

Universität Stuttgart



Image Rectification to Improve Tie Point Matching in Mobile Mapping Applications

Author: Bikram Koirala

Supervisors: Prof. Dr. -Ing. N. Haala

Dipl.-Ing. Stefan Cavegn

1. Introduction

Image based mobile mapping is one of the options available for capturing geodata. For an application requiring geodata in the cm range, the high accuracy requirement of captured geodata in a given reference frame can be achieved by using high-quality georeferencing of the multi-view image sequences. The captured image sequences can be directly georeferenced by using position and orientation information obtained from integrating GNSS and high-grade inertial navigation equipment. The high accuracy requirements cannot be achieved by direct georeferencing in urban canyons due to poor GNSS coverage. Image based georeferencing (indirect) can be one of the alternatives for the challenging urban environment. It needs suitable numbers of tie points between images from different look angle cameras with different geometries.

2. Contributions

Feature matching plays a vital role in image-based georeferencing. It delivers corresponding points in two images of the same scene. The problem of matching images from mobile mapping platforms is the large viewpoint difference of images captured from different cameras. Most of the feature matching algorithms are able to find good matches only if viewpoint difference between image pairs is low. Tie point matching for large viewpoint difference images can be improved by modifying the SIFT implementation and using iterative Homography and Fundamental matrix based geometric transformation.

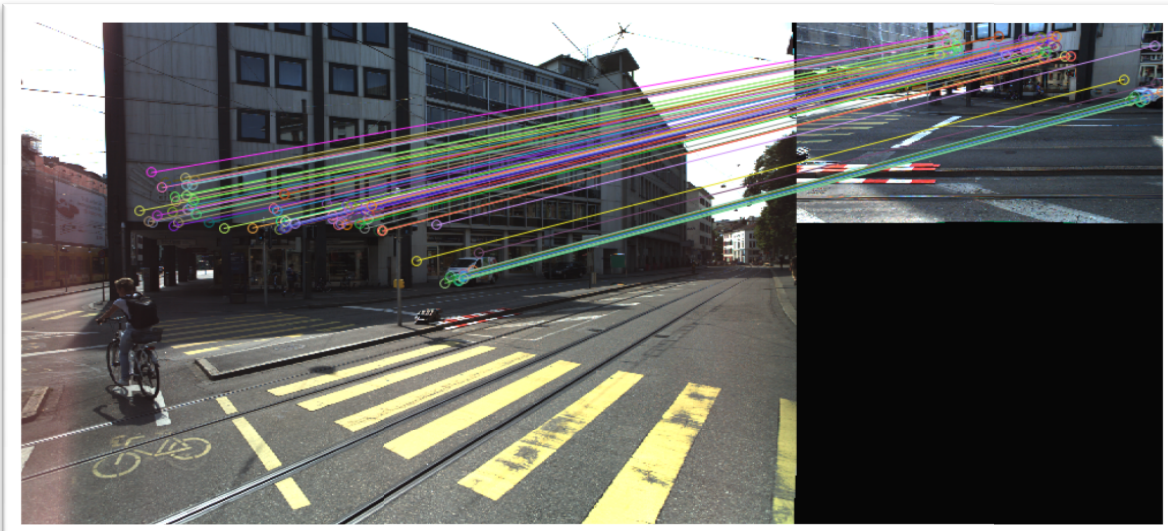


Figure 2.1: The matching result for the image pair captured from forward and left looking cameras. In this figure image from left looking camera looks small due to low resolution compare with a resolution of image from the forward-looking camera.

RANSAC-based geometric transformation is estimated to eliminate outliers for matches provided by the SIFT algorithm from stereo image pairs. If in the image pair one image is a spherical image, then we have to modify epipolar geometry. This modified epipolar geometry is called hybrid epipolar geometry. The geometry can be seen in figure 2.2. The matches provided by the SIFT algorithm should be linearized by lifting homogeneous coordinates of spherical images.

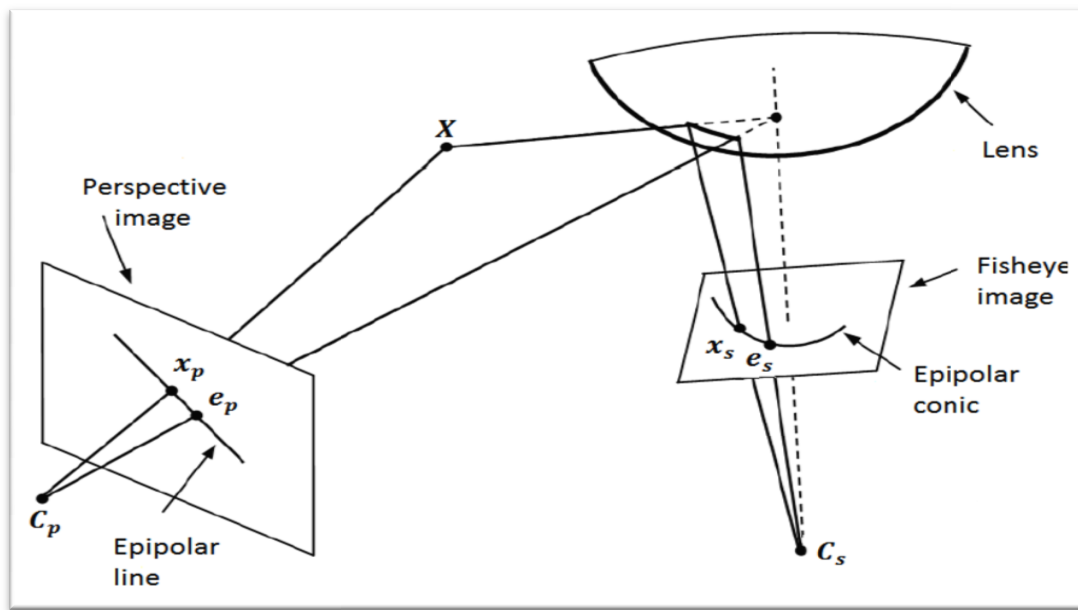


Figure 2.2: Hybrid epipolar geometry between a perspective and a spherical image. x_p and x_s are the projections of a 3D point X on perspective and spherical images respectively. e_p and e_s are the epipoles in the perspective and spherical images respectively. (Source: Bastanlar, Y. et al., 2012)

Lifting can be performed for points from the spherical image with coordinates $x_s = (x, y, 1)^T$ to $\hat{x}_s = (x^2 + y^2, x, y, 1)^T$, and then it satisfies

$$\hat{x}_s^T F_{sp} x_p = 0$$

where F_{sp} is a hybrid fundamental matrix with a size 4×3 .

For bundle block adjustment, matching with only stereo image pairs is not enough. Bundle block adjustment presumes a scene graph which connects tie points and cameras. In this graph, images are considered as nodes and verified pairs of images as edges (Schönberger, J. L. *et al.*, 2016). Verification of all image pairs by using the matching algorithm is very time-consuming. An algorithm is required to reduce this processing time. This is solved via the implementation of point tracking.

The basic idea of point tracking for images captured from still cameras is to find matching results without the use of the matching algorithm. If there are two sets of matching points from the image matching, we can track points if and only if both sets contain matching points from one same image.

Let, $A = \{a, b\}$ and $B = \{c, d\}$ where a and b are matching points obtained from matching image 1 and image 2, similarly, c and d are matching points obtained from matching image 2 and image 3. We can track points from image 1 to image 3 using the same features from image 2 which matches with image 1 and image 3 as shown in figure 2.3. By tracking we have a new set of matching pairs $C = \{a', d'\}$ from image 1 and image 3. Here a' and d' are a subset of points a and d respectively.

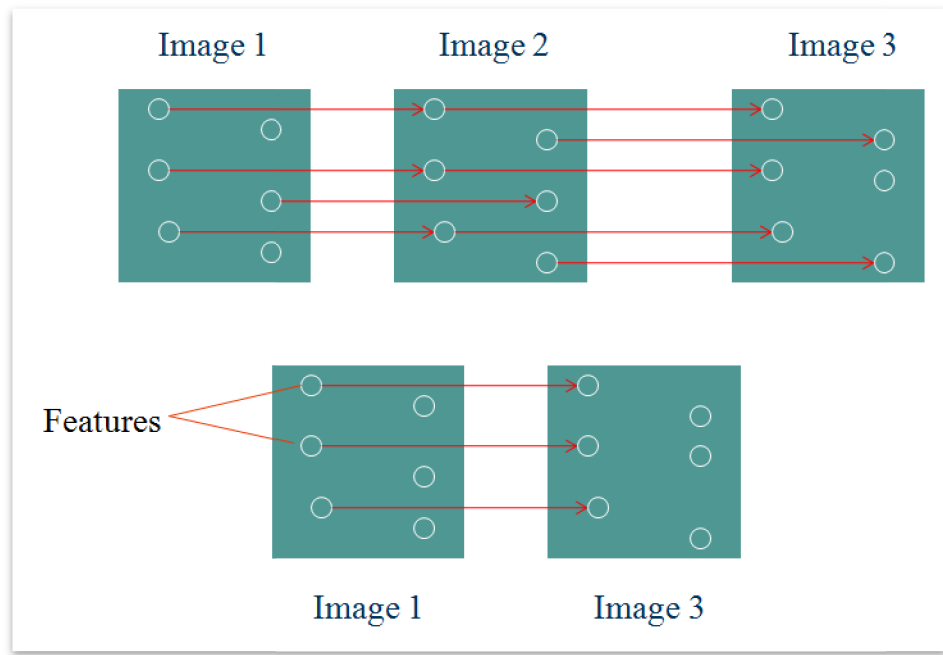


Figure 2.3: Upper part: It describes matching results for image 1 and image 2 and image 2 and image 3. Bottom part: It shows tracking results by using the same feature from image 2 which match with features of two different images.

3. Accuracy analysis

We are able to improve tie point matching for large viewpoint difference images and now want to evaluate our improved tie points by bundle block adjustment using the open source software COLMAP (Schönberger, J. L. *et al.*, 2016). COLMAP provides 3D sparse cloud points and their re-projection error. The re-projection error provided by COLMAP defines internal quality. For external quality check and reliability we have to perform bundle adjustment using ground control points (GCPs) established in mapping areas. Some GCPs can be used as checkpoints for independent accuracy checks. These checkpoints do not contribute to adjustment but obtain residuals. The residual of checkpoints reflect the true errors, which can be used to estimate the absolute accuracy of the adjustment.

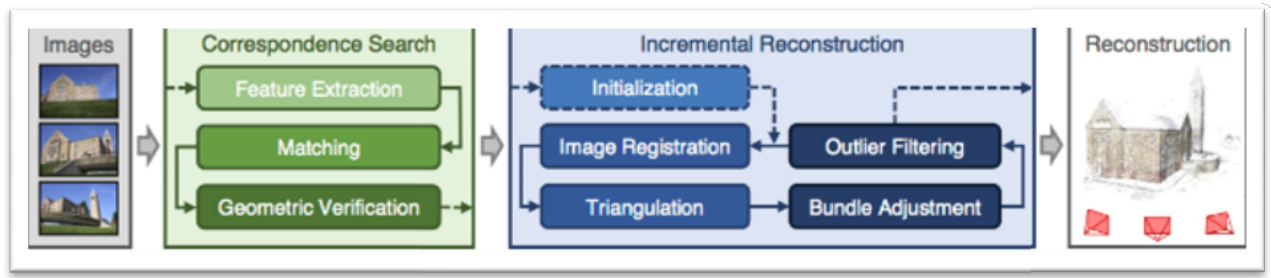


Figure 3.1: COLMAP's incremental Structure-from-Motion pipeline
(Source: Schönberger, J. L. *et al.*, 2016)

COLMAP has two important steps; the first one is correspondence search and the other incremental reconstruction. The output of correspondence search is a scene graph with images as nodes and verified pairs of images as edges. The output of correspondence search is used in incremental reconstruction to have an output as a pose estimate of registered images and reconstructed scene structure as a set of points.



Bundle adjustment report

Residuals:	270266
Parameters:	78821
Iterations:	12
Initial cost:	0.28 [px]
Final cost:	0.28 [px]
Termination:	Convergence

Figure 3.2: Reconstruction results using images from the forward-looking perspective camera.

Figure 3.2 shows the reconstruction result with “Final Cost” of 0.28 pixels. From the COLMAP provider, the “Final Cost” is the re-projection error. “Residuals” in the bundle adjustment report denote image observations (2 counts for each image observation) while “Parameters” denote estimated intrinsic camera parameters, a number of 6DOF (degree of freedom) camera poses, and a number of 3D points (counts 3 for each 3D points).



Bundle adjustment report

Residuals:	546736
Parameters:	144266
Iterations:	4
Initial cost:	0.383 [px]
Final cost:	0.382 [px]
Termination:	Convergence

Figure 3.3: Reconstruction results using images from perspective cameras (perspective image) and Ladybug cameras (spherical image)

From figure 3.3 it can be seen that the re-projection error for this case is 0.38 pixels.

Nebiker, S. *et al.*, 2015 showed that cm level accuracy of 3D cloud points can be achieved by mobile mapping if the projection error is in the range of sub-pixel.

4. Conclusion and future work

In this work, a comparison using most of the feature detectors and feature descriptors for large viewpoint difference of images and image pairs containing very low matches were performed. We initially realized that image rectification helps the SIFT algorithm find good matches for large viewpoint difference of images. We later replaced image rectification by modifying the SIFT implementation and using iterative Homography and Fundamental matrix based geometric transformation. Occlusion due to traffic and advertising boards reduces suitable matches in image pairs captured from built-up environments.

Point tracking is possible in images captured from still cameras by using the matching results of the neighbor image pairs. We achieved almost similar results by directly matching image pairs using the matching algorithm or by tracking the matches. The tracking results are better for the region with repetitive patterns.

The Fundamental matrix defined for perspective image pairs cannot remove outliers properly for hybrid image pairs. The matches provided by the SIFT algorithm should be linearized by lifting homogeneous coordinates from spherical images. This set of lifted coordinates now needs a 4×3 Fundamental matrix for removing outliers if the spherical image is assumed to be the second image in image pairs. For the case of spherical image pairs a 3×3 Fundamental matrix is enough, but a minimum of 9 matching points is needed for removing outliers.

The open source software COLMAP is a very powerful tool in order to obtain 3D sparse reconstructions for the provided tie points and features. This software is user-friendly and allows the user to customize the database for its purpose. Verification of our tie points by the bundle block adjustment using COLMAP attained re-projection errors in the range of 0.2-0.4 pixels.

Future work includes investigations on external quality by providing ground control points on the bundle block adjustment. Mobile mapping imagery contains large regions with repetitive patterns; investigation into how to make features unique enough for matching is needed. The analysis contained with this thesis was performed on a limited number of images. This investigation might be insufficient if we have to reconstruct an entire city.

Bibliography

Bastanlar, Y., A. Temizel, Y. Yardimci, and P. Sturm, 2012: Multi-view Structure-from-Motion for Hybrid Camera Scenarios, March 16, 2012.

Nebiker, S., S. Cavegn, and B. Loesch, 2015: Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. In: ISPRS International Journal of Geo-Information, 2015.

Schönberger, J. L. and J. M. Frahm, 2016: University of North Carolina at Chapel Hill. Structure-from-Motion Revisited.