Multi-Sensor Fusion of GNSS receivers, inertial sensor and cameras for Precise and Reliable Positioning

Patrick Henkel (*,**) and Patrick Burger (***)
* Technische Universität München, Munich, Germany
** Advanced Navigation Solutions (ANavS), Munich, Germany
*** Universität der Bundeswehr, Neubiberg, Germany
patrick.henkel@tum.de, patrick.burger@unibw.de

***Abstract*—A precise and reliable position and attitude information is needed with high availability in many applications.**

In this paper, we describe a multi-sensor fusion of two GNSS receivers, a virtual reference station, an inertial sensor, a camera and geo-referenced satellite/ aerial images. The visual positioning requires a pre-processing of the images: First, the camera images are projected to bird view by a homogeneous projection. A color transformation and correction, a morphological operation, and an edge and corner detection are subsequently applied to extract distinctive image points. Features (e.g. street markers) are determined by matching subsets of these distinctive image points with a template. This matching implies a search of the optimum scaling, rotation and translation of the subset of distinctive image points with respect to the center of gravity of the point cloud. The matching is substantially simplified by using a principal component analysis to determine the orientation of the features.

Finally, the features from the pre-processed camera and satellite/ aerial images are matched with an iterative closest point algorithm. As the satellite/ aerial images are geo-referenced, the absolute position of the camera can then be derived. This absolute position estimate is then integrated in the sensor fusion.

We show that the vision based position information substantially simplifies the GNSS carrier phase ambiguity resolution. The multi-sensor fusion was also verified in a test drive. We observed a substantial improvement of the positioning accuracy compared to a GPS/ INS-only solution.

***Index Terms*—Sensor fusion, Tight coupling, Satellite Navigation, Inertial Navigation, Visual Navigation, SLAM.**

I. Introduction

The autonomous driving of vehicles is coming within the next few years. The navigation of autonomous cars is challenging as a both *precise* and *reliable* position and attitude information is needed in *any environment* at *all times*. A multi-sensor fusion will be performed to achieve the necessary performance. The use of GNSS receivers is attractive as GNSS enables an unbiased absolute position determination with centimeter-level accuracy.

However, the carrier phase integer ambiguities need to be resolved to achieve this high accuracy. This is a non-trivial task as the sum of all ranging errors needs to be smaller than a small fraction of the wavelength of 19 cm. As GNSS signals are often shadowed by buildings, trees, bridges or tunnels, the phase tracking loops frequently loose lock and the ambiguities need to be re-adjusted. Moreover, the use of *low-cost* GNSS receivers/ antennas and inertial sensors implies some additional challenges:

Challenges of low-cost GNSS receivers/ antennas:

- code multipath of several tens of metres
- frequent half and full cycle slips
- lack of timing input and precise synchronization
- single frequency receivers, i.e. no elimination or estimation of ionospheric delays

Challenges of low-cost inertial sensors:

- biases of gyroscope and accelerometer and their variation

The ambiguity resolution remains challenging also over multiple epochs as both position and code multipath are changing over time and, thus, there is only a small redundancy. The ambiguity resolution could be substantially simplified if some additional *independent* position information is available.

This information could come from visual positioning with cameras and geo-referenced satellite/ aerial images.

This paper has two objectives: The first objective is to show the achievable positioning accuracies with a tight coupling of low-cost GNSS and INS. The second objective is the determination of an absolute position from camera images and geo-referenced satellite/ aerial images, and its integration into the multi-sensor tight coupling.

II. Tight coupling of GNSS and INS

In this section, we describe the tight coupling of GNSS and INS. A Kalman filter [6] is widely used since it uses both the measurements and a state space model (describing the vehicle dynamics). It performs a state prediction and state update such that the variance of the a posteriori state estimate is minimized. A Kalman filter is attractive for real-time applications as it performs an epoch-by-epoch processing with moderate memory and processing power requirements.

Measurements:

- Pseudorange, carrier phase and Doppler measurements of two GNSS receivers mounted on vehicle

- Pseudorange and carrier phase measurements of a Virtual Reference Station (VRS)
- 3D acceleration and angular rate measurements of vehicle

Estimated parameters:

- Absolute position, velocity and acceleration of vehicle
- Attitude (roll, pitch, heading) and angular rates of vehicle
- Single and double difference carrier phase ambiguities
- Single difference code multipath parameters
- biases of accelerometer and gyroscope

We use an extended Kalman filter [6] due to the non-linear relationship between the GNSS measurements and the attitude angles. For a detailed description of the tight coupling, we would like to refer to [1], [2] and [3].

The measurement set-up includes

- 2 u-blox LEA 6T GPS receivers (5 Hz)
- 1 MPU 9150 inertial sensor from Invensense (100 Hz)
- 1 Virtual reference station from Axionet (1 Hz)

The arrangement of the hardware is shown in Fig. 1. Two GNSS antennas are mounted on top of the vehicle to obtain its attitude. The antennas are aligned with the longitudinal axis of the vehicle such that no additional corrections need to be applied. The Virtual Reference Station (VRS) serves as a reference station. It is considered as third receiver in our notation. Thus, the geometry of the three receivers is fully described by the attitude baseline $\vec{b}_{12}$ and the RTK baseline $\vec{b}_{13}$ between the VRS and the first (front) GNSS antenna.

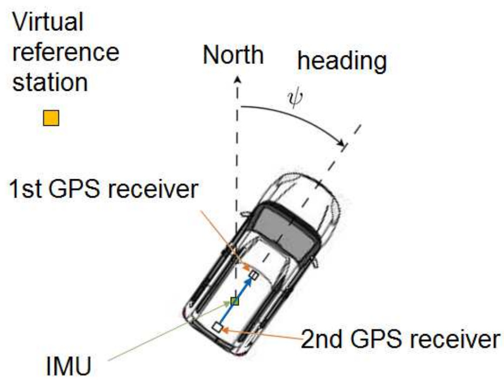

Fig. 1. Measurement set-up at vehicle

Fig. 2 shows the achievable absolute positioning accuracy for the tightly coupled, ambiguity fixed RTK baseline estimation. The enlarged section shows that our (ANavS) position solution differs by less than 10 cm from the geodetic reference (Applanix) solution. Obviously, this error is below the image resolution, which explains the discontinuous curves for both solutions. The float GPS-only solution is also shown and deviates by ~ 75 cm from the reference.

Fig. 3 shows a similar result for a more challenging urban environment: The car drives very close to a multi-story building. An offset of only 35 cm between our (ANavS) and the reference (Applanix) solution can be observed, which is a quite impressive result for this environment and the used low-cost hardware. The offset is rather constant, which indicates a certain error in the ambiguity resolution. The discontinuities in the enlarged sections of the ANavS and Applanix solutions are again an artifact of the limited image resolution.

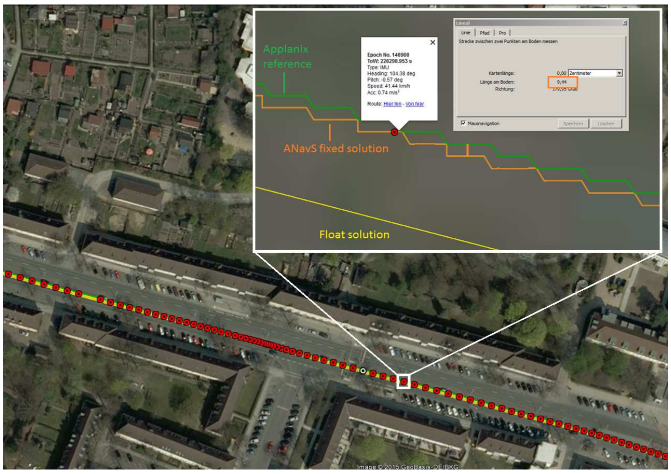

Fig. 2. Comparison of ANavS position solution with geodetic reference system (Applanix) in a relatively ”easy” urban environment. The positions differ by less than 10 cm, which is below the resolution of the images. The float GPS-only solution has an error of ~ 75 cm.

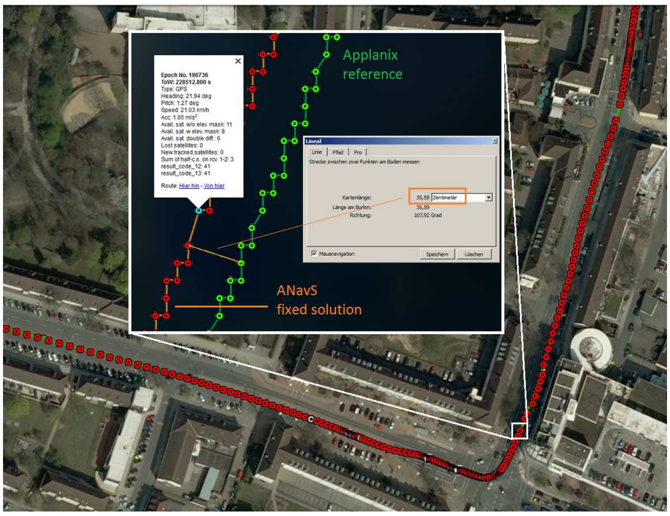

Fig. 3. Comparison of ANavS position solution with geodetic reference system (Applanix) in a ”challenging” urban environment. The positions differ by only ~ 35 cm, which corresponds to the resolution of the images.

Fig. 4 shows the cumulative distribution of the horizontal position error of our low-cost solution for a 30 minutes drive in the city of Wolfsburg. One can observe that in 68% (1σ) of the time the horizontal position error is less than 37.6 cm. This is sufficient for keeping a vehicle on its lane.

However, the cumulative distribution also shows that the position error exceeds 60 cm in 10% of the time. This indicates that the ambiguity fixing, cycle slip correction and IMU bias estimation are sometimes erroneous, and another independent sensor (e.g. camera) is required.

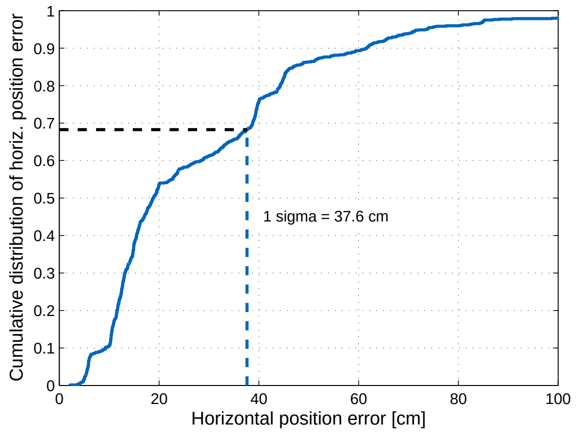

Fig. 4. Cumulative distribution of horizontal position error: The horizontal position error is less than 37.6 cm in 68 % of the time, which is an excellent performance for low-cost GNSS and INS. However, the cumulative distribution becomes quite flat for errors of more than 50 cm. This indicates that the ambiguity fixing, cycle slip correction and IMU bias estimation are sometimes erroneous, and another independent sensor (e.g. camera) is required.

III. Tight coupling of GNSS, INS and Visual Positioning

In this section, we describe the visual positioning and its integration in our tightly coupled sensor fusion. The visual positioning is based on a matching of camera images and geo-referenced satellite/ aerial images (i.e. where the position of each pixel is known) [4]. It is sufficient to use some characteristic image points (e.g. corners of street markers) for visual positioning.

Fig. 5 shows a functional diagram for the sensor fusion. We perform a tight coupling with an extended Kalman filter. It predicts the state vector with a movement model and subsequently updates the state prediction with the measurements either from the GNSS receivers, Virtual reference station, inertial sensor or Visual Positioning Unit, or any combination of these measurements.

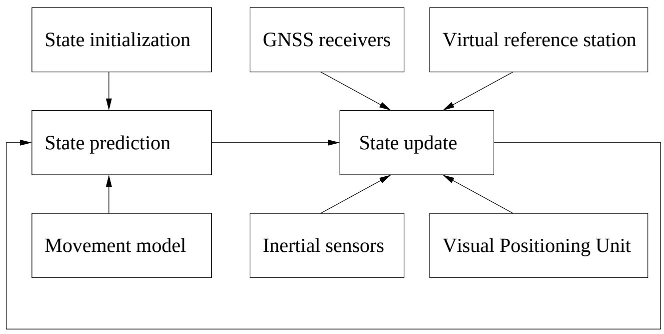

Fig. 5. Functional diagram of sensor fusion with Kalman filter: First, the state vector is initialized and predicted with a movement model. Subsequently, it is updated with measurements. This state update is performed with the measurements either from the GNSS receivers, Virtual reference station, inertial sensor or Visual Positioning Unit, or any combination of these measurements.

A. Image pre-processing

The extraction of these feature points from the camera images involves the following steps: First, a homogeneous projection (Inverse Perspective Mapping - IPM) is applied to transform the camera images from the driver's perspective to bird view. Thereby, we reconstruct a linear relationship between distances in the image and respective distances in the real world.

Fig. 6 was taken from Burger [4] and shows a street marker as taken from the camera's perspective (left subfigure) and from bird view (right subfigure). It is also enlarged to show the high resolution of the street marker. The relationship between the vector of world-frame coordinates $\tilde{\vec{p}}$ and the vector of image-plane coordinates $\vec{p}$ is given by

$$W \underbrace{\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}}_{\tilde{\vec{p}}} = \underbrace{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}_{\vec{p}}, \qquad (1)$$

with the scaling factor W and the homogeneous transformation matrix $\boldsymbol{H}$. The knowledge of $\boldsymbol{H}$ is of essential importance to eliminate the optical distortion and to reconstruct the original shape and angles of an object. In our application, we consider a set of known 2D points $\vec{p}$ of a calibration object (parking lot) of the camera image and the respective set of 2D points $\tilde{p}$ of the real-world. The distances of the latter one were determined with a laser as described in [4].

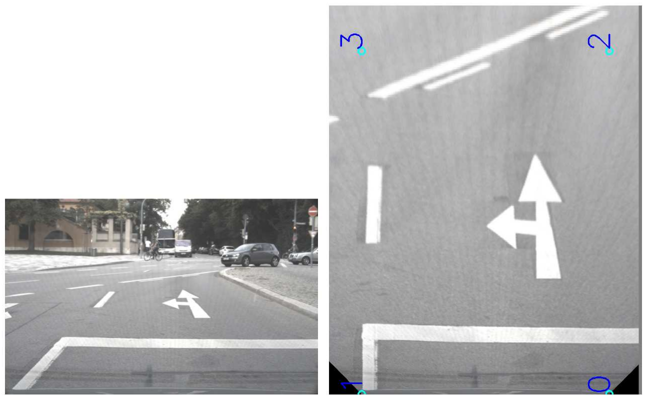

Fig. 6. Homogenous projection: The left subfigure includes the 2D camera image taken from the driver's perspective. The right subfigure shows the enlarged 2D street marker after transformation to bird view using the homogenous projection.

We will briefly derive the determination of $\boldsymbol{H}$. Dividing Eq. (1) by W yields

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \frac{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}{\begin{bmatrix} g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}, \qquad (2)$$

with $W = gx + hy + i$. The denominator describes the effect of the non-linear perspective transformation. The X and Y

components follow from Eq. (2) as

$$\begin{aligned} X &= \frac{ax+by+c}{gx+hy+i} \\ Y &= \frac{dx+ey+f}{gx+hy+i}. \end{aligned} \tag{3}$$

Multiplying X and Y by $gx + hy + i$ and re-arranging gives

$$\begin{aligned} Xi &= ax + by + c - gXx - hXy \\ Yi &= dx + ey + f - gYx - hYy, \end{aligned} \tag{4}$$

which can also be written in matrix-vector notation as

$$\begin{bmatrix} X \\ Y \end{bmatrix} i = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -Xx & -Xy \\ 0 & 0 & 0 & x & y & 1 & -Yx & -Yy \end{bmatrix} \begin{bmatrix} a \\ b \\ \vdots \\ g \\ h \end{bmatrix}. \tag{5}$$

The variable i is a scaling factor that can be absorbed by the other elements $\{a, b, \ldots, h\}$ of $\boldsymbol{H}$, i.e.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -Xx & -Xy \\ 0 & 0 & 0 & x & y & 1 & -Yx & -Yy \end{bmatrix} \begin{bmatrix} \tilde{a} \\ \tilde{b} \\ \vdots \\ \tilde{g} \\ \tilde{h} \end{bmatrix}, \tag{6}$$

and the homogenous projection becomes

$$\tilde{\boldsymbol{H}} = \begin{bmatrix} \tilde{a} & \tilde{b} & \tilde{c} \\ \tilde{d} & \tilde{e} & \tilde{f} \\ \tilde{g} & \tilde{h} & 1 \end{bmatrix}. \tag{7}$$

For n matched feature points, we obtain a system of $2n$ equations with 8 unknowns. This means that we need at least $n = 4$ image points to obtain an exact solution. If more feature points are available, the transformation matrix $\tilde{\boldsymbol{H}}$ is estimated by a least-squares adjustment.

Subsequently, we detect the street markers by performing (a) a color transformation and reduction with adaptive brightness correction to focus on relevant features as shown in Fig. 7, (b) a subsequent morphological operation to enhance the structure recognition, (c) an edge and corner detection to extract feature points, and (d) a point matching of the corner points with a template to recognize the street markers [4].

The extraction of the characteristic image points from the satellite/ aerial images is in principal more challenging due to their lower resolution. Therefore, we use a Maximally Stable Extremal Region (MSER) detector and a Haar feature-based classifier in this step. The matching of the characteristic image points with the satellite/ aerial images can be simplified if the orientation of the camera images corresponds to the orientation of the satellite images. We use a Principal Component Analysis (PCA) to estimate the orientation of point clouds and to determine and compare the magnifying ratio of objects. In particular, it provides robust heading information of distinctive point pattern especially for traffic arrows as shown in Fig. 8.

Street markings are used to determine the absolute position. The big advantage is that they are standardized [7] in size and shape, day and night visibility, chromaticity point etc. As a result, they are distinct, easy to detect and robust in lighting

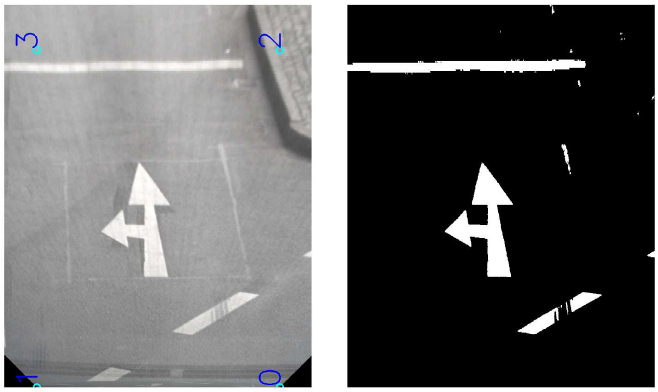

Fig. 7. Camera image in bird-view before and after color transformation: The conversation to gray-scale and black-white is performed to eliminate irrelevant information.

Fig. 8. Principal Component Analysis (PCA) of street markings. The center of gravity is depicted in white and the first principal component vector in red. The 5 point clouds are highlighted in different colors.

and resistant to weather changes. The shape of these patterns is used to detect and classify objects of our camera image. Each pattern is saved in a database with detailed information such as [4]:

- horizontal and vertical length
- center of gravity - COG
- first and second principal component vector
- description.

The absolute heading is obtained from the first principal component/ eigenvector as

$$\psi_k^{ae} = \pi/2 - \arctan\left(y_{\text{pca}}/x_{\text{pca}}\right). \tag{8}$$

B. Localization and Mapping

In this section, we introduce our Visual Simultaneous Localization and Mapping (V-SLAM) which continuously estimates position and heading information of a vehicle in a previously 'unknown' environment. We build a map with available object features from our camera. Additionally, we build a map available from online services like Google maps with a number of landmarks, i.e. street and road markings. Based on the absolute position of these keypoints, the camera position can be estimated. If new street markings become visible, they can be added to our map to improve the accuracy and to keep the point cloud of street markings updated. For real time capability, our system needs a rough absolute position to keep the search time for corresponding point clouds as short as possible.

In our approach, the position and attitude of the vehicle are first estimated in a separate vision system. We call it the Visual Positioning Unit - VPU. In a further step, we provide the output data as an input to our extended Kalman filter with GPS and INS as additional measurements. In addition, we use a key-point based approach for localization and mapping with the following notation [4]:

- positions of n street marking corners are represented in the camera frame as: $\vec{x}^{\,c}_{1,g} \dots \vec{x}^{\,c}_{n,g}$ [pixel] for the camera image I^c_g, $g \in \{1,\dots,f\}$,
- landmark representation by cloud of key points with coordinates $\boldsymbol{p}^{c,l}_n \in R^2$ in camera frame c, with n being the number of corners and l being the landmark/ point cloud id,
- landmark center of gravity $\boldsymbol{cog}^{c,l} \in R^2$ in camera frame c for point cloud id l,
- positions of street marking corners are represented in the aerial frame as: $\vec{x}^{\,ae}_{1,g} \dots \vec{x}^{\,ae}_{n,g}$ [pixel] with the corresponding Google Maps images I^{gm}_g, $g \in \{1,\dots,f\}$.
- landmark representation by cloud of key points with coordinates $\boldsymbol{p}^{ae,k}_n \in R^2$ in aerial frame ae, with n being the number of corners and k being the point cloud id,
- landmark representation by cloud of key points with coordinates $\boldsymbol{p}^{gm,k}_n \in R^2$ in geographic coordinates with $i = 1\dots n$, $j = 1\dots m$ as the horizontal and vertical tile of our Google Maps map and id k.
- landmark center of gravity $\boldsymbol{cog}^{gm,k} \in R^2$ in geographic coordinates for point cloud id k.

In section III-A we have introduced a method to transform our camera image into a so called bird-view where features on the street surface will keep their magnification after transformation and the distances of world planes will be calculated from their perspective images to anticipate the uncertainties of the measurement. In the next step we determine and extract relevant features in each image and evaluate them with a prior known point pattern.

Therefore, the extracted point clouds with coordinates $\boldsymbol{p}^{c,l}_n$ of the camera images $I^c_1 \dots I^c_f$ are stored in a local map. Our point cloud database for aerial images $I^{gm}_1 \dots I^{gm}_f$ is composed of the clouds of points with coordinates $\boldsymbol{p}^{ae,k}_n$ for street marking k.

A GPS-based code-only position estimate is sufficient to load the closer surrounding representation of street marking coordinates $\boldsymbol{p}^{c,l}_n$ as shown in Fig. 9.

The coordinate transformation from $\boldsymbol{p}^{ae,k}_n$ into $\boldsymbol{p}^{c,l}_n$ can be described by a rotation matrix $\boldsymbol{R}$, a scaling factor s and the translation vector $\vec{t}$. These parameters are determined by a least-squares estimation, i.e.

$$\{\hat{s}, \hat{\mathrm{R}}, \hat{\vec{t}}\} = \mathrm{argmin}_{s,\mathrm{R},\vec{t}} \sum_n \left\|\boldsymbol{p}^{c,l}_n - s \cdot \boldsymbol{R} \cdot \boldsymbol{p}^{ae,k}_n - \vec{t}\,\right\|^2. \quad (9)$$

The translation vector can be eliminated by taking the center of gravity of the camera object $cog^{c,l}_n$ (PCA) into account: $\tilde{\boldsymbol{p}}^{c,l}_n = \boldsymbol{p}^{c,l}_n - cog^{c,l}_n$. Eq. (10) simplifies to

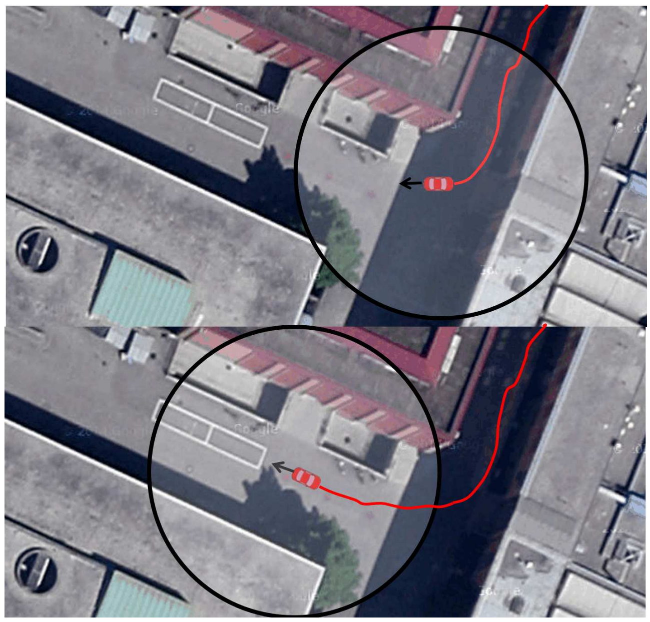

Fig. 9. VPU for street marking localization in closer surrounding to the car.

$$\{\hat{s}, \hat{\mathrm{R}}\,\} = \mathrm{argmin}_{s,\mathrm{R}} \sum_n \left\|\tilde{\boldsymbol{p}}^{c,l}_n - s \cdot \boldsymbol{R} \cdot \boldsymbol{p}^{ae,k}_n\right\|^2, \quad (10)$$

which gives

$$\hat{s} = \left\|\tilde{\boldsymbol{p}}^{c,l}_n\right\|_2 / \left\|\boldsymbol{p}^{ae,k}_n\right\|_2. \quad (11)$$

The minimization with respect to R is performed with a singular value decomposition (SVD), i.e.

$$\mathbf{SVD}\left(\hat{\boldsymbol{p}}^{ae,k}_n \cdot (\hat{\boldsymbol{p}}^{c,k}_n)^T\right) = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \quad (12)$$

and the rotation matrix follows as

$$\boldsymbol{R} = \boldsymbol{V}^T\boldsymbol{U}. \quad (13)$$

Again, we use the principal component analysis to determine an approximation of the difference in orientation angles $\Delta a = |a^{c,pca}_n - a^{ae,pca}_n|$. Accordingly, we rotate the camera object point clouds by the rotation matrix $\boldsymbol{R'}_n(\Delta a)$ to an angle similar to the aerial one and, thereby, to simplify the search space of $\boldsymbol{R}$, i.e.

$$\{\hat{s}, \hat{\boldsymbol{R}}\} = \mathrm{argmin}_{s,\boldsymbol{R}} \left\|\mathbf{R'}_n \tilde{\boldsymbol{p}}^{c,l}_n - s\,\boldsymbol{R} \cdot \boldsymbol{p}^{ae,k}_n\right\|^2. \quad (14)$$

The relative rotation angle $\hat{a}$ is then obtained by $\hat{a} = a + \Delta a$.

Finally, the absolute camera position is determined using the geometric model of Burger [4] given by

$$\vec{p}^{\,g}_{cam} = \boldsymbol{p}^{gm}_n + \boldsymbol{S}^{\,g}_n(\phi')\boldsymbol{R}(\psi)\Delta\vec{p}^{\,n}_{cam} \quad (15)$$

with the geographic coordinates $\vec{p}^{\,g}_{cam}$ of the camera, the geographic coordinates $\boldsymbol{p}^{gm}_n$ of the n-th street marking corner, the rotation matrix $\boldsymbol{R}(\psi)$ with heading ψ, the scaling matrix $\mathbf{S}^{\,g}_n(\phi)$ depending on latitude ϕ' and the known relative position $\Delta\vec{p}^{\,n}_{cam}$ of the camera w.r.t. the street marking in Cartesian coordinates. The scaling matrix $\boldsymbol{S}^{\,g}_n(\phi)$ was introduced because the geographical coordinate system is represented in spherical coordinates (WGS84) and a flat earth model with linear scaling is not representative. For instance, 1° of longitude at the equator is equivalent to approximately 111.321 km, while 1° of longitude at a latitude of 45° is approximately equivalent to 78.849 km. Thus, longitude decreases to zero as the meridians converge at the poles. Thus, the length of one degree of longitude is dependent on the latitude. Additionally, the flattening of the Earth causes a slight variation of the latitude spacing. We use the approach of Sanchez [8] to calculate the scaling matrix $\boldsymbol{S}_n^g(\phi')$ in degrees per meter at a given latitude.

C. GNSS/ INS/ VPU Sensor Fusion

The absolute position information $\vec{p}_{\mathrm{cam}}$ of the camera obtained from the street marker is included in the sensor fusion by augmenting the measurement vector z_n, i.e.

$$z_n = \left[\begin{array}{c} \lambda\varphi_1(t_n) \\ \lambda\varphi_2(t_n) \\ \lambda\varphi_3(t_n) \\ \rho_1(t_n) \\ \rho_2(t_n) \\ \rho_3(t_n) \\ \hline \vec{p}_{\mathrm{cam}} + \vec{b}_{1,\mathrm{cam}} \end{array}\right], \tag{16}$$

with the carrier wavelength λ, the phase measurement φ_r and pseudorange measurement ρ_r of receiver r, and the a priori known baseline $\vec{b}_{1,\mathrm{cam}}$ between the camera and 1st GPS antenna. The measurement covariance matrix is extended respectively, i.e.

$$\Sigma_{z_n} = \left[\begin{array}{cc|c} \Sigma_{\lambda\varphi}(t_n) & 0 & 0 \\ 0 & \Sigma_\rho(t_n) & 0 \\ \hline 0 & 0 & \Sigma_{\vec{x}_1}(t_n) \end{array}\right]. \tag{17}$$

Clearly, the choice of $\Sigma_{\vec{x}_1}(t_n)$ describes the impact of the visual navigation on the sensor fusion. The measurements are used to update the state prediction $\hat{x}_n^-$, i.e.

$$\hat{x}_n^+ = \hat{x}_n^- + K_n\left(z_n - h_n(\hat{x}_n^-)\right), \tag{18}$$

with K_n being the Kalman gain and $h_n(\cdot)$ being the mapping of the state space into the measurement space.

IV. Simulation of potential benefit of visual navigation

In this section, we would like to quantify the potential benefit of visual navigation for GNSS carrier phase ambiguity resolution. Therefore, we set-up an enhanced RTK simulation with absolute position a priori information.

The double difference carrier phase and pseudorange measurements of satellites k and l are modeled as

$$\begin{aligned} \lambda\varphi_{12}^{kl} &= \vec{e}^{\,kl}\vec{b}_{12} + \lambda N_{12}^{kl} + \varepsilon_{12}^{kl} \\ \rho_{12}^{kl} &= \vec{e}^{\,kl}\vec{b}_{12} + \Delta\rho_{\mathrm{MP}_{12}}^{kl} + \eta_{12}^{kl}, \end{aligned} \tag{19}$$

with the sat.-sat. differenced line of sight vector $\vec{e}^{\,kl}$, the baseline $\vec{b}_{12}$, the wavelength λ, the double difference (DD) integer ambiguities N_{12}^{kl}, the DD phase noise ε_{12}^{kl}, the DD code multipath $\Delta\rho_{\mathrm{MP}_{12}}^{kl}$, and the DD code noise η_{12}^{kl}.

We also introduce a state space model: The position and code multipath are simulated as Gauss-Markov processes, i.e.

$$\begin{aligned} \vec{b}_{12}(t_n) &= \vec{b}_{12}(t_{n-1}) + \eta_{\vec{b}_{12}}(t_n) \\ \Delta\rho_{\mathrm{MP}}^{kl}(t_n) &= \Delta\rho_{\mathrm{MP}}^{kl}(t_{n-1}) + \eta_{\Delta\rho_{\mathrm{MP}}^{kl}}(t_n), \end{aligned} \tag{20}$$

and the DD ambiguities are assumed to be constant.

The baseline $\vec{b}_{12}$, the DD ambiguities N_{12}^{kl} and the DD code multipaths $\Delta\rho_{\mathrm{MP}_{12}}^{kl}$ are estimated in a Kalman filter [6] as float parameters. An ambiguity fixing can then be performed based on the estimate of N_{12}^{kl} and its covariance matrix. For a sequential conditional least-squares adjustment, the success rate can be determined in closed form as the conditional ambiguity estimates are uncorrelated. It is given by

$$\begin{aligned} P_{\mathrm{suc}} &= \prod_{k=1}^{K} P_{\mathrm{suc}}^k \\ &= \prod_{k=1}^{K} \int_{-0.5}^{+0.5} \frac{1}{\sqrt{2\pi\sigma^2_{\hat{N}_{k|1,\ldots,k-1}}}} \\ &\quad \cdot \exp\left(-\frac{\left(\varepsilon_{\hat{N}_{k|1,\ldots,k-1}} - b_{\hat{N}_{k|1,\ldots,k-1}}\right)^2}{2\sigma^2_{\hat{N}_{k|1,\ldots,k-1}}}\right) d\varepsilon_{\hat{N}_{k|1,\ldots,k-1}}, \end{aligned} \tag{21}$$

with the conditional variances $\sigma^2_{\hat{N}_{k|1,\ldots,k-1}}$. Teunissen has derived these conditional variances from the triangular decomposition $\Sigma_{\hat{N}_n^+} = LDL^{\mathrm{T}}$ in [5]. The conditional variances correspond to the diagonal elements of D, i.e.

$$\sigma_{\hat{N}_{k|1,\ldots,k-1}} = \sqrt{D_{k,k}}. \tag{22}$$

We make the following assumptions for the measurement and process noises to quantify the benefit of visual a priori information.

Measurement noise assumptions:

- undifferenced phase measurements:
  $[1, 10]$ mm according to satellite elevation
- undifferenced code measurements:
  $[0.5, 1.0]$ m according to satellite elevation
- visual a priori information on position: 25 cm

Process noise assumptions:

- baseline: 1 m
- code multipath: $[2.0, 5.0]$ m according to sat. elevation

Fig. 10 shows that the visual position information enables a substantial reduction of the probability of wrong fixing. Thus, the convergence of the Kalman filter is significantly improved and a reliable fixing becomes feasible within a few seconds even in several multipath environments.

V. Measurement analysis

In this section, the benefit of a tightly coupled GPS/ INS/ VNS (Visual Navigation System) is analyzed with real-data.

We use a 2D Point Grey monocular camera with Full HD resolution and with up to 160 frames per second, and the GPS/ INS hardware as described in Section II. The measurements were taken at the Königsplatz in Munich, Germany, i.e. the vehicle was driving several rounds around the *Propyläen* at

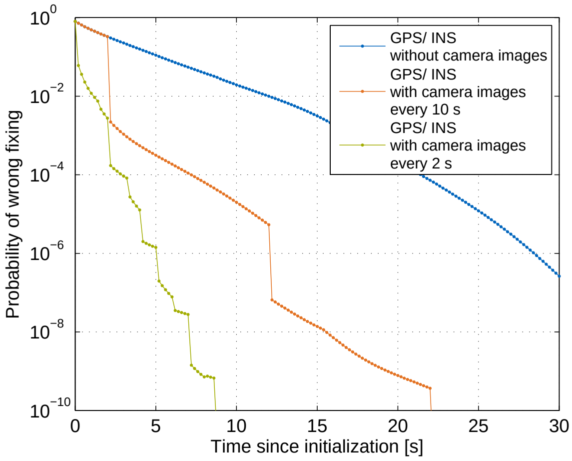

Fig. 10. Reliability of GNSS carrier phase ambiguity resolution: The probability of wrong fixing reduces with increasing time as the Kalman filter converges. The use of camera and geo-referenced satellite images provides an independent absolute position information, which enables a significant reduction of the probability of wrong fixing. A reliable fixing becomes achievable with a few camera images.

Königsplatz. The sensor fusion is performed according to Eq. (16) - (18). In the following figures, we show the GPS-only solution in *red*, the VPU-only solution in *blue* and the tightly coupled solution in *white*.

At the beginning of the trajectory all solutions are very close. Once a street marking is detected, a VPU position information is instantaneously available and the tightly coupled float solution is corrected as shown in Fig. 11. Obviously, the VPU has a relatively high weight in the sensor fusion as it enables an instantaneous correction. The high weight is a consequence of the *float* solution, which depends to a certain extend on code measurements and, thus, has a lower accuracy than the fixed solution.

Fig. 11. Benefit of VPU - first approach at Königsplatz: The VPU-based positioning enables an instantaneous correction of the biased GPS-only and tightly coupled solutions.

The second approach towards the same street marking is visualized in Fig. 12. At the beginning of this trajectory segment (in the top right corner), we can see that the trajectory of the tightly coupled float solution is shifted w.r.t. the GPS-only solution because of the VPU-based correction during the first approach. We can observe that the VPU is tracking a street marking over several epochs which results in a continuous position information. Only the first epoch of the VPU-based position is erroneous as the street marking is still quite far away.

Fig. 12. Benefit of VPU - second approach at Königsplatz: The VPU is tracking a street marker over several epochs which results in a continuous position information. Only the first epoch of the VPU-based position is erroneous as the street marking is still quite far away.

Fig. 13 shows the third approach towards the street marking. One can observe a very continuous position track of the VPU, which improves the tightly coupled solution.

Fig. 13. Benefit of VPU - third approach at Königsplatz: The VPU provides a continuous position information over multiple epochs.

Fig. 14 shows a comparison of the trajectory with and without the VPU of the complete track. The trajectory without VPU is significantly biased and partially lies off the road due to an erroneous ambiguity resolution and/ or cycle slip correction. The integration of the VPU-based position information into the tight coupling corrects for this error and results in an almost unbiased trajectory. A similar benefit can be obtained for the attitude estimation.

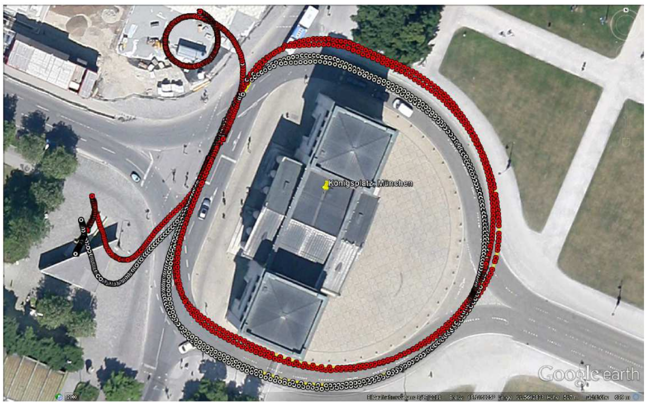

Fig. 14. Benefit of tightly coupled GPS/ INS/ VPU over a tightly coupled GPS/ INS without VPU The integration of VPU enables a correction of erroneous ambiguity fixes and, thus, a precise, unbiased position determination.

VI. Conclusion

The classical GPS/ INS tightly coupled position determination becomes ill-conditioned if an additional code multipath parameter needs to be estimated for every satellite. This is typically needed for precise positioning with low-cost GNSS receivers and patch antennas, which can not suppress the code multipath. We developed and integrated an additional Visual Positioning Unit to solve this problem.

The use of two GPS receivers per car enables an attitude determination. The use of an additional Virtual Reference Station allows also an estimation of the absolute position of a car with centimeter accuracy as atmospheric errors can be corrected. Inertial sensors provide 3D acceleration and angular rate measurements, which enable a seamless position and attitude determination also below trees and bridges but drift over time. Vision-based navigation with cameras and geo-referenced satellite/ aerial images enable an unbiased position and attitude determination based on characteristic street markings. For autonomous driving, precise and reliable position and attitude information is essential. A fusion of all considered sensors is required.

In this paper, we described a tight coupling of GPS/ INS with a Visual Positioning Unit using camera images and geo-referenced satellite/ aerial images. The Visual Positioning Unit of our system detects street markings in the camera and satellite images, and uses extrinsic camera calibration and feature locations in a road marking on an inverse perspective mapped image to estimate the vehicle position with respect to the corner features. We showed that the vision-based position information reduces the probability of wrong GNSS carrier phase ambiguity fixing by several orders. The proposed method was also verified in a test drive. The measurement results show that a lane keeping and tracking is feasible in an urban environment with substantial multipath.

References

[1] P. Henkel and M. Iafrancesco, *Tightly coupled position and attitude determination with two low-cost GNSS receivers*, Proc. of 11th Intern. Symp. on Wireless Communication Systems (ISWCS), pp. 895–900, Barcelona, Spain, Aug. 2014.

[2] P. Henkel, P. Berthold and C. Günther, *Tightly coupled position and attitude determination with two low-cost GNSS receivers, a gyroscope and an accelerometer*, Proc. of Intern. Symp. on Certification of GNSS Systems and Services (CERGAL), DGON, Dresden, Germany, Jul. 2014.

[3] P. Henkel, *Tightly coupled precise point positioning and attitude determination*, IEEE Transactions on Aerospace and Electronic Systems, submitted, 2014.

[4] P. Burger, *Tightly coupled sensor fusion with camera images, satellite images, GPS and INS*, Master Thesis, Technische Universität München, Institute for Communications and Navigation, Dec. 2014.

[5] P. Teunissen, *The least-squares ambiguity decorrelation adjustment: a method for fast GPS ambiguity estimation*, J. of Geodesy, vol. 70, pp. 65-82, 1995.

[6] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman filtering*, John Wiley, 4th edition, 2012.

[7] Deutscher Verkehrssicherheitsrat e.V. and E.V, für Straßenmarkierungen, *Deutscher Verkehrssicherheitsrat e.V., E.V - 2014 - Leitfaden Fahrbahnmarkierung 17*,2014.

[8] J. Sanchez and M. P. Canton, *Space Image Processing*, Vol. 1, 1998.