CO-REGISTRATION OF PANORAMIC MOBILE MAPPING IMAGES AND OBLIQUE AERIAL IMAGES

PHILLIPP FANTA-JENDE* (p.l.h.jende@utwente.nl)
FRANCESCO NEX (f.nex@utwente.nl)
GEORGE VOSSELMAN (george.vosselman@utwente.nl)
Faculty ITC, University of Twente, Enschede, The Netherlands
MARKUS GERKE (m.gerke@tu-bs.de)
Technische Universität Braunschweig, Braunschweig, Germany

*Corresponding author

Abstract

Mobile mapping relies on satellite-based positioning, which suffers from line-of-sight and multipath issues. As an alternative, this paper presents a fully automatic approach for the co-registration of mobile mapping and oblique aerial images to introduce highly accurate and reliable ground control for mobile mapping data adjustment. An oblique view of a scene introduces similarities as well as challenges regarding co-registration with mobile mapping images, which is supported by mutual planes in both datasets. Façade planes from a sparse point cloud are used as projection surfaces for the mobile mapping and aerial datasets, overcoming large perspective differences between them to simplify the registration. The performance of the procedure indicates an inlier rate of around 80%.

Keywords: image registration, mobile mapping, oblique aerial imagery

INTRODUCTION

MOBILE MAPPING (MM) has become an established technique to acquire large, high-resolution datasets, predominantly in urban areas, at relatively low cost. As MM can be regarded as a complementary acquisition technique to aerial photogrammetry, it enables intriguing applications from a terrestrial point of view in computer vision, autonomous driving and robotics. With high-precision positioning equipment for absolute and relative localisation, MM accuracy is comparable to survey-grade technologies.

Since the absolute position of these platforms is provided by Global Navigation Satellite Systems (GNSS), such as GPS or GLONASS, MM campaigns are directly affected by typical implications of these systems. Multipath and line-of-sight deficiencies impede fixing an accurate absolute position by GNSS, particularly in urban areas where, for instance, high-rise buildings may obstruct the direct line of sight to GNSS satellites. Consequently, this leads to wrong positional estimations of the MM platform and thus an
unknown error in the data product. In order to correct or compensate for this issue, various techniques have been used. For instance, surveyed ground control points (GCPs) (Cavegn et al., 2016) or digital maps (Levinson et al., 2007) have been introduced into the adjustment of MM data products. In any case, external references are required to either verify, or increase the accuracy of, the dataset.

The method presented in this paper utilises oblique aerial imagery as a tertiary dataset. This can be regarded as a logical continuation of previous efforts, where aerial nadir images were used within the same context (Jende et al., 2018). Exclusively utilising aerial nadir images is limited to ground-based correspondences only, which has certain disadvantages with respect to the availability of correspondences and an adverse geometry within the adjustment solution. Utilising aerial oblique images, however, enables finding correspondences in narrow alleyways on façades or other vertical objects, and hence increases the geometrical stability within an adjustment due to a higher number of correspondences.

Although, a registration of MM perspective and oblique aerial images poses similar challenges, especially with respect to overcoming perspective differences, the approach discussed in this paper is different from the vertical aerial registration approach in many instances. For example, a registration with vertical aerial images is based on the assumption of a flat ground plane, which allows for a remapping of the MM panoramic image. In the present case of oblique aerial images, however, mutual planes between the image datasets need to be identified first.

The structure of this paper is as follows. First, an overview of related and previous work is given. The next section discusses all relevant strategies and methods to register MM perspective and oblique aerial images. The following section comprises an extensive set of experiments. Lastly, special cases and possible advancements are discussed.

**Related Work**

In this work, MM data are registered with aerial image data to introduce independent ground control for data adjustment. Two areas of research overlap in such an approach, namely wide-baseline/non-standard geometry image registration as well as localisation and positioning. Image registration is a research topic that is widely covered in the literature. Although the field is moving towards machine learning and is able to cope with wide-baseline problems (Shan et al., 2014; Lin et al., 2015; Zagoruyko and Komodakis, 2015; Melekhov et al., 2017; Tian et al., 2017), these approaches are still very experimental and cannot provide the reliability and registration accuracy required for the task at hand. Alternatively, authors rely on synthesised views to homogenise the datasets to be registered (Morel and Yu, 2009; Roth et al., 2017). Similarly, in the authors’ previous work, MM panoramic images were reprojected onto an artificial ground plane to achieve a higher resemblance to a vertical aerial (nadir) dataset (Jende et al., 2018). The image registration problem can be constrained using guided matching strategies, as the absolute orientation (and thus the positioning error) is within an expected margin of error. Consequently, the challenge is identifying accurate and salient features in both datasets rather than sensor-specific correspondences. This translates the problem to finding salient, repeatable corner features by feature detection algorithms, such as the adaptive and generic accelerated segment test (AGAST) proposed by Mair et al. (2010). Since these corners are used to register MM images with each other to obtain a sparse point cloud, as well as serving as the basis for a registration with oblique aerial images, repeatable recognition of distinctive façade elements, such as window frames or gutters, is the most important property.
Regarding the localisation problem in GNSS-denied areas, available approaches can be coarsely separated in two different categories:

1. methods which focus on the correction of the data product itself (as in the current approach); and
2. methods which aim for the correction of the platform’s trajectory for the purpose of correct localisation.

Whereas, the latter category is mainly suitable for real-time applications, correcting the data product in (1) allows for post-processing. This difference has an influence on the methodology. Although a method could be designed to correct data directly during acquisition, post-processing enables the factoring out of GNSS effects on the positioning to a greater extent.

Real-time capable applications, such as simultaneous localisation and mapping (SLAM) or visual odometry, allow for the introduction of external references by map-aiding (Schindler, 2013; Gruyer et al., 2014; Gu et al., 2016; Roh et al., 2016) or shadow matching (Groves et al., 2013; Irish et al., 2015; Strode and Groves, 2016); the latter classifies GNSS signals with respect to their angle of incidence. It should be noted that there are a multitude of approaches which do not utilise external data, but rather optimise the trajectory using onboard sensor systems (Zhang and Singh, 2015; Balazadegan Sarvrood et al., 2016; Carlone and Karaman, 2017). These approaches, however, cannot achieve the desired accuracy.

On the other hand, post-processing approaches traditionally rely on GCPs, which are introduced into the adjustment (Cavegn et al., 2016; Hofmann and Brenner, 2016; Han and Lo, 2017; Molina et al., 2017). This proved to be accurate but is labour intensive and thus costly. Javanmardi et al. (2017) and Ji et al. (2015), for instance, also utilised aerial images to introduce an external reference into the solution. These approaches only utilise vertical aerial images, and thus have similar limitations to the current authors’ previous approach as far as the availability of potential correspondences is concerned. Merging oblique aerial images, as well as terrestrial images, for creating complete representations of a scene has been pursued by Wu et al. (2018). Although this is a different task with the image data already co-registered with respect to their orientation, a method similar to the one used in the Wu et al. paper is used, as those authors relied on plane priors extracted from 3D meshes. Furthermore, the necessary point cloud was derived from an aerial dataset. In contrast to this, the approach in this paper relies on a point cloud derived from MM images, which is cheaper to compute and potentially yields a more accurate representation of the scene and thus a more reliable registration result. Among other major differences, feature matching between rectified images is utilised rather than template matching using image patches.

**METHODOLOGY**

This section outlines the registration algorithm in detail (see Fig. 1 for an overview). It is designed to work in a fully automatic fashion, where MM images are processed sequentially with respect to their location in the trajectory. There are four main steps:

1. MM images are registered with each other to yield a sparse point cloud. This step is also a requirement for the adjustment procedure, as it allows for propagating a position update across multiple MM recording locations, and updating of positions of recording locations without direct correspondences with the aerial dataset.
(2) The sparse point cloud is used for identifying planar surfaces along the trajectory, particularly building façades. If predefined criteria for a plane are met, such as visibility from both datasets and orientation, the respective oblique aerial and MM images are projected onto patches coinciding with that plane.

(3) Patch-based registration is conducted, determining the translation between both patch centres.

(4) A consensus based on the median translation between the MM and oblique aerial patches is computed to remove outliers.

**Sparse Point Cloud from Mobile Mapping Images**

A prerequisite to the MM-to-oblique-aerial registration procedure, as well as the adjustment of the MM dataset (which is not within the scope of this paper), is the generation of a sparse point cloud. The procedure is designed to identify correspondences, particularly on
building façades, which are assumed to be approximately parallel to the trajectory of the MM platform. A subsequent plane fitting identifies those points of the sparse point cloud, which coincide with a putative façade. Moreover, inlying points are used for patch creation, which is required for the actual registration between the MM and oblique datasets. MM panoramic images are encoded in a spherical equirectangular projection. This projection entails strong distortions in comparison to common perspective images (see Fig. 2).

To enable reliable feature matching across multiple panoramic images while ensuring a similar coverage of the scene, panoramic images are transformed to perspective images with specific yaw deviations from the driving direction. For each iteration, three perspective images for each side (left/right of the trajectory) are derived from three panoramic images. The principle of differing yaw deviations is depicted in Fig. 3(a). For the left-hand side of the trajectory, for instance, the first perspective image in the trajectory is created with a yaw deviation of 300°, the second one with 270° and the third one with a yaw deviation of 240°. Employing that configuration, the orientation of the perspective images coincides with the position of recordings in the trajectory in such a way that the same scene in object space can be depicted from different angles (Fig. 3(b)). In order to capture building façades, perspective images are generated with a fixed pitch angle of 30° above the horizon (Fig. 4).

Feature matching is conducted using a combination of the AGAST detector (Mair et al., 2010) and the DAISY descriptor (Tola et al., 2010). As a corner detector, AGAST proved to be useful for façade matching since it identifies features where image gradients intersect, rather than distinctive image areas by blob detection. Because these correspondences are reused in a later step for patch creation and for the registration with oblique images, these image patches are, for instance, derived at and around window frames, which potentially returns more salient features than patches from arbitrary image areas.

![Fig. 2. Equirectangular panoramic image.](image)

![Fig. 3. (a) Yaw deviations for perspective image creation. (b) Principle of sparse point cloud generation based on image triplets.](image)
The registration procedure is based on perspective image triplets, such as images 1, 2 and 3 in Fig. 4. Feature matching is conducted between images 1 and 2, and between images 2 and 3. The feature set of image 2 is used to link images 1 and 3 as well. In contrast to panoramic images, these artificial images employ a pinhole camera model. In combination with known extrinsic (exterior) elements and rotations, it allows for the derivation of projection matrices and hence a trifocal tensor (three-photo relative orientation). Employing the trifocal point-transfer constraint enables reliable outlier removal. First, the focal length (principal distance) \( f \) for artificial perspective views needs to be determined, starting from an ideal opening angle (field of view):

\[
f = \tan \left( 90^\circ - \frac{\text{vfov}}{2} \right) \times w/2
\]  

where \( \text{vfov} \) is the vertical field of view and \( w \) is the image width.

The camera matrix \( \mathbf{K} \) only has to be determined once, as the intrinsic (interior orientation) parameters of artificial views do not change:

\[
\mathbf{K} = \begin{bmatrix} f & 0 & w/2 \\ 0 & f & w/2 \\ 0 & 0 & 1 \end{bmatrix}.
\]

Subsequently, respective rotations \( \mathbf{R}_i \) for every view of the triplet need to be computed. The rotations into the camera frame are already included. (Note the sequence of the rotations in this four-term matrix multiplication.)

\[
\mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(90^\circ) & -\sin(90^\circ) \\ 0 & \sin(90^\circ) & \cos(90^\circ) \end{pmatrix} \begin{pmatrix} \cos(\omega) & 0 & \sin(\omega) \\ 0 & 1 & 0 \\ -\sin(\omega) & 0 & \cos(\omega) \end{pmatrix}
\]

\[
\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(-\varphi) & -\sin(-\varphi) \\ 0 & \sin(-\varphi) & \cos(-\varphi) \end{pmatrix} \begin{pmatrix} \cos(\kappa) & -\sin(\kappa) & 0 \\ \sin(\kappa) & \cos(\kappa) & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

Fig. 4. Example of a perspective image triplet for sparse point cloud generation with respective correspondences. From left to right, images 1, 2 and 3.
where $\phi$ is the pitch angle for the camera (fixed to $30^\circ$ to capture façades), $\kappa$ is the yaw rotation including defined yaw deviations (see Fig. 3(a)) and $\omega$ is the roll angle (fixed to $0^\circ$).

Now, the projection matrices $M_i$ can be determined using:
\[
M_i = K[R_i|t_i]
\]
where $t_i$ are the world coordinates (object coordinates) of the respective panoramic image.

After the projection matrices for each perspective view of the two triplets have been obtained, two trifocal tensors can be computed, one for each side of the trajectory. The computation and further details can be found in Hartley and Zisserman (2004). In order to remove outliers, the trifocal point transfer (equation (8) in Hartley (1997)) is used. By selecting a putative correspondence in two views, the trifocal tensor is used to determine the correspondence’s location in the third image. After comparing the actual image coordinate from feature matching to the computed one, a wrong correspondence can be removed unless the configuration is degenerate, thus ensuring points lie on the same epipolar line.

Subsequent to feature matching and outlier removal, the remaining correspondences are translated into the original equirectangular projection for the triangulation (intersection) of object points. Although intersection would be possible using the perspective geometry, for subsequent steps the equirectangular image coordinates are required.

**Plane Fitting to Identify Façades**

Plane fitting is used to determine whether object points coincide with a building façade. Other objects, such as trees, cars and traffic lights, are unfit for registration and can be thus rejected. In this approach, only façades that are parallel to the trajectory are considered (thus building frontages facing the street along which the MM vehicle is travelling). To prevent instabilities of the process at the expense of flexibility, a collinearity criterion with respect to the recording locations is employed. Subsequently, a vector perpendicular to the driving direction is defined as a reference vector for plane fitting. Thus, an angular threshold determines the maximum offset between the reference vector and the plane’s normal vector. In the “Experiments” section, it will become evident that this threshold has a great impact on the number of inlying object points, and hence possible correspondences.

As façades may have some projections, such as balconies or oriels, a plane can only approximate its geometry. To account for this, a maximum deviation of 0.5 m for object points is defined. Plane fitting is based on maximum likelihood estimation sample consensus (MLESAC; Torr and Zisserman, 2000). After a plane is identified, the number of object points between the plane and the recording location is used to determine whether the plane is occluded. Although, a plane may have been identified properly, occluded planes may impede the remapping of image information on a façade later, and consequently complicate the registration task (see Fig. 5).

**Visibility Hypothesis**

Visibility of a façade is an important prerequisite for the registration between MM and oblique aerial images. Especially in urban areas, high-rise buildings may lead to occlusions with respect to oblique aerial images. Although no surface or building model is used in the procedure, a series of methods can be employed to determine if a set of inlying points is visible from respective oblique images. Three criteria have to be fulfilled:
Two angles, $\varphi$ and $\theta$ (Fig. 6), are computed to ascertain the orientation of a façade with respect to an individual oblique aerial camera (Fig. 6). Therefore, inlying points are averaged to determine their centre (centroid) coordinate ($p_{\text{avg}}$) to obtain the perpendicular foot ($c_f$) of the oblique camera ($c_{\text{obl}}$). Subsequently, $\theta$ can be computed as the angle between the plane’s normal vector $n$ and the vector $c_f p_{\text{avg}}$, as well as $\varphi$ between $c_f p_{\text{avg}}$ and $c_{\text{obl}} p_{\text{avg}}$.

$$\theta = \arctan2\left(\frac{\|c_f p_{\text{avg}} \times c_f p_{\text{avg}}\|}{\|c_f p_{\text{avg}}\| \cdot c_f p_{\text{avg}} \cdot n}\right)$$

$$\varphi = \arctan2\left(\frac{\|c_f p_{\text{avg}} \times \bar{n}\|}{\|c_f p_{\text{avg}}\| \cdot c_f p_{\text{avg}} \cdot \bar{n}}\right).$$

A threshold of 70° for both angles has been set. For instance, if $\theta$ were larger than 70°, the plane may be still visible, but its representation in the oblique image would be too low for patch creation.

The theoretical resolution ($t$) at the centroid ($p_{\text{avg}}$) with respect to an individual oblique aerial image, based on its focal length, is determined. A threshold of 8 cm is defined as minimum admissible spacing between the oblique image’s pixels, projected onto the wall plane. If the resolution is too low and may impede a solution, the registration is determined later in the patch creation step.

The centroid $p_{\text{avg}}$ is back-projected onto the candidate oblique aerial images, to check if the point lies within the aerial image’s extent. Although the created image patch will be incomplete in cases where its border does not lie within an oblique image’s extent, a registration may be still successful as the registration is conducted with respect to the patch’s centre.

These methods can reduce the number of points considered for registration considerably. However, due to the lack of a 3D model, ray tracing or similar techniques are
not possible, and thus false positives may occur. In most of these cases, a registration will fail. To analyse the actual registration performance properly in the “Experiments” section, however, these false positives will be labelled “not matchable” (see Fig. 7 for an example). In a subsequent adjustment, these observations would, most likely, be identified as outliers, given the model is strong enough. A detailed evaluation of this problem is subject to future work.

**Point Cloud Thinning**

The algorithm to derive a sparse point cloud is based on feature matching techniques. Non-maximal suppression is used during feature detection, which prevents detection of adjacent corners. Since this is a local operation, corners can still be detected in close proximity to one another. Consequently, parts of the point cloud may be dense. This may lead to the generation of almost identical image patches, which only differ a little in translation. This affects both MM and oblique aerial patches to the same extent. On the other hand, more object points lying on a potential façade lead to better and more reliable plane fitting. To this end, thinning of the point cloud is only applied after a plane has been identified. The input of the thinning process is all object points which coincide with a plane (inlying points). A non-uniform box grid filter is used to reduce the number of points (Pomerleau et al., 2013). A maximum number of points per box is defined, from which one point is randomly retained. This process is adaptive, as points with high proximity, rather than isolated points, will be removed. This is a useful property, since the general distribution of the original inlying points can be maintained while cluttered points are removed.

**Image to Plane Projection**

Panoramic MM and oblique aerial images depict a scene from different perspectives. In order to simplify the registration problem, mutual planes that are potentially visible from both datasets have been identified. In order to register both datasets with each other, image patches are created. As mentioned earlier, a plane does not necessarily comply with the real
geometry of a façade. As already noted, object points with a maximum distance from the plane of up to 0.5 m are still labelled as *inlying* to compensate for deviations from a true plane in real-world façades. In order to account for these offsets during registration, image patches are created with an object point as its centre (centroid). To discretise an image patch in 3D space, a local coordinate system constituting the respective patch is defined. The corresponding object point is defined as its origin. The coordinate system is derived as follows. All inlying points \( p_{\text{abs}} \) are normalised with respect to their centroid \( p_{\text{avg}} \) to obtain a covariance matrix \( C \):

\[
C = \sum_{i=1}^{m} (p_i - p_{\text{avg}}) \cdot (p_i - p_{\text{avg}})^T, \quad p_i \in p_{\text{abs}} 
\]  

where \( m \) is the entire set of inlying points \( p_{\text{abs}} \). Since a covariance matrix is symmetric, its eigenvectors are orthogonal to each other and can be used to define a coordinate system in conjunction with the plane’s normal vector (Fig. 8). In order to discretise a grid, the sampling rate, as well as the extent of the image patch, are used. The sampling rate corresponds to the theoretical resolution \( t \) of centroid \( p_{\text{avg}} \) with regard to an individual oblique image. The extent of the image patch is, by default, defined as 8 m by 8 m. In the case of hierarchical matching (discussed in the next section), the extent is reduced to 6 m by 6 m in the second iteration.

Subsequently, the resulting discretised grid in object space is filled with red/green/blue (RGB) values from both image sources via a back-projection mechanism. This results in two image patches per inlying object point (Fig. 9). Although the theoretical resolution has been computed earlier, it does not necessarily correspond to the real resolution. This would only be the case if the grid was parallel to the oblique aerial image’s image plane. In the scenario in this research, a grid is usually coplanar with the building façades. In order to obtain the real resolution of a grid in a certain oblique image, an intuitive solution would be a projection of the grid into the image itself. However, a simpler solution in this context is to relate the number of assigned individual oblique image’s RGB values to the pixels in the grid. In other words, if the theoretical resolution corresponded to the real resolution while the grid is parallel to an oblique image’s image plane, the ratio between assigned RGB values and the number of pixels in the grid would be one, since a one-to-one assignment would be possible. In a real-world scenario, however, a single RGB value is retrieved, more often by the back-projection mechanism reducing the ratio and thus the real resolution of

![Fig. 7. An example of a façade which is not matchable, since the awning above the windows in the left image is occluding most of the patch.](image_url)
the image patch. A threshold for the ratio of \(0.6\) prevents a low real resolution and hence blur in the oblique image patch, which may impede the registration (Fig. 10).

**Registration**

Image patches simplify the registration scenario considerably, as variance between both datasets can be overcome. For instance, the image patches share a similar (but not exactly the same) scale, perspective and rotation. Hence, the registration problem is simplified to finding the correct translation between a MM and an oblique aerial image patch. As the registration is based directly on image patches, the translation is interpreted as the shift between the patch centres, which are the respective image observations of the object point previously labelled as an inlier in the plane fitting procedure.
Although the image patches share a number of similarities, a patch-based registration requires context to find the correct transformation. To this end, an optional hierarchical registration approach was implemented which firstly finds an initial, but coarse, transformation between larger patches with a lower resolution. In a second iteration, the initial transformation is used to constrain the registration for a more accurate result. Besides minor computational speed improvements, the main benefit of hierarchical matching is a higher inlier rate. More details comparing hierarchical and non-hierarchical matching are analysed in the next section. In both cases, template matching techniques are employed. A coarse registration is conducted using normalised cross correlation, whereas a fine registration is performed using mutual information as a similarity criterion (Mattes et al., 2001). For mutual information, a one-plus-one evolutionary algorithm is used to find the best parametrisation (Droste et al., 2002). This algorithm modifies its set of parameters every iteration until it converges to the best registration result.

In general, utilising mutual information proved to be a powerful method for the scenario in this paper, as it can overcome illumination details while ignoring differing image content. Even though outliers may be still present, it performed better for fine registration than, for example, normalised cross correlation, which is more sensitive to noise (see experiment 8/9 in the next section).

**Outlier Removal**

Since a correspondence between a pair of image patches can be regarded as either the translation between image observations or a translation in object space, two outlier removal techniques have been devised: a 2D and a 3D case. A set of correspondences is defined as all correspondences between one panoramic image and one oblique aerial image. The difference between the two outlier removal techniques is their dimension. In the 2D case, image observations (row/column) are fed into the procedure. In the 3D case, a MM image correspondence coincides with an inlying object point \((p_{\text{obs}})\), whereas an oblique aerial image correspondence is converted into 3D by relating it to the grid in object space. In both cases, a consensus based on the median translation is used. If a correspondence deviates from the median value by a certain threshold, it is labelled as an outlier. This technique is based on the assumption that most correspondences are inliers and follow the same pattern.

Fig. 10. MM image patch (a) with corresponding oblique aerial patch (b). Due to the low resolution of the oblique aerial patch, this pair cannot be used for registration.
The 2D case is only used when using hierarchical matching to reject outliers after the first iteration, whereas the 3D case is used after fine registration and consensus tracking. Since image patch generation can account for minor deviations in the façade plane, such as small building projections, a 3D-based outlier removal is more robust in such a case. For hierarchical matching, image patches have a lower resolution and a translation between image patches is only approximated. Therefore, 2D-based outlier removal is sufficient.

Moreover, a technique to track a consensus across multiple panoramic images along the trajectory was implemented. The requirement for this technique, however, is that adjacent MM images share correspondences with the same oblique image. Certainly, this tracking technique requires a 3D-based translation to relate registration results of adjacent MM images due to different scenes and thus façades. Experiments are presented in the next section.

**EXPERIMENTS**

This section presents results and discusses multiple experiments with respect to various aspects of the registration algorithm. The MM dataset used in this section consisted of 108 panoramic images and depicts a narrow road in the city centre of Rotterdam, the Netherlands (Fig. 11). The majority of the buildings have three to four storeys and range from about 10 to 20 m in height. The width of the road is mostly around 10 m, though in some places narrows to 7 m and widens up to 20 m. The absolute accuracy of the dataset cannot be disclosed due to proprietary knowledge of the data provider. MM panoramic images were recorded every 5 m along the trajectory. A panoramic image was encoded in a spherical equirectangular projection, where each pixel corresponded to the same angular resolution. An image covered $360^\circ \times 180^\circ$ degrees with a resolution of $0.075^\circ$/pixel, resulting in an image dimension of $4800 \times 2400$ pixels. The entire test area was covered by oblique aerial images, which have been acquired in a pentacam fashion (four oblique plus one nadir images) at an altitude of 450 m with a ground sample distance (GSD) of 5 to 15 cm. For the procedure presented in this paper, the nadir images of the pentacam were not used. Depending on the individual MM recording location, two to six oblique aerial images qualified for registration.

The processing was implemented to work in a sequential fashion, treating the left and right parts of the trajectory independently (Fig. 3(b)). Later in the adjustment, when correspondences are converted into the original image geometries, both sides of the trajectory are combined. In total, 14 different experiments with different processing parameters were conducted (Table I). Entries in bold in Table I depict differences from the default setting. For all the other tables (Tables II to VIII), bold entries represent the best result. In the experiments, four different inlier ratios are provided. The inlier ratios were distinguished in two ways: first, before/after outlier removal; and second, with and without considering non-matchable correspondences.

**Maximum Angular Distance to Reference Vector for Plane Fitting**

The first four experiments (1 to 4 in Table I) focused on the maximum angular distance for plane fitting. The angular distance of a plane is defined by the difference of its normal vector to the reference vector in degrees in two dimensions (yaw and pitch). Since the reference vector for plane fitting corresponds to the vector perpendicular to the trajectory, relaxing this parameter enables fitting façades that are not parallel to the trajectory. The results for these four experiments and the default are shown in Fig. 12.
It became evident that relaxing the threshold for a plane’s deviation from the reference vector has rather negative consequences on the result, with a decrease of about 20% with respect to the inlier rate (Table II). This is mostly because the plane does not fit properly to the façade points, which leads to a skewed image patch and thus a wrong registration result (Fig. 13). Certainly, additional correspondences may be identified which were not possible on the assumption of parallelism between MM vehicle trajectory and façade surface. If the angular distance is reduced to 2° (parameter set for experiment 1), a slightly higher number of inliers, compared to the default parameter set, could be identified. Although more inlying correspondences were returned, a stricter threshold reduces the flexibility of the approach. In this case, the default parameter set returned a higher number (four in total) of individual recording locations with direct correspondences to the aerial images than parameter set 1. Moreover, a more tolerant maximum angular distance threshold does not necessarily result in the same set of estimated planes. This is only the case if a façade is represented well by the sparse point cloud.

Minimum Number of Points for Plane Fitting

Another threshold in the plane fitting process is to adjust the minimum number of points required for plane fitting (Fig. 14). This parameter determines whether defined criteria for plane fitting can conform to a required minimum number of points per plane. The number and distribution of points which can be used for plane fitting depends on the actual...
Table I. Parameters for different experiments. NCC is normalised cross correlation; MI is mutual information. Entries in **bold** represent differences from the default setting.

| Experiment | Maximum angular distance (°) | Coarse/fine registration | Minimum number of points for a plane | Hierarchical matching | Patch extent for registration (coarse/fine) (m) |
|------------|-----------------------------|---------------------------|-------------------------------------|----------------------|-----------------------------------------------|
| Default    | 5                           | NCC/MI                    | 10                                  | Yes                  | 8/6                                           |
| 1          | 2                           | NCC/MI                    | 10                                  | Yes                  | 8/6                                           |
| 2          | **10**                      | NCC/MI                    | 10                                  | Yes                  | 8/6                                           |
| 3          | **15**                      | NCC/MI                    | 10                                  | Yes                  | 8/6                                           |
| 4          | **20**                      | NCC/MI                    | 10                                  | Yes                  | 8/6                                           |
| 5          | 5                           | NCC/MI                    | **5**                               | Yes                  | 8/6                                           |
| 6          | 5                           | NCC/MI                    | **20**                              | Yes                  | 8/6                                           |
| 7          | 5                           | NCC/MI                    | **30**                              | Yes                  | 8/6                                           |
| 8          | 5                           | MI/MI                     | 10                                  | Yes                  | 8/6                                           |
| 9          | 5                           | NCC/NCC                   | 10                                  | Yes                  | 8/6                                           |
| 10         | 5                           | –                         | 10                                  | **No**               | 8                                             |
| 11         | 5                           | –                         | 10                                  | **No**               | **10**                                         |
| 12         | 5                           | NCC/MI                    | 10                                  | Yes                  | **10/6**                                       |
| 13         | 5                           | –                         | 10                                  | **No**               | 12                                             |
| 14         | 5                           | NCC/MI                    | 10                                  | Yes                  | **12/6**                                       |

Table II. Results for the default and experiments 1 to 4 [different threshold values for maximum angular distance for plane fitting]. **Bold** entries represent the best result.

| Parameter set (see Table I) | Default | 1 | 2 | 3 | 4 |
|-----------------------------|---------|---|---|---|---|
| Maximum angular distance (°)           | 5       | 2 | 10 | 15 | 20 |
| Inlier percentage                      | 54%     | 56% | 45% | 40% | 43% |
| Inlier percentage without non-matchable correspondences | 67% | 68% | 54% | 50% | 51% |
| Inlier percentage after outlier removal | **70%** | **70%** | 47% | 60% | 58% |
| Inlier percentage after outlier removal without non-matchables | **80%** | **80%** | 54% | 66% | 64% |

Table III. Parameters for the default and results of experiments 5 to 7 [different minimum number of points for plane fitting]. **Bold** entries represent the best result.

| Parameter set (see Table I) | Default | 5 | 6 | 7 |
|-----------------------------|---------|---|---|---|
| Minimum number of points per plane       | 10      | 5  | 20  | 30  |
| Number of MM recordings with correspondences (out of 108) | 69 | **78** | 53  | 42  |
| Inlier percentage                      | 54%     | 53% | 61% | **63%** |
| Inlier percentage without non-matchable correspondences | 67% | 67% | 76% | **77%** |
| Inlier percentage after outlier removal | **70%** | **66%** | **70%** | 63% |
| Inlier percentage after outlier removal without non-matchables | **80%** | **85%** | 83% | 78% |

Table IV. Statistics of correspondences between perspective views of the entire trajectory.

| Side of the trajectory | Right | Left |
|------------------------|-------|------|
| Number of correspondences | 7094 | 4050 |
| Median correspondences per view | 22 | 16.5 |
| Standard deviation of correspondences per view | 107.92 | 66.45 |
Table V. Results for the default and experiments 8 and 9 [initial transformation/fine registration with MI or NCC only]. Bold entries represent the best result.

| Parameter set (see Table I)                  | Default | 8   | 9   |
|-----------------------------------------------|---------|-----|-----|
| Initial transformation                        | NCC     | MI  | NCC |
| Fine registration                             | MI      | MI  | NCC |
| Inliers                                       | 578     | 421 | 679 |
| Outliers                                      | 287     | 1124| 271 |
| Non-matchable correspondences                 | 202     | 1347| 189 |
| Inlier percentage                            | 54%     | 15% | 60% |
| Inlier percentage without non-matchable corres | 67%     | 27% | 71% |
| Inlier percentage after outlier removal       | 70%     | 18% | 73% |
| Inlier percentage after outlier removal without non-matchables | 80%     | 33% | 79% |

Table VI. Results for the default and experiments 10 to 14 [different image patch sizes and hierarchical matching]. Bold entries represent the best result.

| Parameter set (see Table I)                  | Default | 10  | 11  | 12  | 13  | 14  |
|-----------------------------------------------|---------|-----|-----|-----|-----|-----|
| Initial transformation                        | NCC     | –   | –   | NCC | –   | NCC |
| Patch extent (m)                              | 8/6     | 8   | 10  | 10/6| 12  | 12/6|
| Inlier percentage                            | 54%     | 38% | 59% | 61% | 72% | 53% |
| Inlier percentage without non-matches         | 67%     | 56% | 74% | 77% | 82% | 67% |
| Inlier percentage after outlier removal       | 70%     | 45% | 68% | 71% | 61% | 77% |
| Inlier percentage after outlier removal without non-matches | 80%     | 66% | 82% | 82% | 75% | 80% |

scene (Fig. 15). Table IV provides an insight into the absolute number of correspondences used for the plane fitting process. Fig. 16 depicts the distribution of correspondences across all perspective views.

The results among different parameter sets are almost identical, returning a similar number of inliers, outliers and non-matchable correspondences. However, the number of recording locations with correspondences decreased with the number of points required for a plane, indicating that some planes along the trajectory are constituted by only 10 or fewer points (see Fig. 17).

Mutual Information to Compute Initial Transformation

By default, hierarchical matching is used to retrieve an estimate of the translation between the image patches by normalised cross correlation. Subsequently, a fine registration computes the precise offset between both patches using mutual information. In this experiment, mutual information is used to obtain an initial transformation instead of normalised cross correlation using the same set of parameters as for a subsequent fine registration (see parameter set 8 in Table V). For comparison, the same experiment has been conducted using normalised cross correlation for fine registration (parameter set 9 in Table V).

Although the total number of identified correspondences is higher if mutual information is used for coarse registration, the inlier rate drops significantly. On the other hand, if normalised cross correlation is used for both steps – finding an initial transformation as well as fine registration – the results remain comparable to the results with the default parameter.
set. While normalised cross correlation is suitable for coarse registration and its fine registration capabilities are good, it is not on par with mutual information, which is able to locate correspondences more accurately (Fig. 18).

### Table VII. Comparison of inlier rate before and after outlier removal. **Bold** entries represent the best result.

| Experiment (parameter set) | Number of inliers | As a percentage (%) | Number of inliers after outlier removal | As a percentage (%) | Ratio of number of inliers before and after outlier removal | Relative increase in inlier ratio (%) |
|----------------------------|-------------------|----------------------|----------------------------------------|----------------------|--------------------------------------------------|-------------------------------------|
| Default                    | 578               | 66.82                | 137                                    | 79.65                | 23.70                                            | 119.20                              |
| 1                          | 610               | 68.00                | 163                                    | 79.51                | 26.72                                            | 116.92                              |
| 2                          | 498               | 54.01                | 102                                    | 54.26                | 20.48                                            | 100.45                              |
| 3                          | 416               | 49.17                | 133                                    | 66.17                | 31.97                                            | **134.57**                          |
| 4                          | 468               | 51.77                | 122                                    | 64.21                | 26.06                                            | 124.03                              |
| 5                          | 594               | 67.04                | 167                                    | **84.77**            | 28.11                                            | 126.44                              |
| 6                          | 577               | 76.32                | 159                                    | 82.81                | 27.56                                            | 108.5                               |
| 7                          | 587               | 76.93                | 153                                    | 78.06                | 26.06                                            | 101.47                              |
| 8                          | 421               | 27.25                | 117                                    | 33.24                | 27.79                                            | 121.98                              |
| 9                          | 679               | 71.47                | 193                                    | 78.46                | 28.42                                            | 109.77                              |
| 10                         | 484               | 56.15                | 274                                    | 66.34                | 56.61                                            | 118.16                              |
| 11                         | 367               | 74.44                | 232                                    | 81.69                | 63.22                                            | 109.74                              |
| 12                         | 724               | 77.19                | 168                                    | 81.55                | 23.2                                             | 105.66                              |
| 13                         | 216               | **81.51**            | 135                                    | 75.00                | 62.5                                             | 92.01                               |
| 14                         | 568               | 67.06                | 142                                    | 79.78                | **75.00**                                         | 118.96                              |

### Table VIII. Rejection threshold for default parameter set. The last column shows results with consensus-tracking activated. **Bold** entries represent the best result.

| Rejection threshold | 0.5 (default) | 1   | 0.3 | 0.5 (tracking) |
|---------------------|---------------|-----|-----|----------------|
| Number of inliers after outlier removal | 137           | **173** | 101 | 145 |
| Number of outliers  | 35            | 81   | 21  | 41  |
| In per cent         | 79.65         | 68.11 | **82.79** | 77.96 |

**Fig. 12.** Number of correspondences (ordinate) with different maximum angular distance (abscissa) for plane fitting.
Hierarchical Matching and Patch Size

Important properties of the registration procedure are the definition of the image patch size and utilizing hierarchical matching. A first iteration derives a coarse registration between the image patches utilizing contextual information, whereas a second iteration conducts the fine registration. In this experiment, both properties (patch size and hierarchical matching) are analysed.

Comparing the results in Table VI and Fig. 19 of the default parameter set with set 10 (see Table I for details), it becomes evident that hierarchical matching improves the matching result considerably. The increase in inliers if hierarchical matching (HM) is used is also observable between set 11 (no HM) and set 12 (with HM), as well as between set 13 (no HM) and set 14 (with HM). The relationship between the inlier/outlier ratio is, however, different. Whereas hierarchical matching seems to improve this ratio in the first two set pairs (default and 10), the second set pair (11 and 12) retains the same ratio while the relation in the third set pair (13 and 14) flips (if outlier removal has not been used). The patch size itself also has a great impact on the results. Set 12 returned most inliers, and uses an initial patch size of 10 m. Increasing the patch size to 12 m, however, leads to a reduction in the total number of correspondences. With fewer than a third of the inliers of set 12, set 13 returns the least number of correspondences, as large patches may involve parts of the scene which are not part of the physical plane (Fig. 20).

A drawback of larger patches is also related to processing time, as the patch creation method is one of the most computationally expensive steps in the procedure. In fact, a patch with 12 m sides has an area of 144 m², which is more than twice the area (64 m²) of a patch with 8 m sides.

Fig. 13. Wrong plane estimation leads to skewed image patches.

Fig. 14. Number of correspondences with different number of points per plane.

Hierarchical Matching and Patch Size

Important properties of the registration procedure are the definition of the image patch size and utilising hierarchical matching. A first iteration derives a coarse registration between the image patches utilising contextual information, whereas a second iteration conducts the fine registration. In this experiment, both properties (patch size and hierarchical matching) are analysed.

Comparing the results in Table VI and Fig. 19 of the default parameter set with set 10 (see Table I for details), it becomes evident that hierarchical matching improves the matching result considerably. The increase in inliers if hierarchical matching (HM) is used is also observable between set 11 (no HM) and set 12 (with HM), as well as between set 13 (no HM) and set 14 (with HM). The relationship between the inlier/outlier ratio is, however, different. Whereas hierarchical matching seems to improve this ratio in the first two set pairs (default and 10), the second set pair (11 and 12) retains the same ratio while the relation in the third set pair (13 and 14) flips (if outlier removal has not been used). The patch size itself also has a great impact on the results. Set 12 returned most inliers, and uses an initial patch size of 10 m. Increasing the patch size to 12 m, however, leads to a reduction in the total number of correspondences. With fewer than a third of the inliers of set 12, set 13 returns the least number of correspondences, as large patches may involve parts of the scene which are not part of the physical plane (Fig. 20).

A drawback of larger patches is also related to processing time, as the patch creation method is one of the most computationally expensive steps in the procedure. In fact, a patch with 12 m sides has an area of 144 m², which is more than twice the area (64 m²) of a patch with 8 m sides.
Outlier Removal

The method to remove outliers is based on the metric (3D case) or pixel (2D case) distance from a median consensus of all correspondences. Table VII depicts the number of inliers before and after outlier removal for the 3D case. It is apparent that a higher relative number of inliers (114% on average) can be achieved at the expense of removing a large number of correct correspondences (67% on average). In only one instance did the relative number of inliers decrease after outlier removal (set 13), due to a high inlier rate and a low number of correspondences. The ratio between the numbers of inliers before and after outlier removal depicts how many inliers in the original set could be preserved after outlier removal. In some cases (for instance, the default set), fewer than a quarter of the original inlier points remain. The last column of Table VII expresses this relationship as the increase of inliers with respect to the absolute number of correspondences.

A rejection threshold defines if a set of correspondences conforms to the consensus. A default of 0.5 m (in 3D space) has been selected for the experiments above. Table VIII depicts different rejection thresholds for the default parameter set. Certainly, the number of inliers increases with a more tolerant rejection threshold (1.0) while it decreases with a less tolerant threshold (0.3). However, the inlier ratio will decrease if the rejection threshold is too tolerant.

Moreover, the last column in Table VIII represents a result with consensus-tracking activated. In this case, a median consensus of adjacent MM images sharing correspondences with the same oblique image can be tracked. This technique is, however, only useful if rather long, uninterrupted stretches of MM images are present which share tie information.

Fig. 15. Example of correspondences between two triplets of the same location. (Top: right-hand side of trajectory; bottom: left-hand side.)
Fig. 16. Distribution of triplet correspondences across the entire trajectory for both sides. Top row: correspondences on the right-hand side of the trajectory. Bottom row: correspondences on the left-hand side.

Note the equal distribution as well as the slanted pattern on the side views representing façades.

with the same oblique image. In the test dataset, this was only the case with a maximum sequence of five MM images. Hence, this technique will have to prove itself to be useful in future experiments.

**DISCUSSION AND CONCLUSION**

This paper has presented a fully automatic procedure to register panoramic MM images with oblique aerial imagery. An array of different techniques and heuristics is needed to retrieve accurate correspondences. Based on mutual planes in object space, both image datasets can be homogenised to allow for a registration at pixel-level accuracy.

As far as the preprocessing steps are concerned, it became evident that the influence of plane fitting is very significant on the entire outcome of the procedure. For example, the parameter “minimum number of points per plane” translates to a constraint, which can reduce the number of outliers at the expense of façades to be integrated into the process. In the “Experiments” section, all results have been compared to the outcome using the default parameter set, which the authors consider best practice. Currently, parameters are defined such as an occlusion threshold, a maximum distance of a point to a plane, a maximum angular distance to the reference vector and a minimum number of points contributing to a
plane. For instance, the size of a plane, the distribution of points constituting it and a tracking of planes across multiple images, could further improve the detection rate. Certainly, the integration of additional parameters may complicate the procedure.

Another interesting property of the procedure is the patch size. The higher its value, the larger the grid will become. Choosing the right size for a patch is of great significance, as a template registration technique is used. If the grid is too small, lack of context will impede the registration; if the grid is too large, it may project above the actual plane in object space, leading to distortions. In the “Experiments” section, it is apparent that a large patch size returns a higher inlier rate, as the registration becomes more stable. This requires, however, that the scene is depicted similarly in both datasets, as the registration would not converge otherwise. Hierarchical matching takes advantage of this and can improve the registration result considerably. Although, an iterative hierarchical matching with more than two iterations seems a logical continuation, additional computational costs due to the recomputation of patches with different sampling rates, as well as the nature of template matching, may render additional iterations superfluous. Further endeavours should focus on utilising registration results from previous recording locations to constrain the search space and enable implicit quality control of correspondences.

| Minimum number of points per plane | 10  | 5  | 20 | 30 |
|-----------------------------------|-----|----|----|----|
| Number of recordings with correspondences | 62  | 60 | 45 | 35 |

Fig. 17. Influence on minimum number of points per plane; a comparison of recording locations with (green) and without (red) correspondences. From left to right: 10 points (default), 5 points, 20 points and 30 points. See also Fig. 11.
With this idea in mind, an outlier removal technique which is able to track consensus has been presented. Although it returns many false negatives in the procedure, it increases the inlier rate in almost every case. Moreover, it is prone to converge incorrectly if the majority of correspondences are false. Thus, an interesting research question is whether a robust bundle adjustment, which includes outlier detection as well, will perform better with or without previous outlier filtering. In general, the procedure works reliably in this non-standard registration scenario. The experiments have shown that, even in difficult scenarios, an inlier rate of 80% is achievable. Future efforts will be directed towards the integration of this procedure into an adjustment solution, as well as to further increase the flexibility of the approach to reduce false positives.

Fig. 18. Comparison between fine registration results of mutual information and normalised cross correlation for three examples (top, middle, bottom). In each example: left image is the MM patch; middle image is the MI registration result in the oblique patch; right image is the NCC registration result in the oblique patch.
ACKNOWLEDGEMENTS

This work is part of the TTW-Open Technology Programme with project number 13589, which is financed by the Netherlands Organisation for Scientific Research (NWO).

REFERENCES

BALAZADEGAN SARBROOD, Y., HOSSEINYALAMDARY, S. and GAO, Y., 2016. Visual-LiDAR odometry aided by reduced IMU. ISPRS Journal of Geo-information, 5(1). 24pages. https://doi.org/10.3390/ijgi5010003

CARLONE, L. and KARAMAN, S., 2017. Attention and anticipation in fast visual-inertial navigation. IEEE International Conference on Robotics and Automation, Singapore. Pages 3886–3893.

CAVEGN, S., NEBIKER, S. and HAALA, N., 2016. A systematic comparison of direct and image-based georeferencing in challenging urban areas. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 41(B1): 529–536.

DROSTE, S., JANSEN, T. and WEGENER, I., 2002. On the analysis of the (1+1) evolutionary algorithm. Theoretical Computer Science, 276(1–2): 51–81.

GROVES, P. D., JIANG, Z., RUDI, M. and STRODE, P., 2013. A portfolio approach to NLOS and multipath mitigation in dense urban areas. 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2013), Nashville, Tennessee, USA. Pages 3231–3247.
Gruber, D., Belaroussi, R. and Reviloud, M., 2014. Map-aided localization with lateral perception. *IEEE Intelligent Vehicles Symposium, Dearborn, Michigan, USA*. Pages 674–680.

Gu, Y. L., Hsu, L.-T. and Kamijo, S., 2016. GNSS/onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon. *IEEE Transactions on Vehicular Technology, 65*(6): 4274–4287.

Han, J. and Lo, C., 2017. Adaptive time-variant adjustment for the positioning errors of a mobile mapping platform in GNSS-hostile areas. *Survey Review, 49*(352): 9–14.

Hartley, R. I., 1997. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision, 22*(2): 125–140.

Hartley, R. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK. 672 pages.

Hofmann, S. and Brenner, C., 2016. Accuracy assessment of mobile mapping point clouds using the existing environment as terrestrial reference. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 41*(B1): 601–608.

Irish, A., Iland, D. and Madhow, U., 2015. Urban localization and 3D mapping using GNSS shadows. *Inside GNSS, September/October 2015: 60–66.*

Javanmardi, M., Javanmardi, E., Gu, Y. and Kamijo, S., 2017. Towards high-definition 3D urban mapping: road feature-based registration of mobile mapping systems and aerial imagery. *Remote Sensing, 9*(10): article 975. 30 pages.

Jende, P., Nex, F., Gerke, M. and Vosselman, G., 2018. A fully automatic approach to register mobile mapping and airborne imagery to georeference the correction of platform trajectories in GNSS-denied urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing, 141*: 86–99.

Ji, S., Shi, Y., Shan, J., Shao, X., Shi, Z., Yuan, X., Yang, P., Wu, W., Tang, H. and Shibasaki, R., 2015. Particle filtering methods for geo-referencing panoramic image sequence in complex urban scenes. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*: 1–12.

Levinson, J., Montemerlo, M. and Thrun, S., 2007. Map-based precision vehicle localization in urban environments. *Robotics: Science and Systems Atlanta, Georgia, USA*. 8 pages.

Lin, T.-Y., Yin, C., Belongie, S. and Hays, J., 2015. Learning deep representations for ground-to-aerial geolocalization. *IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA*. Pages 5007–5015.

Mair, E., Hager, G. D., Burschka, D., Suppa, M. and Hirzinger, G., 2010. Adaptive and generic corner detection based on the accelerated segment test. European Conference on Computer Vision. *Lecture Notes in Computer Science, 6312*: 183–196.

Mattes, D., Haynor, D. R., Vesselle, H., Lewellyn, T. K. and Eubank, W., 2001. Nonrigid multimodality image registration. *SPIE, 4322*: 1609–1621.

Melekhov, I., Ylioinas, J., Kannala, J. and Rahtu, E., 2017. Relative camera pose estimation using convolutional neural networks. *International Conference on Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science, 10617*: 675–687.

Molina, P., Blazquez, M.,ucci, D. A. and Colomina, I., 2017. First results of a tandem terrestrial-unmanned aerial mapKITe system with kinematic ground control points for corridor mapping. *Remote Sensing, 9*(1): article 60. 13 pages.

Morel, J.-M. and Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences, 2*(2): 438–469.

Pomerleau, F., Colas, F., Siegwart, R. and Magnenat, S., 2013. Comparing ICP variants on real-world data sets. *Autonomous Robots, 34*(3): 133–148.

Roh, H., Jeong, J., Cho, Y. and Kim, A., 2016. Accurate mobile urban mapping via digital map-based SLAM. *Sensors, 16*(8): article 1315. 24 pages.

Roth, L., Kuhn, A. and Mayer, H., 2017. Wide-baseline image matching with projective view synthesis and calibrated geometric verification. *PFG Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 85*(2): 85–95.

Schindler, A., 2013. Vehicle self-localization with high-precision digital maps. *IEEE Intelligent Vehicles Symposium (IV), Gold Coast, Queensland, Australia*. Pages 134–139.

Shan, Q., Wu, C., Curless, B., Furukawa, Y., Hernandez, C. and Seitz, S. M., 2014. Accurate georegistration by ground-to-aerial image matching. *IEEE 2nd International Conference on 3D Vision, Tokyo, Japan. 1*: 525–532.

Strode, P. R. R. and Groves, P. D., 2016. GNSS multipath detection using three-frequency signal-to-noise measurements. *GPS Solutions, 20*(3): 399–412.
La cartografía móvil utiliza el posicionamiento por satélite, que está limitado por los problemas de masqueo y de multi-traje. Este artículo presenta una aproximación totalmente automatizada para el recuadrido de las imágenes de cartografía móvil con imágenes aéreas oblicuas. El uso de puntos de referencia ortogonales en las dos imágenes permite una precisión y fiables. La vista oblicua de una escena presenta similitudes y desafíos en la correspondencia con las imágenes de cartografía móvil y las fotografías aéreas, lo que permite simplificar el recuadrido evitando las importantes diferencias de perspectiva. La performance de la procedura indica un taque de conformidad d’environ 80%.

Zusammenfassung
Mobile-Mapping basiert auf satellitengestützten Positionierungsverfahren, welche grundsätzlich Okklusions- und Multipatheffekten ausgesetzt sind. Der vorliegende Aufsatz behandelt ein vollautomatisiertes Koregistrierungsverfahren zwischen Mobile-Mapping- und Schrägluftbildern zum Zwecke der Einbringung einer hochgenauen sowie zuverlässigen Referenz für die Korrektur von Mobile-Mapping-Daten und -Trajektorien. Schrägluftbilder besitzen zwar einerseits perspektivische Gemeinsamkeiten mit Mobile-Mapping-Bildern, stellen andererseits jedoch eine Herausforderung für eine genaue Koregistrierung derselben dar. Um die großen perspektivischen Unterschiede zwischen den beiden Bilddatensätzen für eine Registrierung zu kompensieren, dienen aus Punktwolken im Objektraum extrahierte Fassaden als Projektionsflächen. 80% der resultierenden Korrespondenzen dieses Verfahrens können als korrekt gewertet werden.

Resumen
Los sistemas móviles de cartografía se apoyan en sistemas de posicionamiento satelitales, afectados de occlusiones y propagación multiruta de la señal. Este documento presenta una aproximación completamente automática para el registro de imágenes de sistemas móviles de cartografía e imágenes aéreas oblicuas para introducir puntos de control en tierra de alta precisión y fiabilidad para el ajuste de datos de plataformas terrestres móviles. La vista oblicua de una escena presenta similitudes y desafíos en la correspondencia con imágenes captadas por sistemas móviles, que se apoya en planos captados en ambos conjuntos de datos. Los planos de fachada extraídos de una nube de puntos poco densa se utilizan como superficies de proyección para
las imágenes del sistema móvil y de las imágenes aéreas, salvando las grandes diferencias de perspectiva entre ellos y simplificando el registro. El procedimiento muestra una tasa de acierto alrededor del 80%.

摘要

移动测绘依赖于卫星定位，而卫星定位受到对空通视与多路径问题的影响。本文提出了一种全自动的移动测绘影像与倾斜航空影像匹配方法，为移动测绘数据平差提供高精度且可靠的地面控制。倾斜影像与移动测绘影像之间具有相似性，但是其匹配亦具有挑战性。本文之方法利用建筑物外墙平面的稀疏点云作为移动测绘影像和航空影像的共同面，克服其间的大视角差异，简化了匹配问题。本方法的匹配结果能达到约 80% 的正确率。