Precise RTK Positioning
with GNSS, INS, Barometer and Vision

Patrick Henkel *,**, Alexander Blum *,** and Christoph Günther *,**,***
Technische Universität München (TUM), Munich, Germany
**ANavS GmbH - Advanced Navigation Solutions, Munich, Germany*
***German Aerospace Center (DLR), Oberpfaffenhofen, Germany*

BIOGRAPHIES

Patrick Henkel received his Bachelor, Master and PhD degrees from the *Technische Universität München*, Munich, Germany. In 2010, he graduated with a PhD thesis on reliable carrier phase positioning (summa cum laude) and is now working towards his habilitation in the field of precise point positioning. He visited the Mathematical Geodesy and Positioning group at TU Delft in 2007, and the GPS Lab at Stanford University in 2008 and 2010. Dr. Henkel received the Pierre Contensou Gold Medal in 2007, the 1st prize in Bavaria at the European Satellite Navigation Competition in 2010, and the Vodafone Award for his dissertation in 2011. He is one of the founders and currently also the managing director of Advanced Navigation Solutions - ANavS.

Alexander Blum studied Electrical Engineering and Information Technology at the *Technische Universität München*, Munich, Germany. In 2017, he graduated with a master thesis on "Road marking based visual positioning using aerial imagery".

Christoph Günther studied theoretical physics at the Swiss Federal Institute of Technology in Zurich. He received his diploma in 1979 and completed his PhD in 1984. He worked on communication and information theory at Brown Boveri and Ascom Tech. From 1995, he led the development of mobile phones for GSM and later dual mode GSM/Satellite phones at Ascom. In 1999, he became head of the research department of Ericsson in Nuremberg. Since 2003, he is the director of the Institute of Communication and Navigation at the German Aerospace Center (DLR) and since December 2004, he additionally holds a Chair at the Technische Universität München (TUM). His research interests are in satellite navigation, communication and signal processing.

ABSTRACT

The navigation of autonomous ground vehicles and unmanned aerial vehicles requires precise and reliable position and attitude information. A sensor fusion of GNSS, INS, barometer and vision is attractive for unmanned aerial vehicles. The focus of this paper is on high-precision position and attitude determination, i.e. the aiding of vision and barometric height for fast RTK integer ambiguity fixing.

The paper gives an overview of the sensor fusion architecture and the individual processing steps. The visual positioning uses camera images and geo-referenced satellite/ aerial images. An efficient feature matching is described. It uses GNSS/ INS tightly coupled position and attitude information for alignment of the camera and satellite images, and an iterative closest point search algorithm for matching the feature points.

INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are widely used, e.g. for surveying and mapping, for agriculture, for forestry, for security, for monitoring of facilities, and for transport and logistics. A precise and reliable position is needed using sensors of low weight and cost.

A sensor fusion of GNSS, INS and vision is promising and initial work has been performed by a few authors:

30th International Technical Meeting of the Satellite Division of the Institute
of Navigation (ION GNSS+ 2017), Portland, Oregon, September 25-29, 2017

2290

Wang et al. proposed an integrated GPS/ INS/ vision navigation system for UAVs in [1]. The vision system consists of a video camera and a laser range-finder. The velocity and height were estimated by fusion of GPS-based velocity, optical flow, velocity relative to ground from laser range-finder, and of gyroscope angular rates in a Kalman filter. However, no absolute position information was derived directly from vision, and no carrier-phase based RTK positioning was used.

Winkler et al. [2] detected the horizon line with an on-board camera. They estimated pitch and roll angles from the horizon line, and integrated this pitch and roll angle information into GPS/ INS coupling.

Hol analyzed a sensor fusion of inertial sensors, vision, ultra-wideband and GPS in [3]. The fusion of inertial and vision measurements was performed in a smart manner: First, position, velocity and attitude were predicted with a movement model. Similarly, the feature positions were predicted. Subsequently, the true feature positions were searched in the new image around the predicted positions. Once the feature positions were found, the position, velocity and attitude were updated accordingly.

Henkel and Burger developed a sensor fusion of GNSS receivers, inertial sensors and visual positioning in [4]. The visual positioning enabled an instantaneous correction of biased RTK float and fixed solutions. The method was tested with a vehicle at Königsplatz in Munich, Germany.

This paper is organized as follows: In this section, we describe the models for the GNSS, INS and barometric measurements. Section II describes the sensor fusion including the required pre-processing and the GNSS carrier phase ambiguity fixing. Section III includes a detailed description of the visual positioning. Prominent features are extracted from both the camera image and a geo-referenced map, and a feature matching is based on the iterative closest point search algorithm. Section IV describes the re-adjustment of the GNSS/ INS tightly coupled solution with vision. Finally, section V summarizes this paper.

GNSS measurement model

We use three GNSS receivers being mounted on the UAV, and the measurements of a (virtual) GNSS reference station. Double differences (DD) are calculated between the receivers of the UAV and of the reference station for both carrier phase and pseudorange measurements to eliminate orbital errors, clock errors, biases and ionospheric delays. The DD carrier phase measurement of users u and r and satellites k and l is modeled as

$$
\begin{aligned}
\lambda \varphi_{ur}^{kl} \quad := \quad & (\lambda^k \varphi_u^k - \lambda^l \varphi_u^l) - (\lambda^k \varphi_r^k - \lambda^l \varphi_r^l) \\
= \quad & \left(\vec{e}_u^k (\vec{x}_u - \vec{x}_u^k) - \vec{e}_u^l (\vec{x}_u - \vec{x}_u^l) \right) - \left(\vec{e}_r^k (\vec{x}_r - \vec{x}_r^k) - \vec{e}_r^l (\vec{x}_r - \vec{x}_r^l) \right) \\
& + \lambda^k N_{ur}^{kl} + (\lambda^k - \lambda^l) N_{ur}^l + m_{\mathrm{T}}^{kl} \Delta T_{\mathrm{z},ur} + \varepsilon_{ur}^{kl} \\
= \quad & \vec{e}_u^{kl} \vec{x}_{ur} + c_{ur}^{kl} + \lambda N_{ur}^{kl} + m_{\mathrm{T}}^{kl} \Delta T_{\mathrm{z},ur} + \varepsilon_{ur}^{kl},
\end{aligned}
\tag{1}
$$

with wavelength λ^k, normalized line of sight vector $\vec{e}_u^k$ from the k-th satellite to the u-th receiver, the receiver position $\vec{x}_u$, the satellite position $\vec{x}^k$, the DD carrier phase integer ambiguity N_{ur}^{kl}, the single difference carrier phase integer ambiguity N_{ur}^l, the tropospheric mapping function m_{T}^k, the differential tropospheric zenith delay $T_{\mathrm{z},ur} = T_{\mathrm{z},u} - T_{\mathrm{z},r}$, and the DD phase noise ε_{ur}^{kl}. The synchronization correction takes the satellite movement within the differential receiver clock offset into account and is given by:

$$
c_{ur}^{kl} = \left(\vec{e}_u^k (\vec{x}_r - \vec{x}_u^k) - \vec{e}_u^l (\vec{x}_r - \vec{x}_u^l) \right) - \left(\vec{e}_r^k (\vec{x}_r - \vec{x}_r^k) - \vec{e}_r^l (\vec{x}_r - \vec{x}_r^l) \right)
\tag{2}
$$

The DD pseudorange measurement is modeled similarly as

$$
\rho_{ur}^{kl} := (\rho_u^k - \rho_u^l) - (\rho_r^k - \rho_r^l) = \vec{e}_u^{kl} \vec{x}_{ur} + c_{ur}^{kl} + m_{\mathrm{T}}^{kl} \Delta T_{\mathrm{z},ur} + \Delta \rho_{\mathrm{MP},ur}^{kl} + \eta_{ur}^{kl},
\tag{3}
$$

with the DD pseudorange multipath error $\Delta \rho_{\mathrm{MP},ur}^{kl}$ and the DD pseudorange noise η_{ur}^{kl}.

We also use satellite-satellite single difference (SD) Doppler measurements being modeled as

$$
f_{\mathrm{d},u}^{kl} = -f_{\mathrm{c}} \frac{\vec{e}_u^k (\vec{v}_u - \vec{v}_u^k) - \vec{e}_u^l (\vec{v}_u - \vec{v}_u^l)}{c} - \delta \dot{\tau}^{kl} + \eta_{f_{\mathrm{d},u}^{kl}},
\tag{4}
$$

with the carrier frequency f_c, the receiver's velocity $\vec{v}_u$, the satellite's velocity $\vec{v}^k$, the speed of light c, the SD satellite clock drift $\delta \dot{\tau}^{kl}$ and the SD measurement noise $\eta_{f_{\mathrm{d},u}^{kl}}$. As the satellite velocities and clock drifts are known from the broadcast message, the SD Doppler measurements can be rearranged as

$$
\tilde{f}_{\mathrm{d},u}^{kl} = f_{\mathrm{d},u}^{kl} - f_{\mathrm{c}} \left(\vec{e}_u^k \vec{v}_u^k - \vec{e}_u^l \vec{v}_u^l \right) / c + \delta \dot{\tau}^{kl} = -f_{\mathrm{c}} \vec{e}_u^{kl} \vec{v}_u / c + \eta_{f_{\mathrm{d},u}^{kl}},
\tag{5}
$$

which has only the receiver's velocity $\vec{v}_u$ as unknown.

The distances between the GNSS antennas on the UAV can be easily measured with a meter or laser. This baseline length prior information is very valuable for fixing the DD integer ambiguities, and shall be modeled as

$$
l_{ur} = \| \vec{x}_u - \vec{x}_r \| + \eta_{l_{ur}},
\tag{6}
$$

with the measurement noise $\eta_{l_{ur}}$ being typically below 2 cm.

Inertial measurement model

MEMS-based inertial sensors typically include gyroscopes and accelerometers. The gyroscope provides angular rate measurements of the body-fixed (b-) frame with respect to the inertial (i-) frame, being coordinatized in the body-fixed b-frame. These measurements are modeled as

$$\omega_{\mathrm{ib}}^{\mathrm{b}} = R_{\mathrm{n}}^{\mathrm{b}}\omega_{\mathrm{in}}^{\mathrm{n}} + \omega_{\mathrm{nb}}^{\mathrm{b}} + b_{\omega_{\mathrm{ib}}}^{\mathrm{b}} + \eta_{\omega_{\mathrm{ib}}}^{\mathrm{b}}, \tag{7}$$

with R_n^b being the rotation matrix from the local navigation (n-) frame (aligned with north, east and down axes) into the body-fixed (b-) frame, the angular rates $\omega_{\mathrm{nb}}^{\mathrm{b}}$, the measurement bias $b_{\omega_{\mathrm{ib}}}^{\mathrm{b}}$ and the measurement noise $\eta_{\omega_{\mathrm{ib}}}^{\mathrm{b}}$. The rotation matrix $R_{\mathrm{n}}^{\mathrm{b}}$ depends on the roll φ, the pitch θ, the heading ψ, and is given by

$$R_{\mathrm{n}}^{\mathrm{b}} = R_1(\varphi)R_2(\theta)R_3(\psi), \tag{8}$$

with R_i, $i \in \{1, 2, 3\}$, being a rotation matrix for a rotation around the i-th axis. The angular rates $\omega_{\mathrm{in}}^{\mathrm{n}}$ and $\omega_{\mathrm{nb}}^{\mathrm{b}}$ are derived by Jekeli in [5] and are provided for completeness:

$$\omega_{\mathrm{in}}^{\mathrm{n}} = \begin{pmatrix} (\dot{\lambda}_u + \omega_{\mathrm{e}})\cos(\varphi_u) \\ \dot{\varphi}_u \\ -(\dot{\lambda}_u + \omega_{\mathrm{e}})\sin(\varphi_u) \end{pmatrix}, \tag{9}$$

and

$$\omega_{\mathrm{nb}}^{\mathrm{b}} = R_1(\varphi)R_2(\theta)\begin{pmatrix} 0 \\ 0 \\ \dot{\psi} \end{pmatrix} + R_1(\varphi)\begin{pmatrix} 0 \\ \dot{\theta} \\ 0 \end{pmatrix} + \begin{pmatrix} \dot{\varphi} \\ 0 \\ 0 \end{pmatrix}, \tag{10}$$

with the Earth rotation rate ω_{e}, the rate of longitude $\dot{\lambda}_u$, the rate of latitude $\dot{\varphi}_u$, the rate of roll $\dot{\varphi}$, the rate of pitch $\dot{\theta}$ and the rate of heading $\dot{\psi}$.

The acceleration is measured also in the local body-fixed (b-) frame and is modeled as

$$\vec{a}^{\,\mathrm{b}} = R_{\mathrm{n}}^{\mathrm{b}}(\vec{a}^{\,\mathrm{n}} + \vec{g}^{\,\mathrm{n}}) + b_{\vec{a}}^{\mathrm{b}} + \eta_{\vec{a}}^{\mathrm{b}}, \tag{11}$$

with the true acceleration $\vec{a}^{\,\mathrm{n}}$ in the navigation-frame, the gravitational acceleration $\vec{g}^{\,\mathrm{n}} = (0, 0, g)^{\mathrm{T}}$, the measurement bias $b_{\vec{a}}^{\mathrm{b}}$ and the measurement noise $\eta_{\vec{a}}^{\mathrm{b}}$.

Barometric measurement model

We model the barometric air pressure measurement according to Hopfield [6] as

$$p(h) = p_0(h_0) \cdot \left(1 - \frac{\alpha \cdot (h - h_0)}{T_0(h_0)}\right)^{\gamma} + b_p + \eta_p, \tag{12}$$

with the height h, the air pressure p_0 at height h_0, the temperature lapse rate $\alpha = 6.5$ K/km, the temperature T_0 at height h_0, the measurement bias b_p, the measurement noise η_p, and the exponent

$$\gamma = \frac{g}{R_{\mathrm{d}}\alpha} \approx 5.25, \tag{13}$$

where $g = 9.80 \mathrm{m/s}^2$ denotes the gravitational acceleration and $R_{\mathrm{d}} = 287.1\frac{\mathrm{J}}{\mathrm{kg \cdot K}}$ represents the specific gas constant for dry air. The measurement bias can be derived from Eq. (12) using the air pressure $p_0(h_0)$ from a close weather station and the height $h(p)$ of the GNSS/ INS tightly coupled solution.

Visual positioning

Visual positioning with a camera and a geo-referenced map provides an independent position information that is obtained in the following steps: First, an inverse perspective mapping (IPM) is applied to the camera images to transform them from slant view to top view using the UAV's attitude and camera calibration data. In the next step, a feature extraction is applied to focus on the relevant information. Finally, the extracted features of the camera image and map are matched, and a visual position is derived from the position of the matched features of the geo-referenced map.
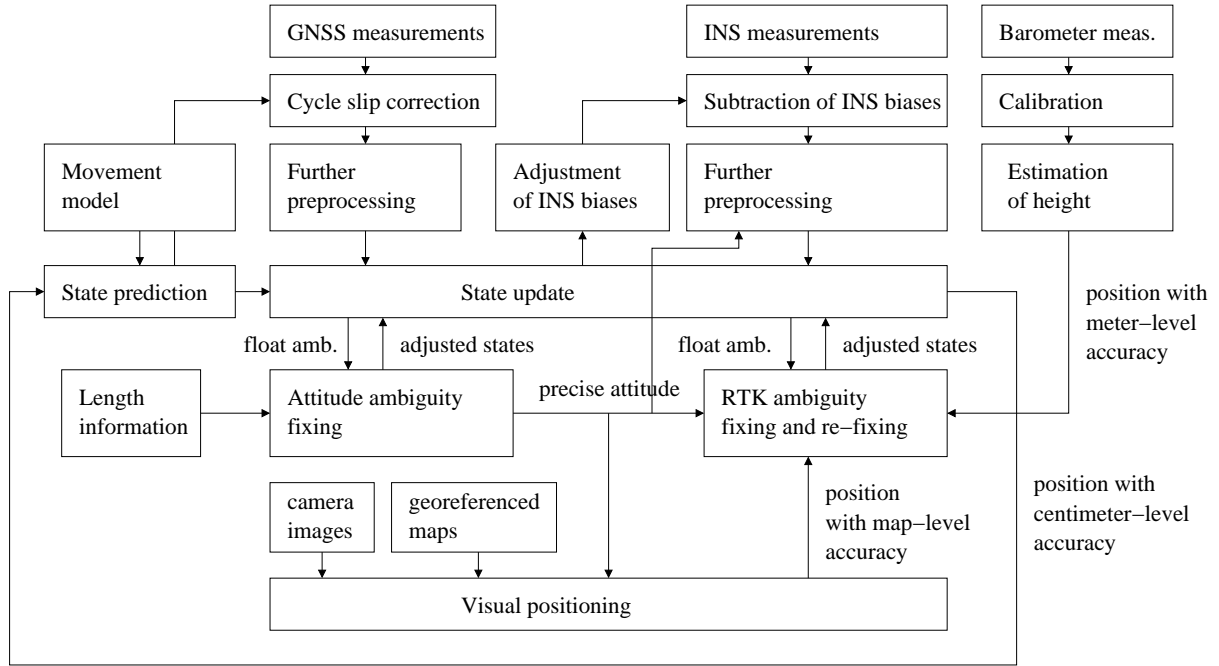
Fig. 1: Sensor fusion with GNSS, INS, barometer and vision: GNSS and INS are tightly coupled. The ambiguities of the attitude and RTK baselines are fixed to achieve centimeter-level positioning accuracy. The barometer and vision provide independent position information with a slightly lower accuracy. Therefore, barometric and visual position information are used to fasten the RTK ambiguity fixing and re-fixing.

SENSOR FUSION

This section describes the sensor fusion of GNSS, INS, barometric height information and visual positioning. We use the measurements from three GNSS receivers to obtain a precise attitude information and additionally use the measurements from a (virtual) reference station for Real-Time Kinematic (RTK) positioning. Fig. 1 shows a block diagram of the sensor fusion. The individual blocks are described in the following paragraph.

We use a Kalman filter (see Brown and Hwang [8]) and perform a tight coupling of GNSS and INS. The Kalman filter performs alternating state predictions and state updates. The state vector includes the absolute position $\vec{x}_u$, velocity $\vec{v}_u$ and acceleration $\vec{a}_u$ of the u-th GNSS receiver, the three-dimensional attitude (roll φ, pitch θ and heading ψ), the angular rates ($\dot{\varphi}$, $\dot{\theta}$, $\dot{\psi}$), the differential tropospheric zenith delay $\Delta T_{z,ur}$, the single and double difference GNSS carrier phase ambiguities N_u^{kl} and N_{ur}^{kl}, and the pseudorange multipath offsets $\Delta_{\rho_r^{kl}}$. The state prediction is performed at every GNSS and INS measurement epoch, and uses the linear movement model

$$\begin{pmatrix} \vec{x}_u(t_n) \\ \vec{v}_u(t_n) \\ \vec{a}_u(t_n) \end{pmatrix} = \begin{pmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{x}_u(t_{n-1}) \\ \vec{v}_u(t_{n-1}) \\ \vec{a}_u(t_{n-1}) \end{pmatrix} + \begin{pmatrix} \frac{\Delta t^3}{6} \\ \frac{\Delta t^2}{2} \\ \Delta t \end{pmatrix} \cdot \dot{a}_u(t_n), \tag{14}$$

with the time step $\Delta t = t_n - t_{n-1}$. The state update uses either pre-processed GNSS or INS measurements. It is important that all state parameters are updated at every epoch as all state estimates are *correlated* and, thus, are affected by an update of a subset of states.

The pre-processing of GNSS measurements includes the following steps:

- selection of satellites based on stability of phase tracking

- selection of common receiver and of reference satellites

- cycle slip correction using predicted state information

- outlier detection and removal

- synchronization of measurements

- estimation of noise statistics

- determination of GLONASS ambiguity transformation (see Henkel et al. [9])

- double differencing using measurements from a (virtual) reference station

The pre-processing of INS measurements includes the following steps:

- subtraction of INS sensor biases

- correction of lever arm effects

- determination and subtraction of the gravity vector (using attitude information) from the acceleration measurements (provided in the body-fixed frame)

The DD carrier phase ambiguities need to be fixed from float to integer numbers to fully benefit from the high accuracy of carrier phase measurements. The DD ambiguities related to the attitude baseline are fixed using prior information on the baseline length as described by Teunissen in [10]. Mönikes et al. [11] and Henkel and Günther [12] used additional prior information on the *attitude* to increase the reliability of integer fixing. Subsequently, the DD ambiguities of the RTK baseline are fixed. The DD measurements of the attitude baseline are also used for RTK ambiguity fixing, i.e. the measurements of the attitude baselines, the measurements of the (virtual) reference station and the fixed attitude are combined and expressed in terms of the RTK baseline as described by Teunissen in [13]. The ambiguity fixing is performed with Teunissen's integer decorrelation and sequential tree search [14]. The fixing decision is based on the *statistics* of the weighted sum of squared ambiguity residuals given by

$$\| \hat{N}_{ur} - \check{N}_{ur} \|^2_{\Sigma^{-1}_{\hat{N}_{ur}}} , \tag{15}$$

and on the drift of the fixed DD phase residuals

$$\Delta r_{ur}(t_n) := (\Delta r_{ur}^{1l}(t_n), \dots, \Delta r_{ur}^{kl}(t_n))^{\mathrm{T}}, \tag{16}$$

with the time-differenced DD fixed phase residual

$$\Delta r_{ur}^{kl}(t_n) = r_{ur}^{kl}(t_n) - r_{ur}^{kl}(t_1), \tag{17}$$

and the DD fixed phase residual

$$r_{ur}^{kl}(t_n) = \lambda \varphi_{ur}^{kl}(t_n) - (\vec{e}^{\,kl})^{\mathrm{T}} \breve{b}_{ur}(t_n) - \lambda \check{N}_{ur}^{kl}. \tag{18}$$

A reliable RTK fixing with single frequency GNSS receivers/ antennas takes a few minutes due to the need for pseudorange multipath estimation (e.g. Henkel and Sperl [7]). Thus, a fixed solution can hardly be kept in urban environments with very limited satellite visibility and frequent losses of lock below bridges and/ or in tunnels. However, a fast RTK fixing can be achieved if an independent prior information on the position with a 1 m accuracy is available.

In this paper, the prior position information is obtained from visual positioning with a camera and a geo-referenced map (satellite/ aerial). The accuracy of the visual positioning is mainly limited by the accuracy of the map but also affected by errors of the feature extraction and matching. The visual positioning uses the tightly coupled position and attitude to align the camera image with respect to the map.

An additional prior information on the height is derived from the barometric air pressure measurement. The barometer is calibrated with the GNSS/ INS tightly coupled height information once an RTK fixed solution is available. The calibrated barometric height information is used for RTK re-fixing.

VISUAL POSITIONING

In this section, the visual positioning with a camera image and a geo-referenced map is described. The section is divided into three parts: the feature extraction of the camera image, the feature extraction of the satellite/ aerial image, and the feature matching.

(a) Camera image: slant perspective of street markings.　　　(b) Camera image, projected to top view.

Fig. 2: Transformation of camera image with Inverse Perspective Mapping.

Feature extraction of camera image

The process steps for feature extraction of the camera image are provided in the following list. The steps are applied to the camera image of Fig. 2a and the results of the most important steps are shown in Fig. 2b - 4 (see also Blum [15]).

- Inverse Perspective Mapping of camera image from slant view (Fig. 2a) to top view (Fig. 2b)
- Edge detection by subtraction of sufficiently blurred projected image from projected image (Fig. 3a)
- Application of threshold on brightness to enhance contrast and to eliminate (some) outliers (Fig. 3b)
- Application of morphological transformations:
    - Erosion and dilatation of image to eliminate small erroneous feature points
    - Dilatation and erosion to join broken feature points
- Polygonisation of features (Fig. 4)
- Selection of features based on size and shape

In principle, the subtraction of the blurred image from the raw image corresponds to an edge detector. However, we used a Gaussian filter with a large standard deviation of 2 m for blurring. In this case, the inner parts of the road markings also remain bright after subtraction of the blurred image.

Feature extraction of satellite/ aerial image

The feature extraction of the map (satellite/ aerial image) is more challenging than the feature extraction of the camera image due its lower image resolution. Nevertheless, we use a similar processing:

- Transformation of map (satellite/ aerial image) to grayscale (Fig. 5a)
- Edge detection by subtraction of sufficiently blurred map from map (Fig. 5b)
- Application of threshold on minimum brightness to increase contrast and to remove outliers (Fig. 6)
- Application of two additional masks (Fig. 7) to reduce noise (Fig. 8)
    - Application of mask for minimum brightness
    - Application of masks for elimination of regions with dark spots
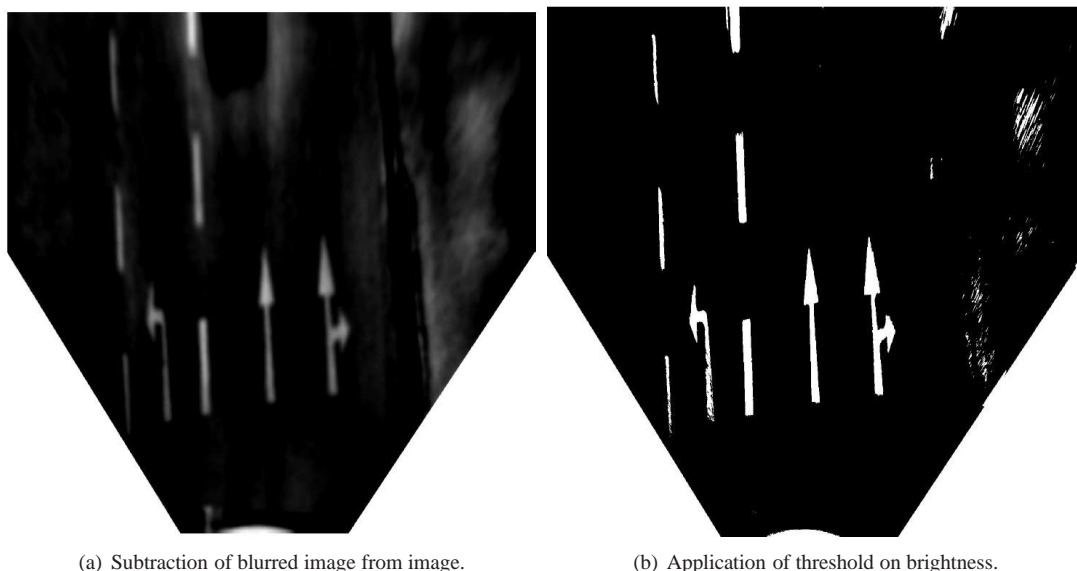
(a) Subtraction of blurred image from image.   (b) Application of threshold on brightness.

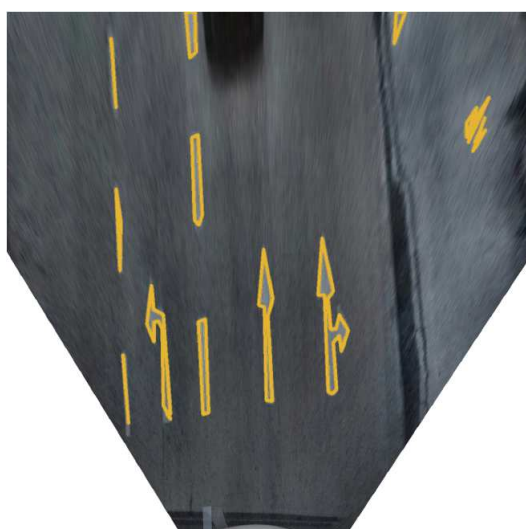Fig. 3: Highlighting of features with edge detection.

Fig. 4: Camera image with extracted road markings

Matching of feature points from camera image and satellite/ aerial image

In this subsection, we describe the matching of the feature points of the camera image and of the geo-referenced satellite/ aerial image. It is assumed that a precise GNSS/ INS tightly coupled attitude and a less precise GNSS/ INS tightly coupled position are available. The different accuracy levels typically occur in areas with very limited or no satellite visibility, in which the GNSS attitude ambiguities are still fixed but the GNSS RTK ambiguities are only float.

The precise attitude information is used to rotate the camera image to be aligned with the geo-referenced satellite/ image. The position information is assigned to the camera image and additionally used to select the relevant geo-referenced satellite/ aerial image. The camera is assumed to be calibrated, i.e. the scaling factor is known and both images are scaled with equal factors.

The feature points of the properly scaled *satellite/ aerial image* differ still in a few aspects from the feature points of the properly rotated *camera* image:

- Translation by common offset due to errors in GNSS/ INS position and errors in position of satellite/ aerial maps

- Additive Gaussian noise due to different pixel sampling dependent on feature extraction

(a) Satellite image in gray scale.

(b) Subtraction of blurred image.

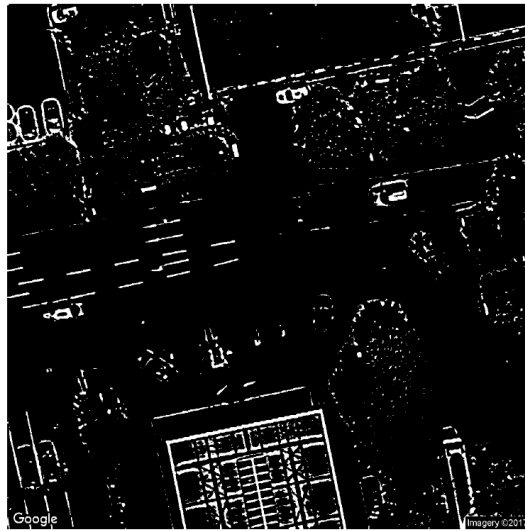Fig. 5: Feature detection in satellite image using edge detection.

Fig. 6: Satellite/ aerial image with hightlighted bright regions, obtained by application of threshold on brightness.

- Unknown relation between feature points of camera image and feature points of the satellite/ aerial image

- Reduced number of feature points in satellite/ aerial image compared to camera image due to lower image resolution

The latter two aspects can be modeled by a permutation matrix P and a selection matrix S.

In this paper, the two-dimensional pixel coordinates of the k-th feature point of the properly rotated (using heading from GNSS/ INS tight coupling) *camera image* are denoted by

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix}. \tag{19}$$

A model for the two-dimensional pixel coordinates of the j-th feature point of the (properly scaled) *satellite/ aerial image* is obtained from the k-th feature point of the (properly rotated) *camera* image by

$$\begin{pmatrix} \bar{x}_j \\ \bar{y}_j \end{pmatrix} = S \cdot P \cdot \begin{pmatrix} x_k + \Delta x + \eta_{\Delta x_k} \\ y_k + \Delta y + \eta_{\Delta y_k} \end{pmatrix}. \tag{20}$$
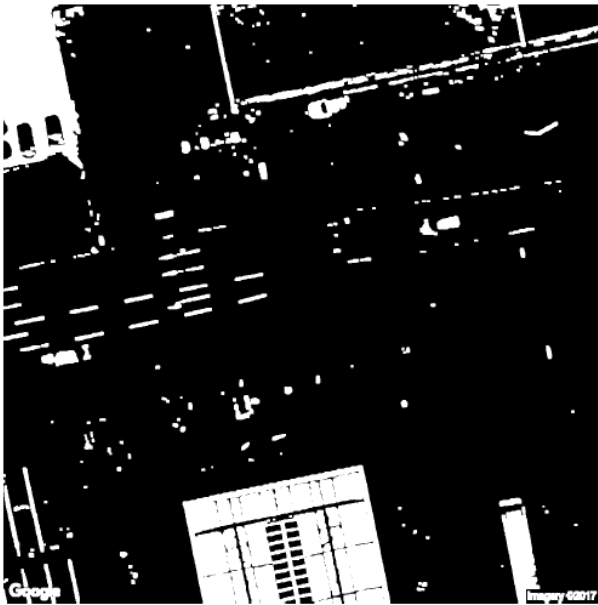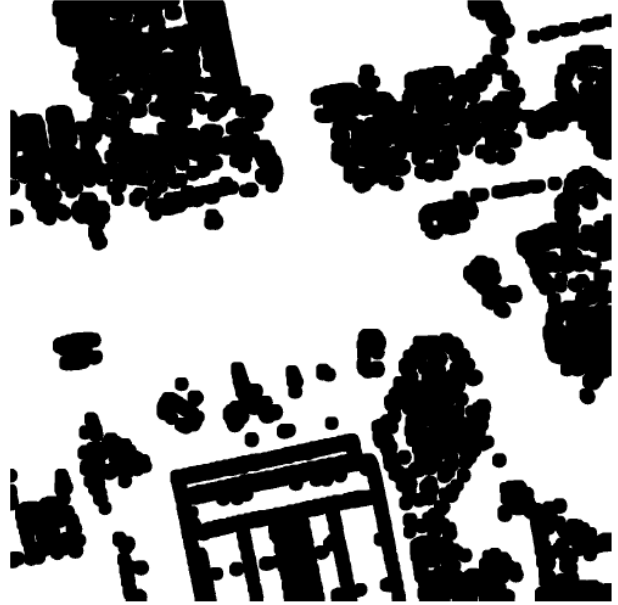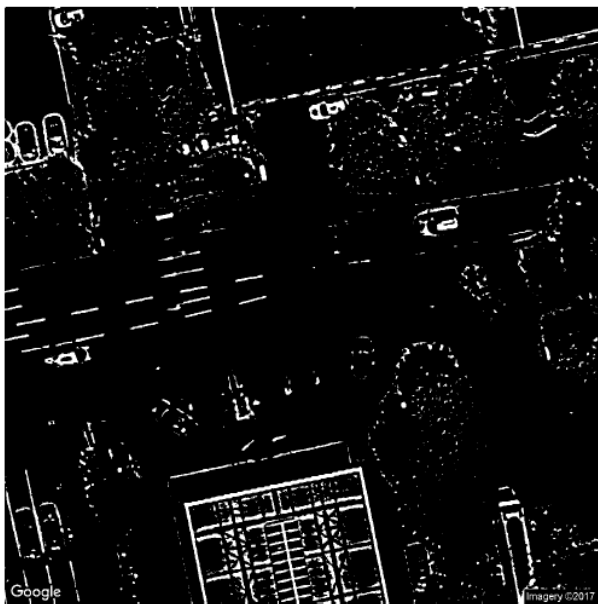
(a) Mask for minimum brightness   (b) Mask for elimination of regions with dark spots
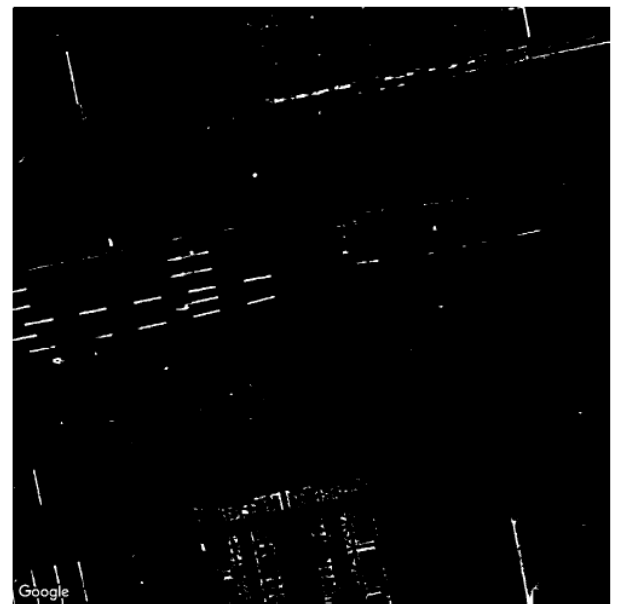
Fig. 7: Masks for improving feature extraction

We estimate the translation Δx and Δy with the Iterative Closest Point Search Algorithm 1.

We simulated a point cloud of the extracted features (arrow and some superposed noise) of the *camera* image. Fig. 9 shows this point cloud with + markers. We also simulated the point cloud of the extracted features of the *satellite* image (x markers) using Eq. (20) and the point cloud of the camera image. A random permutation and a sub-sampling of 80% was applied, and Gaussian noise with standard deviations of $\sigma_{\eta_{\Delta x_k}} = 5$cm and $\sigma_{\eta_{\Delta y_k}} = 5$cm was added. Fig. 9 shows also the point cloud corresponding to the *satellite* image and the mapping of six exemplary feature points. Obviously, the point cloud matching is a

(a) Application of mask for minimum brightness   (b) Application of masks for both minimum brightness and for elimination of regions with dark spots

Fig. 8: Application of masks for improving feature extraction to satellite image

---

Algorithm 1 Point cloud matching

1: Iterative solution:
2: **for** $i = 1$ to n_i **do**
3:     Loop over all satellite/ aerial pixels:
4:     **for** $j = 1$ to n_j **do**
5:         Search of closest camera pixel:
        $j_{\min} := \arg \min_k (\bar{x}_j - x_k)^2 + (\bar{y}_j - y_k)^2$
6:         Determine translation vector:
        $\Delta_j := (\bar{x}_j - x_{j_{\min}}, \bar{y}_j - y_{j_{\min}})^{\mathrm{T}}$
7:     **end for**
8:     Average translation vectors:
    $\bar{\Delta} := 1/n_j \sum_{j=1}^{n_j} \Delta_j$
9:     Scaling of averaged translation vector:
    $\bar{\Delta} := \gamma \cdot \bar{\Delta}$
10:    Apply translation:
    $\bar{x}_j := \bar{x}_j - \bar{\Delta}_1, \bar{y}_j := \bar{y}_j - \bar{\Delta}_2$
11: **end for**

---

non-trivial task since both the mapping/ permutation and the translation are unknown.

Fig. 10 visualizes the point cloud matching with the iterative closest point seach algorithm. The feature points of the satellite image are iteratively shifted towards the feature points of the camera image. These intermediate solutions are indicated with · markers. A scaling factor of $\gamma = 2$ was used to fasten the convergence. The final solution is chosen with o markers and coincides with the feature points of the camera image.
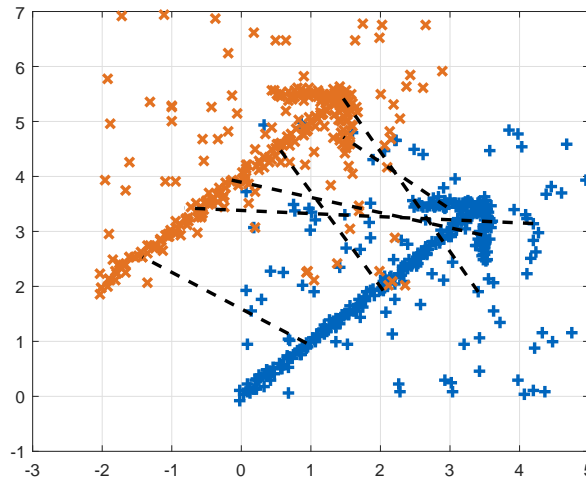

Fig. 9: Point clouds of camera and satellite images: Each point cloud includes an arrow and noise. The point clouds are shifted, permuted, sub-sampled and superposed by additional noise. The matching is shown for six exemplary points, and visualizes the randomness of the mapping.

Fig. 11 shows the position matching error as a function of the number of iterations. The error floor of ~ 5 cm is caused by the additive Gaussian noise in the simulated point cloud of the satellite image. The convergence behaviour depends also on the number of extracted feature points/ image resolution. A large image resolution is not necessarily an advantage since a large number of close feature points might lead to biased solutions (local optima) of the iterative closest point search algorithm. The error floor is only reached with an image resolution of 2 cm/ pixel or lower, and the fastest convergence is achieved for an image resolution of only 0.2 m/ pixel.

The point cloud matching is now applied to the real camera image of Fig. 4 and the real satellite/ aerial image of Fig. 8. Fig. 12 shows the merged features of both images. The features of the camera image were rotated and aligned using the GNSS/ INS tightly coupled position and heading. One can observe that the heading accurately agrees between the feature points of both
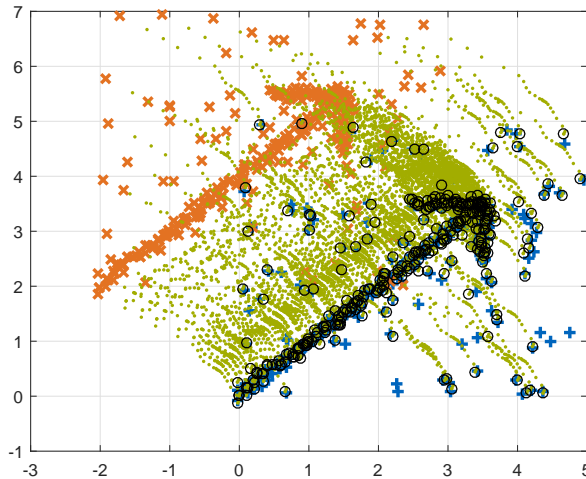
Fig. 10: Matching of camera and satellite/ aerial images: The matching is determined with the iterative closest point search algorithm. It starts with the point clouds described by x-markers. The intermediate solution after each iteration is shown as dots. The final point cloud is shown with o-markers, and accurately matches the point cloud marked with +.
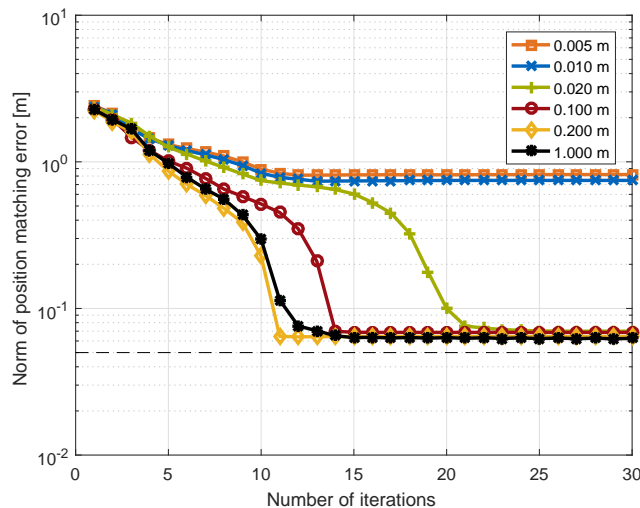

Fig. 11: Convergence of iterative point cloud matching: The matching error reduces with increasing number of iterations until an error floor is reached. This error floor is described by the variance of the additive Gaussian noise. The convergence depends also on the pixel resolution: A very high pixel resolution of 5 mm/ pixel results in a local optimum. A moderate resolution of 0.2 m per pixel enables a faster convergence.

images but that there is a position offset of a few meters. Obviously, the street markings of the satellite image appear much more noisy due to the lower image resolution.

Fig. 13 shows the aligned feature points of the camera and satellite/ aerial image. The alignment was performed with the iterative closest point search Alg. 1. The residual error is less than 50 cm.

RE-ADJUSTMENT OF GNSS/ INS/ BAROMETER TIGHTLY COUPLED POSITION WITH VISION

The GNSS/ INS tightly coupled solution is drifting in environments without satellite visibility, e.g. in tunnels or below bridges. For low-cost sensors, the position error grows above decimeter-level within a few seconds. Fig. 14 shows both the tightly coupled GNSS/ INS RTK position (after a longer GNSS outage) and the vision-based position. The latter one indicates
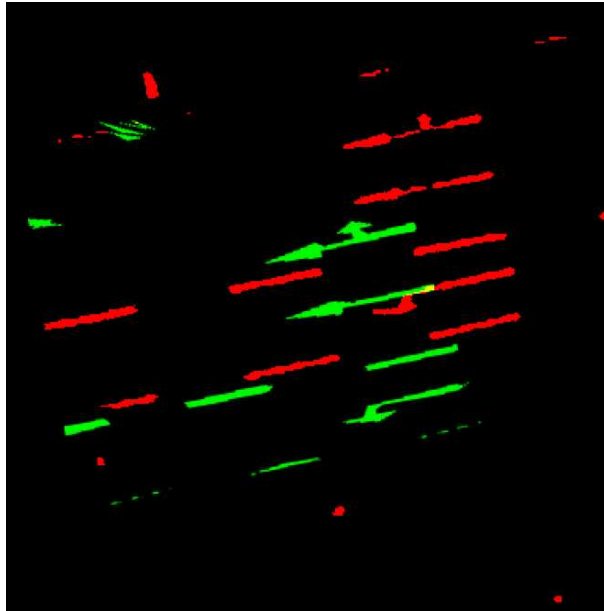
Fig. 12: Merged features of camera image (green) and satellite/ aerial image (red): The features of the camera image were rotated using the GNSS/ INS tightly coupled heading to be aligned with the satellite/ aerial image. One can observe that the street markings of the camera image are much less noisy than the ones of the satellite/ aerial image due to the closer distance.
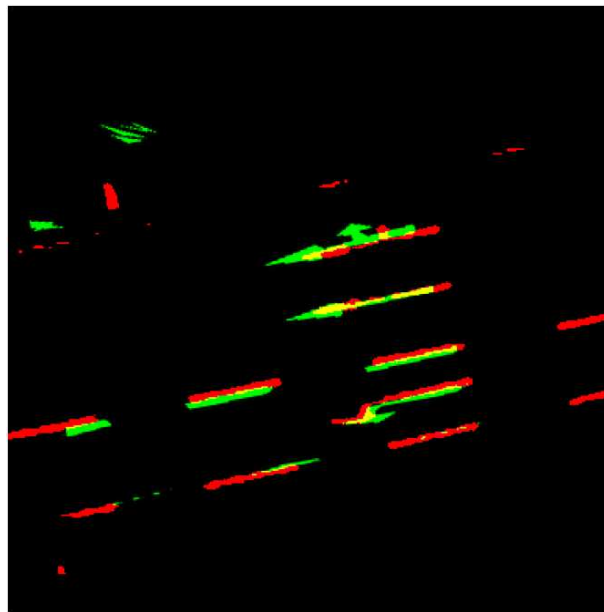

Fig. 13: Aligned features of camera image (green) and satellite/ aerial image (red).

that the car was on the middle lane, which agrees with the camera image of Fig. 2a.

Obviously, the vision-based position information can be used to re-adjust the float GNSS/ INS RTK solution and, thereby, to enable a faster ambiguity re-fixing.
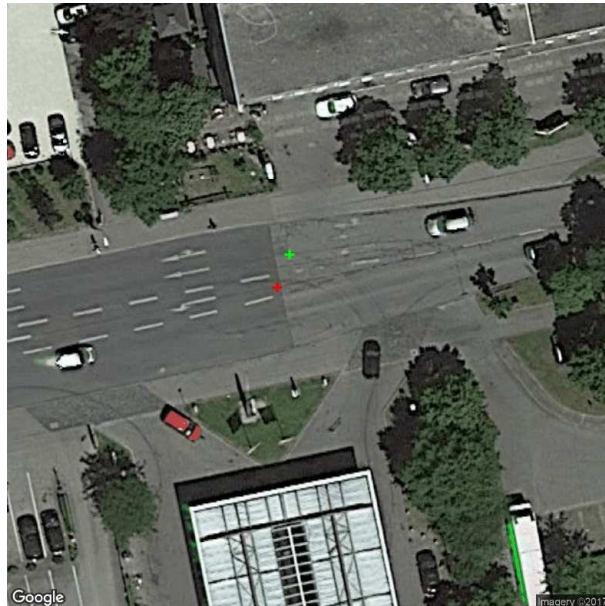
ACKNOWLEDGEMENT

Fig. 14: Re-Adjustment of tightly coupled positioning after sections with GNSS outages (e.g. tunnels): The GNSS/ INS/ barometer coupled solution is shown with a red + and the vision-based position is shown with a green +.

CONCLUSION

In this paper, we presented a fusion of GNSS, INS, barometer and vision for precise position and attitude determination. The GNSS and INS measurements were tightly coupled in a Kalman filter. Carrier phase measurements from a reference station were used to enable ambiguity fixing and RTK positioning. The barometer and vision were used to re-adjust the GNSS/ INS tightly coupled float solution after GNSS outages and, thereby, to enable a faster re-fixing. The paper also provides an iterative closest point search algorithm for efficient matching of the feature points of the camera and map.

REFERENCES

[1] J. Wang, M. Garratt, A. Lambert, J.J. Wang, S. Han and D. Sinclair, *Integration of GPS/ INS/ vision sensors to navigate unmanned aerial vehicles*, Internat. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37 (B1), Beijing, pp. 963 – 970, 2008.

[2] S. Winkler, H.-W. Schulz, M. Buschmann, T. Kordes and P. Vasmann, *Improving Low-Cost GPS/MEMS-based INS Integration for Autonomous UAV Navigation by Visual Aiding*, Proc. of ION GPS, Institute of Navigation, Long Beach, CA, USA, pp. 1069 – 1075, 2004.

[3] J. Hol, *Sensor Fusion and Calibration of Inertial Sensors, Vision, Ultra-Wideband and GPS*, Dissertation, Linköping University, Sweden, 2011.

[4] P. Henkel and P. Burger, *Multi-Sensor Fusion of GNSS receivers, inertial sensors and cameras for Precise and Reliable Positioning*, Proc. of Conference on Future Automotive Technology (COFAT), Fürstenfeldbruck, Germany, pp. 1 – 5, Apr. 2015.

[5] C. Jekeli, *Inertial navigation systems with geodetic applications*, Walter de Gruyter, 2001.

[6] H.S. Hopfield, *Two-Quartic Tropospheric Refractivity Profile for Correcting Satellite Data*, J. of Geophysical Res., vol. 74, pp. 4487 – 4499, 1969.

[7] P. Henkel and A. Sperl, *Real-Time Kinematic Positioning for Unmanned Air Vehicles (UAVs)*, Proc. of IEEE Aerospace Conference, Big Sky, MT, USA, pp. 1 – 7, Mar. 2016.

[8] R.G. Brown and P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons Inc., New York, 1997.

[9] P. Henkel, U. Mittmann and M. Iafrancesco, *Real-Time Kinematic Positioning with GPS and GLONASS*, Proc. of 24-th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 5 pp., Sep. 2016.

[10] P.J.G. Teunissen, *The LAMBDA method for the GNSS Compass*, Artificial Satellites, vol. 41, no. 3, pp. 89 – 103, 2006.

[11] R. Mönikes, O. Meister, J. Wendel and G.F. Trommer, *Yaw Angle Estimation of VTOL-UAVs with the Extended LAMBDA Method and Low Cost Receivers*, Proc. of ION NTM, San Diego, CA, USA, pp. 179 – 186, 2007.

[12] P. Henkel and C. Günther, *Reliable Integer Ambiguity Resolution: Multi-Frequency Code Carrier Linear Combinations and Statistical A Priori Knowledge of Attitude*, Navigation, vol. 59, nr. 1, pp. 61 – 75, 2012.

[13] P.J.G. Teunissen, *A-PPP: Array-Aided Precise Point Positioning with Global Navigation Satellite Systems*, IEEE Transactions on Signal Processing, vol. 60, no. 6, pp. 2870 – 2881, Jun. 2012.

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

[15] A. Blum, *Road marking based visual positioning using aerial imagery*, Master thesis, Technische Universität München, May 2017.