ISSN 2464-4617 (print)
ISSN 2464-4625 (DVD)
Computer Science Research Notes
CSRN 3201
WSCG 2022 Proceedings

Towards an Automated System for Reverse Geocoding of Aerial Photographs

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

[1]{christoph.praschl, michael.stradner, gerald.zwettler}@fh-hagenberg.at

[2]g236s001@s.iwate-pu.ac.jp

[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

ABSTRACT

Aerial photographs of buildings are often used as memorabilia sold by trading companies. Such photographs come with an issue regarding the address of the shown buildings, since the recording location of the camera may be known, but shows a spatial distance to the actual subject of the image. In addition to that, also this recording location is often not known in detail but only roughly in the form of the flight route/area. To address this problem, a methodology for reverse geocoding is proposed, allowing to identify the position of buildings that are photographed from aerial vehicles. This is done using a process for extending recording locations and a second process based on the registration of invariant features within aerial shots compared to maps.

Keywords

Aerial Photography, Reverse Geocoding, Building, Segmentation

1 INTRODUCTION

Aerial shots of private buildings are a kind of memorabilia, that is typical for european countries, especially Austria. For this, trading companies are flying across the country with helicopters, airplanes or even drones making those unique products. Especially, in the case of the first two vehicles, a pilot is accompanied by an additional photographer. This photographer is typically using some system camera in combination with a telephoto lens and is responsible for creating the photos. To do so, the photographer sits in the respective aerial vehicle and tries to capture a good clipping of the target buildings from the vehicle's side window by adapting the recording angle (as far as possible) and especially by changing the zoom level. Depending on the area, respectively the number of buildings at the specific location, multiple clippings may be overlapping and show related information like neighborhoods. After the flight, salespersons are trying to reference the photographs with maps to identify the exact addresses and like this be able to contact the owners of the buildings as basis for a sales pitch. Especially, this manual reference process still requires a lot of time and work due to the situation that although the position of the aerial vehicle is known, the actual viewing direction of the photographer as well as the position and with this the addresses of the focused buildings are unknown. Next to already globally referenced photographs, such trading companies often also have huge archives of never sold images, for which not even the exact recording position but only the rough flying route/area is known. To tackle these problems, the present work introduces a methodology for reverse geocoding aerial shots.

2 MATERIAL

The proposed methodology is developed in reference to the image archive of our project partner. This archive consists of hundreds of images, which partially contain coordinates of the Global Positioning System (GPS) [HWLC12] related to the recording positions. Next to images with GPS information, there are also many images with an artificial annotation, referencing a physical map or a textual protocol, where the flight area/route is recorded. One sample for each case is shown in Figure 1.

3 METHODOLOGY

In this work a methodology is proposed based on two independent sub processes, with the first process used for images, where the recording location is known, and the second process, where only the rough flight

(a) (b)

Figure 1: Samples images of the used aerial photography archive showing (a) a not geo-referenced image from 2004 with a textual note at the left corner (zoomed in for better visibility) and (b) an image from 2018 for which the recording location is known in the form of EXIF data.

route/area is known, which leads to time-intensive calculations. This methodology is shown in Figure 2.

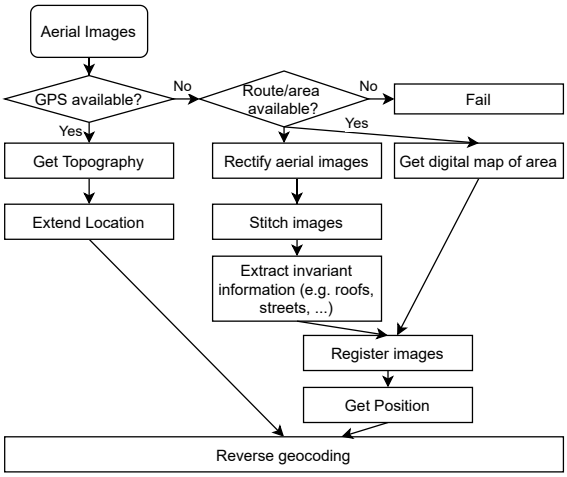


Figure 2: The overall process used for geocoding aerial images. Depending on the situation, if the images contain GPS information, the process either tries to reference the photographed buildings based on a topography model of the area or based on a registration process in reference to digital map services.

3.1 Extending Recording Location

For images, where the recording location is known in the form of GPS information, the reverse geocoding is done based on a ray tracing approach that tries to find the position and with this the address of the building based on the position of the helicopter, its flight height and a topography model of the surrounding area. The topography model is required because of the deviating ground level of the photographed building due to e.g. hills compared to the sea level, which is used as reference of the flight height. This setup is shown in Figure 3.

The GPS positions of the recording locations during a flight allow estimating the route of the helicopter/airplane using splines. In combination with the knowledge, that the photographer either takes the images from

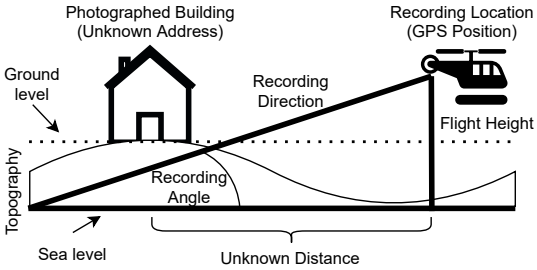


Figure 3: Record situation showing the known record position, the flight height and the unknown distance as well recording angle.

the left or the right side window of the aerial vehicle, the possible recording direction can be estimated orthogonal to the flight route as shown in Figure 4. This recording direction can be represented as global bearing angle β, with $0°$ representing the magnetic north and $180°$ pointing to the south. Finally, the recording angle can also be estimated to around $45° \pm 10°$, due to the situation that the photographer tries to create a clipping, where not only the roof but also the building itself is visible, so a direct overflight or a too flat angle are unsuitable. Knowing both, the flight height fh and the recording angle α, allows calculating the distance d between the aerial vehicle and the targeted position on sea level, like:

$$d = fh/tan(\alpha) \tag{1}$$

Utilizing the earth radius e, the bearing angle β and the recording location defined by its latitude lat and longitude lng the targeted position can be calculated, which is defined by lat_2 and lng_2, as:

$$lat_2 = asin(sin(lat) * cos(d/e) + cos(lat) * sin(d) * cos(\beta)) \tag{2}$$

$$a = sin(\beta) * sin(d/e) * cos(lat) \tag{3}$$

$$b = cos(d/e) - sin(lat) * sin(lat_2) \tag{4}$$

$$lng_2 = lng + atan2(a,b) \tag{5}$$

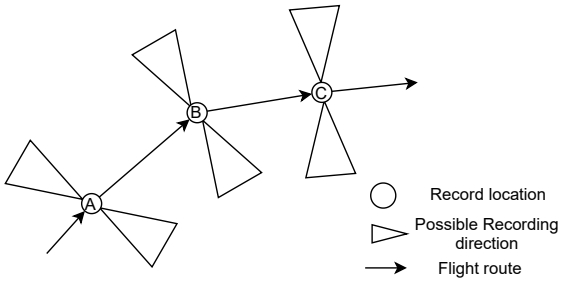


Figure 4: Based on the known recording locations of aerial shots, the flight route of the aerial vehicle can be reconstructed, allowing to estimate the recording direction.

Knowing both the recording position and the targeted position, allows to sample-wise determine the position of the photographed building. This is done by creating n sample points with a distance d/n starting from the recording position towards to the target position. Based on the topography model, the geo-referenced position of the photographed building can be reconstructed by finding the best matching intersection of the sample positions and the trace between recording and target position. Knowing the GPS position of the building, a reverse geocoding service like Google Maps [Sve10] can be used to provide suitable addresses to the salesperson.

3.2 Comparison with digital map services

Next to the tracing based approach, the second process is intended for images where only the rough flight route/area is known. This process is based on the idea that multiple aerial shots show a partially overlapping area with more or less invariant information such as streets, roofs or even natural features such as rivers. Within a registration process these invariant information is compared to a digital map to identify the location of the photographed area.

3.2.1 Rectification

In a first step, the aerial images are rectified by projecting the image plane onto the ground plane, as shown in Figure 5. This step is required to approach a bird's eye view, comparable to the digital map used in the registration step. This is done based on the image meta information in the form of its width w, height h, the estimated recording angle α and the field of view angle δ of the used camera lens. Based on this information, the focal length fl in pixel and the vertical field of view f_v as well as the horizontal one f_h can be calculated with the subsequent formulas:

$$fl = sqrt(w^2 * h^2)/2 \tag{6}$$

$$f_h = atan(w/2/fl) * 2 \tag{7}$$

$$f_v = atan(h/2/fl) * 2 \tag{8}$$

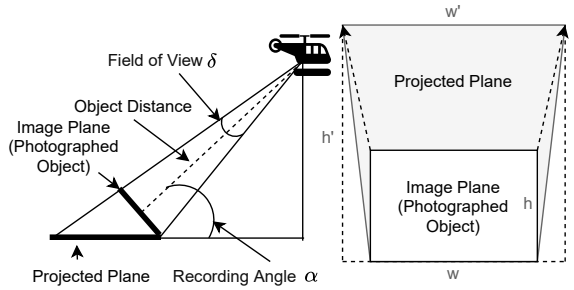


Figure 5: Projection of the image plane showing the photographed building to the ground plane within a rectification process.

Using these values in turn allows to calculate the general projection information like the height h' of the projected image using the following trigonomic equations:

$$\varepsilon = 90 - \alpha \tag{9}$$

$$\lambda = (180 - \delta)/2 \tag{10}$$

$$\mu = 180 - \lambda \tag{11}$$

$$\omega = 180 - \mu - \varepsilon \tag{12}$$

$$p = h * sin(\varepsilon)/sin(\omega) \tag{13}$$

$$h' = sqrt(h^2 + p^2 - 2 * h * p * cos(\mu)) \tag{14}$$

Using this information, also the width w' of the projected image can be calculated like:

$$e = a/sin(\delta/2) \tag{15}$$

$$d = p + e \tag{16}$$

$$w' = sin(f_h/2) * d * 2 \tag{17}$$

Knowing both the target height and width allows calculating the respective projection angle Ω for every row defined by y within the range [0, h] like:

$$g = abs(y - h/2) \tag{18}$$

$$\phi = atan(g/fl) \tag{19}$$

$$\Omega = \left\{ \begin{array}{ll} \phi + (f_v/2), & \text{for } y > h/2 \\ f_v/2, & \text{for } y == h/2 \\ (f_v/2) - \phi, & \text{for } y < h/2 \end{array} \right\} \tag{20}$$

Based on Ω the actual pixel mapping $I(x,y) \rightarrow T(x',y')$ for the source image I to its projection T can be applied using multiple sub steps as shown below:

$$\zeta = 180 - \Omega - \lambda \tag{21}$$

$$\gamma = 180 - \zeta \tag{22}$$

$$\rho = 180 - \gamma - \varepsilon \tag{23}$$

$$p = y * sin(\varepsilon)/sin(\rho) \tag{24}$$

$$y' = sqrt(y^2 + p^2 - 2 * y * p * cos(\mu)) \tag{25}$$

$$f = sqrt(e^2 + y^2 - 2 * e * y * cos(\lambda)) \tag{26}$$

$$w'' = sin(f_h/2) * (p + f) * 2 \tag{27}$$

$$u = sqrt((p + f)^2 - (w''/2)^2) \tag{28}$$

$$o = abs(x - w/2) \tag{29}$$

$$\Phi = atan(o/fl) \tag{30}$$

$$i = u * tan(\Phi) \tag{31}$$

$$x' = \left\{ \begin{array}{ll} (w''/2) - i, & \text{for } x < w/2 \\ (w''/2) + i, & \text{for } x >= 2/2 \end{array} \right\} \tag{32}$$

3.2.2 Image stitching and extraction of invariant visual features

After the rectification of the spatially connected aerial shots, the images are stitched together to create a virtual map. To do so, a SIFT [Low99] operator is used to extract scale invariant features in the images that are compared using a brute force feature matcher [Nob16]. Based on the retrieved information of this feature matching process, a homography matrix is determined, that is used to warp the individual images, so they can be stitched together [BL07].

Using the stitched image, invariant visual features such as streets, roofs or rivers can be extracted. To do so, one or multiple segmentation models such as U-Nets [RFB15] can be used to separate the background pixels from the desired foreground pixels, representing the mentioned invariant information.

3.2.3 Registration

Using both the segmentation mask for the stitched aerial shot as well as a digital map of the flight area, a registration process can be used to identify the geo-referenced position of the photographed buildings. This registration is done using an euclidean distance map $\mathscr{D}_{euclid}$ with P and P'' representing the set of pixels of invariant features. The rigid registration problem is defined using the mean-squared error (MSE) metric as

$$P' = Trans(P, \theta_{best}, T_{x_{best}}, T_{y_{best}}, Sc_{x_{best}}, Sc_{y_{best}}, Sk_{x_{best}}, Sk_{y_{best}}) \tag{33}$$

where the best transformation parameters lead to minimal squared distances between P' and P''. These best parameters are determined in a discrete search space with 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-skweness $Sk_x \in [-Sk_x; Sk_x]$ and y-skweness $Sk_y \in [-Sk_y; Sk_y]$.

$$\theta_{best}, T_{x_{best}}, T_{y_{best}}, Sc_{x_{best}}, Sc_{y_{best}}, Sk_{x_{best}}, Sk_{y_{best}} = \underset{\theta, T_x, T_y, Sc_x, Sc_y, Sk_x, Sk_y}{\operatorname{argmin}} \sum_{i=1}^{|P|} (\mathscr{D}_{euclid}(P'')[Trans(P, \theta, T_x, T_y, Sc_x, Sc_y, Sk_x, Sk_y)[i]])^2 \tag{34}$$

The discrete search space thereby comprises $k = 11$ steps, i.e. radius $r = 5$, for each of the seven variables $(\theta, T_x, T_y, Sc_x, Sc_y, Sk_x, Sk_y)$ and for each of the $m = 10$ optimization runs according to the scale factor s_i with $[-r*s_i, -(r-1)*s_i, ..., 0, ..., (r-1)*s_i, r*s_i]$ as search offset for globally optimal parameters from the last entire run. To move from global to local search with increasing number of optimization runs performed, the search space scale factor is reduced with $s_i = s_{i-1} * 0.9$ and initial s_1 defined from image resolution.

4 IMPLEMENTATION

The presented methodology is implemented using Python [vR95] and the OpenCV library [Bra00]. For the extraction of invariant landscape features in the form of streets, a U-Net model has been trained using Tensorflow [AAB+16]. Due to the lack of a suitable training dataset for segmented aerial photographs, this convolutional neural network is trained using 7500 RGB satellite images from a digital map service. Despite the different perspectives of satellite images compared to aerial photographs, both capture methods represent bird's eye views of the recorded area and are intended as interchangeable in the sense of a transfer learning methodology [WKW16] for the proposed area of application. As ground truth for the streets, the corresponding map views of these satellite images are used. The used training dataset contains map sections with a size of 1920×1080 px from different zoom levels. Based on this base dataset, multiple augmentation strategies are applied to further increase the amount of images and with this the variety of image properties regarding e.g. the brightness. For this task, a random selection of up to five augmentation methods per image is applied, allowing to vertically and/or horizontally flip, scale, rotate, translate, blur the image or changing its brightness, contrast, saturation or hue.

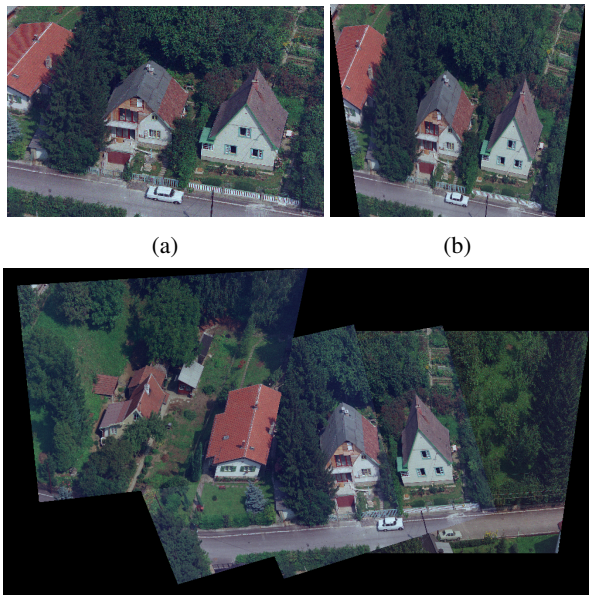
5 RESULTS

For the evaluation of the first sub process based on the extension of the recording location, an image subset of 36 subsequent aerial shots is used from our project partner's archive. Using the GPS coordinates of these images allows reconstructing the flight route and to retrieve the address of the photographed buildings. This process is tested using Google Maps as digital map service for the final reverse geocoding step and the Open Elevation API[1] for the topography model. This sub process is manually evaluated by comparing the proposed address with its real world counterpart. The addresses are identified correctly for 14 of the 36 shots. For the remaining images, the reversed geocoded address deviates from the real one with in a range of some streets to completely wrong positions.

Next to that, the individual steps of the second approach are also evaluated. First, the rectification and the image stitching steps are evaluated using a sequence of four partially overlapping aerial shots, as shown exemplary in Figure 6. Additionally, also the extraction of invariant features is tested using a U-Net model, that allows to segment aerial photographs as proposed. Applying this segmentation model allows to extract invariant information in the form of streets, as shown in Figure 7. Based on such invariant features, the registration approach can

[1] https://open-elevation.com/

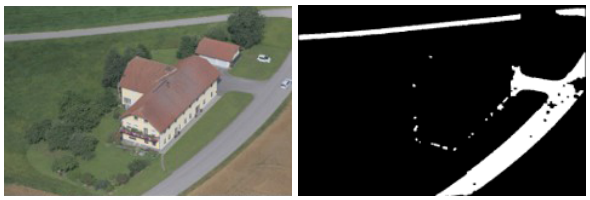
be used to automatically find the best matching position within multiple map sections in questions. Such reference images can for example be retrieved from Google Maps for the known flight area, as shown in Figure 8.



(a) (b)

(c)

Figure 6: (a) One sample aerial image that is (b) rectified using the proposed approach and (c) stitched together with three additional rectified images.



(a) (b)

Figure 7: (a) An aerial shot for which the U-Net segmentation model is applied to create (b) a street binary mask

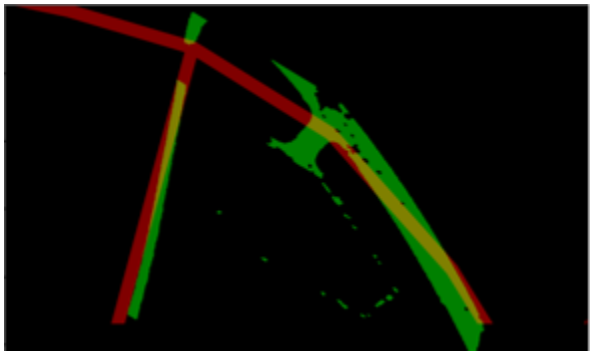


Figure 8: Overlayed binary masks showing streets for a requested area from a digital map service (red) and the registered segmentation result (green) for the aerial shot shown in Figure 7b.

6 RELATED WORK

Jaimes and Castro [JC18] describe a comparable method for the rectification of aerial images. In comparison to our approach, the authors also take the exact orientation (pitch, roll and yaw) of the aerial vehicle into account, which is unknown in our use case.

Nowak [Nov92] compares methods such as rectification for the creation of digital orthophotos. As in our approach, the rectification is done based on the projection of the image plane onto the photographed ground plane and its correction based on the landscape of the photographed area. In contrast to our work, the author also includes exterior orientation parameters (c.f. Jaimes and Castro [JC18]).

Additionally, also Cheng et al. [CYT00] compare approaches in the context of rectifying remote sensing images utilizing polynomial trend mappings, multiquadric interpolation functions, and their own proposed ordinary kriging technique. By taking the spatial structure variation of the terrain into account, the authors are able to outperform the two other approaches. In contrast to our rectification method, the accuracy and with this also the calculation complexity of this anisotropic spatial modeling approach is not required and would require exact knowledge of the photographed area, which is not known.

Allison and Muller [AM93] describe a method for geocoding aerial images. In comparison to our work, the authors have registered multi-spectral images and focus on a pixel-wise registration process. This accuracy is not required for the reverse geocoding of aerial images of buildings used by salesperson.

Karel et al. [KDV+13] present a method for georeferencing aerial photographs in the context of archaeological applications. To do so, the authors propose a rectification approach. Like in the image stitching step of our approach, the authors try to automatically find the relative orientation of the aerial photographs using scale invariant keypoints to combine the available visual information. The so oriented images are then geo-referenced for reconstructing a sparse point cloud. In contrast to our work, the authors are using their approach for creating a 3D reconstruction in the form of an orthophoto map of the photographed scenery and are not doing a reverse geocoding, since the position of the region of interest is known for such archaeological applications.

Long et al. [LJH+15] present a generic framework for the rectification in the context of remote sensing. Like in our registration approach, the idea of this framework is the utilization of invariant, landscape features. The authors present a feature extraction method allowing to find visual features in landscapes including points, straight lines, free-form curves and areal regions, that

are used for the rectification process. In contrast to our work, the authors neither use invariant landscape features for the registration of images, nor for a reverse geocoding process, but for the automatic rectification.

7 CONCLUSION AND OUTLOOK

The proposed methodology shows promise for a semi-automated reverse geocoding process for aerial shots allowing to decrease the required amount of time for salespersons to identify the address of buildings shown in such images. In the future, we plan to extend our tests and improve the current results, especially for the recording location extension approach. Additionally, a retraining of the created segmentation model and the training of additional models for other invariant information such as roofs or rivers is planned.

ACKNOWLEDGMENT

Our thanks to our research partners of the Amido HandelsgesmbH, especially the owner Andreas Ratzenberger, for the close cooperation and the efforts. Next to that, we also thank the province of Upper Austria for facilitating the research project with the easy2innovate funding program as well as the Austrian Research Promotion Agency with the Innovationscheck+ program.

REFERENCES

[AAB+16] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.

[AM93] David Allison and Jan Muller. An automated system for sub-pixel correction and geocoding of multi-spectral and multi-look aerial imagery. Int. Archives of Photogrammetry and Remote Sensing, 1993.

[BL07] Matthew Brown and David G Lowe. Automatic panoramic image stitching using invariant features. International journal of computer vision, 74(1):59–73, 2007.

[Bra00] G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[CYT00] Ke-Sheng Cheng, Hui-Chung Yeh, and Chang-Hsuan Tsai. An anisotropic spatial modeling approach for remote sensing image rectification. Remote Sensing of Environment, 73(1):46–54, 2000.

[HWLC12] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger, and James Collins. Global positioning system: theory and practice. Springer Science & Business Media, 2012.

[JC18] B Jaimes and C Castro. Perspective correction in aerial images. 10.13140/RG.2.2.34885.29926, 2018.

[KDV+13] Wilfried Karel, Michael Doneus, Geert Verhoeven, Christian Briese, Camillo Ressl, and Norbert Pfeifer. Oriental: Automatic geo-referencing and orthorectification of archaeological aerial photographs. In XXIV International CIPA Symposium, volume 2, pages 175–180, 2013.

[LJH+15] Tengfei Long, Weili Jiao, Guojin He, Zhaoming Zhang, Bo Cheng, and Wei Wang. A generic framework for image rectification using multiple types of feature. ISPRS Journal of Photogrammetry and Remote Sensing, 102:161–171, 2015.

[Low99] David G Lowe. Object recognition from local scale-invariant features. In Proc. of 7th IEEE int. conf. on comp. vision, 1999.

[Nob16] Frazer K Noble. Comparison of opencv's feature detectors and feature matchers. In 2016 23rd Int. Conf. on Mechatronics and Machine Vision in Practice (M2VIP), 2016.

[Nov92] Kurt Novak. Rectification of digital imagery. Photogrammetric engineering and remote sensing, 58:339–339, 1992.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Int. Conf. on Medical image computing and computer-assisted intervention, 2015.

[Sve10] Gabriel Svennerberg. Beginning google maps API 3. Apress, 2010.

[vR95] Guido van Rossum. Python reference manual. Department of Computer Science [CS], (R 9525), 1995.

[WKW16] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. Journal of Big data, 3(1):1–40, 2016.