Proceedings of the 34th European Modeling & Simulation Symposium (EMSS), 029
19thInternational Multidisciplinary Modeling & Simulation Multiconference

2724-0037 © 2022 The Authors.
doi: 10.46354/i3m.2022.emss.028

Strategies for Semi-Automated Registration of Historic Aerial Photographs Utilizing Street and Roof Segmentations as Durable Landmarks

Gerald A. Zwettler[1, 3*], Yuta Ono[2], Michael Stradner[1] and Christoph Praschl[1]

[1]Research Group Advanced Information Systems and Technology, Research and Development Department, University of Applied Sciences Upper Austria
[2]Graduate School of Software and Information Science, Iwate Prefectural University Japan
[3]Department of Software Engineering, School of Informatics, Communications and Media, University of Applied Sciences Upper Austria

[*]Corresponding author. Email address: gerald.zwettler@fh-hagenberg.at

Abstract

Historical and current aerial photographs are only of great value if the geolocation or address of the photographed areas is also available. In Western Europe, especially Austria, Germany and Czech Republic, there is a market for the sale of aerial photographs of one's own private residential building. Automated geolocation is a feasible way to enable the sales agents to assign the addresses for the sale more quickly. In the course of this research work, a process chain is modeled that allows the assignment of aerial photographs to residential addresses using machine vision. After model-based rectifying the aerial images to compensate for perspective distortions, larger image blocks get assembled using image stitching. The assignment to a 2D reference map, such as satellite imagery via Google Maps, is done by applying a U-Net CNN after extracting durable image features such as roads or buildings. The mapping of aerial imagery to two-dimensional cartography is either automated via registration approaches or based on manually placed corresponding landmarks and homography. Test runs on imagery between the years 1969 and 2020 show that the labor-intensive process of geolocation of aerial imagery can be solved by the proposed process model in a hybrid way.

Keywords: Georeferencing; Aerial images; Feature-Based Image Registration; SIFT; CNN Street Segmentation

1. Introduction

Historical and current aerial photographs are only suitable for commercial exploitation if it is possible to find out the specific georeferenced position on common maps. While the geolocation of recent images can be determined by evaluating the GPS data (Praschl et al., 2022), this is not possible for archive images from earlier decades. For older recordings, there are different reasons why georeferencing may be of great interest. For example, historical and georeferenced imagery can allow alignment and comparison of buildings, landscapes, roads, and rivers across decades. Furthermore, it is also of great interest for owners of private buildings to acquire aerial photographs that show their own residential property centrally and in high gloss. Especially in the central European area (Germany, Austria, Czech Republic) there is a market for such aerial photographs. Special providers fly over individual cities and create a sequence of individual shots of private homes. However, these images can only be efficiently distributed to private households via salespersons if an ideally automated assignment of the images to the respective geolocations can take place. Otherwise, the manual assignment or search by the selling agents for the correct house number of the photographed contents is a tedious but necessary step in order to be able to aim for a successful sales pitch.

To create these aerial photos, a photographer is transported over the target area via aerodyne (helicopter or airplane or, in the future, possibly also drones) and thereby photographs the individual houses in a professional manner. While the image acquisition process can be carried out quickly, the subsequent manual geolocation as a basis for a possible distribution of the images is a much more labor-intensive process, especially if comprehensive historical archives are to be processed.

Consequently, in the course of this work, procedures for semi-automatic georeferencing are presented, which should significantly accelerate and automate the distribution and processing of image archives in the future.

1.1. State of the Art

Image registration always necessitates a set of robust features being present in both, the moving image *A* and the reference image *B*. Based on these features, image registration can then be applied with either rigid/affine, elastic or perspective transformation of the pixel coordinates as in-depth delineated by Sonka and Fitzpatrick (2000). However, for aerial images due to the acquisition angle and perspective distortion, simple affine transformations such as translation, rotation, scale or sharing are insufficient. Instead, rectification of the images as pre-processing step is highly recommended prior to the particular image registration. Thereby, various aspects of the imaging system can be taken into account.

Jaimes and Castro (2018) introduce an algorithm for aerial image rectification that takes the exact orientation (pitch, roll and yaw) of the aerodyne into account which is generally unknown if no Inertial Measurement Unit (IMU) sensors are attached to the camera device. With an IMU unit being attached for the image acquisition process, aerial image rectification can be achieved in a robust and accurate way as described by Popescu et al. (2015). In rectification of aerial images, incorporation of the geodesic information of the terrain, allows to improve quality of results as delineated by (Cheng et al., 2000), in cases where 3D topography information of the terrain is available. Nevertheless, even with only the FOV angle of the camera and a rough approximation of the recording angle being available, aerial image rectification can help improve the results of subsequent image registration as analyzed by Praschl et al. (2022).

Pixels can be directly utilized as input features only with rigid registration and if the perspective of images *A* and *B* are perfectly matching, allowing for sum of squared errors (SSE) or mutual information (MI) metrics being applied according to the nature of the imaging modalities according to Hajnal and Hill (2001).

To reduce run-time complexity compared to pixel-wise metrics for image registrations, a broad range of feature detectors is applicable. With Hessian and Harris corner detectors, the images get restricted to relevant and discriminable areas, see Harris and Stephens (1988) allowing for registration based on the iterative closest points algorithm as delineated by Arun et al. (1987). If instead of sparse positional features, contour segmentations are present in both of the images, then Chamfer Matching can be applied utilizing an Euclidean Distance Map, see Barrow et al. (1977), which can lead to a massive speed-up compared to ICP. In case of using contours of segmentations as criterion for Chamfer matching image registration, the pixel count can be reduced by extracting the geometric inner object path modelled as a graph rather than the segmentation mask itself. This can be achieved by Hough transformations, cf. Chmielewski (2004), or skeletonization / medial axis extraction, as proposed by Zheng et al. (2010), for example.

If corresponding landmarks are present in both the images *A* and *B*, then the registration process can be solved as numerical problem rather than heuristic search problem. Thereby, corresponding landmarks can result from explicit domain-specific landmark extraction (Seim et al., 2009) or feature descriptors that allow to describe a particular image location in a numerically solvable way such as KAZE (Alcantarilla et al., 2012), ORB (Rublee et al., 2011) or the highly prominent SIFT feature detector (Lowe, 1999) If none of these feature descriptors allows to derive corresponding landmarks in a robust and automated way, manual inspection and placement of the markers can serve as fast, yet manual, fallback strategy allowing for direct image warping based on the homography matrix, see Magalhães (2022) and Szeliski (2006).

1.2. Related Work

Registration of images and in particular aerial image has a broad range of applicability. Exemplarily, image stitching is the process of combining several slightly shifted and/or rotated images to assemble a larger, often panorama photo, (Mann and Picard, 1994; Szeliski, 2006). With the particular images largely overlapping and a narrow perspective variability, these algorithms nowadays allow for panorama image stitching on conventional smartphones.

If the pose of the camera, heavily varies during image acquisition, then conventional image stitching will not be applicable due to perspective warping. Nevertheless, such video sequences of a moving camera recording and a rather static scene are the perfect setup for photogrammetric reconstruction with structure from motion (SfM) (Westoby et al., 2012). Thereby, perspective image transformations and feature-trajectories are incorporated for a 3D surface reconstruction, as available from many applications nowadays (Bianco et al., 2018). In case of aerial images, this reconstructed 3D topography approximates a bird's eye perspective and thus can get registered for geolocalization, too (Aicardi et al., 2016).

In the field of aerial image geocoding, the work of (Allison and Muller, 1993) introduces a highly precise method incorporating multi-spectral images used for a pixel-wise registration process. This high level of accuracy is not necessitated for the reverse geocoding of aerial images to ease the image-to-building assignment for the salespersons as it is the focus of this paper.

Geometric visual features such as lines, curves and areal regions derived from aerial images are used for a landscape and topography-aware image rectification process in the work of Long et al. (2015). In contrast, similar features are used for image registration in our work, only.

For segmentation of durable landmarks, such as rivers and land areas (Pai et al., 2019), buildings (Liu et al., 2020) or streets (Henry et al., 2018) several approaches exist with satellite imagery. Nevertheless, such deep-learning based approaches heavily depend on the nature of the input images. If the recording angle is much smaller than 90° bird's eye perspective of satellite, models might fail, cf railroad surveillance (Wang et al., 2019).

1.3. Strategies for Semi-Automated Registration of Historic Aerial Photographs

In this research work, we aim at geo-referencing of historic aerial photos from image registration. As the sequences are quite sparse, conventional SfM photogrammetry is not possible. Thus, only semi-automated approaches are applicable for this kind of input data. Besides, due to the lack of knowledge regarding the camera pose, conventional approaches for image rectification are not applicable due to the lack of information. Thus, a very simple rectification model for aerial images needs to be applied in this paper, cf. Praschl et al. (2022). In this paper we address the following research questions:

- Is it possible to assemble larger 2D image patches from sparsely overlapping aerial images?
- Can durable landmarks such as streets and building roofs be utilized for automated image registration?
- Can a combination of automatically registered aerial images and manually mapped images allow for reverse geocoding when registering them with a bird's eye map, e.g. from a satellite image?

The article is structured as follows: In Section 2 the image material utilized for this research work is addressed. In Section 3 the methodologies for semi-automated and automated geocoding of the heterogeneous image sequences are delineated, while in Section 4 details regarding the implementation are provided. The practical applicability of the proposed process models is evaluated in Section 5 with quantitative and qualitative results discussed in Section 6. An outlook on future work and the general applicability in the industrial and modelling context completes this article with Section 7.

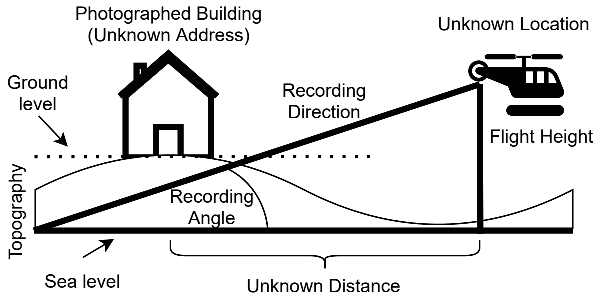

Figure 1. Illustration of the image acquisition. When recording the aerial images, the aerodyne flies along a predefined route with the on-board photographer taking pictures of the buildings at a recording angle $\alpha \approx 40°$. For historic images without GPS the flight height is unknown.

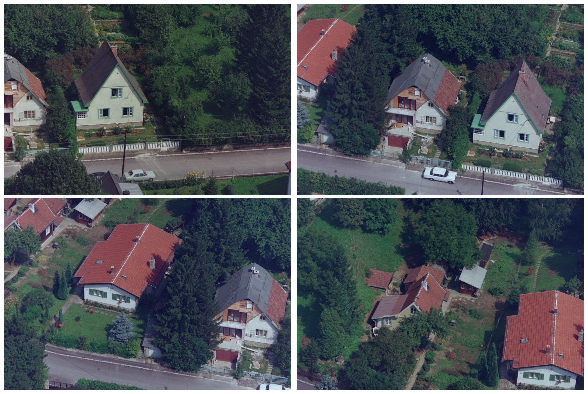

Figure 2. Sample images from year 2012, area code "A 7 058 11". The aerial images 6, 10, 11 and 12 are partially overlapping allowing for subsequent image stitching after rectification.

2. Material

The aerial image data set comprises photograph sequences acquired between the years 1969 and 2020. The older image sequences (years 1969-1972) show a size of 3245×2165 pixels at 24bit gray scale while 24bit RGB images of the years 1979 to 1984 show an iso-image size of 1393×1393 pixels; all with a recording angle around 40°, see Fig. 1.

For the recent years (2017-2020) available GPS data is not utilized for the algorithms. For the most recent photographs starting in 2021, a *GoPro* Cam was statically installed on the aerodyne to provide a video sequence during the aerial image acquisition process.

Image sequences comprise sequentially aligned photographs from a single flight with the trajectory planned prior to the flight, see Fig. 2. The image sequences referring to a single flight thereby comprise $\mu = 84 \pm 50.15[34;162]$ images. Images can either be registered with other images of the sequence to incrementally produce larger conglomerates or be registered with a bird eye reference image, e.g. a satellite image. Inter sequence image registration is considered possible if union of two images A and B covers at least $\sim 5\%$ of the image area. This is the case for ratios $\mu = .2866 \pm .2017[.0857;.6626]$ of the images within the sequences.

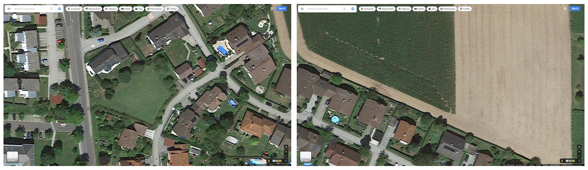

Figure 3. Satellite images at $scale$ = 20 from neighbouring tiles 0_0_1 and 0_0_2 of the mapped target area.

Besides, satellite images are required to develop and test registration for geolocalization. Therefore, Google Maps is utilized, cf. 3. Overall, n = 1000 32bit RGB satellite tiles at pixel size 1920 × 1080 at $scale$ = 17 are prepared to cover the acquisition area of the testing images.

3. Methods

An overview of the proposed automated and semi-automated registration approaches presented in this paper is provided in Fig. 4. For the image sequence to process, a preliminary manual verification is necessary to check, weather the location of the image acquisition is known and if a reference map (e.g. google maps or 3D photogrammetry) is available for this area. Otherwise, the process is terminated as besides partial image stitching no reverse geocoding would be possible.

The aerial imagery is recorded at an angle $\alpha \approx 40^\circ$. This leads to an approximate trapezoid distortion of the images compared to bird's eye perspective and necessitates for correction.

Subsequently, the rectified images get clustered by utilizing automated image stitching. The final step is then to register the clusters of aerial images with the 2D reference map. This can be achieved at two different ways. Either durable and static landmarks get extracted from the images to allow for automated registration or manually placed markers allow the image superposition based on image warping. With the aerial image clusters placed on the georeferenced 2D map, a method for reverse geocoding is given.

3.1. Image Rectification

The aerial imagery is recorded at an angle $\alpha \approx 40^\circ$, cf. Fig. 5. Thus, the subsequent registration with 2D reference maps in bird's eye perspective, such as satellite images, would lack from a mismatch in perspective. Consequently, a rectification process is required that corrects these geometric issues as in-depth delineated in Praschl et al. (2022) by projecting all pixels onto the virtual ground plane utilizing bilinear interpolation. For rectification, only the parameters $width$, $height$, the diagonal FOV angle δ and the approximation for the recording angle α are known. Based on these values, the focal length fl in pixel space as well as the vertical and horizontal FOV angles f_v and f_h respectively can be calculated.

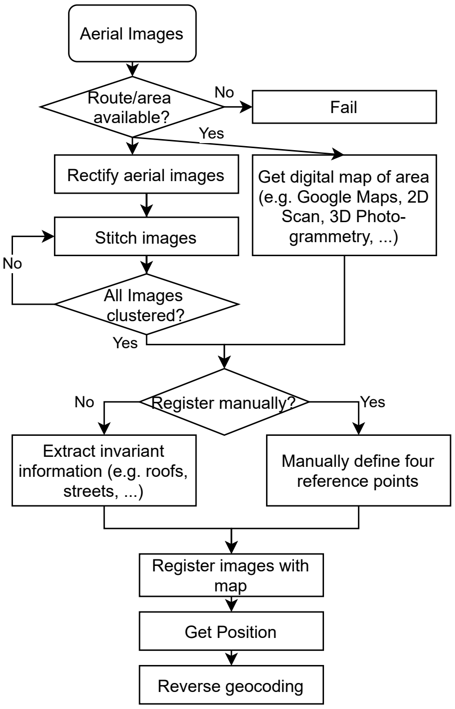

Figure 4. Overview of the Process Model. Input images need to show at least a hint for the rough geolocation to execute the proposed process model. After rectification and stitching of the aerial images, they get registered with the 2D reference map for geocoding by either using automatic segmentations of durable landmarks or by a sufficient number of manually placed landmarks. The final process step of the reverse geocoding model aggregates the clustered images with a 2D reference map in a semi-automated way.

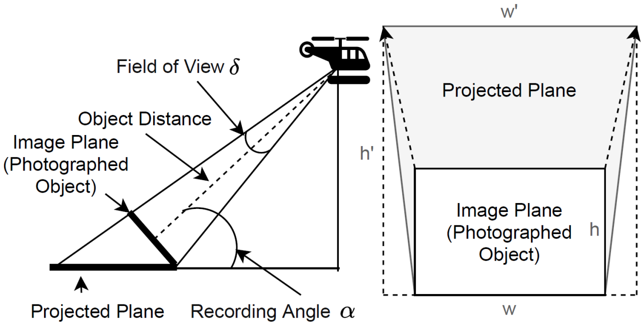

Figure 5. Illustration of the image rectification process. With unknown height and known viewing angle α and FOV angle δ the projected 2D images can get rectified.

3.2. Stitching of Aerial Images

After the rectification process, the aerial images can get stitched if they are partially overlapping. Thus, the input image sequence is transformed into clusters of aerial images manifesting a virtual 2D map. For image stitching, an algorithm based on SIFT (Lowe, 1999) feature mapping is utilized. With the scale- and orientation invariant image features derived from the SIFT operator, prominent areas of the images can get described by a 128-element real-valued vector. With a particular feature and it's neighbour features being present in two images of the sequence, these congruent areas can get matched utilizing a brute force feature matching algorithm available from OpenCV, cf. Noble (2016). Based on the corresponding features, a homography matrix is determined, thereby eliminating potentially outliers, i.e. invalid feature trajectories. Based on this homography matrix, the images get stitched together (Brown and Lowe, 2007).

The described stitching process leads to a reduction in images, as the sequence of aerial image frames gets reduced to locally coherent clusters. Due to smooth border transitions of the stitched images, application of U-Net based segmentation of durable landmarks such as streets or buildings can be processed on the input aerial images and the image clusters after stitching, too.

3.3. Segmentation of Durable Image Features

In general, databases for aerial images with attached reference segmentations are hardly available or result from a totally different image acquisition setup. Thus, our U-Net CNN is trained by utilizing $n = 7500$ RGB satellite images available from a digital map service. Due to the fact that our aerial images get rectified, a very similar bird's eye perspective compared to common satellite imagery is given anyways. Thus, paradigms of transfer learning are applicable, cf. Weiss et al. (2016). Thanks to the digital map service, the ground truth for the streets can get derived from the particular map views from map sections of size 1920×1080 with varying zoom levels. To significantly enrich the data set used for training, several image augmentation strategies get applied to increase the variability regarding orientation, brightness and noise level, cf. Zwettler et al. (2020). In out work, we utilize the following augmentations strategies that are performing well on aerial and satellite images, namely *horizontal/vertical flip*, *scale*, *rotation*, *blur and noise* as well as adaptions on *brightness, contrast, saturation and hue*.

3.4. Automated and Manual Registration With The 2D Reference Map

For automated registration, both the segmentation mask derived from the rectified aerial image as well as the 2D map of the area are necessitated as input. To allow for a Chamfer matching registration, an Euclidean distance map $\mathcal{D}_{euclid}$ is calculated based on the chosen features, e.g. street segmentations, denoted as P and P'' for the aerial image and the reference map, respectively. Based on $\mathcal{D}_{euclid}$, calculated from P'', and P of the moving aerial image a sum-of-squared error (SSE) metric is evaluated as

$$P' = F(P, \theta_{best}, Tx_{best}, Ty_{best}, Scx_{best}, Scy_{best}, Skx_{best}, Sky_{best}) \quad (1)$$

with $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ as affine transform. The overall best transformation F_i is identified by minimal SSE error, see Equation 2. The multi-dimensional and real-valued search space is getting iterated at permutations of the discrete ranges of rotation $\theta \in [-\theta_{min}; \theta_{max}]$, translation along x-axis $T_x \in [-T_x; T_x]$, translation along y-axis $T_y \in [-T_y; T_y]$, as well as x-scaling $Sc_x \in [-Sc_x; Sc_x]$, y-scaling $Sc_y \in [-Sc_y; Sc_y]$, x-skewness $Sk_x \in [-Sk_x; Sk_x]$ and y-skewness $Sk_y \in [-Sk_y; Sk_y]$.

$$\begin{aligned} \theta_{best}, Tx_{best}, Ty_{best}, Scx_{best}, Scy_{best}, Skx_{best}, Sky_{best} &= \\ \underset{\theta, T_x, T_y, Sc_x, Sc_y, Sk_x, Sk_y}{\arg\min} \sum_{i=1}^{|P|} (\mathcal{D}_{euclid}(P'')[F(P, \theta, T_x, T_y, & \quad (2) \\ Sc_x, Sc_y, Sk_x, Sk_y)[i]])^2 & \end{aligned}$$

The continuous search space is discretized at $k = 11$ steps, leading to a search radius $r = 5$ around the global best F_{best} after each entire iteration. To incrementally adjust from global to local search, for the $m = 10$ optimization runs, the search area is scaled with factor $s_i = 0.9^i$ for each of the subsequent iterations.

In case of the images to register being acquired from totally different viewing directions or no robust markers/landmarks being visible in the particular aerial images that could be detected, the automated registration process will fail. Consequently, a proper fallback strategy is necessitated, allowing to finish the image stitching and registration in a semi-automated way. To do so, at least 4 corresponding landmarks need to get placed in both of the images taking care, that they refer to positions in 3D space of approximately the same altitude. Based on defined reference positions, homography is calculated for automated affine warping of the images, see Magalhães (2022); Szeliski (2006).

4. Implementation

The method presented in this research work are implemented with Python 3.7 and OpenCV 4.5.4 (Bradski, 2000). For training the U-Net on street and building segmentation, Tensorflow 2 (Abadi et al., 2016) in version 2.8 is utilized.

Image warping based on homography is utilized via `cv::warpAffine()` method instead of `cv::warpPerspective()` as the aerial images are already rectified at the time of calling.

For the street segmentation U-Net, the solution as proposed by Ronneberger et al. (2015) is re-implemented with Tensorflow. Thereby, $n = 5$ up and down-sampling layers are implemented with input tensor size of $(3, 112, 112)$ at a batch size of 16 and smallest image size $(3, 8, 8)$ during U-Net down-sampling. For training, *Adam* optimizer with a learning rate of $l = 0.003$ and 200 epochs is utilized.

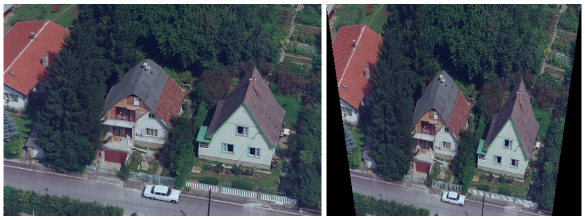

Figure 6. Test runs on image 10 of sequence "A 7 058 11" from 2012. The original aerial image and the result after rectification.

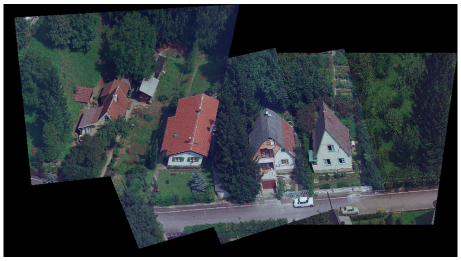

Figure 7. Test runs on images of sequence "A 7 058 11" from 2012. With rectified images 6, 10, 11 and 12 image stitching of the particular street of houses becomes possible.

5. Results

In this chapter, results on the particular process steps of the processing pipeline for semi-automated registration of aerial photographs are presented.

5.1. Results from Image Rectification

Tests on the image rectification process show that even with a rough and thus potentially inaccurate approximation of the recording angle α the geometric properties of the aerial images get improved. Analyzing potentially linear and parallel structures such as roof tops shows an increased level of straightness compared to the input data when analyzing the Hough space (Chmielewski, 2004).

Tests on the image sequences show that rectified images are perfectly applicable for the delineated image stitching process as long as a high enough ratio of the particular image areas is overlapping, see Fig. 6.

5.2. Results from Stitching of Aerial Images

Based on the rectified images as seen in Fig. 2, the subsequent process of image stitching is boosted as all the images to assemble are now approximately in bird's eye perspective. Results of stitching four neighbouring and partly overlapping images can be found in Fig. 7.

5.3. Results from Segmentation of Image Features

Exemplary results from U-Net based street segmentation as a robust and durable image feature can be found in Fig. 8. Thanks to the chosen augmentation process for the training, streets can be segmented in a solid way for all areas of the image. In case of occlusions, the mask result is of course missing. The TP areas of the streets are compact and free from notable artifacts. In contrast, FP areas are formed from small bulks of artifacts.

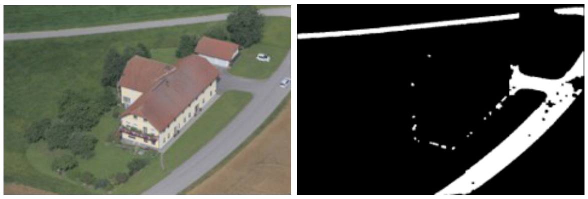

Figure 8. U-net based segmentation allows solid identification of the streets as target structures. Despite occlusions, the particular street fragments are provided at high quality with only few FP scatter present.

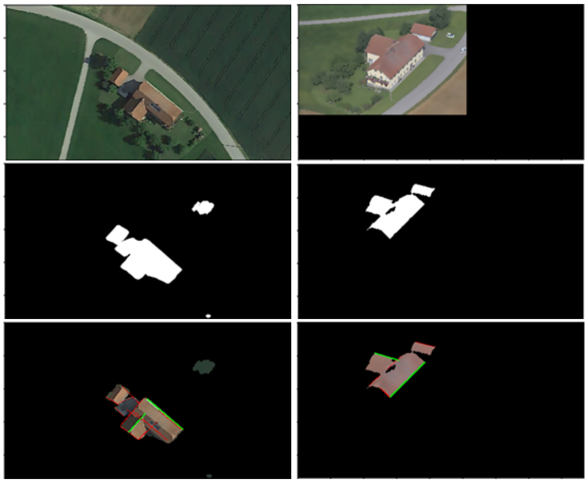

Figure 9. Roof segmentation from reference map (first column) and aerial image (second column) allows for roof-based registration after Hough line detection.

Besides for street segmentation, another U-Net is trained for building (roof) segmentation. Based on the detected roof sections, the outer contours can get detected utilizing Hough lines to serve as registration features, see Fig. 9.

5.4. Results on Registration with the 2D Reference Map

Test runs on the automated registration approach proof that a) the quality of results from the subsequent image segmentation is sufficient and b) that the fully-automated registration approach leads to correct results even if the perspective of the images is differing and only parts of the pre-segmented image features are congruent. See Fig. 10 for results on the image registration process. Even with both streets showing a highly different shape, orientation, scale and segmentation quality, the chosen search strategy leads to a very good match of the image areas.

Another registration example is presented in Fig. 11. With parameter settings $-45 \leq skew \leq 45$ and $0.8 \leq scale \leq$ after 15,000 evaluated positions the error regarding orientation is 1.245° only. Results are definitely ac-

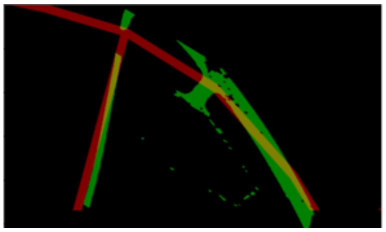

Figure 10. Based on Euclidean distance maps derived from the street segmentations derived from the previous process the image registration is performed. Even with a clear mismatch in street characteristics, i.e. missing parts, street thickness, the local optimum is approached in a robust manner.

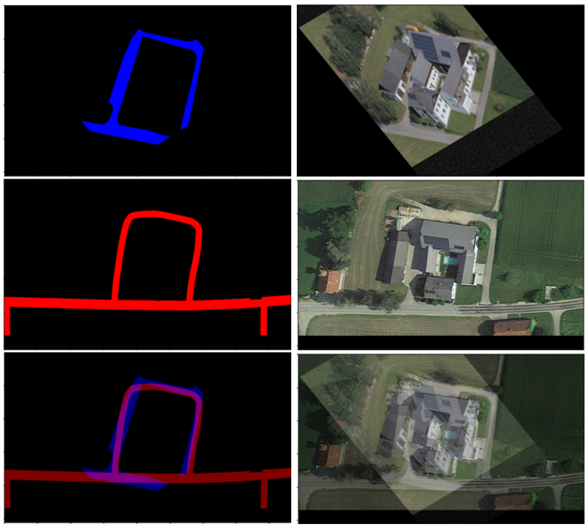

Figure 11. Registration of aerial images (top row) with Google maps satellite (middle row) leading to precise overlap and thus geolocalization.

curate enough to determine the address of the particular buildings.

Registration based on manual markers is a solid backup-approach. Nevertheless, the markers should be a) defined close to the ground level and b) spanning up a larger area. Otherwise, image warping with homography might lead to inaccurate results necessitating for registration-based post-processing, see Fig. 12.

6. Discussion and Conclusion

Regarding the described image processing pipeline, initial rectification of the aerial images proofs to be of high importance. For the subsequent image stitching it is inevitable, that due to the rectification the reduced image area is overcompensated by a better geometric match of the image content. Result quality heavily depends on the local landscape. The flatter the surrounding region is, the better our model assumptions allow to describe the real word. The same aspect is true w.r.t. buildings and other objects of height. While the rectification allows to correct ground areas very well, objects showing a noteworthy height cannot be correctly transformed into bird's eye perspective.

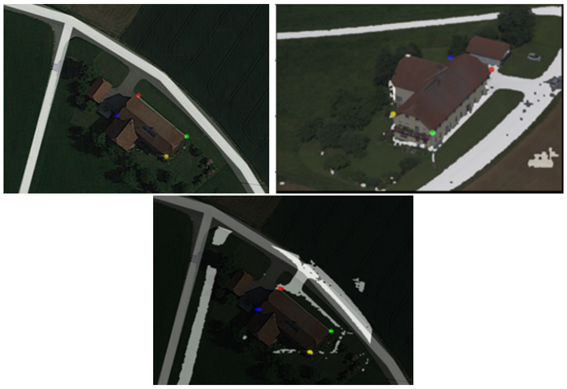

Figure 12. Registration based on 4 manual landmarks precisely defined on the building roof.

The automated process of registration leads to robust results even if the street segmentations are not of highest quality. With the chosen registration approach based on the Euclidean distance map, partial occlusions and mismatch in object thickness and perspective can get compensated quite well. The chosen search strategy allows to smoothly transition from global to local search at linear run-time w.r.t. the target number of search iterations. Features derived close to the ground level and spanning a larger area are to be preferred. Thus, street segmentations lead to better results compared to roofs, cf. Fig. 12.

Regarding the search strategy, the current approach features a strictly deterministic search through the continuous search space which leads to a solid level of quality. Nevertheless, incorporating aspects of gradient descent search or simulated annealing would allow to introduce some stochasticity and thus increase the chance to further approach the maxima of the search space.

7. Outlook

Based on the presented methods and tool sets for semi-automated aerial image registration, the step-wise integration into the business workflow of AMIDO trading ltd. can be achieved. Thereby, the focus will be laid on a semi-automated strategy allowing for the best balance between automation and expert-user guidance.

The registration of aerial images based on the road trajectories can be further refined in the future by using only the inner course, cf. skeleton or medial axis, instead of the cross-sectional area of the roads. By applying mixed-adjacency, the road course can be transformed into a vectorized graph data structure. This has the advantage that geolocations can be found from distinctive reference streets as rough registration. Furthermore, it is expected that the restriction to the logical road course can lead to advantages in terms of performance as well as accuracy.

Besides the discussed street and roof segmentations being utilized as landmarks for the registration process, the applicability of rivers with their characteristic and durable course will be evaluated in future, too. Not all local villages and urban entities will show a stream course. Nevertheless, in the case of a river being present, it is highly expected to serve as a strong and prominent landmark detection system further allowing to identify the rough geolocation from graph analysis of the river course.

Furthermore, the landmarks from street, river courses as well as from roofs can then be combined in a multi-criteria registration approach to strengthen robustness of the proposed algorithms for semi-automated geolocation.

In addition to the practical application of the proposed process model for the industrial domain, the individual process steps, such as image rectification or even the extraction of durable image properties, are a basis for future modelling and simulation. For example, if test data is to be synthesized, the same intentional bias model can be applied to prepare realistic scenarios, such as for machine learning applications. Furthermore, the evaluation and modeling of the decades-long landmarks opens up targeted strategies for image registration in the future.

8. Fundings

Our thanks to the provincial government of Upper Austria, for funding the project "Amido: Research Prototypes for Image-Guided Georeferencing of Aerial Images" with funding program number Wi-2021-287054/25.

9. Acknowledgments

Many thanks to Mr. *Andreas Razenberger*, company director of the AMIDO Handelsges.m.b.H and project initiator, for the provided aerial images and valuable discussion.

References

Abadi, M. et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, abs/1603.04467.

Aicardi, I., Nex, F., Gerke, M., and LINGUA, A. (2016). An image-based approach for the co-registration of multi-temporal uav image datasets. *Remote Sensing*, 9.

Alcantarilla, P. F., Bartoli, A., and Davison, A. J. (2012). Kaze features. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012*, pages 214–227. Springer Berlin Heidelberg.

Allison, D. and Muller, J. (1993). An automated system for sub-pixel correction and geocoding of multi-spectral and multi-look aerial imagery. *Int. Archives of Photogrammetry and Remote Sensing*.

Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE Trans. on Pat. Analysis and Machine Intelligence*, PAMI-9:698–700.

Barrow, H. G., Tenenbaum, J. M., Bolles, R. C., and Wolf, H. C. (1977). Parametric correspondence and chamfer matching: Two new techniques for image matching. In *IJCAI*, pages 659–663.

Bianco, S., Ciocca, G., and Marelli, D. (2018). Evaluating the performance of structure from motion pipelines. *Journal of Imaging*, 4:98.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brown, M. and Lowe, D. (2007). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74:59–73.

Cheng, K.-S., Yeh, H.-C., and Tsai, C.-H. (2000). An anisotropic spatial modeling approach for remote sensing image rectification. *Remote Sensing of Environment*, 73(1):46–54.

Chmielewski, L. (2004). Choice of the hough transform for image registration. In *Proc. SPIE*, volume 5505.

Hajnal, J. and Hill, D. (2001). *Medical Image Registration*. Biomedical Engineering. CRC Press.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151.

Henry, C., Azimi, S., and Merkle, N. (2018). Road segmentation in sar satellite images with deep fully convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*.

Jaimes, B. and Castro, C. (2018). Perspective correction in aerial images. 10.13140/RG.2.2.34885.29926.

Liu, Z., Chen, B., and Zhang, A. (2020). Building segmentation from satellite imagery using u-net with resnet encoder. In *2020 5th Int. Conf. on Mech., Control and Comp.r Eng. (ICMCCE)*, pages 1967–1971.

Long, T., Jiao, W., He, G., Zhang, Z., Cheng, B., and Wang, W. (2015). A generic framework for image rectification using multiple types of feature. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102:161–171.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ICCV '99, page 1150, USA. IEEE Computer Society.

Magalhães, M. (2022). Accurate image alignment and registration using opencv.

Mann, S. and Picard, R. W. (1994). Virtual bellows: constructing high quality stills from video. *Proc. of 1st Int. Conf. on Image Processing*, 1:363–367 vol.1.

Noble, F. K. (2016). Comparison of opencv's feature detectors and feature matchers. In *2016 23rd Int. Conf. on Mechatronics and Machine Vision in Practice*.

Pai, M. M., Mehrotra, V., Aiyar, S., Verma, U., and Pai, R. M. (2019). Automatic segmentation of river and land in sar images: A deep learning approach. In *2019 IEEE AIKE*, pages 15–20.

Popescu, G., Balota, L., and Iordan, D. (2015). Section photogrammetry and remote sensing direct georeferencing application of aerial photogrammetry using a gnss/imu/sensor system. In *In Proc. of the SGEM Conf., Book 2 Vol.1*, pages 1043–1050.

Praschl, C., Stradner, M., Ono, Y., and Zwettler, G. A. (2022). "towards an automated system for reverse geocoding of aerial photographs". In *Proc. of the 30th Int. Conf. WSCG2022*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *2011 Int. Conf. on Computer Vision*, pages 2564–2571.

Seim, H., Kainmueller, D., Heller, M., Zachow, S., and Hege, H.-C. (2009). Automatic extraction of anatomical landmarks from medical image data: An evaluation of different methods. In *2009 IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, pages 538–541.

Sonka, M. and Fitzpatrick, J. M. (2000). *"Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis"*. SPIE - Reprint edition (June 14, 2000).

Szeliski, R. (2006). *Image Alignment and Stitching*, pages 273–292. Springer US, Boston, MA.

Wang, Y., Wang, L., Hu, Y. H., and Qiu, J. (2019). Railnet: A segmentation network for railroad detection. *IEEE Access*, 7:143772–143779.

Weiss, K., Khoshgoftaar, T., and Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3.

Westoby, M., Brasington, J., Glasser, N., Hambrey, M., and Reynolds, J. (2012). 'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179:300–314.

Zheng, Q., Sharf, A., Tagliasacchi, A., Chen, B., Zhang, H., Sheffer, A., and Cohen-Or, D. (2010). Consensus skeleton for non-rigid space-time registration. *Comput. Graph. Forum*, 29:635–644.

Zwettler, G., Holmes III, D., and Backfrieder, W. (2020). Strategies for training deep learning models in medical domains with small reference datasets. *Journal of WSCG*, 28(1-2):37–46.