2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Automatic Lane-level Map Generation for Advanced Driver Assistance Systems using Low-cost Sensors

Chunzhao Guo, Jun-ichi Meguro, Yoshiko Kojima, and Takashi Naito

Abstract— **Lane-level digital maps can simplify driving tasks for robotic cars as well as enhance performance and reliability for advanced driver assistance systems (ADAS) by providing strong priors about the driving environment. In this paper, we present a system for automatic generation of precise lane-level maps by using conventional low-cost sensors installed in most of current commercial cars. It mainly consists of two modules, i.e. road orthographic image generation and lane graph construction. First, we divide the global map into fixed local segments based on the road network topology. According to the local map segments, we accumulate the bird's eye view images of the road surface by fusing GPS, INS and visual odometry, and subsequently integrate them into synthetic orthographic images with the reference of the local map segments. Furthermore, the information of the driving lanes is extracted from the orthographic images and a large amount of vehicle trajectories, which is used to construct the lane graph of the map based on the lane models we proposed. Such a system can offer increased value as well as promote the automation level for today's commercial cars without being supplemented additional sensors. Experiments show promising results of the automatic map generation of the real-world roads, which substantiated the effectiveness of the proposed approach.**

I. INTRODUCTION

Nowadays digital maps benefit a wide variety of driving applications from providing road information for route planners in today's navigation systems to navigating mobile robot cars across various terrains and roadways. The knowledge of the driving environment in the maps is essential for robotic vehicles to understand the situation, comply with traffic rules, and achieve high system reliability. However, the high cost of the precise digital map generation systems nowadays prevents their benefits to normal commercial cars as they usually depend on specialized and expensive sensors and need great manual efforts for data analysis. In this paper, we present a system for automatic lane-level map generation that makes use of conventional low-cost sensors installed in today's standard commercial cars, such as GPS, INS and cameras. It is based on the concepts of probe cars in server-based intelligent transportation systems (ITS). The use of the low-cost sensors makes standard vehicles possible to sample the environment, which can ensure a large fleet of probe vehicles that provides a larger and faster coverage of the accessible area as well as up-to-date information.

The flow diagram of the proposed map generation system is shown in Fig. 1. The GPS/INS-based localization is firstly performed by tightly coupled integration of the raw INS data and GPS Doppler shift frequency measurement, followed by a bundle adjustment over the entire GPS pseudorange data recorded while driving. The results are then used to initialize an optimization-based visual odometry, which improves both vehicle localization and the pose estimation. Furthermore, the global map is divided into fixed local segments based on the road network topology. According to the local map segments, the bird's eye view road images are accumulated to generate local road image segments, which are subsequently integrated into synthetical orthographic images of the road. Finally, the lane graph is constructed by extracting the information of the driving lanes from the orthographic images and a large amount of vehicle trajectories.

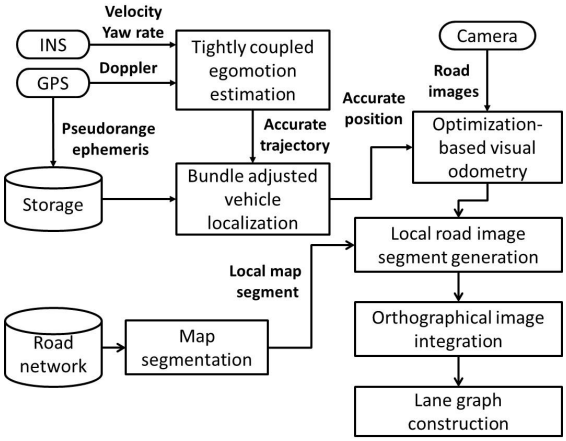

Fig. 1. Flow diagram of the proposed map generation system using conventional low-cost sensors.

II. PREVIOUS RELATED WORK

Current navigator maps typically have a precision of several meters [1], which are mainly used for route planners. In recent years, great efforts have been made to generate high-precision digital maps [2]. This commonly involves driving the entire road network, recording the car position and road data with specialized and expensive sensors, and huge manual post-processing efforts. While the volume of detailed information and precision can be achieved, the high cost of the probe cars and the low efficiency of manual processing make such map data hard to update, which prevents their benefits to normal commercial cars. Another category of work generates the map using aerial images exclusively [3], sometimes in combination with Lidar for SLAM (simultaneous localization and mapping) [4]. Such methods can obtain the map automatically; however, the cost of the accurate data acquisition is still too expensive to generate the map in large scales.

In contrast of the above map generation systems, we focus in this work on low-cost sensors and extract the lane-level information of the road from the synthetical orthographic images of the road, which are generated based on the ego-

C. Guo, J. Meguro, Y. Kojima and T. Naito are with Toyota Central R&D Labs., Inc., 41-1, Yokomichi, Nagakute 4801192, Aichi, Japan. {czguo, meguro, yoshiko-k, naito} at mosk.tytlabs.co.jp

positioning using conventional GPS, INS and camera sensors. In recent years, there have been a lot of researches in the construction of road orthographic images from commodity sensors [5]-[8]. Most of these approaches focused on the use of vision exclusively. For example, the methods in [5] and [6] used monocular front cameras to perform visual odometry for road image mosaicing. Although visual odometry offers an accurate and smooth local estimation for both localization and image alignment, it may be affected dramatically by a number of factors, such as road appearances, illumination conditions, etc., or even fail in some extreme situations. For instance, visual odometry cannot work well in a newly built asphalt road or the conditions with heavy shadow/strong backlight since there is no enough texture in the road area. It also cannot recover from the situation in which the camera is occluded by a nearby vehicle. Furthermore, visual odometry would accumulate error and cannot provide absolute positions of the vehicle/road image without reference landmarks while such information is essential for map generation.

Others utilized GPS/INS sensors for absolute positioning, for example, in our previous work we generated road orthographic images based on the vehicle trajectory purely estimated by a GPS and a gyro [8]. It works well in most of the cases, however, the accuracy may drop when the vehicle's velocity (speed and direction) changes dramatically (e.g., the vehicle makes a turn in an intersection). In order to overcome this problem, in this paper, we fuse the GPS/INS measurements with visual odometry for vehicle localization and road orthographic image generation, which improve the accuracy greatly, as shown in the comparison result in Section III.C. The method in [7] also performed the vehicle localization by fusing visual odometry from a rear parking camera with a U-blox automotive type GPS receiver. However, they just used the conventional GPS positioning technique, which estimate the absolute position of the receiver based on the GPS pseudorange measurements epoch by epoch. The low resolution of the GPS pseudorange measurements, insufficient number of visible satellites and multi-path reflections of the signal would generate great variations between adjacent estimation. Whereas we proposed a novel positioning technique by tightly coupled integration of the GPS Doppler with INS sensors, followed by a bundle adjustment with the GPS pseudorange measurements over a period of epochs, which is much more accurate, smooth and reliable than the conventional technique.

Furthermore, there have recently been attempts to use computer vision techniques for estimation of the driving lanes or understanding the scene layout, such as machine learning [9], image segmentation [10], and semantic models [11]. These methods are designed for online systems that need to infer the environment based on the input images while they are not suitable for map generation since the input information is limited and the uncertainty is quite large. On the contrary, we extract the information of the driving lanes from the orthographic images of the road together with a large amount of vehicle trajectories that normal cars traveled in the real road environment, which is much more accurate and reliable compared with the estimation/inference from images alone.

In summary, the proposed system is distinguished from the previous ones in the following ways:

1) We focus on the use of conventional low-cost sensors for automatic lane-level map generation. With such a technology, the standard cars can sample the environment, which ensures the amount of detail, coverage of accessible area and up-to-date information of the generated map due to the large amount of probes.

2) In order to cope with the noisy observations of low-cost sensors, we develop an ego-positioning and image integration system by fusing a novel GPS/INS-based positioning technique and visual odometry. Compared with the conventional GPS techniques, the proposed technique is much more accurate and stable due to the high resolution of the GPS Doppler measurement and the sophisticated design of tightly coupled integration and bundle adjustment. Thereby, the fusion of visual odometry with our GPS/INS-based positioning system outperforms the conventional fusion systems.

3) We present an efficient approach for automatic extracting the fundamental properties of the driving lanes, including the lane centerline, endpoints, lane width, lane curvature and lane connectivity, etc., based on the integrated orthographic images and a large amount of real vehicle trajectories. Particularly, we proposed an efficient transition lane model to generate "virtual" driving lanes at intersections for guiding robot cars and ADAS systems.

III. ORTHOGRAPHIC IMAGE GENERATION BASED ON FUSION OF LOW-COST SENSORS

In order to obtain the road information for generating maps, we accumulate and integrate the road images, acquired by a rear parking camera while driving, into synthetic images on the ground plane in the East-North-Up (ENU) world coordinate. We perform vehicle localization by fusing GPS/INS and vision sensors, since the benefits of them are complimentary. GPS provides absolute drift-free position measurements fixed to a global map in the world coordinate whereas visual odometry offers an accurate and smooth local estimation for both localization and image alignment. Accuracies of both systems are paramount to the final result, especially in urban environments where a number of factors will affect the observations of both sensors.

A. Global Map Segmentation

In order to process the road images and construct the lane graph locally, we firstly divide the global map into fixed local segments based on the road network topology. As shown in Fig. 2, for a map of interest in the global world map, we define it by the ENU coordinate of its upper left corner and the ranges in the east and north directions, which is then converted into a map image with a given scale. Subsequently, the road network topology of the map of interest is imported automatically from the Open Street Map (OSM) [12], which provides not only the positions and shapes of the intersections/road ends (node) and roadways between them (links), but also the connectivity information of the nodes and links. For a node in the map of interest, the local segment is a square with a fixed given range (magenta boxes). For a link in the map of interest, the ranges of its endpoints and waypoints are firstly computed. Considering the width of the roadways and possible offsets of the waypoints, a fixed given margin is then added to the

Fig. 2. Example of the map segmentation. The left images show the global world map and the map of interest (Cyan box). The right image shows the segmentation result of the map of interest, including the local node segments (magenta boxes) for intersections/road ends (magenta dots) and the local link segments (white boxes) for roadways between intersections/road ends (yellow lines). The green circles are the waypoints of the roadways exported from the OSM.

computed range to generate the local link segment (white boxes), which can also ensure the overlapping between the connected node segment and link segment. Note that both the local node segment and the local link segment are also defined by the ENU coordinate of their upper left corner and their ranges in the east and north direction. With such a map segmentation, the road images and vehicle trajectories can be locally processed and the generated orthographic images and lane graph of the road can be constructed and represented locally for any map of interest in the global world.

B. Precise Vehicle Localization with Standard In-car GPS and INS Sensors

Two types of signals for positioning are transmitted from the GPS satellites, i.e. the Pseudo-random (PN) code signal and the carrier signal. The GPS pseudorange and the GPS Doppler correspond to the relative distance and 3D velocity between the GPS satellite and the receiver, respectively. The former one is measured using the PN code signal with the resolution of 300m while the latter one is measured using the frequency shift of the GPS carrier signal with the resolution of 0.2m. Therefore, the carrier signal is often used for precise positioning, such as the RTK GPS which computes the accurate position of the receiver with the reference carrier signal from a base station. The GPS Doppler itself cannot measure the distance independently; however, it is able to measure changes of the distance accurately. Our vehicle localization algorithm takes this advantage to achieve the precise localization results based on the calculation of the velocity from the GPS Doppler measurements.

The relationship between the 3D velocity of the vehicle and the GPS Doppler shift frequency can be described by the following equation,

$$V_v = V_{si} + D_i \cdot C/f_1 + Cbv \tag{1}$$

where V_v is the vehicle velocity in the Earth-Centered Earth-Fixed (ECEF) coordinate system. V_{si} and D_i are the velocity and GPS Doppler shift frequency of the i^{th} visible satellite, respectively. Cbv is the clock bias variation in the GPS receiver. C and f_1 are two constants, representing the velocity of light and the frequency of the carrier signal $L1$. The unknown variables include the 3D velocity of the vehicle and the clock bias variation, which can be solved with the measurements from 4 or more visible satellites.

The GPS Doppler shift is not affected by the ionosphere or troposphere, and the resolution is high, thus, the vehicle velocity can be accurately calculated. However, it is well-known that GPS-based localization often suffers from two problems in urban environments, i.e. insufficient number of visible satellites and multi-path reflections of the signal due to the existence of tall buildings. We alleviate the influences of these problems by the following two techniques.

Tightly coupled integration of INS and GPS In the case of insufficient number of satellites, INS data is often used for integration with GPS systems since it can provide additional information [13]. In the proposed system, the value of the vehicle velocity $\parallel V_v \parallel$ and the yaw rate ω from the in-car INS sensor are used, which can be described by

$$\begin{aligned} \parallel V_v \parallel &= \sqrt{V_{ve}^2 + V_{vn}^2 + V_{vu}^2} \\ \varphi_{t-1} + \omega \cdot \triangle t &= \tan^{-1}(V_{ve}/V_{vn}) + Gb \cdot \triangle t \end{aligned} \tag{2}$$

where (V_{ve}, V_{vn}, V_{vu}) is the 3D velocity vector of the vehicle in the ENU coordinate, which can be transformed from the ECEF coordinate. $\triangle t$ is the time interval, and φ_{t-1} is the vehicle's heading angle calculated from the vehicle velocity at the previous time step. Gb is the yaw rate bias of the INS sensor. With these constraints, only 3 satellites are needed to calculate the 3D velocity vector together with the unknown Cbv and Gb. Moreover, the clock bias variation Cbv is almost constant in the physical property and the yaw-rate bias Gb does not change rapidly, therefore, the vehicle velocity can be computed with only one GPS Doppler signal in the extreme situation by using the values of Cbv and Gb calculated previously. Note that these constraints are only used when the number of visible satellites is not adequate.

Bundle adjustment Given the vehicle's velocity estimated by the above method, the accurate trajectory of the vehicle can be obtained while the absolute positions cannot, since only the relative positions are computed. In the proposed system, the raw GPS pseudorange data is used to provide a "virtual" base station by using the bundle adjustment technique. The conventional positioning methods using GPS pseudorange measure the absolute position of the receiver epoch by epoch, thus, the problems mentioned above may generate great variations between adjacent measurements. However, such variations can be compensated by optimization with the pseudorange data over a period of epochs. Therefore, we preserve a number of the pseudorange data, and then estimate the global position adjustment of the accurate local trajectory in an optimization procedure. Suppose the local trajectory $L = (X^t, Cbv^t)$ is represented in the world coordinate system with the origin at the start point of the trajectory, which can be derived from the precise velocity vectors over the time period T. The absolute positions of the vehicle can subsequently be obtained with a translation vector $J = (\triangle X, \triangle Cbv)$ relative to the "virtual" base station, which is optimized by minimizing the following cost function $E(J)$,

$$E(J) = \sum_{t \in T} \sum_{i \in N(t)} (d_{i,L+J}^t - \rho_i^t)^2 \tag{3}$$

where $N(t)$ is the number of visible satellites at time t. $d_{i,L+J}^t$ is the distance between the satellite i and the estimated absolute location of the vehicle at time t. ρ_i^t is

the distance between the satellite i and the vehicle obtained from the raw GPS pseudorange data. Such a cost function measures the overall alignment differences between the local trajectory and the global absolute locations, therefore, it can greatly reduce the errors caused by the two problems mentioned above in urban environments despite of the coarse resolution, compared with the conventional epoch-by-epoch measuring.

C. Visual Odometry Fused with GPS/INS Localization

Given the vehicle trajectory and absolute positions, the visual odometry is performed to improve the vehicle localization in terms of local accuracy and robustness. The images recorded from a rear parking camera while driving will largely image the ground behind the vehicle. Although many point features can be detected from the road surface in the image, they are individually highly ambiguous. Therefore, rather than feature matching based methods, we choose to perform the visual odometry between consecutive frames as an optimization problem over the road region in the image.

As shown in Fig. 3, in the projective geometry of a moving camera and the ground plane, the map from the point p to p' can be described by a 3×3 homography H induced by the plane π and there $p' = Hp$ [14]. The projection between the image planes actually can be decomposed into three individual projections, including the inter-frame vehicle motion

$$[R|t] = \begin{bmatrix} \cos\alpha & -\sin\alpha & \triangle x \\ \sin\alpha & \cos\alpha & \triangle y \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

the projection from ground to camera

$$T^{gc} = \begin{bmatrix} \cos\phi & -\sin\phi\cos\theta & -\sin\phi\sin\theta h \\ \sin\phi & \cos\phi\cos\theta & \cos\phi\sin\theta h \\ 0 & \sin\theta & -\cos\theta h \end{bmatrix} \tag{5}$$

and the projection from camera to image, i.e. the intrinsic matrix

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where α, θ, ϕ are the yaw, pitch and roll, $\triangle x, \triangle y$ are the vehicle translation, h is the camera height, (f_x, f_y) are the focal lengths, and (c_x, c_y) is the principal point. In the proposed system, we obtain T^{gc} and K by the camera calibration beforehand so that the image pixels can be projected to the body-fixed vehicle coordinate with a given scale defined in the camera calibration.

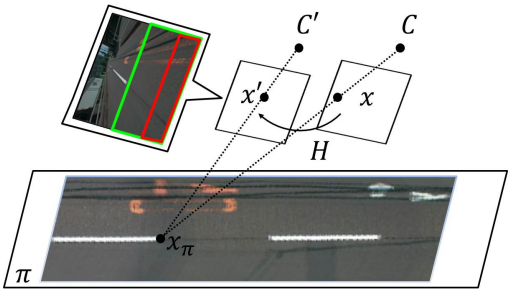

Fig. 3. Relationship of a point on the ground plane between consecutive frames acquired from a rear parking camera while driving. C and C' are the current and the adjacent previous camera locations.

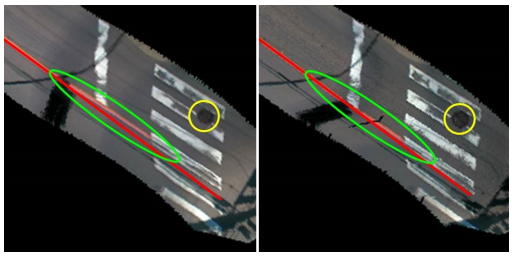

Fig. 4. Example of the comparison between our previous system (left) and the proposed system (right).

Generally, The intrinsic parameters and the camera height barely change during the vehicle driving, therefore, we parameterize the inter-frame vehicle motion and the changes of pitching angle by the vector $\beta = (\triangle x, \triangle y, \triangle\alpha, \triangle\theta)^T$, and optimize it by minimizing the compatible cost $E(\beta)$ in (7) with the Levenberg−Marquardt algorithm (LMA) algorithm between the consecutive frames, which corresponds to the maximum likehood estimation of β.

$$E(\beta) = \sum_{p\in\tilde{P}} \sum_{k} \lambda_k (\Phi_k^t(p) - \Phi_k^{t-1}(p'))^2 \tag{7}$$

where p indicates each road pixel in the computational region of the current frame, and p' is the pixel projected from the same road point in the adjacent previous frame using K, T^{gc} and β. $\tilde{P}$ is the computational region, which is defined as the free road space in a predefined region of interest (ROI) of image, as indicated by the green box in Fig. 3. Here, we adopted our previous work in [15] for free space detection. Here, the ground plane-induced homography between consecutive images is used, instead of the homography between a binocular stereo camera in the original work. $\Phi(\cdot)$ is the feature vector we construct that consists of the intensity value plus a two dimensional gradient vector in the horizontal and vertical axes. The initial values of β are derived from the GPS/INS-based localization results for each and every single frame. A relatively accurate initial values are very important for the optimization in the visual odometry in urban environments due to the possible low-textured roads or occlusion of nearby cars.

Based on the optimization results, we subsequently project the image patch in a predefined ROI, as indicated by the red box in Fig. 3, to the ground plane and then align them into a local bird's eye view image for each local map segment. Note that each road pixel is associated with a weight, which is defined as the inverse proportion of its distance to the camera, since the image calibration result is more accurate in the near range of the camera. Therefore, in the overlapped areas during the image alignment and integration, the pixel with highest weight will be preserved. An example result with the images acquired during a right turn of the vehicle is shown in Fig. 4. For comparison, the result generated by our previous system, which used GPS/INS only, is also given. We can see that the lid in the yellow circle is not round and the orange lane marking is not straight in the result obtained by the previous system. Whereas in the result of the proposed system, these problems are improved greatly. Specifically, the error between the lane marking and the reference straight red line, shown in the green ellipse, was reduced from $0.4m$ to $0.15m$.

Subsequently, the generated local images in the same local

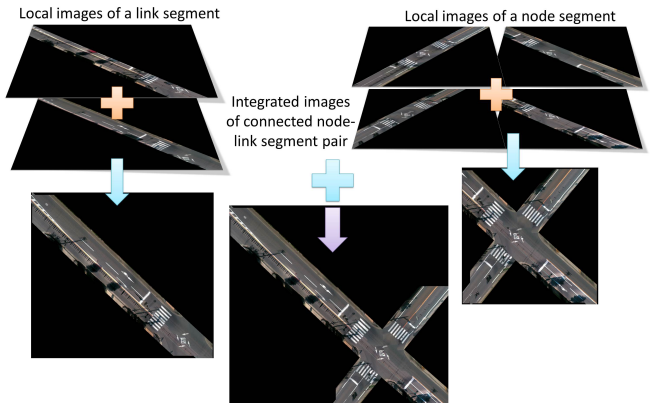

Fig. 5. Example of the orthographic image generation from the local bird's eye view images of the road.

map segment will be integrated. For each pair of local images that overlap partially, a vector $\beta' = (\triangle x', \triangle y', \triangle \theta')^T$ is used to define the transformation between the image pair, which will be optimized by using (7) with the overlapped parts as computational region. In the same way, the integrated local images of the connected node segment and link segment will be further combined to generate the orthographic images of the road. Fig. 5 shows an example that include the image integration inside a map segment as well as the combination between connected segments. Note that the resultant locations of the integrated orthographic images are actually the averaged ones of all the local image in the map segment.

IV. LANE GRAPH CONSTRUCTION BASED ON ORTHOGRAPHIC IMAGES AND TRAJECTORIES

In the proposed system, a driving lane is represented by its centerline associated with additional information such as lane width, lane curvature, etc. Such a representation is more suitable for vehicle navigation and ADAS applications, rather than the conventional representation in which the lane boundaries are used. We model the lane centerlines as smooth curves, instead of dense center point clouds, since their amount of data for storage and processing would be immense. A number of curve representations have been employed for map generation, such as polylines [16], circular arc splines [17], clothoid splines [18], etc. Most of the related works focus on modeling the well-defined roadways, which have specific lateral boundaries, with one curve representation, however, there actually exists another type of the driving lanes for vehicle navigation, which we call transition lanes. Typical transition lanes are the paths that pass through intersections. These driving lanes actually have different requirements for the curve representation, since they usually involve specific traffic rules for the vehicles to follow. Therefore, in the proposed system, we use an approximated clothoid spline to model the normal roadways in a link segment, and employ cubic splines to generate the "virtual" transition driving lanes for intersections in a node segment.

A. Lane Graph Construction of Link Segment

A clothoid is a spline with constant curvature change as a function of arc length. We select it for the normal lane model since it is widely used in road design for constructing the lane boundaries. We use a Taylor series representation of a clothoid, which is a third-order polynomial, to describe the lane centerline in a link segment as

$$y(x) = y_0 + \tan(\varphi_0)(x - x_0) + C_0(x - x_0)^2/2 + C_1(x - x_0)^3/6 \quad (8)$$

where, (x_0, y_0) and φ_0 represent the the origin of the map segment and the orientation (tangent) of the base of the curve, respectively. C_0 is the curvature of the lane, and C_1 is the rate of curvature. W is the lane width.

As shown in Fig. 6, given a local image in a link segment, the lane boundaries are firstly detected. In the proposed system, we adopted our previous work in [19] for detecting lane markings/boundaries, in which, an elongated filter was used in the region-of-interest (on both sides of the vehicle trajectory) to measure the response of difference of oriented (same as the trajectory's direction) means of the image intensity. The large response corresponds to the consistent boundaries that are parallel to the driving direction. Furthermore, a neural network detector was trained for detecting painted lane markings, and another neural network detector was trained for detecting edges by shadows so that such boundaries could be removed. Subsequently, the vehicle trajectories are clustered based on the traveling direction and the Euclidean distance of the trajectory points inside the same lane boundary pairs. Note that the trajectories which cross the lane boundaries, e.g. during the lane change or obstacle avoidance behaviors of the vehicle, will be abandoned since they do not correspond to a single lane. With the lane boundaries and vehicle trajectories, the centerline points inside the lane boundaries are extracted as follows. Let $o = (x, y)$ be a trajectory point inside the driving lane area and r be the rib at (x, y) , then the (x, y, r) defines a circular domain D. Let the cost of a point q in the rib $I(q) = 0$ if the line segment $\overline{oq}$ does not touch the lane boundary, otherwise $I(q) = 1$. To localize the point (x, y) to the center of the lane and find its rib r, we use a deformable circle whose behavior is determined by the following energy function[20]:

$$\varepsilon(x, y, r) = \iint_D (r - \sqrt{u^2 + v^2}) I(x+u, y+v) du dv - r^\rho/\rho \quad (9)$$

where, the first term attempts to keep the circle inside the lane while the second term tries to maximize the radius of the circle. The constant ρ is used to adjust the contribution of the radius r to the energy function. By minimizing (9), we can find a point (x, y) on the centerline of the driving lane with the lane width being $2r$.

Once the centerline points are obtained by the above process, the RANdom Sample Consensus (RANSAC) algorithm is applied to fit the lane centerline based on the clothoid lane model in the link segment. The endpoints and waypoints of link are firstly used to fit the lane model, which provides the rough estimation of the lane curvature C_0 and rate of curvature C_1. Such estimation is subsequently used as the constraints to fit the lane model with extracted centerline points. The use of RANSAC algorithm can remove outliers effectively, thereby enabling accurate and robust centerline generation, and the rough lane parameter estimation with waypoints can prevent overfitting of the lane centerline.

After fitting the clothoid lane model, it is crucial to determine the entry and exit points for the driving lanes. Considering the host vehicle enters a driving lane from

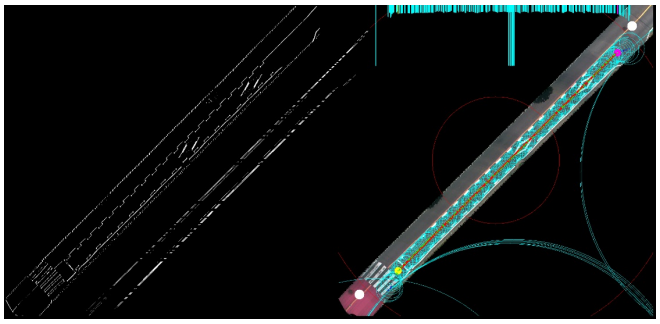

Fig. 6. Example of the lane graph construction of a driving lane. From left to right: detected lane markings, detected lane boundaries, and the constructed lane graph. White dots are endpoints of the link. Cyan circles are the deformable circles and cyan bars on the top of the image show the sizes of the circles. The red line is the fitted lane model. Green dots are inlier centerline points and blue dots are outliers. The inliers between the two red circles are the candidates of entry and exit points. The magenta dot is the determined entry point and the yellow dot is the determined exit point.

an intersection where there are no boundaries, the sizes of the deformable circles along the vehicle trajectory will decrease until the vehicle is inside the lane boundary of the driving lane, and the sizes of the deformable circles will be consistent (for dashed lane markings, the circle sizes keep being consistent for at least a certain of distance). Therefore, we determine the entry point by analyzing the sizes of the deformable circles. For an inlier centerline point i, which is from the first inlier centerline point near the intersection to the rest of inliers between the two red circles in Fig. 6, the mean and variance of the sizes from inlier i to inlier $i+k$ are computed. The first inlier whose mean is in the range of $[t_{r1}, t_{r2}]$ and variance is smaller than t_c will be determined as the entry point. Here, k, t_{r1}, t_{r2}, t_c are predefined parameters. In the same manner, the exit point can also be determined.

For the local map segment with multiple clustered trajectories, the above procedures are performed for each individual trajectory. Finally, the lane graph in the link segment will be stored as $(N_1, N_2, O, W, H, ppm, n_{lane}, en_k, ex_k, l_k)$, which represent the two endpoints (nodes) of the link, ENU coordinate of the upper left corner, ranges in the each and north directions, number of lanes, the entry point, exit point and lane model of k^th lane, respectively.

B. Lane Graph Construction of Node Segment

As shown in Fig. 7, there is no definition for the safe path from lane i (yellow line) to lane j (purple line), however, it does not mean that the vehicle can drive freely in this area. According to the traffic rules, vehicles cannot drive from the exit point ex_i (yellow dot) to the entry point en_j (magenta dot) directly but pass through an inter point m_{ij} (cyan dot) to reduce the chances of collisions. Therefore, We generate a "virtual" lane which is a smooth path (green spline) that passes through m_{ij} from ex_i to en_j while maintaining the continuity at the joint points, ex_i and en_j, between these three lanes.

We employ the cubic Catmull-Rom spline with 5 control points to model the transition driving lane since its control points are actually on the curve and it has local control. These characteristics enable fast fitting of the curve that passes through ex_i, m_{ij} and en_j. Considering the requirements of the transition lane model, we set $\mathbf{P}_1^{ij} = ex_i$, $\mathbf{P}_2^{ij} = m_{ij}$, $\mathbf{P}_3^{ij} = en_j$. In addition, the tangent vector at internal control points of the cubic Catmull-Rom spline is defined by the control points on either side of it, i.e., $T(\mathbf{P}_k^{ij}) = (\mathbf{P}_{k+1}^{ij} - \mathbf{P}_{k-1}^{ij})/2$. Therefore, the continuity at ex_i $(\mathbf{P}_1^{ij})$ can be ensured as long as we set $\mathbf{P}_0^{ij}$ to a location such that line $\overline{\mathbf{P}_0^{ij}\mathbf{P}_2^{ij}}$ is parallel to lane i. So is $\mathbf{P}_4^{ij}$. Note that the cubic Catmull-Rom spline is a third-order polynomial, whose geometric properties, such as curvature, can be easily computed.

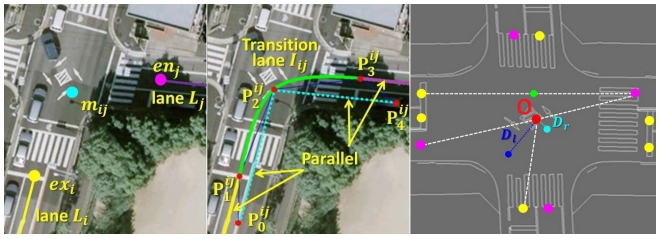

Fig. 7. Illustration of the transition lane model based on the cubic Catmull-Rom splines.

Given the transition lane model to construct the lane graph in a node segment, the entry points and exit points of the driving lanes in its connected link segments are firstly imported. The transition lanes will be generated from each exit point to its feasible entry point in other links. The vehicle trajectories are used to determine both the lane connectivity and its category, i.e. left turn, straight ahead or right turn. Suppose there are N trajectories that pass through the exit point ex_i, for each entry point en_j in other links, if the number of trajectories that travel from ex_i to en_j is more than $N/10$, a transition lane will be generated between them, and its category can be obtained from the change of yaw angle of the trajectories.

Subsequently, the inter point will be determined to complete the transition lane generation. As shown in Fig. 7, the center point O of the intersection is firstly computed by averaging all of the entry and exit points. For the left turn transition lane, the inter point is determined by a point m_{ij} in the bisector of the angle between $\overrightarrow{ex_iO}$ and $\overrightarrow{Oen_j}$ with the distance $\overline{Om_{ij}}$ equals to D_l. In the same manner, the inter point for a right turn lane can be determined with a smaller D_r, since the right turn should be closer to the center of the intersection (for the left-hand traffic). For the straight ahead lanes, the inter point is simply determined as the middle point of $\overline{ex_ien_j}$.

Finally, the lane graph in a node segment will be stored as $(N, O, W, H, ppm, n_{link}, en_k^i, ex_k^i, P[5]_{ks}^{ij})$, which represent the ENU coordinate of the node, ENU coordinate of the upper left corner, ranges in the each and north directions, number of links, the entry point, exit point and control points of the transition lanes, respectively.

It should be clarified that we assume that the neighboring lanes in the lane graph, going in the same direction, are implicitly connected at all points, including merging/splitting lanes. Therefore, it is not necessary to generate transition lanes between them, and the driving behaviors, such as lane changes, merging/splitting maneuvers, etc., are allowed between such neighboring lanes at all points.

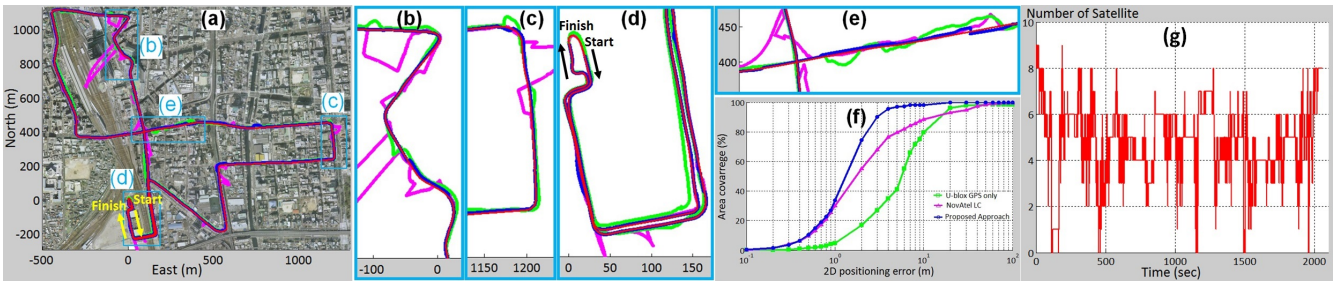

Fig. 8. Evaluation and comparison of the localization accuracies of U-blox GPS only (green), loosely coupled GPS/INS (magenta), the proposed approach (blue) and the ground truth (red). (a) shows the localization results with the data collected in the center of Nagoya city (near Nagoya Station). (b)-(e) show the corresponding portions of the trajectories, indicated by cyan boxes in (a), in greater detail. (f) shows the quantitative evaluations of the localization results against the ground truth. (g) shows the number of visible satellites during the vehicle driving.

V. EXPERIMENTAL RESULTS

In the experiments, two vehicles, each equipped with a conventional low-cost GPS receiver (Novatel OEMV ProPak-V3) and a commonly used rear parking cameras (Kenwood CMOS-200), have been used to collect the road data in urban streets from multiple runs within six month. For reference, a high-precision GPS (Applanix POSLV610) was used to provide the ground truth of the vehicle's position, and another conventional GPS receiver (Ublox LEA-4T) was used for comparison.

A. Vehicle Localization

Fig. 8 shows the evaluation and comparison results between the methods using U-blox GPS only, Loosely Coupled GPS/INS (LC, fusing Novatel GPS and INS observations by Kalman filter) and the proposed approach with the data acquired in the center of Nagoya City, where there are many tall buildings. From Fig. 8 (a)-(e) we can see that the localization results obtained by the proposed approach are in good agreement with the ground truth while other methods have large errors occasionally. Furthermore, the number of visible satellites is less than 4 for over 40% of the driving (Fig. 8 (g)), however, the proposed approach can cover over 90% of the whole trajectory/area with the absolute position error smaller than 3 meters, while the coverage rates with the same error are 35% and 77% for the GPS only and GPS/INS loose coupling methods, respectively (Fig. 8 (f)). The experimental results substantiated effectiveness of the proposed approach in terms of both accuracy and robustness.

It should be noted that the proposed approach has meter-level errors for the absolute positioning in urban environments, since the absolute positions can only be obtained by the GPS sensor and we use a conventional low-cost one. However, the relative positioning is much more accurate. In case of ADAS system with such a map, the map matching technique with the input observations of cameras can correct the offset easily. In future, we will also integrate landmarks in the Layer 3 of the proposed map to achieve higher absolute accuracy.

B. Orthographic Image Generation

We firstly evaluate the image alignment based on the fusion of GPS/INS and camera. Fig. 9 shows an example of the aligned bird's eye view images of the road in a $1.636km$ closed-loop driving. As we can see that the aligned image coincides with the map very well. Furthermore, the enlarged image patch shows the overlapped part of the images and the error between the same point (indicated by the red circle) is only $1.3m$.

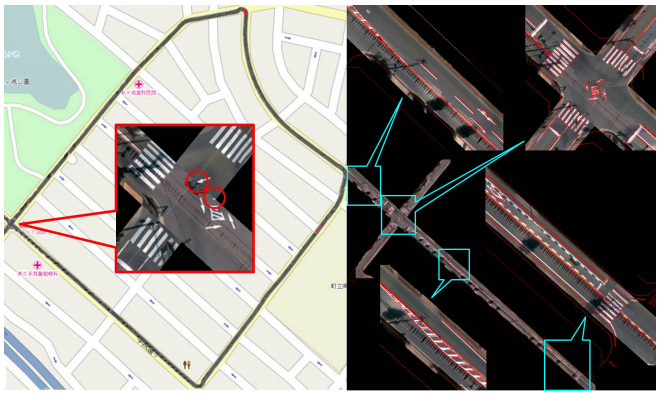

Fig. 9. Quantitative evaluation of the orthographic image generation. Left: image alignment. Right: integrated local image.

Furthermore, we evaluate the spatial accuracy of the map generation by comparing the synthetical orthographic image against the manually labeled ground truth, as shown in Fig. 9. We align the two images at the center of the intersection, and the error at the other end of the longest link (about 260m) is about 40 centimeters. We can also see that the generated orthographic image is in good agreement with the manually labeled ground truth in other links. From the Fig. 9 we can see that the vehicle trajectories/egomotion estimation for synthetical orthographic image generation is accurate due to the fusion of the proposed precise GPS/INS-based localization and optimization-based visual odometry.

C. Lane Graph Generation

An example lane graph at an intersection is shown in Fig. 10. From the example lane graph we can see that the proposed approach can extract the fundamental properties of the lane segments efficiently and generate safe and reasonable paths for driving through the intersections. Furthermore, a quantitative evaluation of the lane graph generation was given in Table I. The success rate of centerline point extraction is a little bit low since the lane boundary detection may not find the correct boundaries in some challenging scenarios. An example is shown in Fig. 11, in which, the left boundary is out of the image range. However, our lane model fitting can remove the outliers of the centerline points, and generate the correct centerline of the driving lane, thanks to the constraint from the waypoints. The errors of the entry/exit point determination are also mainly due to the bad lane boundary detections. Thanks to the lane fitting

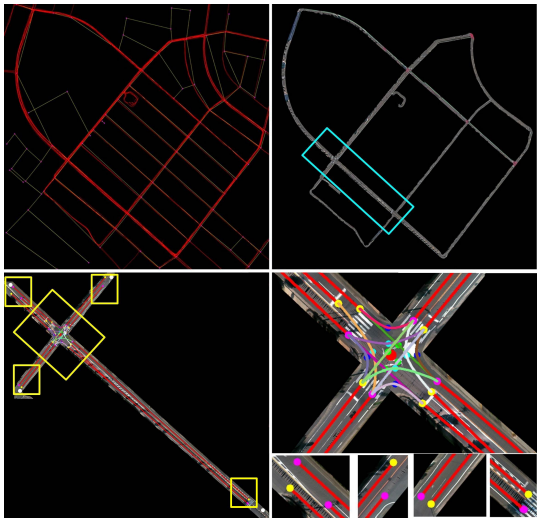

Fig. 10. Example results of the lane graph. Top-left: vehicle trajectories. Top-right: generated local images. Bottom-left: enlarged image of the cyan box. Bottom-right: Enlarged images of the yellow boxes.

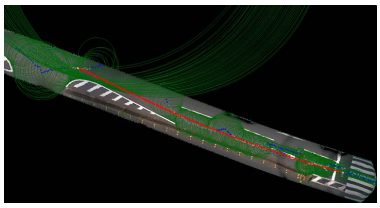

Fig. 11. An example result of centerline point extraction and lane model fitting.

procedure that removed most of the outliers, the success rate of entry/exit point determination is improved compared with the centerline point extraction, since we only compute the mean and variance of the inlier centerline points. As for transition lane generation, it is accurate and robust as long as the errors of the imported entry/exit points are not very big. As shown in Fig. 10, the position of the up-most exit point has some errors, however, the proposed approach can still generate reasonable lanes for it.

It should be noted that the proposed system is not designed for online detection but off-line map generation, therefore, we can just abandon the poor observations/images and only use the good ones, which will improve the system performance.

TABLE I

QUANTITATIVE EVALUATION OF THE LANE GRAPH GENERATION

	Success rate
Centerline point extraction	83%
Clothoid lane model fitting	95%
Entry/exit point determination	92%
Transition lane generation	90%

VI. CONCLUSIONS

In this paper, an approach for automatic generation of lane-level map was addressed for the applications of robotic cars and ADAS systems. Our first contribution is that we focus on the normal vehicles equipped with conventional low-cost sensors in the scope of server-based ITS services, in which, the standard cars could be both service users and the probes. Such a system can offer increased value as well as promote the automation level for today's commercial cars without being supplemented additional sensors. Our second contribution is the precise vehicle localization algorithm by fusing conventional GPS, INS and vision sensors, which outperforms the existing systems in terms of both accuracy and robustness. Our third contribution is the efficient approach for extracting the driving lane information and generating the lane-level digital map structures for both roadways and intersections. Experiments show promising results of efficient precise map generation of the real-world roads, which substantiated the effectiveness of the proposed approach.

Future works will be focused on the following two topics: one is to integrate the visual odometry into the tightly coupled GPS/INS system as the IMU, and the other is to generate 3D maps using these conventional low-cost sensors.

REFERENCES

[1] S. T'Siobbel, et al., Safety digital maps requirements, PReVENT Consortium 2004, Tech. Rep., Sept 2004.

[2] V. Blervaque, et al., PreVENT MAPS&ADAS final report, ERTICO-ITS Europe Std., 2008.

[3] O. Pink, and C. Stiller, Automated map generation from aerial images for precise vehicle localization, Proc. of 13th IEEE Conf. on Intelligent transportation systems, 2010, pp:1517-1522.

[4] R. Kummerle, et al., Large scale graph-based SLAM using aerial images as prior information, Proc. of Robotics: Science and Systems, 2009.

[5] A. Geiger, Monocular road mosaicing for urban environments, Proc. of Intelligent Vehicles Symposium, 2009, pp:140-145.

[6] A. Napier, and P. Newman, Generation and exploitation of local orthographic imagery for road vehicle localisation, Proc. of Intelligent Vehicles Symposium, 2012, pp:590-596.

[7] S. Lovegrove, et al., Accurate visual odometry from a rear parking camera, Proc. of Intelligent Vehicles Symposium, 2011, pp:788-793.

[8] J. Meguro, et al., Road ortho-image generation based on accurate vehicle trajectory estimation by GPS Doppler, Proc. of Intelligent Vehicles Symposium, 2012, pp:276-281.

[9] A. Geiger, et al., Joint 3d estimation of objects and scene layout, Advances in Neural Information Processing Systems, 2011, pp:1467-1475.

[10] A. Ess,et al., Segmentation-based urban traffic scene understanding, Proc. of 20th British machine vision conference-BMVC 2009.

[11] G. Singh, et al., Acquiring semantics induced topology in urban environments, Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA), 2012, pp: 3509-3514.

[12] Open Street Map, Available at http://www.openstreetmap.org/

[13] S. Godha, and M. Cannon, GPS/MEMS INS integrated system for navigation in urban areas, GPS Solutions, 2007, Vol.11, No. 3, pp. 193-203.

[14] R. Hartley, and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge, UK:Cambridge University Press, 2003.

[15] C. Guo, S. Mita, and D. McAllester, Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning. IEEE Transaction on Intelligent Transportation Systems, 2012, 13(3), pp. 1338-1354.

[16] N. Mattern, R. Schubert, and G. Wanielik, High-accurate vehicle localization using digital maps and coherency images, IEEE Intelligent Vehicles Symposium, 2010, pp. 462-469.

[17] A. Schindler, G. Maier, and F. Janda, Generation of high precision digital maps using circular arc splines, IEEE Intelligent Vehicles Symposium, 2012, pp. 462246-251.

[18] K. Baass, The use of clothoid templates in highway design, Transportation Forum 1, 1984, pp. 47-52.

[19] S. Zhu, and A. Yuiller, FORMS: a flexible object recognition and modeling system, Proc. IEEE ICIP94, pp. 465-472.

[20] C. Guo, S. Mita, and D. McAllester, Lane detection and tracking in challenging environments based on a weighted graph and integrated cues, Proc. of 2010 IEEE/RSJ Intel. Conf. on Intelligent Robots and Systems, 2010, pp. 5543-5550.

[21] S. Zhu, and A. Yuiller, FORMS: a flexible object recognition and modeling system, Proc. IEEE ICIP94, pp. 465-472.