Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

Localisation in urban environment using GPS and INS aided by monocular vision system and 3D geographical model

Cindy Cappelle, Maan El Badaoui El Najjar, Denis Pomorski
LAGIS - CNRS UMR 8146
USTL, Polytech'Lille, Avenue Paul Langevin
59655 Villeneuve d'Ascq Cedex, France
Cindy.Cappelle@polytech-lille.fr
Maan.E-elnajjar@univ-lille1.fr
Denis.Pomorski@univ-lille1.fr

François Charpillet
LORIA - INRIA Lorraine - MAIA Team
Campus Scientifique, BP 239
54506 Vandoeuvre-lès-Nancy, France
francois.charpillet@loria.fr

Abstract— **In this paper, a geo-localisation method was proposed, using GPS, INS, monovision camera and a new geo-information source, which is the 3D cartographical model. A 3D-GIS (Geographical Information System) has been developed to manipulate and navigate in a precise 3D cartographical model database. To have a continuous pose estimation, an EKF has been implemented to fuse GPS and INS. To integrate the 3D cartographical model information, a 3D cartographical observation has been constructed using 2D/3D images matching between the real image captured by the embedded camera and the virtual images provided by the 3D-GIS. A real data acquisition platform has been developed to test and validate the proposed method.**

I. INTRODUCTION

Autonomous navigation vehicles usually used two types of sensors : the absolute sensors and the dead-reckoning sensors. The following table presents a comparison of these two kinds.

Absolute sensors	Dead-reckoning sensors
Absolute localisation	Relative localisation
Acquire information from the exterior of the vehicle	Acquire information from the vehicle itself
GPS, Cellular phone, Laser Telemeter ...	INS (Gyrometer and Accelerometer), Odometer ...
Low acquisition frequency	High acquisition frequency
Average accuracy in short term	Good accuracy in short term
Good accuracy in long term	Bad accuracy in long term : High position drift

TABLE I
ABSOLUTE AND DEAD-RECKONING SENSORS

These kinds of sensors are complementary. So, data fusion methods are often used to combine measurements provided by different sensors. An intuitive solution is the use of an absolute sensor to correct the data provided by a dead-reckoning sensor.

Among these sensors, we used in the paper the GPS (Global Positioning System) and the INS (Inertial Navigation System), which are two popular sensors for navigation.

An INS is composed of three accelerometers and three gyrometers which measure the angular velocities and the accelerations of the vehicle along three orthogonal axes.

Low-cost INS are constructed in a strapdown configuration, that is to say the gyroscopes and accelerometers are fixed on a common platform (unlike the gimballed INS which align themselves). The angular rotation rate is used to convert the acceleration in an appropriate frame. Then by integration, we can obtain the position of the vehicle. But strapdown INS presents its drawbacks, for example, its high drift with time.

The GPS is usually the external sensors used to correct INS. GPS satellites are equally spaced on six orbits in such a manner that, an user can see at least four satellites (without consideration of surrounding structures). Each satellite transmits navigation and range data simultaneous on two frequencies : L1 (1575,42MHz) and L2 (1227,60MHz) [4].

In this work, we fuse GPS and INS with an Extended Kalman filter. The GPS signal quality can be damaged by several phenomena like satellites masks and wave multipathes in urban environment, tunnels. So the availability of GPS measurements is not always guaranteed. For this reason, another information source is needed to correct the INS drift. We introduce in this paper a new source of information which exploits a 3D geo-referenced model of the environment, managed with a 3D-GIS (Geographic Information System) and a monocular vision system. This information can make up for the non-availability of the GPS measurements.

The idea is to compute an observation of the geo-position by matching two images, one provided by the virtual camera of the 3D-GIS managing the 3D cartographical model and the other image is provided by the embedded camera in the vehicle.

The principle of the localisation method, which we propose in this paper, is described in Figure 1.

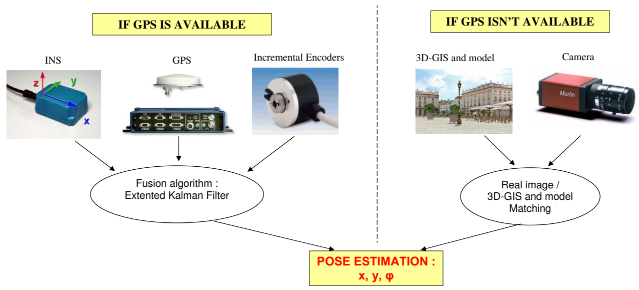

Fig. 1. Localisation method

In the second part of this article, we present INS/GPS data fusion with EKF. Then, in the third part, we explain how we geo-localise the vehicle with the 3D-GIS and the camera. The results of our experiments are detailed on the fourth part. Finally, perspectives and conclusion are presented in the section five.

II. GPS/INS EXTENDED KALMAN FILTERING

The structure of the localisation system proceeds directly from the EKF equations. At the k-th step, the kinematic model of the vehicle is used to compute an inertial prediction $\hat{x}_{k/k-1}$ and an associated covariance matrix $P_{k/k-1}$. An observation prediction $\hat{z}_k$ is formed and compared with the measure z_k provided by the GPS. The results are the innovation term ν_k and its covariance matrix S_k, that are used by the EKF to produce the state estimate $\hat{x}_k$ and the new associated covariance P_k [5]. To fuse INS with GPS, several conversions are needed to pass from INS magnetic terrestrial coordinates system to GPS geographical coordinates system (WGS84). To realize these conversions, we deal with these coordinates systems :

- The INS body frame (b-frame), noted (X_b, Y_b, Z_b) is rigidly attached to the vehicle. It's an orthogonal axes set which is aligned with the roll (forward), pitch (left) and heading (up) axes of the vehicle (Figure 2).
- The navigation frame (n-frame), noted (E, N, U) is defined in the local tangent plan. In this work, the used convention is : ENU (East, North, Up). (Figures 2, 3).
- The inertial frame (i -frame), noted (X_i, Y_i, Z_i) has its origin at the centre of the Earth. Its axes are non-rotating with respect to the fixed stars. Z_i axis is parallel to the spin axis of the Earth, X_i axis points towards the mean vernal equinox, and Y_i axis completes a right-handed orthogonal frame (Figure 3).
- The Earth frame (e-frame), noted (X_e, Y_e, Z_e) has its origin at the centre of mass of the Earth and axes which are fixed with respect to the Earth. Its X_e axis points towards the mean meridian of Greenwich, Z_e axis is parallel to the mean spin axis of the Earth, and Y_e axis completes a right-handed orthogonal frame (Figure 3).

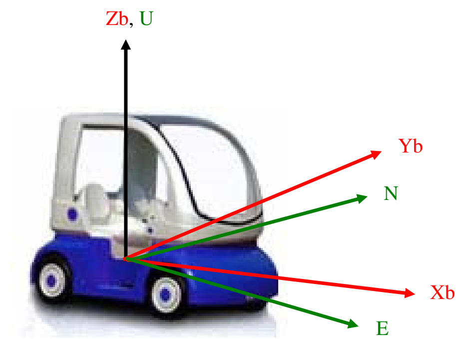

Fig. 2. b-frame and n-frame

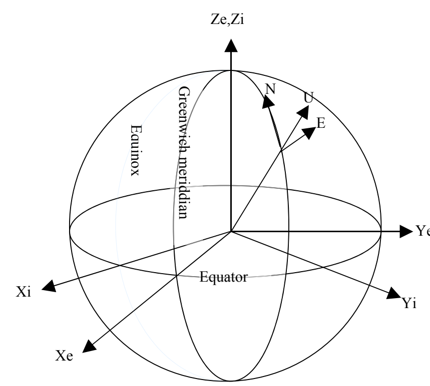

Fig. 3. e-frame, i-frame and n-frame

A. Kinematic vehicle model

The used INS sensor MT9B acquires the 3D accelerometers, 3D gyroscopes, 3D magnetometers and temperature sensors. Using these measurements, an algorithm implemented in the MT9 can provide the Euler angles (Roll, Pitch, Yaw).

Let consider a kinematic vehicle model defined by the state vector :

$$x_k = (e_k, n_k, \phi_k, v_{x_k}, v_{y_k}, b_{x_k}, b_{y_k})^T$$

where $[e, n]^T$, ϕ and $[v_x, v_y]^T$ represent respectively the position on the East and North axes (n-frame), the orientation (i.e. the angle between the North direction (n-frame) and the X_b axis of the vehicle (b-frame) clockwise) and the velocity of the vehicle in the b-frame. $[b_x, b_y]^T$ denotes the accelerometer biases.

The kinematic model is an inertial model which has like inputs :

$$u_k = (a_{x_k}, a_{y_k}, \phi_k)^T$$

where $[a_x, a_y]^T$ is the acceleration on the X_b axis (motion direction) and Y_b axis of the vehicle (b-frame) measured by the INS and ϕ the orientation calculated by the INS.

The inertial model prediction is :

$$\hat{x}_{k/k-1} = \begin{bmatrix} 1 & 0 & 0 & -\Delta t_k \sin\phi_k & -\Delta t_k \cos\phi_k & (\frac{\Delta t_k{}^2}{2})\sin\phi_k & (\frac{\Delta t_k{}^2}{2})\cos\phi_k \\ 0 & 1 & 0 & \Delta t_k \cos\phi_k & -\Delta t_k \sin\phi_k & -(\frac{\Delta t_k{}^2}{2})\cos\phi_k & (\frac{\Delta t_k{}^2}{2})\sin\phi_k \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t_k \sin\phi_k & \Delta t_k \cos\phi_k \\ 0 & 0 & 0 & 0 & 1 & -\Delta t_k \cos\phi_k & \Delta t_k \sin\phi_k \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \hat{x}_{k-1}$$

$$+ \begin{bmatrix} -(\frac{\Delta t_k{}^2}{2})\sin\phi_k & -(\frac{\Delta t_k{}^2}{2})\cos\phi_k & 0 \\ (\frac{\Delta t_k{}^2}{2})\cos\phi_k & -(\frac{\Delta t_k{}^2}{2})\sin\phi_k & 0 \\ 0 & 0 & 1 \\ \Delta t_k & 0 & 0 \\ 0 & \Delta t_k & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} u_k$$

$$= f(\hat{x}_{k-1}, u_k, \Delta t_k)$$

B. Measurement model

- The measurement model is defined as :

$$\hat{z}_k = H\hat{x}_{k/k-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \hat{x}_{k/k-1}$$

- The innovation is computed as :

$$\nu_k = z_k - \hat{z}_k$$

  with z_k : data measured by GPS.
- The covariance matrix of the innovation is :

$$S_k = J_x^h(\hat{x}_{k/k-1})P_{k-1}(J_x^h(\hat{x}_{k/k-1}))^T + R_k$$

  with J_x^h : Jacobian matrix of h with respect to $x_{k/k-1}$ and R_k : measurement noise matrix.

C. Extended Kalman Filter (EKF)

The EKF is a predictor-corrector mechanism [1], [2], [3]. The formulas are :

- Prediction :

$$\hat{x}_{k/k-1} = f(\hat{x}_{k-1}, u_k, \Delta t_k)$$

- The covariance matrix associated to the prediction is :

$$P_{k/k-1} = J_x^f(\hat{x}_{k-1})P_{k-1}(J_x^f(\hat{x}_{k-1}))^T + J_u^f(u_k)C(J_u^f(u_k))^T + Q$$

  - $J_x^f(.)$ et $J_u^f(.)$: Jacobian matrices of $f(.)$ with respect to $\hat{x}_{k-1}$ and u_k,
  - P_{k-1} : covariance matrix at time instant t_{k-1},
  - $C = diag\{\sigma_{ax}^2, \sigma_{ay}^2, \sigma_{\phi}^2,\}$: covariance matrix of the gaussian white noise which corrupts the inputs measures,
  - $Q = diag\{\sigma_e^2, \sigma_n^2, \sigma_\phi^2, \sigma_{v_x}^2, \sigma_{v_y}^2, \sigma_{b_x}^2, \sigma_{b_y}^2\}$: covariance matrix of the gaussian white noise which directly affects the state in the kinematics model.
- Kalman Gain :

$$K_k = P_{k/k-1}(J_x^h(\hat{x}_{k/k-1}))^T S_k^{-1}$$

- Correction :

$$\hat{x}_k = \hat{x}_{k/k-1} + K_k[z_k - \hat{z}_k]$$

- Covariance associated to $\hat{x}_k$:

$$P_k = P_{k/k-1} - K_k S_k K_k^T$$

Subsequently, we consider the pose as : (e, n, ϕ)

III. GEO-LOCALISATION BY VISION AND 3D-GIS

A. 3D-GIS presentation

A geographic information system (GIS) is a computer system capable of integrating, storing, editing, analyzing, sharing, and displaying geographically-referenced information. In a more generic sense, GIS is a tool that allows both to create interactive queries (user searches) and to analyze the spatial information, and edit data.

The past 30 years have brought many new technologies to developing GIS related software's [7]. A traditional GIS provides only 2D representation of the spatial entities using simple primitives of points, lines and polygons. In early-1980's, 2D-based visualization and analysis technologies for mapping were introduced. From the mid-1980's to the present time, many technologies related to 3D-GIS implement different kinds of information technologies such as 3D terrain visualization and analysis, virtual city, virtual GIS, 3D-GIS, multimedia GIS and so on.

Actually, 3D geographical models are generated automatically using laser profiler data, 2D digital map and aerial images. All major Japanese cities have been covered since 2002 and are updated every six months. Outdoors mobile robot systems are a typical example of such an application that make use of the real-world data organised in a GIS [6], [8], [9], [10].

The used 3D-GIS : the developed method uses an accurate 3D geographic model managed with a 3D-GIS. The used 3D model (Figure 4), provided by the society Tecnomade[1], has a metric accuracy.

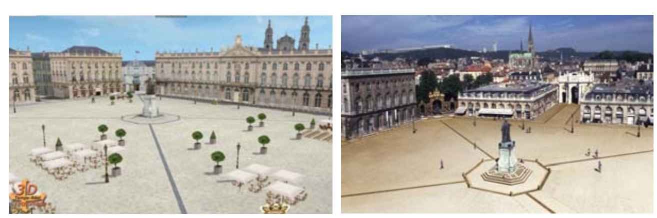

Fig. 4. A 3D image of the Stanislas Square in Nancy

We used the Tecnomade 3D engine in order to develop our 3D-GIS, adapted for our application in robotics and

[1]http://www.tecnomade.fr/

intelligent vehicle. The Tecnomade 3D engine is adapted for real-time applications using the Tecnomade 3D geographical model format.

Several functionalities have been added and integrated to the 3D-GIS (see Fig. 5):

- Virtual images extraction in Bitmap format, by giving to the 3D engine of the GIS, the 6 degrees of liberty in the local coordinates frame.
- Depth image extraction of each virtual image in Bitmap format, which is obtained by the ZBuffer of the video card.
- Extraction of the depth of virtual image pixels in a binary text file.
- Extraction in a XML file of the visible 3D segments which shape the virtual image.

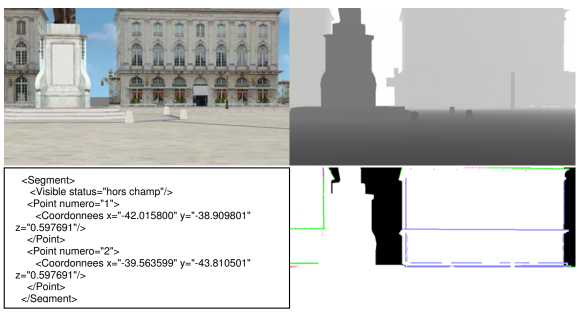

Fig. 5. Illustration of the 3D-GIS output functions

The Tecnomade 3D model is designed in a local coordinates frame. We developed therefore an automatically georeferencing method, which transforms the coordinates from the local system to the French geographical system Lambert 93. Then, we transform the GPS coordinates from the WGS84 system to Lambert 93 system. So, we can navigate in real-time in the 3D geographic model with the GPS geo-position.

B. The 3D cartographical observation

When a GPS position is available, a correction of the INS estimation is performed using an Extended Kalman Filter (EKF). A GPS position is considered available when the GPS quality index is greater or equal than two ($Q_{gps} \geq 2$), that is to say, the GPS operates in the DGPS or LRK mode, respectively metric or centimetric accuracy (LRK is an improved RTK mode specific to the Thales Navigation Sagitta02, which converges faster than RTK mode). If the GPS measurements error is more than 3 meters or not available ($Q_{gps} < 2$), the evolution model provides a dead-reckoned pose. This rough pose is used to compute the 3D cartographical observation.

The 2D/3D matching stage in Figure 6 has 3 inputs at each instant k :

- the predicted pose X_k,
- a real image at the same instant k,
- a set of virtual images in the neighbourhood of X_k.

To create the set of virtual images at the instant k, a spatial discretisation in e, n and ϕ of the position uncertainty zone,

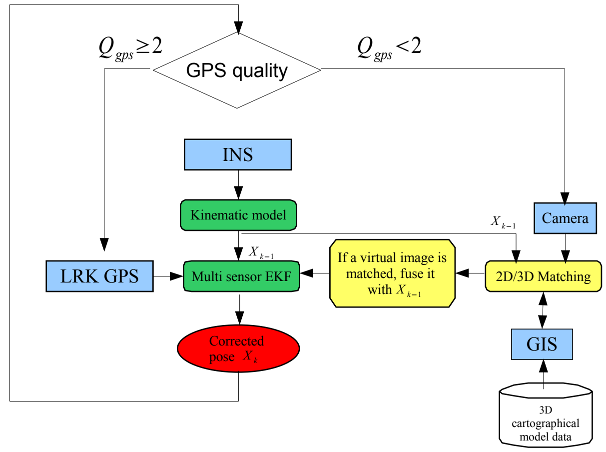

Fig. 6. Geo-localisation method

given by the EKF prediction variance-covariance matrix, is realised. This obtained set of poses is given like an input to the 3D-GIS, which provides the desired set of virtual images.

The following task is to find the most likely virtual image which corresponds the real image captured at the instant k.

Denoted R the real image, V the virtual image, the Zero mean Normalized Cross Correlation coefficient is then :

$$ZNCC(R,V) = \frac{\sum_i \sum_j (R(i,j) - \overline{R})(V(i,j) - \overline{V})}{\sqrt{\left[\sum_i \sum_j (R(i,j) - \overline{R})^2\right]\left[\sum_i \sum_j (V(i,j) - \overline{V})^2\right]}}$$

with $\overline{R}$ (resp. $\overline{V}$) the mean of the grey level of the real image R (resp. virtual image V). We select then the virtual image which gives the maximum correlation.

A way to fuse the geo-information related to the 2D/3D matching with the others sensors is to treat it like an observation.

We can illustrate the presented 2D/3D matching approach with an example.

At instant k, the real image is given in Figure 7. The virtual image of the same figure corresponds to the snapshot from the pose estimated with EKF after an INS drift due to a GPS non-availability. One can remark there is a deviation between the virtual image and the real captured one.

Fig. 7. Real image (left) and initial virtual image (right)

The extracted set of virtual images in the neighbourhood of the estimated pose is given in Figure 8. For reason of implementation simplicity, we modelled the neighbourhood of the pose estimation with a rectangle form which includes the uncertainty ellipse of semi-major axis equal to 3 meters.

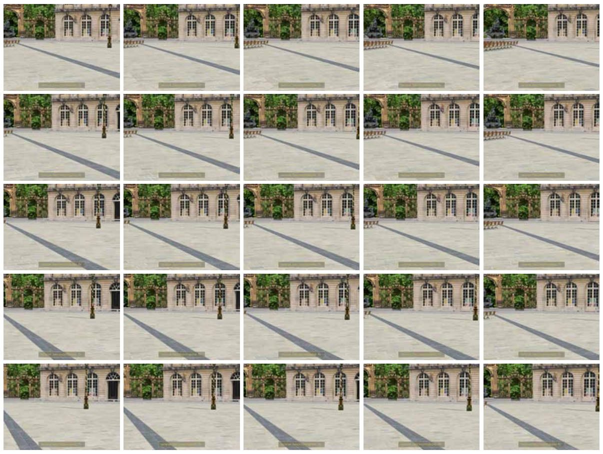

Fig. 8. A sample of virtual images database for a point at instant k

After using the proposed 2D/3D matching approach, one can see in the Figure 9 that the matched virtual image is more corresponding to the real captured image.

Fig. 9. Real image (left) and associated virtual image (right)

Finally, the geo-information related to the obtained virtual image is used to correct the drift of the INS using EKF.

IV. EXPERIMENTAL RESULTS

A. Data acquisition and synchronisation

In order to develop, test and validate our applications, methods and algorithms, real sensor data are needed. We develop our sensor data acquisition platform. This platform is developed under Windows and is integrated on the experimental vehicle CyCab developed by the society Robosoft[2]. This platform can acquire, date and record measurements of several type of sensors. Drivers were developed for the camera we used (AVT MARLIN F-046C connected on the FIREWIRE IEEE 1394 SVGA bus), the GPS, and the inertial sensor.

The used INS is XSens MT9B. This sensor is composed of 3D accelerometers, 3D gyroscopes, 3D magnetometers and temperature sensors. Using these measurements, an algorithm implemented in the sensor provides the Euler angles (Roll, Pitch, Yaw). The frequency is 100 Hz (100 measures per second).

[2]http://www.robosoft.com/eng/

The used GPS is Sagitta02 from Thales Navigation, which has a centimetric precision in LRK mode. Its acquisition frequency is 20 Hz.

B. GPS/INS fusion

In the Figure 10, the results of the INS/GPS fusion by an Extended Kalman Filter are presented.

The doted blue line is the centimetric LRK GPS positions. The continued red line represents the EKF pose estimation.

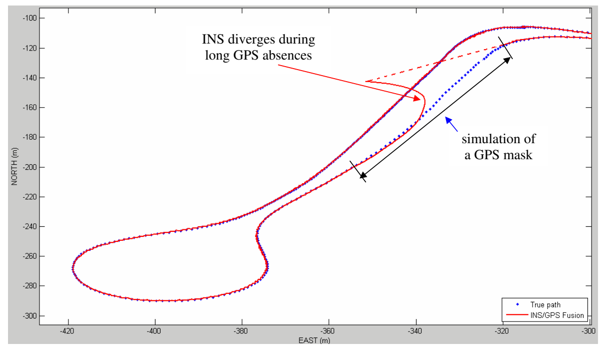

Fig. 10. Results of the INS and GPS fusion

When the GPS is available, we obtain satisfying results. But, when we simulate a GPS measurement non-availability, we notice that the pose estimation diverges. The only INS can't provide an accurate location for a long time.

C. Geo-localisation with 3D-GIS, 3D model and vision

In the Figure 11, we show the obtained pose estimation using the proposed method of geo-localisation. But, in this part of experimental trajectory, the correction of the predicted pose is realised with the vision and 3D cartographical model (GPS measurements are not usable).

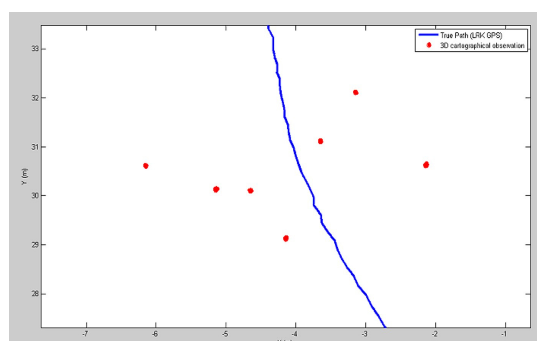

Fig. 11. Results of the 3D cartographical observation

The red points represent these 3D cartographical observations. One can refer to the blue continuous line (LRK GPS positions) to evaluate the accuracy of correction, which we can obtain with the vision and 3D cartographical model. The maximum of the error between the estimated pose and the real position, for this example, is about 3 meters.

V. CONCLUSIONS AND FUTURE WORKS

We have present in this paper a new method of geo-localisation exploring a new kind of information source which is the 3D geographical model managed by a 3D-GIS. It seems that the GPS measurements are not necessary available all the time. But the fusion of INS and 3D cartographic data can provide a good estimation of the position.

Nevertheless, we have noticed that sometimes this estimation can diverge. This is due to the fact that the 2D/3D matching strategy for virtual and real images matching, presented in this paper can be ameliorated. If a wrong virtual image is selected than the good one, the method will diverge, because the GPS is not available to correct this wrong choice. Several ways can be considered to ameliorate the 2D/3D matching method. A method based on the Belief Theory is actually under study to provide a robust and real time implemented 2D/3D matching method. We think that is the main perspective of this research.

ACKNOWLEDGMENTS

The present research work has been supported by ”Science and Technology for Safety in Transportation”, and founded by the ”European Union”, the ”Délégation Régionale à la Recherche et à la Technologie”, the ”Ministère Délégué à l’Enseignement Supérieur et à la Recherche”, the ”Région Nord Pas de Calais” and the ”Centre National de la Recherche Scientifique”.

REFERENCES

[1] R.E. Kalman, ”A new approach to linear filtering and prediction problems”, *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, March, 1960.

[2] G. Welch and G. Bishop, ”An Introduction to the Kalman Filter”, *University of North Carolina, Chapell Hill, TR 95-041* March, 2002.

[3] M.S. Grewal, L.R. Weill and A.P. Andrews, ”Global Positioning Systems, Inertial Navigation and Integration”, *John Wiley and Sons*, ISBN 0471-35032-X, 2001.

[4] E.J. Krakiwsky, C.B. Harris and V.C. Wong, ”A Kalman Filter for Integrating Dead Reckoning, Map Matching and GPS Positioning”, *Proceedings of IEEE Position Location and Navigation Symposium*, pp. 39-46, 1988.

[5] S. Panzieri, F. Pascucci and G. Ulivi, ”An outdoor navigation system using GPS and inertial platform”, *IEEE/ASME Transactions of Mechatronics*, vol. 7, n°2, pp. 134-142, 2002.

[6] S. Persa and P. Jonker, ”Real-Time Computer Vision System for Mobile Robot”, *Proc. SPIE Vol. 4572, p. 105-114, Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision, David P. Casasent; Ernest L. Hall; Eds.* , pp. 105-114, October, 2001.

[7] K. Sung-Soo, K. Kyong-Ho, L. Seong-Ho and L. Jong-Hun, ”High Resolution Earth Imaging for Geospatial Information”, *Video Navigation System using the Geographic Hypermedia*, ISPRS Hannover Workshop, 2005.

[8] S. Zlatanova, A. Rahman and M. Pilouk, ”Present status of 3D GIS”, *GIM International*, vol. 16, pp. 41-43, June, 2002.

[9] S. Zlatanova and F. Van den Heuvel, ”3D GIS for outdoor AR applications”, *Proceedings of the Third International Symposium on Mobile Multimedia Systems & Applications, R.L.Lagendijk, R.Heusdens, K.G.Langendoen, W.A. Serdijn (eds.)*, ISBN 90-9016467-7, Delft, The Netherlands, pp. 117-124, 6 December, 2002.

[10] Y. Takase, N. Sho, A. Sone, K. Shimiya, ”Generation of Digital City Model”, *Journal of the Visualization Society of Japan*, vol. 23, no. 88, pp. 21-27, 2003.