Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Intelligent Vehicle Localization in Urban Environments Using EKF-based Visual Odometry and GPS Fusion

Lijun Wei * **Cindy Cappelle** * **Yassine Ruichek** *
Frédérick Zann *

* *Laboratoire Systèmes et Transports, Université de Technologie de Belfort-Montbéliard, 90010 Belfort, France.(e-mail: lijun.wei@utbm.fr, cindy.cappelle@utbm.fr, yassine.ruichek@utbm.fr, frederick.zann@utbm.fr).*

Abstract: We present a method for intelligent vehicle localization and 3D mapping in urban environments using integration of stereovision and Real-Time Kinematic GPS (RTK-GPS). In our approach, stereoscopic system is used to recover the scene geometry and to predict the camera motion. Accurate GPS positions are integrated to correct the vehicle positions and orientations using Extended Kalman Filter (EKF), and to rectify the global positions of reconstructed 3D landmarks based on the positions of reference stereo frames. The method was tested using data obtained by a real electrical GEM vehicle equipped with stereoscopic system and RTK-GPS in urban environments. The results show that the trajectory obtained by GPS/Vision integrated method can fit the ground truth better than the vision-only method, and can also avoid the position jumps aroused by GPS signal failures.

Keywords: localization, stereovision, GPS, multisensor data fusion, intelligent vehicle

1. INTRODUCTION

Localization is one of the key problems for achieving autonomous abilities of mobile robots. Satellites based navigation systems (e.g. GPS, Galileo) have been the most popular tools for outdoor vehicle global localization and navigation. They can provide accurate absolute positions in long term, but as the GPS signals are affected by atmospheric conditions, satellite positions, radio signal noises, etc, the accuracy in short term is only to a few meters. Although the Real-Time Kinematic GPS (RTK-GPS) can deliver position up to centimeter accuracy, it does not work well anymore in some particular dense urban environments (e.g. urban canyons), as the satellite signals might be blocked or reflected by tall buildings. Insufficient satellite numbers or the multi-path problem will decrease the position accuracy. Furthermore, the non-stationary noise of GPS might affect the GPS observation model.

Another kind of localization method is called dead-reckoning. It can obtain the current position based on its relative motion from the previous one. For example, the wheel encoder based odometry can localize the vehicle by measuring the traveling distance and the elementary rotation. Nevertheless, odometry based localization suffers from the wheel slippage in rock areas or muddy areas on bad wheel radius estimation. IMU (Inertial measurement unit) achieves the purpose by measuring the acceleration and orientation at every instant. Computer vision based visual odometry method (laser or camera) Nister et al. (2004) can help to percept the environment, then assist vehicle localization and navigation by continuously mapping the real world and estimating the relative vehicle motion. Though all these dead-reckoning methods can provide good accuracy in short term, the trajectory might drift in long term as errors accumulate from point to point.

An efficient solution is to integrate GPS and other sensors together, and to take advantages of the best characteristics of every sensor. Such as Najjar et al. (2005) integrate GPS with odometry; Sukkarieh et al. (1999) combine GPS with IMU. As in dense urban environments, there are amount of visual landmarks, the visual odometry method can supply both localization and mapping information. We propose to integrate GPS with visual odometry together using the well known sensor fusion technique Extended Kalman Filter (EKF)(Thrun et al. (2005)). Cameras are employed because they are less expensive and lighter than laser scanning system.

In outdoor environment, vehicle dynamic computation and scenery modeling make the use of images very challenging, especially for large scale cases. Royer et al. (2007) use monocular and structure from motion method to build an accurate 3D map and then use this map to locate the vehicle in real-time. But as the baseline between two instants is unknown, the scale of reconstruction is ambiguous and should be estimated with GPS trajectory as the last step. So stereovision method is adopted in our system because the baseline between left and right cameras is already known by calibration, and the scale of Euclidean reconstruction is directly provided.

An overview of the proposed method is shown in Fig. 1. It is composed of three principal parts: 1) Stereovision is used to perform 3D reconstruction and to produce an incremental motion estimates. The predicted incremental distance

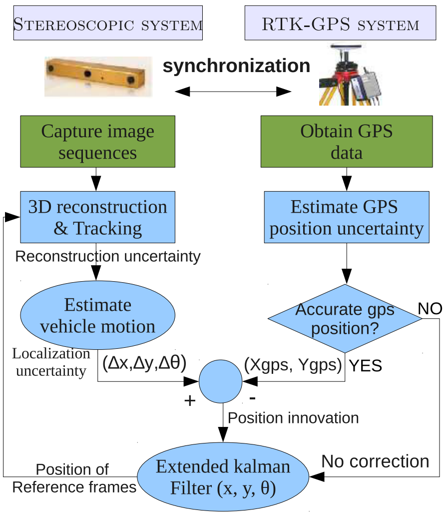

Fig. 1. Overview of the proposed localization method

and yaw angle are used as the inputs of EKF. 2) Accurate GPS positions are used as the position measurements. GPS positions have to be checked if they are accurate or fault by calculating the covariance matrices of GPS positions according to NMEA GST sentences. If accurate GPS positions are available, they are applied as position measurements to correct the vision based prediction with EKF. 3) Finally, the global positions of reconstructed 3D landmarks are rectified by the corrected positions of reference frames.

This paper is organized as follows: section 2 introduces the system configuration. Section 3 presents stereovision based visual odometry(VO) method. Section 4 details the localization mechanism by GPS/VO integration using Extended Kalman Filter. Section 5 tests the proposed localization method with data obtained by a real electrical GEM vehicle. Finally some conclusions and future perspectives are presented in section 6.

2. SYSTEM CONFIGURATION

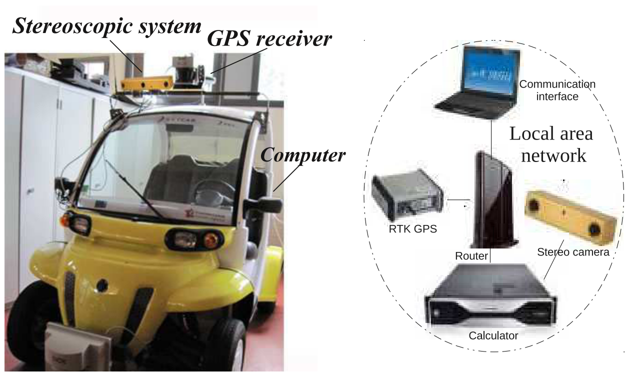

Fig. 2. Left:Experimental vehicle equipped with RTK-GPS and stereoscopic system; Right:Hardware architecture

As shown in Fig. 2, the experimental GEM vehicle is equipped with a stereoscopic Bumblebee XB3 camera($5Hz$) and a Magellan ProFlex500 RTK-GPS($10Hz$) receiver. They are mounted on the roof of the vehicle. The RTK-GPS can achieve up to 1cm accuracy in an horizontal plane. Cameras are calibrated and images are rectified. The Bumblebee XB3 has three cameras on the same line, with $0.12m$ distance apart from each other. Every image has a resolution of $1280 * 960$. GPS and image data are stored under the same computer together with their stored times. Then the synchronization of the two sensor systems are achieved by associating their saving times.

3. STEREOVISION BASED VISUAL ODOMETRY

Stereovision based visual odometry method is composed of 3D reconstruction (3.1) and vehicle ego-motion estimation (3.2). At first, features are detected, matched, and reconstructed from reference pair, then tracked across frames till the reference updates. After that with the reconstructed 3D points, the relative camera motion can be estimated with RANSAC based least square method.

3.1 3D Landmarks Reconstruction

Feature extraction Since SURF features (speeded up robust features) have the advantages of repeatability, distinctiveness, robustness, and can be computed and compared fast (Bay et al. (2008)), they are extracted from every reference stereo pair, and then used during the other procedures of the proposed method.

Feature matching The extracted SURF features in the left and right images are matched by descriptors under several geometric constraints, including: epipolar constraint, disparity constraint, threshold of $ZNCC$ (Zero-mean normalized cross correlation) score, uniqueness constraint, and inverse matching (mutual checking) constraint.

3D landmark reconstruction in a reference pair When the geometric arrangement of the stereoscopic system is known, the local 3D position $Q(Q_x, Q_y, Q_z)$ of an object relative with the camera center can be recovered based on its corresponding image features. The left and right rays passing through camera centers and corresponding features are estimated separately, then the shortest segment that connects these two rays is found, the middle point of this segment is considered as the corresponding 3D position of the feature (Cheng et al. (2006)).

Let r_1 and r_2 be the unit vector that connects the left(right) camera center $C_1(C_2)$ and corresponding left(right) image feature $q(q')$, Q_1 and Q_2 be the endpoints of the shortest line segment connecting these two rays, T is the baseline between C_1 and C_2, then:

$$r_1 = \{q_x, q_y, f_x\}/\|C_1 q\|, r_2 = \{q'_x, q'_y, f'_x\}/\|C_2 q'\| \quad (1)$$

and the relative distance between the 3D point and the two camera centers can be written as:

$$Q_1 = C_1 + r_1 m_1, Q_2 = C_2 + r_2 m_2 \quad (2)$$

while $m_1 = \|Q_1 C_1\|$ and $m_2 = \|Q_2 C_2\|$, then

$$m_1 = \frac{T{\cdot}r_1 - (T{\cdot}r_2)(r_1{\cdot}r_2)}{1 - (r_1{\cdot}r_2)^2}, m_2 = (r_1{\cdot}r_2)m_1 - T{\cdot}r_2 \quad (3)$$

then the coordinates of the 3D point Q can be obtained as: $Q = (Q_1 + Q_2)/2$. In order to achieve accurate navigation over long distances, reconstructed 3D points with depth exceeded 50 meters are eliminated. The other points are stored as landmarks together with the reference frame where they are reconstructed.

2D feature tracking and outlier removing When a new stereo pair is captured, the previous matched key features are separately tracked in the left and right images by Kanade-Lucas-Tomasi feature tracker (Lucas et al. (1981)). During the tracking process, three constraints are applied to discard false tracking: SAD (Sum of absolute differences) of tracked feature intensity across frames, the tracked point cannot move out of the window, $ZNCC$ of tracked image points in the left and right images. Furthermore, a relative depth constraint is added. When the vehicle moves in rigid and static environment, the changes of estimated depths relative with the camera frames t and $t+1$ should be approximately the same for all 3D points according to:

$$\Delta Z = Z(t) - Z(t+1) = f * T * (1/d(t) - 1/d(t+1)) \quad (4)$$

The mean and standard deviation of relative depth changes are calculated for all the tracked couples, and the couples whose depth deviations are more than 3 times of the standard deviation are discarded.

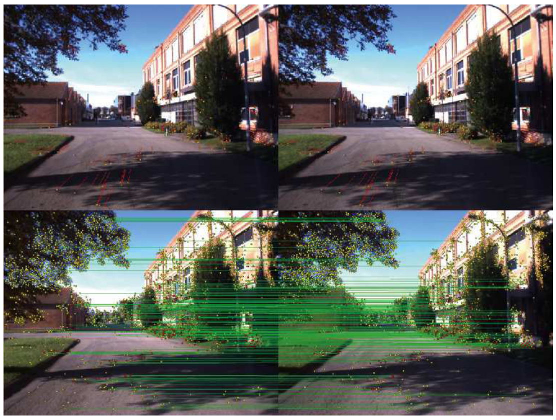

Fig. 3. Feature Matching and Tracking. Bottom: reference pair; Above: tracked pair; Green line: left-right matched points; Red line: tracked points between reference-tracked frames

Reference stereo pair updating As the camera moves, some features may move out of the field-of-view, only features that can be tracked by the previous frame will be tracked sequentially. For accurate pose estimation, enough feature number and spatial distribution should be ensured. If the number or distribution is less than predefined threshold, the previous stereo pair is selected as new reference stereo pair, then the features are detected and the previous procedure are repeated. At the same time, the detected features in the new reference stereo pair are compared with the previous reference pair, the different features are added into the 3D model as new appeared landmarks.

Uncertainty in 3D reconstruction Based on the previous reconstruction part, the uncertainty in 3D reconstruction is discussed here. The partial derivative m_1' and m_2' of m_1 and m_2 with respect to the 2D image coordinates (q_x, q_y, q_x', q_y') in the left and right images are calculated. Then Q', the $3*4$ Jacobian matrix of the 3D point Q with respect to the 2D image coordinates $\{q_x, q_y, q_x', q_y'\}$ in the left and right images, is estimated as in Cheng et al. (2006). The position uncertainty of image features can be separately modeled as an uncorrelated zero-mean Gaussian noise as in Matthies et al. (1987). They can be respectively written as $2*2$ covariance matrix:

$$P_{left} = \begin{bmatrix} \delta_x^2 & 0 \\ 0 & \delta_y^2 \end{bmatrix}, P_{right} = \begin{bmatrix} \delta_x'^2 & 0 \\ 0 & \delta_y'^2 \end{bmatrix} \quad (5)$$

Where δ_x^2, $\delta_x'^2$, δ_y^2, and $\delta_y'^2$ are the standard deviations of the pixel q_x, q_x', q_y, and q_y' coordinates. The covariance matrix of stereo pair P_{pair} can be written as a $4*4$ diagonal matrix with the vector $\{\delta_x^2, \delta_y^2, \delta_x'^2, \delta_y'^2\}$. Then the covariance matrix of the reconstructed 3D points is given by $P_{reconstruction} = Q' P_{pair} Q'^T$. It can approximately measure the uncertainty of the 3D points obtained at every time instant by the stereoscopic system.

3.2 Vision based Vehicle Ego-motion Estimation

Ego-motion estimation As we assume that the ground is plane, the camera motion can be represented by translation distance d on X-Z plane and yaw angle change θ. For two corresponding point sets Q_i^t and Q_i^{ref} ($i = 1 : N$, while N is the number of corresponding points) obtained at the current camera coordinate system t and the reference coordinate system, the two point sets can be related by:

$$Q_i^t = R * Q_i^{ref} + T + V_i \quad (6)$$

with R: 2*2 rotation matrix, $T(T_x, T_z)$: 2D translation vector, and V_i: the noise vector. To find the optimal transformation $[R, T]$ that transforms the points on reference frame onto current local one, it requires to minimize the residual error:

$$\varepsilon^2 = \sum_{i=1}^{N} \|Q_i^t - RQ_i^{ref} - T\| \quad (7)$$

The best rotation matrix in the least squares sense can be found by SVD (singular value decomposition) as the solution of Arun et al. (1987), together with the complements of Umeyama (1991) for some degenerated cases. Then the translation vector T can be obtained by the centroids of two point sets: $T = \bar{Q}^t - R \cdot \bar{Q}^{ref}$. Dynamic RANSAC is used for a more precise estimation. To ensure that the randomly selected three points distribute well in the image, the distance between every two image features should be more than 20 pixels. The other RANSAC parameters are dynamically chosen according to Hartley et al. (2004).

Uncertainty in vehicle localization As described in previous section, we define $\{MO_{ref}^t\}$ as the random motion vectors that transform the vehicle from $reference$ frame to instant t, they can be obtained by the corresponding landmark sets, which are separately reconstructed at the two time instants:

$$MO_{ref}^t = f(Q_i^{ref}, Q_i^t) = [R|T] \quad (8)$$

Then,

$$Q_i^t = [R|T] \begin{bmatrix} Q_i^{ref} & 1 \end{bmatrix}^T = d * \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} + Q_i^{ref} \quad (9)$$

$i = 1, ..., m$, where m is the number of landmarks used for motion estimation. Assume that every landmark Q_i^{ref} follows a Gaussian distribution with mean Q_i^{ref} and covariance matrix P_i^{ref}, $Q_i^{ref} \sim N(Q_i^{ref}, P_i^{ref})$, and landmark Q_i^t follows a Gaussian distribution with mean Q_i^t and covariance matrix P_i^t, $Q_i^t \sim N(Q_i^t, P_i^t)$. So under the error propagation(Moreno et al. (2007)), MO_{ref}^t should follow a Gaussian distribution with mean $\{d_{ref}^t, \theta_{ref}^t\}$, covariance matrix Q_{ref}^t can be estimated by:

$$P_t^{ref^{-1}} = \mathcal{S}^T P_{MO}^{-1} \mathcal{S} = \sum_i (\mathcal{S}_i^T P_{MO_i}^{-1} \mathcal{S}_i) \tag{10}$$

While $\mathcal{S}$ is the Jacobian matrix of equation (9) with respect to d and θ, P_{MO} is the uncertainty of all landmarks, $\mathcal{S}_i$ is the Jacobian matrix of i-th point with respect to d and θ, and P_{MO_i} is the uncertainty of the landmarks i at the $reference$ instant and instant t, $P_{MO_i} = P_i^{ref} + P_i^t$.

4. GPS & VISUAL ODOMETRY INTEGRATION USING EXTENDED KALMAN FILTER

4.1 Vehicle Model

For vision system, the body frame is attached to the car. We take the initial vehicle position as the origin of the global system W, and the initial forward orientation as the positive direction z. The mobile frame M is chosen with its origin attached to the center of the rear axle as shown in Fig. 4. At time instant t, the vehicle position

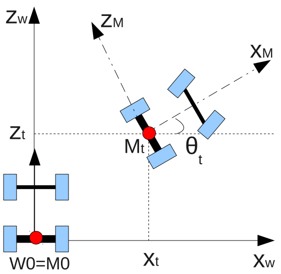

Fig. 4. Vision frame attached to the mobile vehicle

M is represented by (x_t, z_t) in the world frame W. The heading orientation is θ_t.

4.2 State Prediction

Prediction Assume that the road is plane, the evolution vehicle model in Fig. 4 can be expressed as:

$$\begin{cases} x_{t+1} = x_t + d_t^{t+1} cos(\theta_t + \omega_t) \\ z_{t+1} = z_t + d_t^{t+1} sin(\theta_t + \omega_t) \\ \theta_{t+1} = \theta_t + \omega_t \end{cases} \tag{11}$$

While d_t^{t+1} is the circular arc followed by M from instant t to instant $t+1$, ω_t is the rotation angle of the mobile vehicle. Let $X_t = [x_t, z_t, \theta_t]^T$ denote the vehicle state vector, $u_t = (d_t, \omega_t)^T$ is the control data for time interval $(t : t+1)$, the vehicle motion prediction equation (11) can be rewritten as: $X_t^{t+1} = f(X_t, u_{t+1}) + \alpha_t$, where α_t is the model noise. In our system, we choose the rotation angle and translation distance provided by visual odometry as u, and the GPS observation from RTK-GPS as position measurement.

Prediction noise The covariance matrix R_t^{t+1} of motion prediction can be represented by:

$$R_t^{t+1} = \mathcal{J}_{X_t}^f R_t (\mathcal{J}_{X_t}^f)^T + \mathcal{J}_{u_{t+1}}^f C (\mathcal{J}_{u_{t+1}}^f)^T + \mathcal{H} \tag{12}$$

With $\mathcal{J}_{X_t}^f$ and $\mathcal{J}_{u_{t+1}}^f$ are the Jacobian matrices of $f(X_t, u_{t+1})$ with respect to X_t and u_{t+1}; R_t is the covariance matrix of estimation at time instant t; $C = P_t^{t+1}$ is the covariance matrix of VO inputs; $\mathcal{H} = (\delta_d^2, \delta_\omega^2)$ is the covariance matrix of the gaussian white noise that directly affects the state of the vehicle model.

4.3 Measurement Model

GPS uncertainty The GPS observation from RTK-GPS is used as position measurement. As in the urban environment, GPS suffers from multi-path problems, and the non stationary noise of GPS measurement noise affects the observation model, the linear observation equation of GPS positions is:

$$X_{t+1}^{gps} = \begin{bmatrix} x_{t+1}^{gps} \\ z_{t+1}^{gps} \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t+1} \\ z_{t+1} \\ \theta_{t+1} \end{pmatrix} + \beta_{gps} \tag{13}$$

while the GPS observation $(x_{t+1}^{gps}, z_{t+1}^{gps})$ is provided by GPS position measurement and β_{gps} is the measurement noise. As GPS measurements are affected by many independent noise sources, the measurement noise of every GPS position error can be estimated by:

$$Q_{t+1}^{gps} = \begin{pmatrix} \delta_{x,gps}^2 & \rho \cdot \delta_{x,gps} \cdot \delta_{z,gps} \\ \rho \cdot \delta_{x,gps} \cdot \delta_{z,gps} & \delta_{z,gps}^2 \end{pmatrix} \tag{14}$$

while $\delta_{x,gps}$ and $\delta_{z,gps}$ are the standard deviations of the estimation error of x and z as observed in the x-z plane, ρ is the spatial correlation coefficient, and φ is the orientation of semi-major axis of error ellipse in degrees from true North. $\delta_{x,gps}$, $\delta_{z,gps}$, and φ can be obtained by the Standard National Marine Electronics Association (NMEA) sentence "GST", ρ can be calculated according to Najjar et al. (2005). If the standard deviation of x and z are less than 3 $meters$, the GPS positions are used as position measurements, then the visual odometry based vehicle position can be corrected. Otherwise, only visual odometry is used to estimate the vehicle motion.

4.4 Motion Update

Computation of innovation VO based position is:

$$\widehat{X}_t^{t+1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} X_t^{t+1} \tag{15}$$

The position innovation v_{t+1} is computed as:

$$v_{t+1} = X_{t+1}^{gps} - \widehat{X}_t^{t+1} = (x_{t+1}^{gps}, z_{t+1}^{gps})^T - (x_{t+1}, z_{t+1})^T \tag{16}$$

The covariance matrix of innovation is:

$$S_{t+1} = \mathcal{J}_x R_t^{t+1} \mathcal{J}_x^T + Q_{t+1}^{gps} \tag{17}$$

with $\mathcal{J}_x$: the Jacobian matrix of $\widehat{X}_t^{t+1}$ with respect to X_t^{t+1}; Q_{t+1}^{gps}: the covariance matrix of GPS measurement.

Motion Correction The standard form of EKF is used in this part. Kalman gain can be computed by: $K_{t+1} = R_t^{t+1}(\mathcal{J}_x)^T S_{t+1}^{-1}$. Correction is: $X_{t+1} = X_t^{t+1} + K_{t+1}v_{t+1}$. Noise in EKF-based estimated position can be represented by the covariance matrix associated to X_{t+1} as: $R_{t+1} = R_t^{t+1} - K_{t+1}S_{t+1}(K_{t+1})^T$.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of our proposed method, video sequences with ground-truth are tested. One sequence was captured in September, 2010, at Belfort, France. The vehicle was driven in an industrial area where there are buildings around as shown in Fig. 8. The sequence comprises 1880 stereo pairs. Trajectory distance measured by RTK-GPS was about $674.5m$, with direct lines and four big turns with about 90^o. As only RTK-GPS was provided for this experiment, it was both served as the ground-truth and measurements in this experiment, while some parts of GPS failure were simulated.

GPS transformation As GPS provides longitude and latitude information in Earth frame (East, North), the GPS positions obtained from NMEA sentences are projected from WGS84 system to Extended Lambert II that covers the region of Belfort. Then they are translated into the vision frame by transforming the corresponding initial GPS position to $\{0, 0\}$. After that, based on the initial vehicle orientation obtained by VO and GPS, the vision based trajectory is rotated such that its initial direction is the same as the GPS trajectory.

Estimated localization results The vehicle trajectory obtained by vision method is shown in Fig. 5(left). For every 20 GPS instants, the corresponding vehicle positions provided by GPS and VO are linked by a blue line. The total traveling distance obtained only by stereovision is $679.39m$, with a 0.72% difference with ground truth. We note that at first the vision based trajectory fits GPS trajectory very well, then it drifts to the left side gradually. Then we tested the same sequence with the integration method of GPS and stereovision. Results are shown in Fig. 5(right). It shows that the corrected vehicle trajectory can fit the ground truth better. The total distance is $683.81m$, the difference with ground truth is 1.29%.

Table 1. Comparison of traveling distance(/m)

Method	GPS	Estimated dist.	Error%	Mean	Std.
VO	674.5	679.39	0.72	27.92	11.86
GPS/VO	674.5	683.18	1.29	1.23	1.05

In the zoom view of the trajectory within the circle in Fig. 5(right), GPS jumps are particularly visible. Due to the simulated GPS mask, GPS information were not used for correction, only visual odometry is used to estimate the trajectory. It shows that the trajectory obtained by visual odometry can smoothly align with the trajectory when GPS jumps occur. Together with the visual odometry method, the EKF can continuously estimate the vehicle state vector when the GPS signals are blocked or obstructed.

Yaw angle differences We also compute the yaw angle of the vehicle at every GPS instant. As no IMU sensor was incorporated in our system, the ground truth of yaw angle was approximated by GPS according to the GPS NMEA VTG sentences. The parameter COG (orientation with respect to the True North) is used to calculate the angle changes. The results are shown in Fig. 6. The differences of

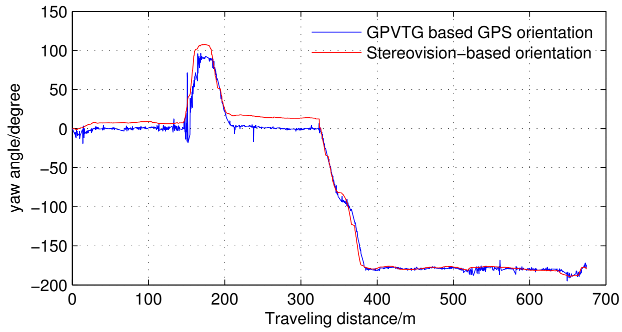

Fig. 6. Yaw angle of GPS and Vision-based trajectory

orientations separately obtained by GPS and VO methods are small, except for some positions where the GPS signals jump.

Position differences For a more accurate comparison, we compare the position error between all the GPS positions and their corresponding VO positions (Fig. 7(above)) and EKF-based positions (Fig. 7(bottom)). For vision

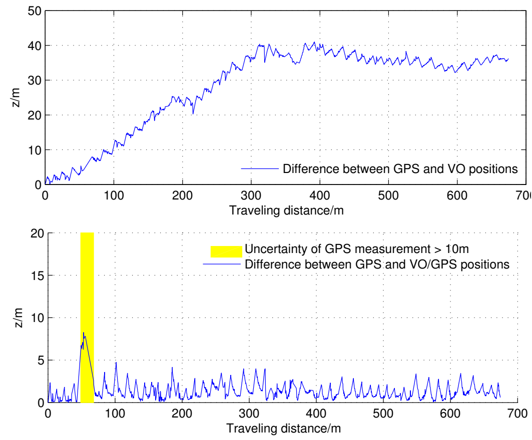

Fig. 7. Above: Position difference between GPS and vision based trajectory; Bottom: Position difference between GPS and EKF-based trajectory

based method, as errors accumulate gradually, the position differences increase continuously. For the EKF-integration method based trajectory, almost all the position errors are less than 5 *meters*, except for the part where GPS positions jump a lot (positions within the yellow zone). These positions can not be used as ground truth. For the whole sequence, the average of position differences is $1.23meters$, with a deviation of $1.07meters$, as shown in table 1.

Based on the corrected vehicle trajectory, 3D landmarks estimated by stereovision are adjusted according to the reference frame where they were reconstructed. As shown in Fig. 8, the corrected trajectory is shown with red line, and the landmarks are represented with green points.

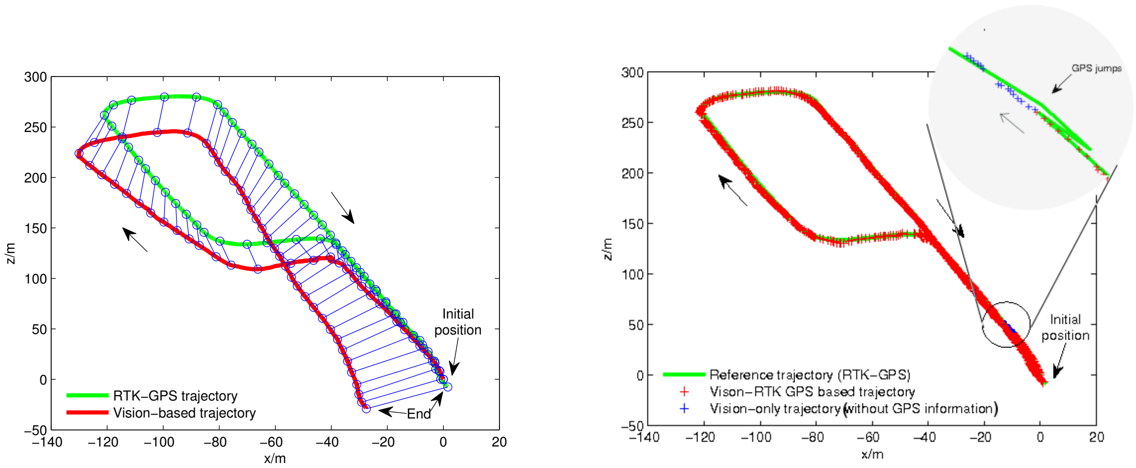

Fig. 5. Left: Vehicle trajectory provided by GPS and Vision; Right: Trajectory based on GPS and vision integration.

Fig. 8. Corrected trajectory & Reconstructed landmarks

6. CONCLUSIONS AND FUTURE WORKS

We presented a vehicle localization method in urban environment using stereovision and GPS. Stereoscopic system is applied to achieve 3D reconstruction and to estimate the camera motion. And accurate GPS positions are used as the measurements to produce more accurate vehicle positions using Extended Kalman Filter. We tested our method with data obtained by a real electrical vehicle in urban environments. Results show that the trajectory obtained by GPS/VO integrated method can fit the ground truth better than the VO method, can reduce the accumulated error of stereovision, and can also avoid the position jumps aroused by GPS outages. The only use of GPS and Kalman filter might be considered enough for localization. However, when the vehicle is in a urban canyon and without GPS during a long time, visual method could be used.

As future works, we plan to incorporate other sensors into our system, such as using normal GPS as measurement while RTK-GPS as ground truth, using IMU to provide orientation and velocity, and using scanning laser range finders to perform mapping in large scale environment. Besides,we assume that the ground is plane without considering the altitude, we will try to solve the 6DOF camera motion.

REFERENCES

K. S. Arun, T.S. Huang and S.D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Trans.Pattern Anal*, 9(5):698–700, 1987.

H. Bay, Andreas. Ess, T. Tuytelaars, and L.V. Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.

Y. Cheng, M. Maimone, and L. Matthies. Visual Odometry on the Mars Exploration Rovers. *IEEE Robotics and Automation Magazine*, 13(2):54–62, 2006.

R. I. Hartley, and A. Zisserman. *Multiple View Geometry in Computer Vision*. pages 117–121, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI'81*, 674–679, 1981.

L. Matthies, and S.A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):239–250, 1987.

F.A. Moreno, J.L. Blanco , J. Gonzalez. An efficient closed-form solution to probabilistic 6D visual odometry for a stereo camera. *ACIVS2007*, 932–942.

M. E. E. Najjar and P. Bonnifait. A road-matching method for precise vehicle localization using belief theory and kalman filtering. *Auton. Robots*, 19:173–191, 2005.

S. Sukkarieh, E.M. Nebot, and H.F. Durrant-Whyte. A high integrity IMU/GPS navigation loop for autonomous land vehicle applications. *IEEE Transactions on Robotics and Automation*, 15(3):572–578, 1999.

D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. *CVPR*, 2004.

E. Royer, M. Lhuillier, M. Dhome, and J. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *Int. J. Comput. Vision*, 74:237–260, 2007.

S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. pages 48–53, The MIT Press, ISBN: 0262201623, 2005.

S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal*, 13(4):376–380, 1991.