Consistent Ground-Plane Mapping: A Case Study Utilizing Low-Cost Sensor Measurements and a Satellite Image

Hang Chu, and Anh Vu

***Abstract*— Vision-aided localization systems are often utilized in urban settings to take advantage of the structured environment, the high availability of unique visual features, as well as complimenting aiding measurements from Global Navigation Satellite System (GNSS). In this paper, we present a case study for roadway texture mapping that combines the low-cost sensor measurements that are already on many production vehicles (e.g. single frequency GPS, wheel odometry, and a forward looking camera) together with a satellite image. The aim of the method presented here is to obtain high resolution texture of the ground plane that is consistent with the low-resolution satellite image through an optimization process that estimates the smooth vehicle trajectory using Maximum-a-Posteriori (MAP). The main benefit of this system comes from the facts that: (1) it utilizes only low-cost sensors and information that are readily available, (2) it can be easily embedded into existing maps. Data and analysis of a drive captured around a block is used in this study.**

I. Introduction

Autonomous vehicle has become an important topic in both robotics and intelligent transportation communities in recent years, accurate estimation of the vehicle state is among one of its core problems that need to be solved. The knowledge of the driving environment is essential for the robotic vehicle to accurately recover its state, hence successfully carry out desired tasks such as complying with traffic rules and ensuring driving safety.

Satellite (aerial) image is an useful resource for providing driving environment information. It has advantages in two aspects. Firstly, taking images from high altitude observing point naturally yields high global consistency. Secondly, satellite image databases are well maintained and some of them are publicly available such as Google maps. However, satellite images, especially the public available ones, suffer from the problem of limited resolution. This makes them less desirable to be directly used in driving assistance systems. Current mainstream precise digital map generation systems provide maps in high quality, but they usually rely on expensive and dedicated sensors as well as great efforts in manual data analysis. These properties limit their benefits to normal commercial cars.

In this paper, we present a study for high resolution lane image generation. We use publicly available low resolution satellite images and widely installed low-cost sensors such as single frequency GPS, wheel odometry in Electronic Stability Programme (ESP) system, and a forward looking camera. Our method produces high quality lane images that are consistent with the satellite image, which makes it easy to be embedded into existing maps. The use of low-cost sensors allows standard vehicles to serve as probe cars, which provides fast and inexpensive coverage of large accessible areas as well as up-to-date information.

*Hang Chu (hc772@cornell.edu) is with School of Electrical and Computer Engineering, Cornell University. Anh Vu (anh.vu@vw.com) is with Electronics Research Laboratory, Volkswagen Group of America.

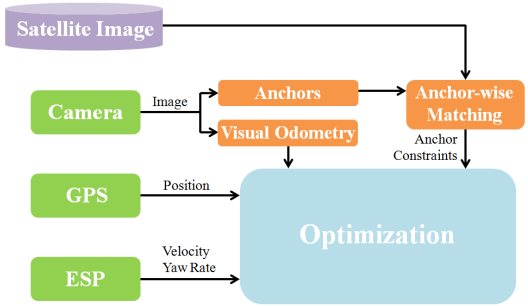

Fig. 1: An overview of our framework.

Fig. 1 shows an framework overview of our method. We propose to identify anchors and match with the satellite image in the unit of anchor. An anchor is a set of consecutive frames that have highly identifiable structured features. Anchors are identified using the orthographical images obtained by the camera. Only identified anchors are matched with the satellite image to compute anchor-wise location constraints. Vehicle states are optimized by summarizing positions measured by GPS, velocity and yaw rate measured by ESP, visual odometry, and the computed anchor constraints. The anchor-based scheme works more effectively than directly correlating camera and satellite images for each single frame due to high uncertainty of single frame matching.

II. Related Work

Great efforts have been made recently to generate high-precision digital maps. However most existing map generation systems commonly require specialized sensors [12,15] and intense manual analysis [1]. This makes them difficult to be applied in large scale and maintain up-to-date. In another category of work such as [2], digital maps are generated using satellite images exclusively. Such method has the disadvantages of expensive data acquisition and unable to recover details that can only be observed in close distance.

In recent years road orthographic image construction systems using only commodity sensors have been developed. The methods in [3] and [4] use vision sensor exclusively. Though visual odometry works fully automatically and produces generally accurate local state estimation, it relies heavily on visual cues and does not work well when insufficient

texture is observed in the image. Furthermore, vision exclusive methods suffer from accumulated errors. This reduces global consistency of their methods by a large margin. The method in [17] uses visual odometry in complement to GPS. However it fails to create road images that are in pixel-level alignment with the satellite image due to limited precision of GPS.

In the work of [5], GPS and INS sensors are coupled and then used for absolute localization. In its recent successor [6], visual odometry and road network topologies are also added to improve localization accuracy. These methods show promising results as well as state-of-the-art accuracy. However, as GPS may suffer from insufficient number of visible satellites and multi-path reflections, it is essentially unable to guarantee high absolute positioning accuracy. Besides, though road network structure information is used in [6] to improve accuracy, sections between road segment intersections are not precisely constrained and thus not accurately mapped.

Methods in [7], [8], and [14] are able to create maps that are consistent with satellite image priors. The satellite image is matched with 3D range scans of outdoor building structures in [7] and [14], and in [8] matched with reconstructed 3D point clouds of the road plane. Although these methods ensure high consistency with the satellite image, but the usage of 3D range scanner or stereo 3D reconstruction systems significantly reduces the scope of available probe vehicles as well as increases their implementation cost. The method in [16] uses only low-cost sensors along with pre-existing map priors. However it requires lane marks labeled digitally, which consumes manual efforts and limits the method's usability in large scale.

The main novelties and contributions of this paper are:

- We present a study that uses conventional low-cost sensors to create high resolution roadway images that are consistent with the publicly available satellite image.
- We propose to use an anchor-based approach that handles camera-satellite image matching effectively and improves mapping accuracy.

III. Approach

In this section we describe the method we use in this study. Our method includes three major components: a preprocessing preparation stage, an anchor-based method for consistent roadway high resolution image generation, and an incremental optimization algorithm that enables real-time map updating.

A. Preprocessing

We perform two preprocessing steps. The first one is computing the GPS-image projection, defined by a homography transformation that transforms a *longitude-latitude* GPS coordinate into its corresponding *x-y* image coordinate. To do this, we survey GPS positions along with image coordinates of 20 points randomly scattered in a $1.5km \times 1.5km$ region. The homography matrix is then computed using Random Sample Consensus (RANSAC).

(a) (b)

Fig. 2: An example of the shadow removing preprocessing, (a) original satellite image, (b) processed image.

The second preprocessing is detecting and removing shadows in the satellite image. Tree or building often project shadows on the lane, which interferes matching camera images because shadow position varies in different time of the day. To detect shadows, we apply watershed segmentation on the image region within a constant distance to the road centreline. Segments with dark color appearances are treated as shadows. We remove the detected shadows by linear interpolation along the road direction. Fig. 2(a) shows an example of the original satellite image, Fig. 2(b) shows an example of processed image after shadow removing.

It should be noted that both preprocessing steps can be easily automated and thus will not impose requirements for extra manual analysis.

B. Anchor-based Method for Consistent Roadway Image Generation

After the preprocessing steps, we compute edges from the processed satellite image and camera inputs. We choose the random-forest-based edge detection method in [9] as it is robust to illumination changes and also computationally efficient. For an initial state, edge maps of transformed orthographic camera input and the nearby satellite image region centered at initial location are computed and matched. The matching correlation is used to compute the optimal position the current frame should be shifted to, thus align the generated map with the satellite image. This is not a trivial task because directly optimize using computed correlations for each single frame will not work. Fig. 3 shows examples of input orthographic images, (a) and (b) show examples that has clear lane marks. These images often produce useful matching results, e.g. Fig. 3(a) produces accurate match result in the direction perpendicular to vehicle heading, Fig. 3(b) provides correct positioning information in both vehicle heading and vehicle heading perpendicular directions. However, not all observed orthographic images contain meaningful textures, such as Fig. 3(c). In that case, matching fails to provide useful alignment information, sometimes even shows strong correlation at false locations.

We further demonstrate why directly matching in single frames fails to produce accurate results by an example shown

(a) (b) (c)

Fig. 3: Examples of transformed road orthographic images. When matched with the satellite image (a) and (b) produce useful location information. In the contrast images like (c) often fail to provide helpful positioning information.

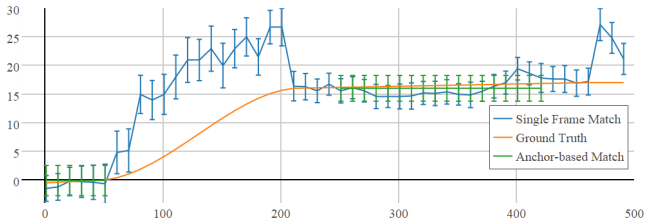

Fig. 4: An illustration of different matching methods. *x-axis*: time as frame number, *y-axis*: adjustment Δx produced by image matching, bars show standard deviations. The anchor-based method automatically turns off bad matching results, also produces smoother and more accurate location constraints than single frame matching within turned on anchor sections.

in Fig. 4. Fig. 4 shows the estimated Δx (the optimal adjust distance in the direction perpendicular to vehicle heading) of different methods from actual data sequence. At frame 80-200 as input images are textureless similar to Fig. 3(c), single frame matching shows strong correlation on wrong locations. To address this problem, we propose to use an anchor-based method. An anchor consists of a set of continuous anchor frames, anchor frames are frames that have highly identifiable structured textures. In the example of Fig. 4 two anchors are found. Frames within an anchor are adjusted uniformly based on all single frame matching correlations. In this way, our method achieves three advantages: (1) it automatically turns off unhelpful matching results, (2) it computes more accurate position constraints by taking into account all correlations within each anchor, and (3) it produces smooth adjustment by moving anchors uniformly, which improves final mapping quality.

In the next, we will describe the method we use for anchor identification, visual odometry, obtaining the optimal state estimations by incorporating anchor constraints and all other measurements , as well as the final image generation using estimated states.

1) Anchor Identification: In this step, our goal is to find whether an orthographic image contains identifiable texture in vertical (vehicle heading or image *y*) or horizontal (perpendicular to vehicle heading or image *x*) directions. First, we sum up edge potentials of the orthographic image along two axes into two 1D signals, i.e.

TABLE I: List of features used in anchor frame identification.

Feature	Definition
mean	$f_1 = \mu\{E_{1D}\}$
variance	$f_2 = \sigma^2\{E_{1D}\}$
peak value	$f_3 = max\{E_{1D}\}$
peak-mean ratio	$f_4 = f_3/f_1$
3 dimensional histogram	$f_5 = 1\{E_{1D} \geq 0.7f_1\}$ $f_6 = 1\{0.3f_1 < E_{1D} < 0.7f_1\}$ $f_7 = 1\{E_{1D} \leq 0.3f_1\}$

$$Ex_j = \sum_{i=1}^{h} E_{ij} \tag{1}$$

$$Ey_i = \sum_{j=1}^{w} E_{ij} \tag{2}$$

where E_{ij} is edge potential of the input orthographic image, h,w are image height and width. To identify anchor frames, we define and extract 7 features from the 1D edge signal. Table 1 lists the features, where E_{1D} can be either Ex or Ey.

We collect two anchor/non-anchor training sets corresponding to the two axes respectively. Then a standard *k*-nearest neighbours (*k*-NN) classifier is applied using the normalized feature vector to determine whether an input should be treated as an anchor frame. We further apply a round-robin buffering scheme to filter the computed labels. This process rejects lone outliers and ensures the continuity of identified anchor sections. Each set of continuous anchor frames forms an anchor, anchor constraints are then computed by averaging matching correlations within anchors.

It should be noted that once a sufficient training set is collected, it can be also used on other areas because road image often show repetitive lane mark patterns.

2) Visual Odometry: As we already have ESP that senses vehicle motion, we want visual odometry process to produce motions that has greater, or at least comparable accuracy to ESP. However most traditional image feature based methods fail to meet this requirement. Thus we use a pixel-level dense stereo optimization approach to achieve high-precision visual odometry.

We follow the method introduced in [6] and [10]. The lower half of the transformed orthographic image is used as the region of interest (ROI) to optimize the 2D vehicle motion $\boldsymbol{\beta}_{vo} = (\Delta x_{vo}, \Delta y_{vo}, \Delta \alpha_{vo})^T$, where $\Delta x_{vo}, \Delta y_{vo}, \Delta \alpha_{vo}$ are the estimated vehicle translation and yaw angle change. The optimal motion vector is obtained by

$$\hat{\boldsymbol{\beta}}^t_{vo} = \underset{\boldsymbol{\beta}_{vo}}{\operatorname{argmin}} \sum_{p \in ROI} \sum_{k} \lambda_k (\boldsymbol{\Phi}^{t+1}_k(p) - \boldsymbol{\Phi}^t_k(p')) \tag{3}$$

where p' denotes the pixel corresponding to p under the current motion $\boldsymbol{\beta}_{vo}$. $\boldsymbol{\Phi}^t_k$ denotes the k-th feature at frame t, with weight λ_k. We use three feature types: image intensity values, edge potential values, and distance transform values

of the edge map. A Nelder-Mead simplex routine is used for minimizing the objective function. Also linear interpolation is used for non-integer pixel coordinates. We use the ESP measurement as initial estimation of $\boldsymbol{\beta}_{vo}$.

3) Optimal State Estimation: At this point, we have derived anchor constraints by matching satellite and camera images at anchor frames, visual odometry derived by matching two consecutive frames, along with GPS and ESP readings. GPS provides noisy absolute location measurement, and it is refined by anchor constraints. Similarly, ESP provides initial relative motion measurement, and it is refined by visual odometry. All measurements and constraints are then used for the final optimization.

The optimization starts with a noisy, non-optimal initial state estimation, e.g. states estimated by using Extended Kalman Filter (EKF) on GPS, ESP, and visual odometry. We then identify anchors using only input images. For a frame inside an anchor, we match its (Ex, Ey) with (Ex', Ey') that are computed similarly but using the satellite image patch centring and oriented at the initial state. Matching correlations are averaged within each anchor to obtain the unified anchor constraints, which give anchor-wise optimal adjustments and covariances. Note that here we assume error propagation inside anchors is negligible.

The optimal trajectory $\boldsymbol{s}^{1:M}$ can be computed iteratively as a nonlinear weighted least squares problem as shown in [11] and [12]. The objective function is defined as

$$\begin{aligned} E(\boldsymbol{s}^{1:M}) = &\sum_{i=1}^{M-1} ({\boldsymbol{p}^i_{esp}}^T \boldsymbol{R}_{esp} \boldsymbol{p}^i_{esp} + {\boldsymbol{p}^i_{vo}}^T \boldsymbol{R}^i_{vo} \boldsymbol{p}^i_{vo}) \\ &+ \sum_{i=1}^{N} {\boldsymbol{p}^i_{gps}}^T \boldsymbol{R}^i_{gps} \boldsymbol{p}^i_{gps} \\ &+ \sum_{i=1}^{K} \sum_{j=1}^{L_i} {\boldsymbol{p}^{ij}_{anchor}}^T \boldsymbol{R}^i_{anchor} \boldsymbol{p}^{ij}_{anchor} \end{aligned} \tag{4}$$

$$\boldsymbol{p}^i_{esp} = \boldsymbol{s}^{i+1} - \boldsymbol{s}^i - \boldsymbol{\beta}^i_{esp} \tag{5}$$

$$\boldsymbol{p}^i_{vo} = \boldsymbol{s}^{i+1} - \boldsymbol{s}^i - \boldsymbol{\beta}^i_{vo} \tag{6}$$

$$\boldsymbol{p}^i_{gps} = \boldsymbol{s}^{u_i} - \boldsymbol{s}^i_{gps} \tag{7}$$

$$\boldsymbol{p}^{ij}_{anchor} = \boldsymbol{s}^{v_{ij}} - \boldsymbol{s}^{v_{ij}}_{init} - \delta \boldsymbol{s}^i_{anchor} \tag{8}$$

where $\boldsymbol{R}_{esp}$, $\boldsymbol{R}^i_{vo}$, $\boldsymbol{R}^i_{gps}$, and $\boldsymbol{R}^i_{anchor}$ are respectively error covariances of ESP measurements, visual odometry at the *i*-th frame, the *i*-th GPS measurement depending on number of visible satellites, and the *i*-th anchor constraint. M, N, and K are numbers of frames, GPS measurements, and identified anchors. L_i is the frame length of the *i*-th anchor. u_i is the time for the *i*-th GPS measurement, and v_{ij} is the time of the *j*-th anchor frame in the *i*-th anchor, all in frame unit. $\delta \boldsymbol{s}^i_{anchor}$ denotes the *i*-th anchor constraint, i.e. unified optimal adjustment of all frames in the *i*-th anchor. We set the yaw dimension in all $\delta \boldsymbol{s}^i_{anchor}$ zero as the 2-axis 1D matching does not count orientation change. We use the version of Levenberg-Marquardt solver implemented in GTSAM [11,13] to optimize the final objective function.

Algorithm 1 The incremental optimization procedure for real-time map updating

```
for t = 1 : M do
  if t == 1 then
    Set s^1 as s^1_gps
  else
    if Detect the end of anchor i, i.e. t == v_{iL_i} then
      if i > k_opt then
        Compute anchor constraints δs^{(i-k_opt):i}_anchor, set
        δs^{i-k_opt-1}_anchor zeros
        Optimize and update s^{v_{(i-k_opt-1)1}:v_{iL_i}}
      else
        Compute anchor constraints δs^{1:i}_anchor
        Optimize and update s^{v_11:v_{iL_i}}
      end if
    else
      Propagate s^M using EKF with s^t_esp, s^t_vo, and
      s^{u_{k'}=t}_gps if has new GPS reading
    end if
  end if
end for
```

4) Final Image Generation: After the optimal states have been estimated, we render all orthographic images onto the initial satellite image background to generate final high resolution roadway image. To deal with overlapping areas, we first compute weights for every pixel in each frame, in inverse proportion to the pixel's distance to the camera. Only the pixel with the highest weight will be shown in the final map. This is because nearer pixels are more accurate as they are less affected by effects such as vehicle bumping. We also find out that keeping only the highest weighted pixel usually yields better image quality than other strategies such as weighted averaging.

C. Anchor-wise Incremental Optimization

The global optimization framework we just described requires all measurement being collected. However this is not desirable in practical mapping applications, especially when the data sequence is long. Therefore we further design an anchor-wise incremental opimization algorithm that allows map updating in real time, as described in Algorithm 1. We combine our above method with EKF. Anchor-based matching and optimization are only triggered at anchor end frames. In addition, we apply optimization and updating only on the most recent k_{opt} anchors, one fixed anchor is also added to ensure continuity of the final trajectory. In our implementation we set k_{opt} as 3.

IV. Experimental Results

In this section we present experimental results of our study. We use an Audi A7 as our test platform, its original GPS, ESP, and an extra added forward looking color camera are used for data collection. Frame rates of the used GPS, ESP, and camera are respectively $0.7fps$, $100fps$, and $30fps$. We stitch image patches downloaded from Google maps as the

TABLE II: Quantitative results of different methods.

	EKF	SFC	CI[6]	**Proposed**
mean error(m)	4.02	2.94	1.16	**0.16**
max error(m)	6.04	12.32	2.74	**0.78**

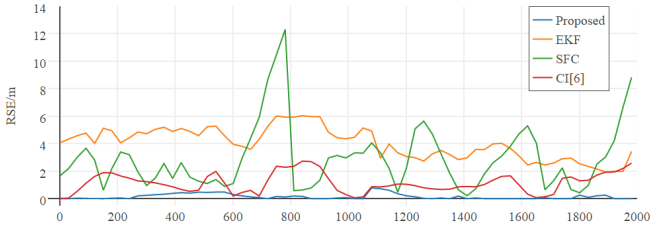

Fig. 5: Diagram of errors over time.

low resolution satellite image prior. The stitched map has resolution of $23.5cm/pixel$. We test our algorithm at Clipper Drive near Belmont, CA. The test sequence has duration of $215s$ and length of $1443m$.

We first evaluate the accuracy of our algorithm. To measure accuracy quantitatively, we select a 2000 frame sub-sequence (duration $66.7s$ and length $439m$) and set one measure point every second. We manually align input orthographic images of measure points with the satellite image, and use them as ground truths. As comparison to our method, we also evaluate 3 other methods using the same data. The compared methods are: (1) Extended Kalman Filtering (EKF): applying EKF on GPS and ESP, without using any image information. (2) Single-Frame Correlation (SFC): using GPS, ESP, visual odometry, and camera-satellite image matching. Rather than anchor identification and anchor-wise constraints, single frame matching correlations are directly used for final optimization. (3) Calibrated Intersections (CI): this is similar to the method described in [6]. GPS, ESP, and visual odometry are used for positioning. The road network is divided into road segments, with the intersections of segments localized accurately. Table 2 and Fig. 5 show quantitative results of all evaluated methods.

From Table 2 and Fig. 5, it can be seen that our method achieves higher accuracy than all three compared methods. In the EKF method no image information is used, thus high accuracy localization cannot be achieved due to limited precision of the low-cost GPS. In the SFC method, in addition to GPS and ESP, visual odometry and camera-satellite image matching are also used. This improves overall accuracy and produces lower mean error than EKF. However as no anchor strategy is used and all single frame correlations are directly sent to optimization, the final estimated states are severely affected by frames that are highly ambiguous and inaccurate. This explains why SFC produces high maximum error. In the CI method [6], 3 road segment intersections (around frame 10, frame 810, and frame 1680) are detected. While intersections are localized accurately, the distance between two intersections are relatively long and drifting error propogates without being refined by any precise location constraints.

Fig. 6: Qualitative results of different methods. Top to fourth row: results produced by EKF, SFC, CI[6], and proposed method. Bottom row: the original satellite image. Areas marked red show that the proposed method has better consistency with the original satellite image.

This leads to error accumulation between intersections and impairs the final performance of CI. Our proposed method outperforms CI as it is able to detect anchors adaptively and calibrate itself using detected anchors. Our proposed anchor strategy automatically turns off inaccurate image matching results, thus avoids the problem of SFC. In this sub-sequence 8 anchors are identified.

In Fig. 6, we show qualitative results of final maps generated by different methods. It can be seen that EKF produces low accuracy and images are not satisfactorily positioned, mainly due to GPS does not have sufficient precision. In SFC sometimes the images are positioned more accurately but sometimes not, also the final map appears to be jagged or twisted. This demonstrated that SFC fails to generate accurate map because inaccurate single frame correlations

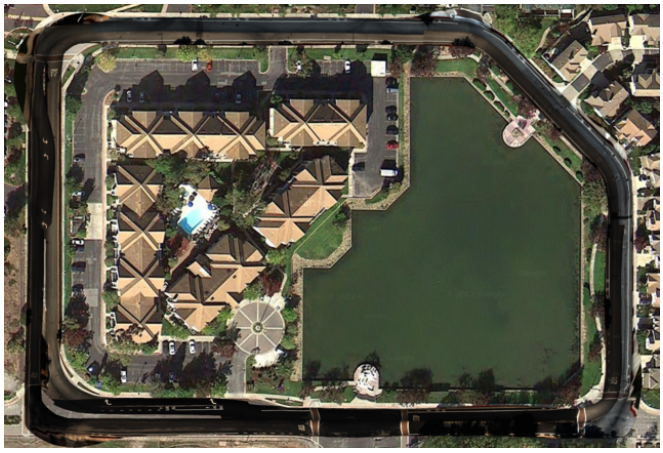

Fig. 7: The final high resolution road image generated in our study.

Fig. 8: Examples from the final generated high resolution road image.

severely interferes final state estimation. Both CI [6] and the proposed method generate high quality maps. The proposed method has better consistency with the original map as aligned anchors impose stronger consistency constraints than calibrated intersections.

Fig. 7 shows the final generated high resolution roadway image and Fig. 8 shows example patches from Fig. 7. In Fig. 8, the first column shows two examples where the generated map shows clear lane marks, the second and third columns show examples that the generated map provides new detailed textures that cannot be observed from the satellite image. These examples imply the usefulness of the generated map in applications such as vehicle localization and intelligent self driving.

Though we are able to create generally satisfactory and useful maps, our method still has a drawback. As shown in the last column in Fig. 8 some parts in the final map are distorted. This is mainly caused by the unflatness of ground and/or sudden vehicle speed change. One way to solve this problem is to estimated full 3D camera poses including pitch and roll angles in visual odometry, instead of only 2D poses in Eq. 3. However we choice not to do so as estimating 3D poses from stereo requires intensive computational power[10], which would significantly upraises the hardware requirement.

An example of our real-time updating incremental optimization procedure is shown in the multimedia attachment of this paper.

V. Conclusions and Future Work

In this paper, a case study for generating high resolution roadway images that are consistent with an existing satellite image was presented. Our method uses publicly accessible map resources and conventional low-cost sensors, indicating more available probe vehicles as well as lower mapping cost. Our generated maps can be easily embedded into existing maps due to its high consistency. An anchor-based method was proposed to interpret and utilize camera-satellite image matching more effectively, hence improve state estimation accuracy and mapping consistency. We conducted experiments and analysis of a drive captured around a block.

Future works will be focused on two topics: investigating the possibility of computationally inexpensive algorithms for full-pose stereo visual odometry, and utilizing the generated high resolution road images for precise vehicle localization.

References

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

[2] 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.

[3] A. Geiger, Monocular road mosaicing for urban environments, *Proc. of Intelligent Vehicles Symposium*, 2012, pp: 140-145.

[4] 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.

[5] 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.

[6] C. Guo, et al., Automatic lane-level map generation for advanced driver assistance systems using low-cost sensors, *International Conf. on Robotics and Automation*, 2014.

[7] R. Kummerle, et al., Large scale graph-based SLAM using aerial images as prior information, *Autonomous Robots*, 30(1), 2009, pp: 25-39.

[8] T. Senlet, and A. Elgammal, A framework for global vehicle localization using stereo images and satellite and road maps, *IEEE International Conf. on Computer Vision Workshops*, 2011, pp: 2034-2041.

[9] P. Dollr, and C. L. Zitnick, Structured forests for fast edge detection, *IEEE International Conf. on Computer Vision*, 2013, pp: 1841-1848.

[10] S. Song, and M. Chandraker, Robust scale estimation in real-time monocular SFM for autonomous driving, *IEEE Conf. on Computer Vision and Pattern Recognition*, 2014.

[11] F. Dellaert, and M. Kaess, Square root SAM: simultaneous localization and mapping via square root information smoothing, *The International Journal of Robotics Research*, 25(12), 2006, pp: 1181:1203.

[12] A. Vu, et al., Centimeter-accuracy smoothed vehicle trajectory estimation, *IEEE Intelligent Transportation Systems Magazine*, 5(4), 2013, pp: 121-135.

[13] F. Dellaert, Factor graphs and GTSAM: a hands-on introduction, *Technical Report*, 2012.

[14] M. P. Parsley, and S. J. Julier, Towards the exploitation of prior information in SLAM, *IEEE/RSJ International Conf. on Intelligent Robots and Systems*, 2010, pp: 2991-2996.

[15] A. Schindler, et al., Generation of high precision digital maps using circular arc splines, *Proc. of Intelligent Vehicles Symposium*, 2012, pp: 246-251.

[16] N. Mattern, et al., High-accurate vehicle localization using digital maps and coherency images, *Proc. of Intelligent Vehicles Symposium*, 2010, pp: 462-469.

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