed approach was implemented and tested on the real electric vehicle called Robucab (see Figure 3). This vehicle is manufactured by the society Robosoft. Its dimensions are 1.50 m of length and 1.20 m of width and its velocity is limited to 5 m/s. Its front and rear wheels can be steered. In our experiments, only front wheels are steered. The vehicle is equipped with an odometric sensor at the wheels. A camera providing 512×384 pixels grayscale images is embedded on the vehicle. The rigid transformation between the camera frame and the vehicle frame (or the control frame) is calibrated manually. For the accuracy checking purpose, the accurate positions (2 cm of accuracy) provided by the RTK-GPS are used as the ground truth.

Fig. 3. Experimental vehicle: Robucab

Two computers are embedded in the vehicle. The first one called low-level deals with the control variables (velocity v and steering angle δ). Proprioceptive sensors are connected to the low-level computer which sends measurements to the second computer called high-level through an ethernet link. The vehicle localization algorithms (vision based and fusion) are implemented on the high-level computer in C++ language using the library AROCCAM presented in [14].

In order to investigate our approach, the experiments are realized over daylight and overnight. In the both lighting conditions, a reference video sequence was recorded and a 3D map was reconstructed. Since the 3D map reconstruction algorithm is based only on vision, the coordinates system is defined by the first camera frame in the reference sequence. Furthermore, the reconstruction is up to an unknown scale factor. However, the position of the vehicle should be provided in the metric coordinate system to be able to control the vehicle. In [1], a global scale factor is obtained using the length of the path. Nevertheless, the scale factor varies differently depending on the vehicle displacement if it goes straight or if it turns. To date and to our knowledge, this variation is not proved but only observed. In our case, the scale factor was set between each two key-frames using the odometric information. Figures 4 shows clearly that the trajectory obtained using our scale factor setting method is better than the one using a global scale factor comparing to the trajectory provided by the RTK-GPS. The vehicle is then localized in real time in autonomous navigation over the 3D map using our approach. The videos of the presented experiments can be found as multimedia material submitted with this paper.

A. Experiment 1

For the first experiment, the reference video sequence was recorded over daylight, preferably on a cloudy day. Indeed, the sun can be sometimes in the field of view of the camera on a clear day which penalizes the quality of the 3D map reconstruction. But during the navigation, the

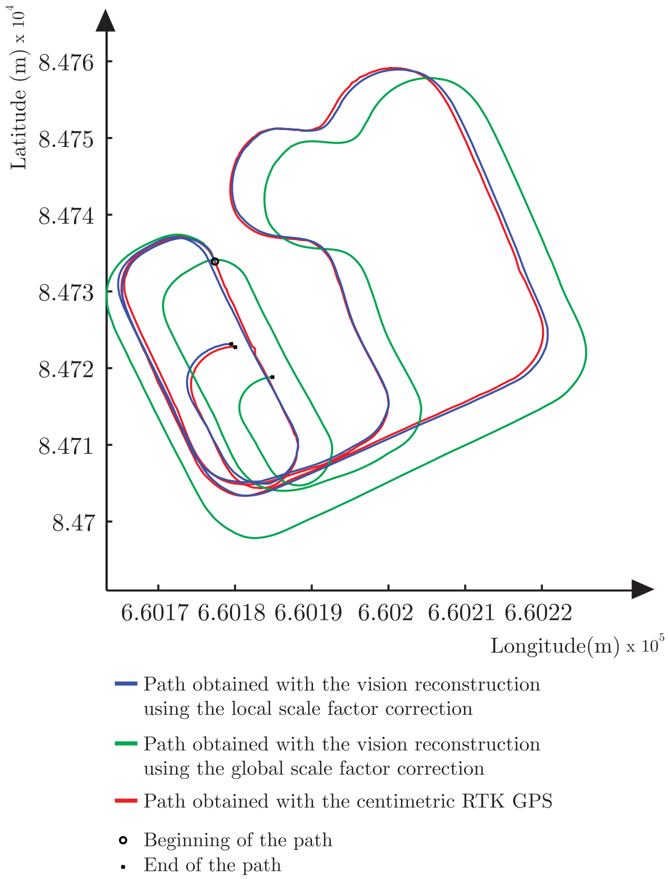

Fig. 4. Reference trajectories obtained by the GTK-GPS and vision with local scale factor correction

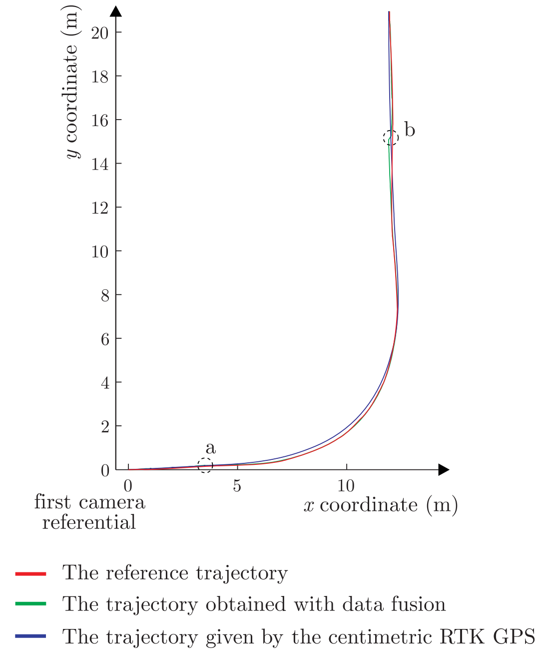

Fig. 5. Reference, ground truth and realized trajectories

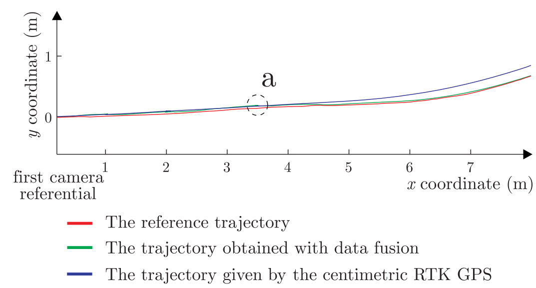

Fig. 6. A local view around the points a

vehicle could be localized under sun blinding or obstacle hiding conditions thanks to the fusion algorithm. To show clearly the robustness of the navigation system under these conditions, we hind totaly the camera field of view along few meters during the autonomous navigation. Figure 5 shows the reference path (drawn in red line), the realized path by the vehicle (drawn in green line) and the trajectory given by the RTK-GPS as the ground truth. In this experiment, the camera is hidden at the position indicated by the point a. The vehicle is navigating from the point a to the point b (thereabouts 21 meters) using only the odometry measurements. Actually, only prediction steps are performed until the point b where the vehicle state is updated since visual observation is available. Noting that when the camera is unmasked, the vehicle is deftly localized since its position is very close to the predicted one through the odometry data. Note that during the camera hiding, the guidance performances are slightly damaged, but are still satisfactory (see Figure 6). Indeed, the lateral deviation estimated using odometric data is well regulated to zero by the control law (8). However, this estimated lateral deviation is different to the real distance of the vehicle with respect to the reference path. Therefore, the vehicle does not lie on the reference path because of the drift occurred by the odometric data. One can observe in Figure 7 that after the vehicle position updating (point b) the real trajectory of the vehicle given by the RTK-GPS comes closer to the reference trajectory. This will be carry out in a few meters depending on the gains K_d and K_p of the control law (about 5 meters in our case).

B. Experiment 2

This experiment is realized overnight. For that, the reference video sequence was recorded in the same condition (overnight). The headlights of the vehicle are used during the map reconstruction and the autonomous navigation for a reasonable visibility. The frame rate is reduced to 1 fps (rather than 15 fps over daylight) since the camera keep the shutter open longer for the successful night video acquisition. The proposed approach can handled these conditions to carry out successfully the navigation system. Indeed, the low frame rate can be viewed as the camera masking during a short time. Example of images acquired overnight and detected features for localization is shown in Figure 8.

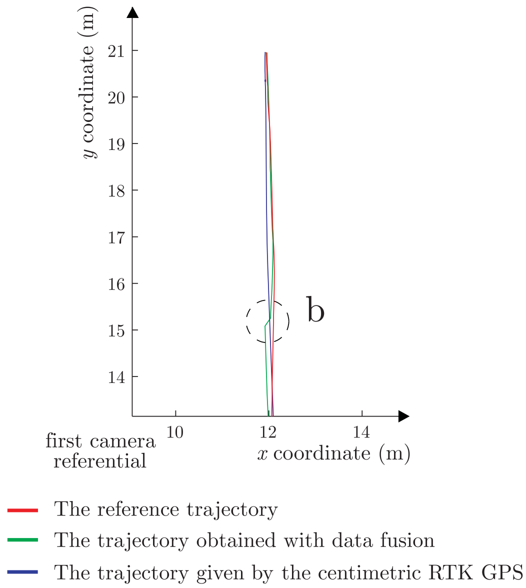

Fig. 7. A local view around the points b

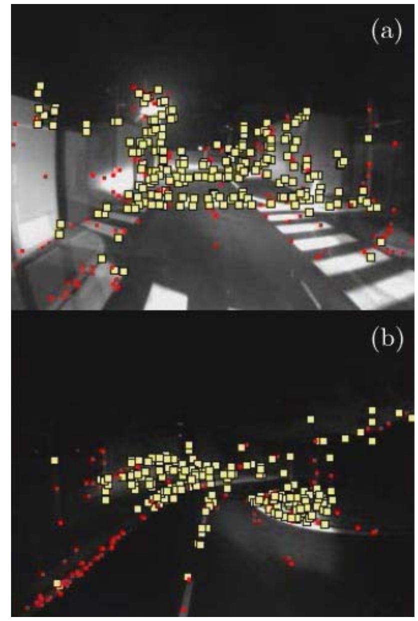

Fig. 8. Example of images acquired overnight with and without streetlights in (a) and (b) respectively. The yellow and the red points are respectively the inlier and the outlier points of the RANSAC

VI. CONCLUSION

Robustness of the autonomous vehicle with respect to the visibility constraints is extremely important for human transport. To improve this robustness, odometric information can be exploited. In this paper, we have addressed the problem of vehicle navigation by fusing information delivered by the odometric sensor (proprioceptive) and the camera sensor (exteroceptive). We have detailed the fusing process using an extended kalman filter where data were integrated at their true acquisition date. The proposed approach can deftly integrate information from various proprioceptive sensors (for example the gyroscope, the accelerometer, etc). We have validate our approach with an urban electric vehicle. Experimental results show clearly the robustness of the proposed navigation system to the temporary camera hiding and overnight navigation. Future work will be devoted to study a vision-based reconstruction and localization using two cameras locking forward and behind the vehicle. This configuration will be advisable to overcome several problems such as long-time sun blinding.

VII. ACKNOWLEDGMENTS

The authors wish to thank Eric Royer from LASMEA laboratory, France who helps us during experiments.

REFERENCES

[1] Eric Royer, Maxime Lhuillier, Michel Dhome, and Jean-Marc Lavest. Monocular vision for mobile robot localization and autonomous navigation. *Int. J. Comput. Vision*, 74(3):237–260, 2007.

[2] Denis Bouvet, Michel Froumentin, and Gatan Garcia. A real-time localization system for compactors. *Automation in Construction*, 10(4):417 – 428, 2001.

[3] Richard Thrapp, Christian Westbrook, and Devika Subramanian. Robust localization algorithms for an autonomous campus tour guide. In *ICRA*, pages 2065–2071, 2001.

[4] Mayhew, David Mcneil Mayhew, and David Mcneil Mayhew. Multi-rate sensor fusion for gps navigation . . . Technical report, Master thesis, Virginia Polytechnic Institute, 1999.

[5] J. Laneurit, R. Chapuis, and F. Chausse. Accurate Vehicle Positioning on a Numerical Map. *International Journal of Control, Automation, and Systems*, 3(1):15–31, 2005.

[6] Amit Adam, Student Member, Ehud Rivlin, and Hctor Rotstein. Fusion of fixation and odometry for vehicle navigation. In *In Proceedings of International Conference on Robotics and Automation*, pages 1638–1643, 1999.

[7] Jonathan Courbon, Youcef Mezouar, and Philippe Martinet. Autonomous navigation of vehicles from a visual memory using a generic camera model. *Trans. Intell. Transport. Sys.*, 10(3):392–402, 2009.

[8] Kiyosumi Kidono Jun, Jun Miura, and Yoshiaki Shirai. Autonomous visual navigation of a mobile robot using a human-guided experience. *Robotics and Autonomous Systems*, 40(2-3):124–132, 2002.

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[10] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Journal of Basic Engineering*, 82:34–45, 1960.

[11] Joseph J. LaViola and Jr. A comparison of unscented and extended kalman filtering for estimating quaternion motion. 2003.

[12] C. Samson. Control of chained system. application to path following and time-varying stabilization of mobile robot. *IEEE Transactions on Automatic Control*, 40(1):64–77, 1995.

[13] B. Thuilot, J. Bom, F. Marmoiton, and P. Martinet. Accurate automatic guidance of an urban vehicle relying on a kinematic gps sensor. In *5th Symposium on Intelligent Autonomous Vehicles*, 2004.

[14] C. Tessier, C. Cariou, C. Debain, R. Chapuis, F. Chausse, and C. Rousset. A realtime, multi-sensor architecture for fusion of delayed observations : Application to vehicle localisation. In *9th International IEEE Conference on Intelligent Transportation Systems*, page On DVDROM, 2006.