

Universität Stuttgart



Deutsches Zentrum
für Luft- und Raumfahrt

Master Thesis

Automatic 3D lane marking reconstruction using multi-view aerial imagery

Chun-Yu Sheu

Course of Study: GEOENGINE

Examiner: apl. Prof. Dr.-Ing. Norbert Haala

Supervisor: Dr.-Ing Franz Kurz

Completed: 15. March 2018

Abstract

The 3D information of road infrastructures are gaining importance with the development of autonomous driving. The exact absolute position and height of lane markings, for example, support lane-accurate localization. Several approaches have been proposed for the 3D reconstruction of line features from multi-view airborne optical imagery. However, standard appearance-based matching approaches for 3D reconstruction are hardly applicable on lane markings due to the similar color profile of all lane markings and the lack of textures in their neighboring areas. This thesis presents a workflow for 3D lane markings reconstruction without explicit feature matching process using multi-view aerial imagery. The aim is to optimize the best 3D line location by minimizing the distance from its back projection to the detected 2D line in all the covering images. Firstly, the lane markings are automatically extracted from aerial images using standard line detection algorithms. By projecting these extracted lines onto the Semi-Global Matching (SGM) generated Digital Surface Model (DSM), the approximate 3D line segments are generated. Starting from these approximations, the 3D lines are iteratively refined based on the detected 2D lines in the original images and the viewing geometry. The proposed approach relies on precise detection of 2D lines in image space, a pre-knowledge of the approximate 3D line segments, and it heavily relies on image orientations. Nevertheless, it avoids the problem of non-textured neighborhood and is not limited to lines of finite length. The theoretical precision of 3D reconstruction with the proposed framework is evaluated.

keywords: 3D line features reconstruction, line fitting, least-squares adjustment

Contents

1	Introduction	11
1.1	Motivation	11
1.2	Autonomous Driving	11
1.3	3D Reconstruction	12
1.4	SGM Generated DSM	13
1.5	Automatic Line Detection and Line Matching	14
1.6	Related Work	14
1.7	Purpose	15
2	Methodologies	17
2.1	Lane markings Properties and Automatic Extraction	19
2.2	Imaging Properties of Aerial Photographs	21
2.3	Line Fitting	23
2.4	3D Line Reconstruction with Nonlinear LS Adjustment	25
2.5	Line Projection on the DSM (Determination of Initial Parameter Estimates) .	34
3	Experimental Results and Evaluations	37
3.1	Materials	38
3.2	Parameter Selection	46
3.3	Simulation	46
3.4	True Data	53
4	Conclusion and Future Work	59
	Bibliography	61

List of Figures

2.1	The work flow	18
2.2	Line types of lane markings [Kep93]	19
2.3	Lane markings Extraction. The extracted long lane-lines are marked in green and the dashed ones are in yellow. Note that both cases are reconstructed into 3D with the same framework; different colors here are only for illustration.	20
2.4	Before reconstruction, the back projection of initial approximate 3D line segment does not best-fit the extracted 2D lines in the covering images. After optimization, the back-projection of the reconstructed line segment should be best fitting the extracted 2D lines in all the covering images.	26
2.5	3D reconstruction of a lane marking by a sliding window.	27
2.6	Measurements collection in image space.	28
2.7	The red circle points out the reconsidered measurements in successive sliding windows.	28
2.8	The projected line on DSM.	35
2.9	Iterative scheme of single projection ray intersecting DSM.	35
3.1	Original Image	38
3.2	Flight trajectory of DLR helicopter visualized on the Google Earth platform. The green polyline shows the flight trajectory. <i>Source: Google Earth 04/01/2017</i>	39
3.3	The impact of uncertain camera attitude in horizontal direction.	41
3.4	Part of the DSM in road area. It is noisy in the center of motorway.	43
3.5	Standard deviations of the height value of the DSM in road area. It has higher values in the center of motorway.	43
3.6	Distribution of stereo pairs used for DSM generation. Lighter color indicates more stereo pairs are used in that area. Maximum 27 stereo pairs are used for one pixel.	44
3.7	Orthorectified Image	45
3.8	Masked Image	45
3.9	The added Gaussian random noise in the observations.	47
3.10	The reconstructed line segments and the true line segments in UTM 32N coordinate system.	49
3.11	The reconstructed line segments and the unrefined DSM profile in UTM 32N coordinate system.	49
3.12	The amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment in simulation case.	50
3.13	The relation between the added random Gaussian noise and the adjusted residuals.	51

3.14 The red area under the probability density function is the rejection region, which is 5%. Since none of the T_{obs} falls in the rejection region, the null hypothesis could not be rejected.	51
3.15 The amount of images, the posterior standard deviation of the estimated parameters and	52
3.16 The reconstructed line segments and the unrefined DSM profile in UTM 32N coordinate system.	54
3.17 Histogram of the distances from the reconstructed line to the unrefined DSM profile.	54
3.18 The amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment in the true data experiment.	55
3.19 The variances of the estimated object coordinates, in horizontal and vertical directions.	56
3.20 The estimated variances of the estimated object coordinates, in horizontal and vertical directions.	56
3.21 All the reconstructed lane-lines in the test area.	57
3.22 All the nodes of reconstructed lines on georeferenced aerial images, in UTM 32N coordinate system.	58

List of Tables

1.1	Summary of level of driving automation [SAE14]. System refers to the driver assistance system, combination of driver assistance systems, or automated driving system.	12
2.1	Widths of lane markings [Kep93]	19
2.2	Lengths of dashed lane markings with ratio 1:2 [Kep93]	19
3.1	Properties of the oblique camera	39
3.2	Viewing geometry	39
3.3	Precisions of Exterior Orientations	40
3.4	Interior Orientations and their precisions	40

1 Introduction

1.1 Motivation

The availability of large-scale, accurate high resolution 3D information of roads with lane markings and road infrastructure plays an important role towards autonomous driving. Aerial imagery is a valuable database to derive 3D information of roads even in areas difficult to access, like on motorways. Compared to optical satellite data, acquiring large-scale 3D lane markings by optical aerial imagery is more efficient and has higher accuracy and spatial resolution. In view of the fact, that in Germany exists no area-wide, high resolution 3D information of the road surfaces including lane markings, new methods to derive this information are demanded.

The standard workflow with aerial images would be to project the images onto a DSM and to derive the information in the projected imagery, but the generation of DSM from stereo images is challenging in the regions with low textures. The lane markings, for example, are the most visible texture on asphalt roads useful for 3D reconstruction. Thus, it is desired to improve the quality of the DSM on the road surfaces by exploiting the line character of the lane markings.

1.2 Autonomous Driving

Autonomous driving is an important part of future transportation systems, and the execution of dynamic driving task is a key issue towards autonomous driving. According to the definition given by Society of Automotive Engineers (SAE) International, dynamic driving task includes the operational (e.g. steering, braking, accelerating, monitoring the vehicle and roadway) and tactical (responding to events, determining when to change lanes, turn, use signals, etc.) aspects. In the information report J3016 [SAE14] proposed by SAE International, driving automation is identified into 6 levels from “no automation” to “full automation” as expressed in Table 1.1.

For autonomous driving, redundant sources of information for the driving environment are useful for increasing the robustness and availability of the system [ARB+15]. It is especially important for level 3, 4 and 5 driving automations where the execution of dynamic driving tasks is totally conducted by the system. 3D lane markings, for example, provide global environment information and can be used on supporting lane-accurate localization.

1 Introduction

SAE level	Name	Execution of Steering and braking	Monitoring of driving environment	Fallback Performance of Dynamic Driving Task	System Capability
Human driver monitors the driving environment					
0	No Automation	Human	Human	Human	n.a.
1	Driver Assistance	Human & System	Human	Human	some driving modes
2	Partial Automation	System	Human	Human	some driving modes
Automated driving system monitors the driving environment					
3	Conditional Automation	System	System	Human	some driving modes
4	High Automation	System	System	System	some driving modes
5	Full Automation	System	System	System	all driving modes

Table 1.1: Summary of level of driving automation [SAE14]. System refers to the driver assistance system, combination of driver assistance systems, or automated driving system.

1.3 3D Reconstruction

Generally, the procedure of 3D objects reconstruction consists of feature extraction in image space and depth information recovery in object space. To reconstruct the depth information at the exposure moment, either multiple projection rays spatial intersection or single projection ray intersection with a surface model can be applied. Spatial intersection, or called *triangulation*, is usually applied in the cases that the correspondences of the extracted features among different views can be established. Alternatively, when Digital Elevation Model (DEM) is available, single extracted feature can be directly projected onto the DEM. In this case, the quality of DEM directly influences the result of 3D object reconstruction.

1.3.1 Feature Extraction and Matching

The aim of feature extraction is to gain the characteristics of the images, through which the stereo correspondence processes. As a result, the characteristics of the images closely link to the choice of matching methods. [Est]

Blob features have properties of being local intensity maximum or minimum in images. Edge features have image brightness discontinuities in the direction perpendicular to the line direction itself. Variance of algorithms have been proposed for different kinds of features detection. The rotation-invariant Harris corner detector, for example, is commonly used to extract corners and infer features of an image.

The features are then matched among different views by comparison of the patches which center on the extracted features. Typically the similarity is measured by taking the Sum of Squared Differences (SSD) or Normalized Cross-Correlation (NCC) between the corresponding pixels of two patches.

Scale-Invariant Feature Transform (SIFT) is proposed by Lowe in 1999 [Low99]. Lowe's approach transforms an image into a large collection of local feature vectors, also known as keypoint descriptors. Each SIFT feature descriptor is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and affine or 3D projection. These SIFT keypoints are then matched by identifying their nearest neighbors.

1.3.2 Dense Image Matching (DIM)

Dense image matching performs matching at the actual image resolution, i.e. pixel-wise correspondence between MVS is to be recovered. Depending on image texture, a per-pixel measure is generally ambiguous. Additional constraints, such as the assumption of a smooth surface, need to be introduced. Proposed by Hirschmüller et al. in 2008, SGM approximates the two-dimensional global aggregation of matching cost by a number of one-dimensional cost path, where the matching step is casted into an energy minimization problem [HB08]. It not only achieves similar accuracy as truly global matching but also significantly reduces computational complexity.

1.4 SGM Generated DSM

High resolution DSM can be generated by laser scanning or dense image matching. Compared to laser scanning, applying dense image matching to produce DSM is of lower cost on data acquisition.

For each pair of stereo images, a disparity map is generated by SGM algorithm. With disparity's property of being inversely proportional to depth, the disparity maps can be used to derive the DSM. In most of the cases, a point in object space is covered by more than two aerial images, resulting in more than one disparity maps. This leads to ambiguities on height value decision during DSM generations. The height value of each DSM pixel is determined by simply taking the median value derived from disparity maps in odd number of disparity maps cases, and the value just below the median in even number cases. Such disparity maps fusion approach may result in systematic errors of having lower height value in some parts of DSM.

1 Introduction

Since SGM is a kind of appearance-based matching algorithm, it is unstable in lowly textured areas. Consequently, the SGM-generated DSM is especially noisy in such poorly textured areas.

1.5 Automatic Line Detection and Line Matching

Automatic detected lines may be used for 3D reconstruction by matching lines in the image space. However, line matching is challenging for several reasons.

Firstly, line segments may be detected inexactly by automatic line detectors or obstructions appear in part of the lines. Consequently the end-points of a line segment often do not correspond to each other in different views.

Secondly, there is no strong disambiguating geometric constraint available [SZ97]. In the case of points, correspondences must satisfy the epipolar constraint. This strong disambiguating constraint helps to efficiently reduce the searching space from the whole image (2D) to a single line (1D) in matching processes. In the case of infinite line matching, however, there is no geometric constraint. For lines of finite length, there is only a weak overlap constraint arising from applying the epipolar constraint to its end-points.

Moreover, the corresponding neighborhoods may well have a very different shape and orientation in different views, or even totally different surroundings when dealing with wiry objects [HWB13].

In the cases of lane markings, they may be partly shaded by vehicles in aerial images. Besides, the continuous lane lines have no endpoints in the images. Worst of all, the asphalt road surface where the lane markings locate on is poorly textured. Therefore, line matching is even hardly applicable on lane markings.

1.6 Related Work

In the following, some works regarding 3D line reconstruction is presented.

First, appearance-based methods are described. For 3D line reconstruction, [BFG05; SZ97; WWH09] have tried to match line segments based on their appearances or some additional geometry constraints.

Schmid et al. exploit the epipolar geometry of line segments and the one-parameter family of homographies to provide point-wise correspondences, allowing cross-correlation of patches around line segments along the candidate lines in the epipolar-beam-region for matching scores evaluation [SZ97].

In the cases of poorly textured or shape-changing neighborhood of line segments in different views, line segments are barely comparable using classical correlation patches yet the color neighborhood along this line segment undergoes only slight changes. Based

on color histogram rather than textures, Bay et al. exploit the appearance similarity of line segment pairs and their topological layout to iteratively increase the correct matches [BFG05]. If region matches are available, they are automatically integrated and exploited in combination. The final coplanar grouping stage allows to estimate the fundamental matrix even from line segments only. While color provides a very strong cue for discrimination, it may fail in the case where color feature is not distinctive, e.g. gray images. Besides, although matching groups of line segments takes more geometric information into account for disambiguation, the disadvantage is the increased computational complexity.

Without resorting to any other constraints or prior knowledge, Wang et al. propose a purely image content-based line descriptor MSLD (Mean-Standard deviation Line Descriptor) for automatic line segments matching. Adapting SIFT-like strategy, MSLD is highly distinctive and robust against image rotation, illumination change, image blur, viewpoint change noise, JPEG compression and partial obstruction [WWH09]

The above appearance-based approaches demand either constant and rich neighboring textures or similar color profile of line segments, they are technically matching the surroundings instead of the lines themselves.

In order to create 3D models without the need of explicit line matching, Jain et al. generate all possible hypothetical straight 3D line segments by triangulating all the detected straight 2D line segments from different views [JKTS10]. They then keep the one whose back projection on the gradient images of neighboring views has the highest score, assuming that line features correspond to high gradient areas in images. Built upon the same principles whilst applying epipolar constraint on the end-points of line segments, Hofer et al. generate less hypothetical 3D line segments and thus increase performance significantly while still creating accurate results [HWB13]. However, both approaches are barely possible in the case of infinite line reconstruction, where the detected 2D lines in different views do not exactly correspond to the same part of a 3D line.

Taylor et al. formulate the Structure from Motion (SfM) problem in terms of minimization of an objective function which measures the total squared distance in the image plane between the observed edge segments and the projections of the reconstructed lines [TK95]. By reconstructing the infinite straight line that supports the observed edge segments rather than the end-points of the line, the algorithm can be used even when multiple edges in a single image correspond to different portions of the same 3D line.

1.7 Purpose

In this thesis, I develop a framework to automatically detect the lane markings in the unprojected aerial imagery, and refine the 3D information of the road surface by exploiting the line character of the lane markings.

The unprojected aerial images with their bundle-adjusted orientations and the DSM are the inputs of my algorithm. The aerial image dataset are acquired with special flight

1 Introduction

configuration at both sides of the motorway. I apply some standard pre-processing steps and a standard line detection algorithm for automatic lane marking detection in image space. By sliding a window of reasonable length and width through the curved long lane lines, I collect all line segments in all covering images assuming the lane markings to be straight in each sliding window.

Similar to Taylor's idea on minimization of the objective function [TK95], I investigate the use of linear regression to optimize the 3D position of each line segments in object space so that its back projection would best fit the detected 2D line in all the covering views, i.e. the position and height of each 3D lane marking segments will be refined in one optimization step. The proposed approach addresses the challenging (quasi) infinite and curved properties of lane markings in the 3D reconstruction.

The framework will be tested on aerial imagery from the German highway A9.

2 Methodologies

The following sections introduce the principles of 3D lane marking reconstruction method of this work, based on the work flow shown in Figure 2.1.

Section 2.1 describes the applied standard line detection algorithm for labeling the lane markings.

To relate the object coordinates of a point with its image coordinates, Section 2.2 introduces the imaging properties of aerial images and their mathematical models, including the collinearity equation and lens distortion correction.

Section 2.3 introduce the principle of line fitting and further presents the orthogonal regression model with line equations in two-point form.

A non-linear LS model is derived and estimated in Section 2.4 to elaborate the usage of line fitting with the combination of collinearity condition for 3D lane marking reconstruction.

In Section 2.5 the generation of approximate 3D line segments is described, as initial values of unknown quantities are required in nonlinear LS estimation.

2 Methodologies

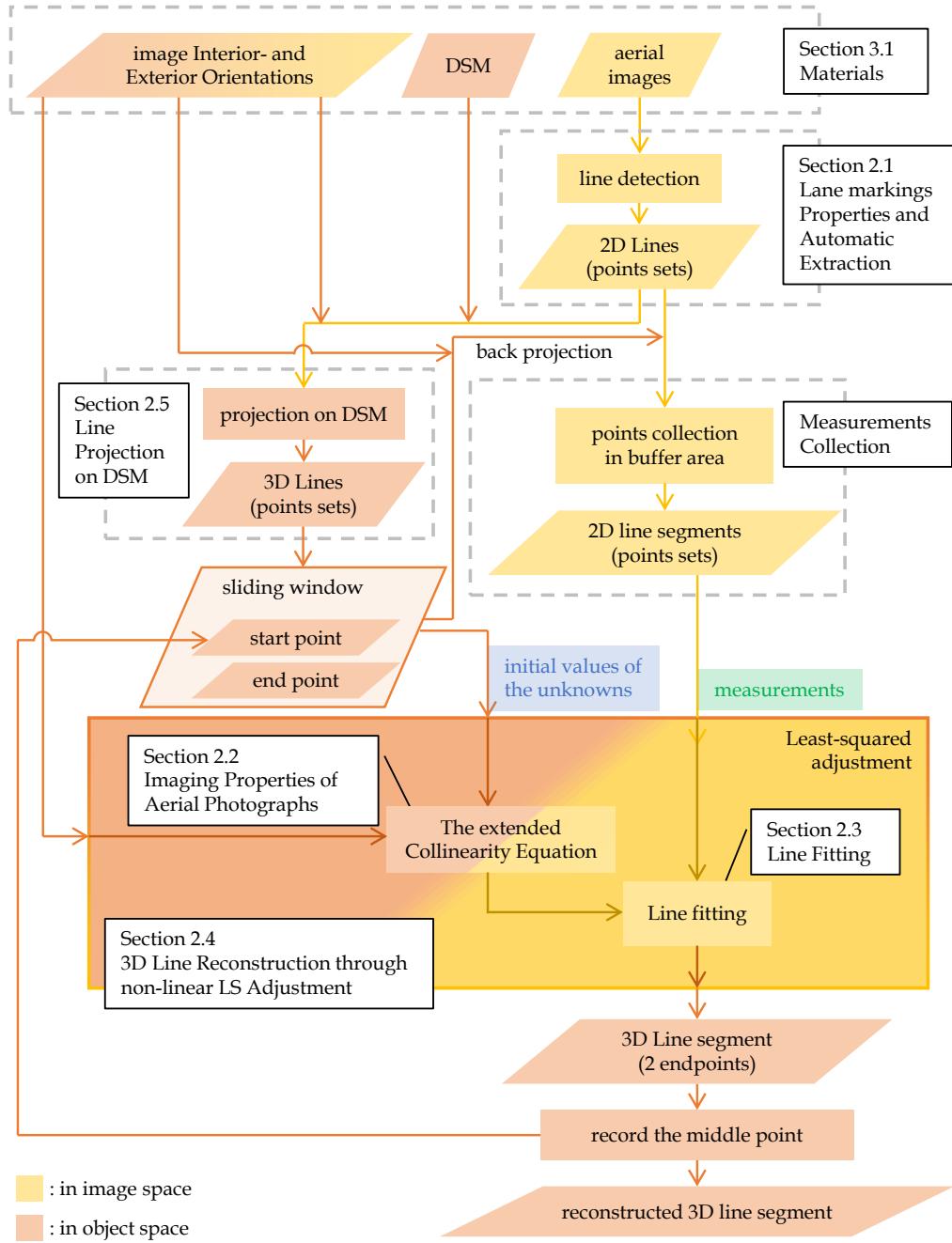


Figure 2.1: The work flow



Figure 2.2: Line types of lane markings [Kep93]

2.1 Lane markings Properties and Automatic Extraction

The appearance of lane markings on German roads including line type, color and width is specified depending on the road type. Different line types of lane markings are shown in Figure 2.2 and their line widths are defined in Table 2.1. As shown in Table 2.2, the dashed lane markings on motorways have 6 meter length.

Because of the appearance, the problem of lane marking detection can be treated as a line detection problem. We restrict the proposed framework to lane markings with single white lines (dashed or continuous) of 0.3 meter width. Other types like in restricted zone, double lines, parking areas, temporal yellow lines in construction sites etc, are excluded.

	motorways ¹	other roads
narrow lines	0.15 [m]	0.12 [m]
wide lines	0.30 [m]	0.25 [m]

Table 2.1: Widths of lane markings [Kep93]

	motorways ¹		other roads	
			in town	out of town
line / gap	6 [m] / 12 [m]	3 [m] / 6 [m]	4 [m] / 8 [m]	

Table 2.2: Lengths of dashed lane markings with ratio 1:2 [Kep93]

¹and corresponding roads in the sense of the VwV-StVO to § 42 to mark 330 (motorway) II

2 Methodologies

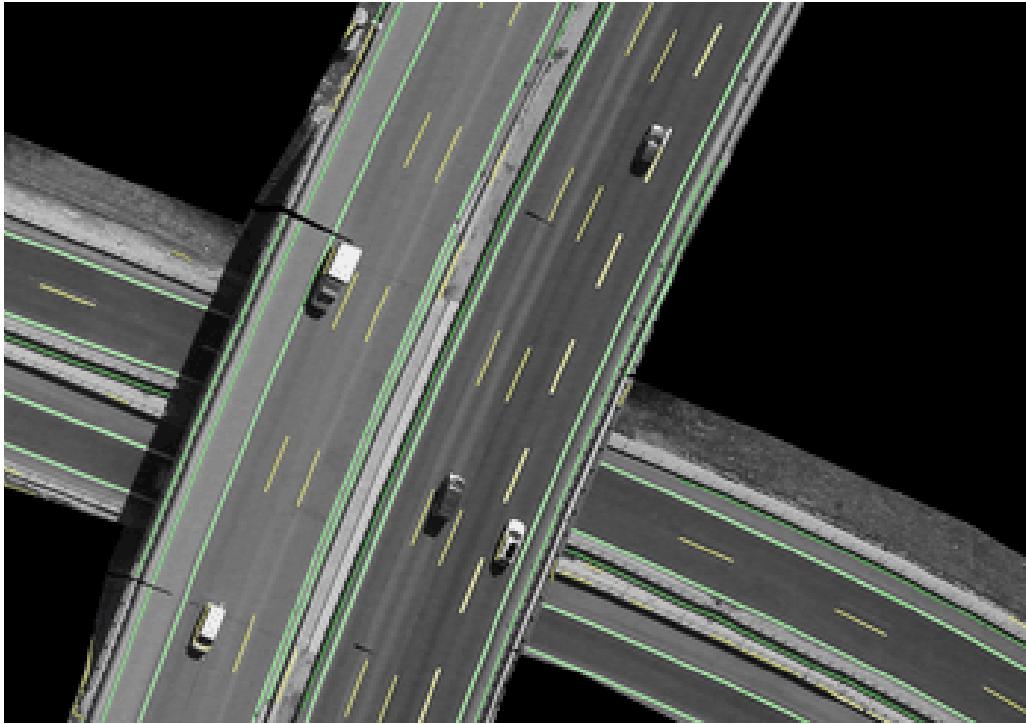


Figure 2.3: Lane markings Extraction. The extracted long lane-lines are marked in green and the dashed ones are in yellow. Note that both cases are reconstructed into 3D with the same framework; different colors here are only for illustration.

There are many algorithms for line detection. Prewitt line detector uses two orthogonal gradient operators, and the pixels in the operators are of same weights. Sobel detector also uses two orthogonal gradients operators, but the weights of pixel in operators are not equal —the closer the pixel to the center of operator, the higher weight it has. Canny edge detector searches local extrema of gradient to locate the positions of line features and is still a state-of-the-art edge detector. Edge-detectors that perform better than the Canny usually require higher computational complexities or a greater number of parameters. Edge drawing [TA12] first spots anchors along rows and columns by Sobel detector, and then joins these anchors to extract line features.

In this work, the principle to extract line features is to firstly derive the line direction for each pixel by using partial derivatives of a Gaussian smoothing kernel. Pixels that have a local maximum in the second directional derivative perpendicular to the line direction are marked as line points. By thresholding their second directional derivative values, the accepted line points are then linked and connected.[Ste98]

The resulting connected points which compose a line are of sub-pixel precision. Figure 2.3 shows the extracted lines on part of the masked original image.

2.2 Imaging Properties of Aerial Photographs

This section describes the geometric model of the projection of 3D points into the image generated by a real camera. We first restrict the discussion in Section 2.2.1 to central perspective projection where the collinearity equation originates from. We then model deviations from this model, addressing real cameras with imperfect lenses, in Section 2.2.2.

2.2.1 Collinearity Equations

We assume frame photography, i.e. photographs exposed on a frame chip in one instant, and assume the central projection model with cameras that have a single viewpoint and a planar sensor and being straight line-preserving. Collinearity indicates the condition that the image point (on the sensor plate of the camera), the observed point (in object space) and the projection center of the camera were aligned at the moment the picture was taken. Every measured point leads to two collinearity equations, describing transformations from object space to image coordinates:

$$\begin{aligned} x &= x_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ y &= y_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \end{aligned} \quad (2.1)$$

where

(x, y) : image coordinates of the point

(x_0, y_0) : image coordinates of principal point

c : principal distance; focal length

(X, Y, Z) : object coordinates of the point

(X_0, Y_0, Z_0) : object coordinates of projection center

r_{11}, \dots, r_{33} : elements of the rotation matrix R (orthogonal 3×3 -matrix from object space to image space, with 3 independent angles ω, ϕ and κ)

2.2.2 Lens Distortion Correction

An original image has some degree of deviations from perspective mapping due to lens distortion, lens refraction or non-planarity of the sensor surface. There are several models to describe these perturbing effects and can be used to undistort the images, resulting in rectified images which are now straight line-preserving.

A subset of physical distortion model [Fra97] is chosen, with two radial symmetric distortion parameters A_1 and A_2 , two asymmetric parameters B_1 and B_2 , and a scaling C_1 and an affine shearing parameter C_2 . Assuming x and y to be the distorted image coordinates, the corrections Δx and Δy are then calculated by the following equations:

$$\begin{aligned} \Delta x &= x_p + A_1 x_* (r^2 - R_0^2) + A_2 x_* (r^4 - R_0^4) + B_1 (r^2 + 2x_*^2) + B_2 2x_* y + C_2 y \\ \Delta y &= y_p + A_1 y (r^2 - R_0^2) + A_2 y (r^4 - R_0^4) + B_1 (r^2 + 2y^2) + B_2 2x_* y \end{aligned} \quad (2.2)$$

2 Methodologies

with $r = \sqrt{x_*^2 + y^2}$, $x_* = \frac{x}{C_1}$ and radius² R_0 being set to a third of the sensor diagonal.

The undistorted image coordinates x' and y' are then calculated by

$$\begin{aligned} x' &= x + \Delta x \\ y' &= y + \Delta y \end{aligned} \quad (2.3)$$

2.2.3 Extended Collinearity Equation

As real cameras only approximate the perspective camera model, lens distortion correction can be additionally included in the collinearity model, attempting to correct the pixel positions so that they obey the perspective model with sufficient accuracy.

By inserting (2.1) and (2.2) into (2.3), the relationship between a 3D point $\mathbf{P}(X, Y, Z)$ and its corresponding distorted image coordinates $\mathbf{p}(x, y)$ can be described as

$$\begin{aligned} x &= x_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ &\quad - (x_p + A_1 x_*(r^2 - R_0^2) + A_2 x_*(r^4 - R_0^4) + B_1(r^2 + 2x_*^2) + B_2 2x_* y + C_2 y) \\ y &= y_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ &\quad - (y_p + A_1 y(r^2 - R_0^2) + A_2 y(r^4 - R_0^4) + B_1(r^2 + 2y^2) + B_2 2x_* y) \end{aligned} \quad (2.4)$$

To express (2.4) shortly, a function \mathcal{G} is defined as

$$\mathbf{p} = \mathcal{G}(\mathbf{q}, \mathbf{P}) \quad (2.5)$$

which takes the interior and exterior orientations as well as the lens distortion parameters of a camera $\mathbf{q}(x_0, y_0, c, X_0, Y_0, Z_0, r_{11}, \dots, r_{33}, A_1, A_2, B_1, B_2, C_1, C_2)$ and the position of a 3D point $\mathbf{P}(X, Y, Z)$, and returns the corresponding distorted image coordinates $\mathbf{p}(x, y)$.

²At the radius R_0 the radial symmetric distortion is zero by definition, which avoids too high distortion values at the edges and reduces the correlation with the focal length.

2.3 Line Fitting

Line fitting is the process of constructing an infinite straight line that has a best fit to a 2D dataset. One of the approaches is linear regression which attempts to find the linear function that "best" predicts the dependent variable values as a function of the independent variable. In this work, "best" predict will be understood as in the Least-Squares (LS) approach: minimization of the sum of squared residuals (differences between the measured and the estimated values of the dependent variable).

In the case of standard linear regression, the regressor x is assumed error free, inconsistencies³ are only for the dependent variable y . Geometrically it means that the vertical distances from observed data to the fitted line is minimized. To minimize the perpendicular distances from the data points to the regression line, a orthogonal regression model is derived in Section 2.3.1.

For a later combination with point-wise extended collinearity equation (2.5) in next section, the aim is to fit the line equation in two-point form to the observed dataset. For such nonlinear functional relation between variables, a nonlinear LS model is derived in Section 2.3.2.

A functional model is unsolvable when the assumed "dependent" variable is indeed not a function of the independent variables, i.e. the assumed functional relation does not really exist. Take an observed set of 2D points with their Cartesian coordinates $\{x_i, y_i\}_{i=1}^n$ on a vertical line $x = \text{constant}$ for example. Their y values have no dependency on their x values, i.e. knowledge of x tells nothing about y . Therefore, for this dataset, the functional model $y = f(x)$ is singular. In such cases, however, x is a function of y (which is actually a constant function) and the equation system which models the dependent variable x being a function of the independent variable y becomes solvable. Regarding the dataset used in this work (which will be described with more details in Section 3.1) where the observed 2D points scatter mainly in column direction in image space, the functional relation between variables x and y will be setup as $x = f(y)$ to avoid weakly solvable equations system.

2.3.1 Orthogonal Regression

A linear regression model describes a dependent variable as a linear function of the regressor (an independent variable). Given a dataset $\{x_i, y_i\}_{i=1}^n$ of n points on a 2D plane, in the case when both dependent variable x_i and regressor y_i are measured with errors, a linear regression model takes the form:

$$x_i - e_{x_i} = a_0 + a_1 \bar{y}_i \quad (2.6)$$

where the regression coefficients a_0 and a_1 are the unknown parameters to be estimated, \bar{y}_i denotes the true but unobserved regressor, and the error variable e_{x_i} is an unobserved

³The word "inconsistencies" indicates the unobserved random errors, also called as measurement errors.

2 Methodologies

random variable that adds noise to the linear relationship between the dependent variable x and true regressor \bar{y}_i . Whereas the true regressor \bar{y}_i is observed with an error e_{y_i} in the pseudo observation equation:

$$y_i - e_{y_i} = \bar{y}_i \quad (2.7)$$

Such models, as the combination of (2.6) and (2.7), that take into account the measurement errors of both dependent variables and regressors, are errors-in-variables models. Furthermore, for the case of equal error variances, i.e. when $\delta = \frac{\sigma_{e_x}}{\sigma_{e_y}} = 1$, it is a orthogonal regression model which minimizes the perpendicular distances from the data points to the regression line.

2.3.2 Orthogonal Regression in Two-point Form

The two-point form of a infinite line in the Cartesian plane passing through the points (x_1, y_1) and (x_2, y_2) is given by:

$$(x - x_1) = \frac{(x_2 - x_1)}{(y_2 - y_1)} \times y - y_1 \quad (2.8)$$

with $y_2 \neq y_1$, where (x, y) is any point on the line.

Let the unknown coordinates of two different points on a line in 2D space be (x_1, y_1) and (x_2, y_2) and the observed 2D points be $\{x_i, y_i\}_{i=1}^n$ with measurement errors e_{x_i} and e_{y_i} in both variables. The orthogonal regression model in two-point form is:

$$x_i - e_{x_i} = (x_1 - \frac{(x_2 - x_1)}{(y_2 - y_1)} \times y_1) + \frac{(x_2 - x_1)}{(y_2 - y_1)} \times \bar{y}_i \quad (2.9)$$

$$y_i - e_{y_i} = \bar{y}_i \quad (2.7 \text{ revisited})$$

To express (2.9) and (2.7) shortly, a function \mathcal{F} is defined as

$$\hat{\mathbf{p}} = \mathcal{F}(\mathbf{p}_s, \mathbf{p}_e, y) \quad (2.10)$$

which takes 2D coordinates of a start-point $\mathbf{p}_s(x_s, y_s)$ and an end-point $\mathbf{p}_e(x_e, y_e)$ that define an infinite line, and takes the measured y-coordinate y of an image point $\mathbf{p}(x, y)$, and returns the estimated image coordinates $\hat{\mathbf{p}}(\hat{x}, \hat{y})$ which lies on the infinite line $\overline{\mathbf{p}_s \mathbf{p}_e}$.

Note that as a combination of (2.9) and (2.7), function \mathcal{F} is actually composed of

$$\begin{aligned} \hat{x} &= \mathcal{F}^x(\mathbf{p}_s, \mathbf{p}_e, y) \\ \hat{y} &= \mathcal{F}^y(\mathbf{p}_s, \mathbf{p}_e, y) \end{aligned} \quad (2.11)$$

2.4 3D Line Reconstruction with Nonlinear LS Adjustment

In this section, the process of refining the position of a 3D line segment in the object space so that its back-projection in each image has a best-fit to the extracted line in the image space is described, as illustrated in Figure 2.4.

Section 2.4.1 introduces the non-linear LS adjustment model with constraints. The observation equations for LS adjustment are set up in Section 2.4.2. They describe the fitting of a straight line to the measurements, which are the extracted lines, in all covering images, where the fitting lines on different images are transformed from a single 3D straight line segment through the extended collinearity equation (2.4). Regarding the fact that the collinearity is a point-wise condition, a line segment is represented by its two endpoints. Correspondingly, the observation equations in the LS model are line equations in two-point form.

By satisfying the collinearity condition, single fitting line on an image spans a plane in \mathbb{R}^3 . The corresponding lines from different views span several planes in \mathbb{R}^3 , which are non-parallel to each other and should intersect in a line in \mathbb{R}^3 . As this infinite line is the solution space of the LS estimation, at least two constraints on the location of the two endpoints of the targeted 3D line segment are necessary to avoid their arbitrary locations on this infinite line in \mathbb{R}^3 . The constraint equations are modeled in Section 2.4.3.

So far the non-linear LS model has been setup. In Section 2.4.4 it is further linearized and the substitute linear LS model is estimated.

To simplify the problem, a long lane-marking segment is partially reconstructed through a sliding window in the object space. Each segment is approximated by a straight line, taking into account the maximum curvature of the highway.

In each sliding window, a segment is reconstructed, i.e. a complete non-linear LS adjustment is performed. Only the middle point of the reconstructed line segment is recorded. The sliding window then moves a stepsize forward, and the process of 3D reconstruction is performed again starting from the recorded middle point of the previous line segment. Another line segment is then reconstructed, with its middle point being recorded, and so on. These recorded middle points are in the end the nodes of the reconstructed line. This process is illustrated in Figure 2.5.

The measurements for each reconstruction process are collected correspondingly, as shown in Figure 2.6 —by back-projecting the initial line segment into image space and buffering 10 pixels width on each side. By this, all the extracted 2D line segments in this region are collected. As shown in Figure 2.7, the reconsideration in overlapping region of successive sliding windows makes the reconstruction more robust.

2 Methodologies



Figure 2.4: Before reconstruction, the back projection of initial approximate 3D line segment does not best-fit the extracted 2D lines in the covering images. After optimization, the back-projection of the reconstructed line segment should be best fitting the extracted 2D lines in all the covering images.



(a) The first line segment of "sliding window length" is reconstructed, with its starting point and its middle point of "step size" from the starting point being recorded.



(b) Starting from the recorded node of last process, another line segment of "sliding window length" will be reconstructed. i.e. the sliding window has moved "step size" forward.



(c) The point of "step size" length from the starting point on the reconstructed segment is recorded.

Figure 2.5: 3D reconstruction of a lane marking by a sliding window.

2 Methodologies



(a) The pink points represent all the extracted lines (in the form of sets of points). The green points are the endpoints of the back projected initial approximate 3D line segment.
(b) The points in the buffering area are collected as the measurements for LS adjustment.

Figure 2.6: Measurements collection in image space.



Figure 2.7: The red circle points out the reconsidered measurements in successive sliding windows.

2.4.1 The Gauss-Markov Model with Constraints

Adjustment theory deals with the optimal combination of redundant measurements/observations, together with the estimation of unknown parameters [Teu00].

Given are the N observations $\mathbf{l} = [l_n]$, $n = 1, 2, \dots, N$, from which the U unknown parameters $\mathbf{x} = [x_u]$, $u = 1, 2, \dots, U$ are to be determined, with generally $U \leq N$.

The Gauss-Markov model with N nonlinear functions $\mathbf{f}(\mathbf{x}) = [f_n(\mathbf{x})]$, $u = 1, 2, \dots, U$ and the H nonlinear constraints $\mathbf{h}(\mathbf{x}) = [h_\eta(\mathbf{x})]$, $\eta = 1, 2, \dots, H$ ($H < U$) between the unknowns can be written as:

$$\mathbf{l} + \hat{\mathbf{v}} = \mathbf{f}(\hat{\mathbf{x}}) \quad \text{or} \quad \hat{\mathbf{l}} = \underset{N \times 1}{\mathbf{f}}(\hat{\mathbf{x}}) \quad (2.12)$$

$$\underset{H \times 1}{\mathbf{h}}(\hat{\mathbf{x}}) = \mathbf{0} \quad (2.13)$$

where the observations \mathbf{l} are explicit functions of the unknowns \mathbf{x} , with the additive residuals \mathbf{v} introduced to the observations \mathbf{l} to achieve consistency.

Assuming that the deviations between the observed values \mathbf{l} and the true values $\hat{\mathbf{l}}$ are of random nature and have normal (or Gaussian) distribution, the uncertain observations \mathbf{l} are modeled with first and second moments:

$$\mathbf{l} \sim \mathcal{N}(\mathbf{f}(\hat{\mathbf{x}}), \Sigma_{ll}) \quad (2.14)$$

where Σ_{ll} is the variance-covariance matrix of observations \mathbf{l} .

The optimal estimate results from minimizing the Mahalanobis distance with given constraints

$$\hat{\mathbf{x}} = \operatorname{argmin}_{x \mid \mathbf{H}^T x = c_h} \mathbf{v}^T(\mathbf{x}) W_{ll} \mathbf{v}(\mathbf{x}) \quad (2.15)$$

Using Lagrangian multipliers λ we aim to minimize

$$\Phi(\mathbf{x}, \lambda) = \frac{1}{2} (\mathbf{l} - \mathbf{f}(\mathbf{x}))^T W_{ll} (\mathbf{l} - \mathbf{f}(\mathbf{x})) + \lambda^T (\mathbf{H}^T \mathbf{x} - c_h) \quad (2.16)$$

with respect to \mathbf{x} and λ .

2.4.2 Observation Equations

Given a start-point $\mathbf{P}_s(X_s, Y_s, Z_s)$ and an end-point $\mathbf{P}_e(X_e, Y_e, Z_e)$ of a line segment L in the object space and the camera parameters \mathbf{q}^j of camera j . Consider the case where there are J images covering this line segment. With the expanded collinearity model (2.4), the start- and end-points of this line segment's back-projection in image j have the image coordinates $\mathbf{p}_s^j(x_s^j, y_s^j)$ and $\mathbf{p}_e^j(x_e^j, y_e^j)$:

$$\begin{aligned} \mathbf{p}_s^j &= \mathcal{G}(\mathbf{q}^j, \mathbf{P}_s) \\ \mathbf{p}_e^j &= \mathcal{G}(\mathbf{q}^j, \mathbf{P}_e) \end{aligned} \quad \forall j = 1, 2, \dots, J \quad (2.17)$$

2 Methodologies

Let l^j be the corresponding line segment of L being extracted (observed) on image j . Given a dataset $\{x_{l,i}^j, y_{l,i}^j\}_{i=1}^{N_l^j}$ of N_l^j points on line segment l^j , their estimated image coordinates $\hat{\mathbf{p}}_{l,i}^j(\hat{x}_{l,i}^j, \hat{y}_{l,i}^j)$ on the infinite line $\overline{\mathbf{p}_s^j, \mathbf{p}_e^j}$ computed from the orthogonal regression model (2.21) are:

$$\hat{\mathbf{p}}_{l,i}^j = \mathcal{F}(\mathbf{p}_s^j, \mathbf{p}_e^j, y_{l,i}^j) \quad \forall i = 1, 2, \dots N_l^j \quad (2.18)$$

Combining (2.17) with (2.18) gives function \mathcal{H} :

$$\begin{aligned} \hat{\mathbf{p}}_{l,i}^j &= \mathcal{F}(\mathcal{G}(\mathbf{q}^j, \mathbf{P}_s), \mathcal{G}(\mathbf{q}^j, \mathbf{P}_e), y_{l,i}^j) \\ &= \mathcal{H}(\mathbf{q}^j, \mathbf{P}_s, \mathbf{P}_e, y_{l,i}^j) \quad \forall i = 1, 2, \dots N_l^j, \quad \forall j = 1, 2, \dots J \end{aligned} \quad (2.19)$$

which takes camera parameters $\mathbf{q}^j(x_0, y_0, c, X_0, Y_0, Z_0, R_{11}, \dots, R_{33}, A_1, A_2, B_1, B_2, C_1, C_2)$, object coordinates of \mathbf{P}_s and \mathbf{P}_e which define a line $\overline{\mathbf{P}_s, \mathbf{P}_e}$, and the observed y-coordinate of the point $\mathbf{p}_{l,i}^j$ in image space, and returns the estimated image coordinates $\hat{\mathbf{p}}_{l,i}^j$ on the back projected line of $\overline{\mathbf{P}_s, \mathbf{P}_e}$.

Corresponding to Equation (2.11), function \mathcal{H} is actually composed of

$$\begin{aligned} \hat{x}_{l,i}^j &= \mathcal{H}^x(\mathbf{q}^j, \mathbf{P}_s, \mathbf{P}_e, y_{l,i}^j) \quad \forall i = 1, 2, \dots N_l^j, \quad \forall j = 1, 2, \dots J \\ \hat{y}_{l,i}^j &= \mathcal{H}^y(\mathbf{q}^j, \mathbf{P}_s, \mathbf{P}_e, y_{l,i}^j) \end{aligned} \quad (2.20)$$

Since the 3D line reconstruction will be done "segment-wise", i.e. for a pair of $P_s(X_s, Y_s, Z_s)$ and $P_e(X_e, Y_e, Z_e)$ of interest, the measurements $(x_{l,i}^j, y_{l,i}^j)$ will be collected correspondingly, the subscription l representing specific line segment will be left out in the followings.

Each image gives $2 \times N^j$ observation equations. These equations are often stacked together and written in vector form as:

$$\begin{bmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_{N^j}^j \\ y_1^j \\ y_2^j \\ \vdots \\ y_{N^j}^j \end{bmatrix} \stackrel{.}{=} \left\{ \begin{array}{l} \mathcal{H}^x(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^j) \\ \mathcal{H}^x(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_2^j) \\ \vdots \\ \mathcal{H}^x(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^j}^j) \\ \mathcal{H}^y(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^j) \\ \mathcal{H}^y(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_2^j) \\ \vdots \\ \mathcal{H}^y(\mathbf{q}^j, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^j}^j) \end{array} \right\}_{N^j} \quad (2.21)$$

where dot equal indicates inconsistencies between the measured values, x_i^j and x_i^j , and the computed values, $\mathcal{H}^x(q^j, P_s, P_e, y_i^j)$ and $\mathcal{H}^y(q^j, P_s, P_e, y_i^j)$.

2.4 3D Line Reconstruction with Nonlinear LS Adjustment

For all covering image $j = 1, 2, \dots, J$, there are $2 \times \sum_{j=1}^J N^j$ observation equations. Being written in the structure of the *Gauss-Markov model*, corresponding to Equation (2.12), they are expressed as:

$$\begin{aligned}
& \left[\begin{array}{c} x_1^1 \\ \vdots \\ x_{N^1}^1 \\ y_1^1 \\ \vdots \\ y_{N^1}^1 \\ \vdots \end{array} \right] + \left[\begin{array}{c} \hat{v}_{x_1^1} \\ \vdots \\ \hat{v}_{x_{N^1}^1} \\ \hat{v}_{y_1^1} \\ \vdots \\ \hat{v}_{y_{N^1}^1} \\ \vdots \end{array} \right] = \left\{ \begin{array}{l} \mathcal{H}^x(\mathbf{q}^1, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^1) \\ \vdots \\ \mathcal{H}^x(\mathbf{q}^1, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^1}^1) \\ \mathcal{H}^y(\mathbf{q}^1, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^1) \\ \vdots \\ \mathcal{H}^y(\mathbf{q}^1, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^1}^1) \\ \vdots \end{array} \right\} 2 \times N^1 \\
& \boldsymbol{l} + \hat{\boldsymbol{v}} = \boldsymbol{f}(\hat{\boldsymbol{x}}) : \\
& \left[\begin{array}{c} x_1^J \\ \vdots \\ x_{N^J}^J \\ y_1^J \\ \vdots \\ y_{N^J}^J \end{array} \right] + \left[\begin{array}{c} \hat{v}_{x_1^J} \\ \vdots \\ \hat{v}_{x_{N^J}^J} \\ \hat{v}_{y_1^J} \\ \vdots \\ \hat{v}_{y_{N^J}^J} \end{array} \right] = \left\{ \begin{array}{l} \mathcal{H}^x(\mathbf{q}^J, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^J) \\ \vdots \\ \mathcal{H}^x(\mathbf{q}^J, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^J}^J) \\ \mathcal{H}^y(\mathbf{q}^J, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_1^J) \\ \vdots \\ \mathcal{H}^y(\mathbf{q}^J, \hat{\mathbf{P}}_s, \hat{\mathbf{P}}_e, y_{N^J}^J) \end{array} \right\} 2 \times N^J
\end{aligned} \tag{2.22}$$

with the amount of observations:

$$N = 2 \times \sum_{j=1}^J N^j \quad (2.23)$$

The unknown parameters in the *Gauss-Markov model* are

$$\boldsymbol{x} : \begin{bmatrix} X_s \\ Y_s \\ Z_s \\ X_e \\ Y_e \\ Z_e \\ y_1^1 \\ \vdots \\ y_{N^J}^J \end{bmatrix} \quad (2.24)$$

with the amount of unknowns:

$$U = 6 + \sum_{i=1}^J N^j \quad (2.25)$$

2 Methodologies

2.4.3 Constraint Equations

There are three constraints on the unknown parameters used in this work:

- Fixing the X-, Y-coordinates of the start-point using the approximate values:

$$\hat{X}_s - X_s^0 = 0 \quad (2.26)$$

$$\hat{Y}_s - Y_s^0 = 0 \quad (2.27)$$

- Fixing the length of the line segment (i.e. constraining the relative location of the end-point):

$$\sqrt{(\hat{X}_s - \hat{X}_e)^2 + (\hat{Y}_s - \hat{Y}_e)^2 + (\hat{Z}_s - \hat{Z}_e)^2} - S = 0 \quad (2.28)$$

Only in the very first line segment reconstruction of a long lane marking, the fixed X_s and Y_s values are from the initial parameter estimates derived in Section 2.5. Starting from the second line segment, the fixed values X_s^0 and Y_s^0 depend on the previously determined values.

The constraint equations (2.26), (2.27) and (2.28) can be stacked together and written in the structure of the *Gauss-Markov model with constraints*, corresponding to Equation (2.13):

$$h(\hat{\mathbf{x}}) = \mathbf{0} : \begin{bmatrix} \hat{X}_s - X_s^0 \\ \hat{Y}_s - Y_s^0 \\ \sqrt{(\hat{X}_s - \hat{X}_e)^2 + (\hat{Y}_s - \hat{Y}_e)^2 + (\hat{Z}_s - \hat{Z}_e)^2} - S \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.29)$$

with the amount of constraints:

$$H = 3 \quad (2.30)$$

2.4.4 Least-Squares Estimation for 3D Line Reconstruction

The nonlinear equation system is approximated to be locally linear with small step size of the unknown quantities. The linearized form is expressed as:

$$\widehat{\Delta l} = \Delta l + \widehat{\mathbf{v}} = \underset{N \times U}{\mathbf{A}} \widehat{\Delta \mathbf{x}} \quad (2.31)$$

$$\mathbf{c}_h = \underset{H \times U}{\mathbf{H}^T} \widehat{\Delta \mathbf{x}} \quad (2.32)$$

where

the $N \times U$ design matrix is the Jacobian of the function evaluated at the approximate values of the unknown parameters

$$\mathbf{A} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}^a}$$

the $U \times H$ constraint matrix is the Jacobian of the constraints

$$\mathbf{H} = \left(\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}} \right)^\top \Big|_{\mathbf{x}=\hat{\mathbf{x}}^a}$$

and the residual constraints are

$$\mathbf{c}_h = -\mathbf{h}(\hat{\mathbf{x}}^a)$$

with the corrections

$$\Delta \mathbf{l} = \mathbf{l} - \mathbf{f}(\hat{\mathbf{x}}^a) =: \hat{\mathbf{v}}^a \quad (2.33)$$

$$\widehat{\Delta \mathbf{x}} = \hat{\mathbf{x}} - \hat{\mathbf{x}}^a \quad (2.34)$$

where $\hat{\mathbf{x}}^a$ is the approximate values for the estimates of the unknown parameters.

In the *linearized substitute model* as shown in (2.31) and (2.32), it turns to solve for the increments of unknowns $\Delta \mathbf{x}$ instead of the unknowns themselves.

As the lines were extracted independently and with the same procedure, the measurements are assumed to be independent from each other and equal-weighted. That is, the weight matrix is an identity matrix:

$$\mathbf{W}_{ll} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (2.35)$$

The unknown parameters $\widehat{\Delta \mathbf{x}}$ of the linearized model can be determined from the extended normal equation system

$$\begin{bmatrix} \mathbf{A}^\top \mathbf{W}_{ll} \mathbf{A} & \mathbf{H} \\ \mathbf{H}^\top & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta \mathbf{x}} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{A}^\top \Delta \mathbf{l} \\ \mathbf{c}_h \end{bmatrix} \quad (2.36)$$

With the iteration index ν and the approximate values in the first iteration $\hat{\mathbf{x}}^{(1)} = \hat{\mathbf{x}}^{(a)}$, we have

$$\hat{\mathbf{x}}^{(\nu+1)} = \hat{\mathbf{x}}^{(\nu)} + \widehat{\Delta \mathbf{x}}^{(\nu)} \quad (2.37)$$

By updating the parameters using Equation (2.37), the LS estimation is applied iteratively until convergence is achieved.

The redundancy of the problem is

$$R = N + H - U = \sum_{j=1}^J N^j - 3 \quad (2.38)$$

The matrix \mathbf{A} does not need to have full rank, but the block matrix $[\mathbf{A}_T, \mathbf{H}]$ must have full rank in order to guarantee the estimation problem has a unique solution.

2 Methodologies

Two kinds of singular cases can happen. First, a configuration defect in object space can happen: as the 3D reconstruction approach still relies on intersection of multiple projection rays from different views, either the cases where there is only one image covering the targeted line segment, or when the targeted line segment lies on all of the epipolar planes of any of its two covering images, in these cases the problem is not solvable. Second, a configuration defect in image space can happen: as mentioned in Section 2.3, when the extracted line lies (nearly) in row direction on all the covering images, the problem is also not solvable. In the cases that the targeted line segment lies only on some of the stereo pairs' epipolar planes, the problem is still solvable whereas those stereo pairs are not contributing to the solution. Or similarly, when only in some of the images the extracted line segments lie (nearly) in row direction, the problem is solvable whereas those images are not contributing measurements to the estimation.

After LS estimation, the estimated variance-covariance matrix of the estimated parameters is

$$\begin{aligned}\hat{\Sigma}_{\hat{X}\hat{X}} &= \hat{\sigma}_0^2((A^T W_{||} A)^{-1} - (A^T W_{||} A)^{-1} H (H^T (A^T W_{||} A)^{-1} H)^{-1} H^T (A^T W_{||} A)^{-1}) \\ &= \begin{bmatrix} \hat{\sigma}_{\hat{X}}^2 & \hat{\sigma}_{\hat{X}\hat{Y}} & \hat{\sigma}_{\hat{X}\hat{Z}} \\ \hat{\sigma}_{\hat{Y}\hat{X}} & \hat{\sigma}_{\hat{Y}}^2 & \hat{\sigma}_{\hat{Y}\hat{Z}} \\ \hat{\sigma}_{\hat{Z}\hat{X}} & \hat{\sigma}_{\hat{Z}\hat{Y}} & \hat{\sigma}_{\hat{Z}}^2 \end{bmatrix}\end{aligned}\quad (2.39)$$

which depends on both the design matrix A (i.e. the configuration) and the posterior standard deviation of the measurements $\hat{\sigma}_0$ (i.e. the posterior measurements quality) under the constraints H .

2.5 Line Projection on the DSM (Determination of Initial Parameter Estimates)

As the target equation 2.16 introduced in Section 2.4 may exhibit multiple local minimum, a "correct" initial approximation of the unknowns is required for convergence to the correct solution. To provide such initial 3D line segment, the extracted line features derived in Section 2.1 can be projected onto a DSM based on the bundle adjusted exterior and interior orientations. An example is shown in Figure 2.8.

Given image coordinates $\mathbf{p}(x, y)$ of a point and (bundle-adjusted) image orientations \mathbf{q} , there is still one degree of freedom in extended collinearity equation (2.5) on solving object coordinates $\mathbf{P}(X, Y, Z)$. Combined with the usage of a DSM, which provides the height information Z_{DSM} given a position (X, Y) , the corresponding object coordinates can be solved iteratively until the increment ΔZ is small enough, i.e. convergence achieved. The iterative scheme is illustrated in Figure 2.9 and Algorithm 2.1.

Considering that the DSM is raster (discrete) whereas X and Y have continuous numerical values, the DSM height Z_{DSM} for the given point (X, Y) is bilinear interpolated.

2.5 Line Projection on the DSM (Determination of Initial Parameter Estimates)

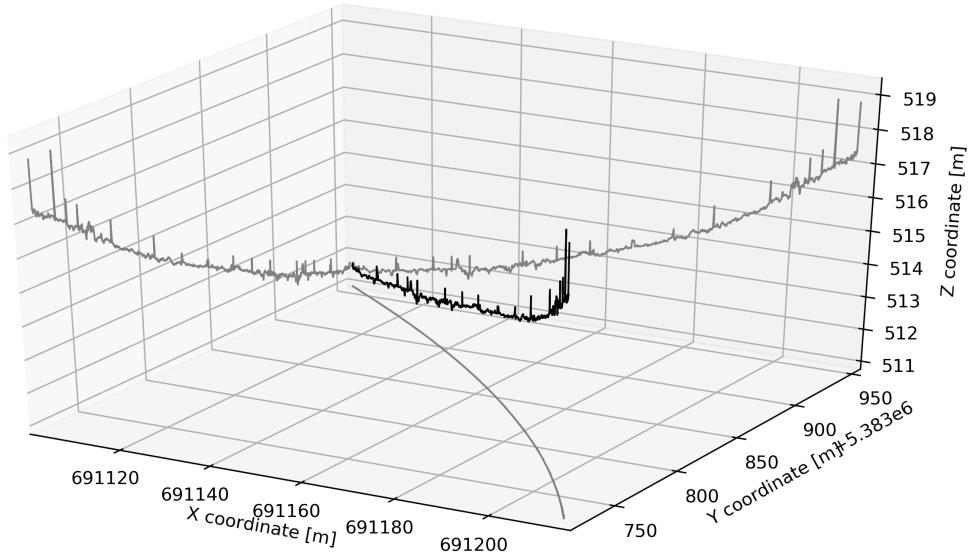


Figure 2.8: The projected line on DSM.

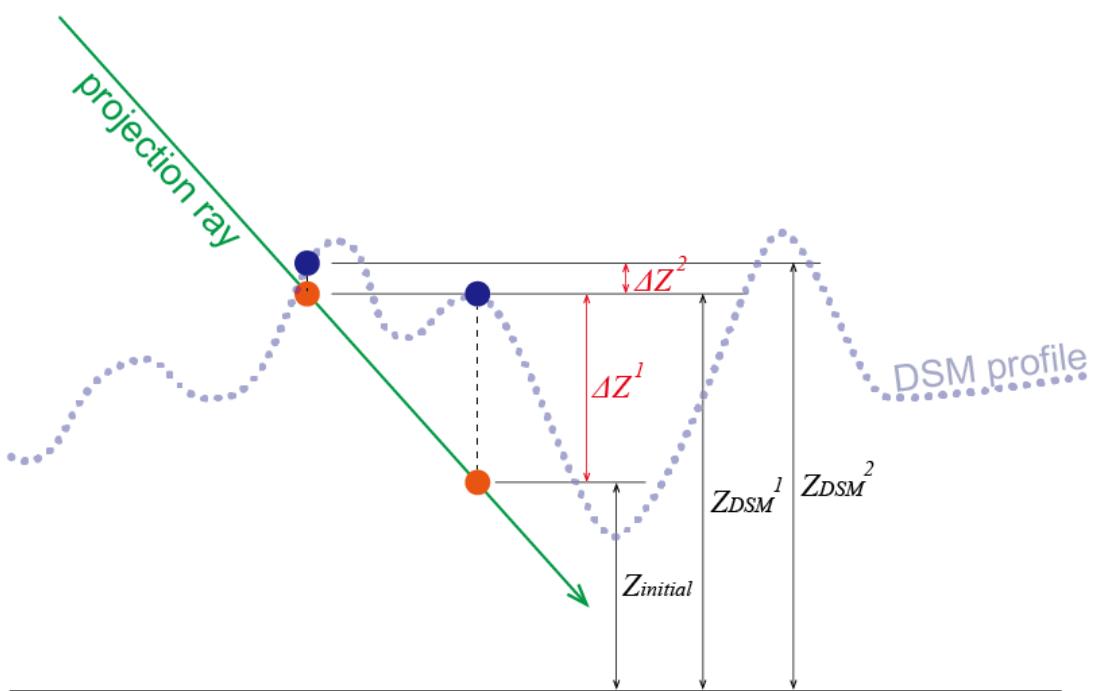


Figure 2.9: Iterative scheme of single projection ray intersecting DSM.

2 Methodologies

Algorithm 2.1 Single Point Projection on DSM

$[X, Y, Z_{DSM}] = \text{PointProjectionOnDSM}(p(x, y), \mathbf{q}, DSM)$

Input: image coordinates of a point $p(x, y)$, camera parameters \mathbf{q} and surface model DSM

Output: object coordinates of the projected point $P(X, Y, Z_{DSM})$ on DSM

```
 $\Theta = 0.5$                                 // unit: [meter], the convergence threshold
 $Z_{initial} = 500$                           // unit: [meter], an initial height value
 $\Delta Z = 9999$                             // unit: [meter], an infinite large value approximation
while  $\Delta Z > \Theta$  do
     $(X, Y) = \mathbf{q}.img2geo(p(x, y), Z_{initial})$ 
     $Z_{DSM} = DSM.\text{GetHeight}(X, Y)$ 
     $\Delta Z = |Z_{DSM} - Z_{initial}|$ 
     $Z_{initial} = Z_{DSM}$ 
end while
return  $(X, Y, Z_{DSM})$ 
```

3 Experimental Results and Evaluations

Section 3.1 provides information about the used dataset.

Section 3.2 describes the chosen parameters for line extraction, line projection on DSM, and window sliding processes.

In Section 3.3, the correctness of the derived LS model for reconstruction is evaluated. Some other properties of the proposed reconstruction approach are also discussed.

Section 3.4 evaluates the true data results. The result of single continuous lane-marking is presented and the theoretical precision is evaluated.

3 Experimental Results and Evaluations



Figure 3.1: Original Image

3.1 Materials

Aerial Images For real-time mapping applications during disasters, mass events and traffic monitoring scenarios, the German Aerospace Center (DLR) has developed a new optical sensor system— the 4k system— operated on a helicopter from DLR. The oblique aerial images used in this work are acquired from a Canon EOS 1D-X camera, one of the three non-metric cameras in the 4k system, with an oblique viewing angle τ of 15° . The pixel size is around $6.9 \mu\text{m}$, with the combination of focal length 50 mm and flying height H_{flight} around 500 m above ground, leading to a GSD of $\sim 6.9 \text{ cm}$.

As described in [FPK+17], this aerial imagery shall be improved to have an absolute geolocation accuracy of better than 30 centimeters if TerraSAR-X geodetic points are included as reference points.

An example aerial image is shown in Figure 3.1. Table 3.1 lists the properties of this camera, and Table 3.2 provides the viewing geometry information.

The images used in this work are acquired with a special flight configuration at both sides of the motorway which guarantees a continuous stereo view perpendicular to the lane marking direction.¹ This is realized by flying at the right-hand side with respect to flying direction along the motorway, with the left oblique camera looking left-down to the motorway, in both forward and backward trips. The flight configuration is shown in Figure 3.2 on the Google Earth platform.

¹The classical photogrammetric approach on flight planning is to have several straight flight lines which cover the whole motorway in a stereo view. This would be possible in this project yet would require more flight costs and would produce many more images.

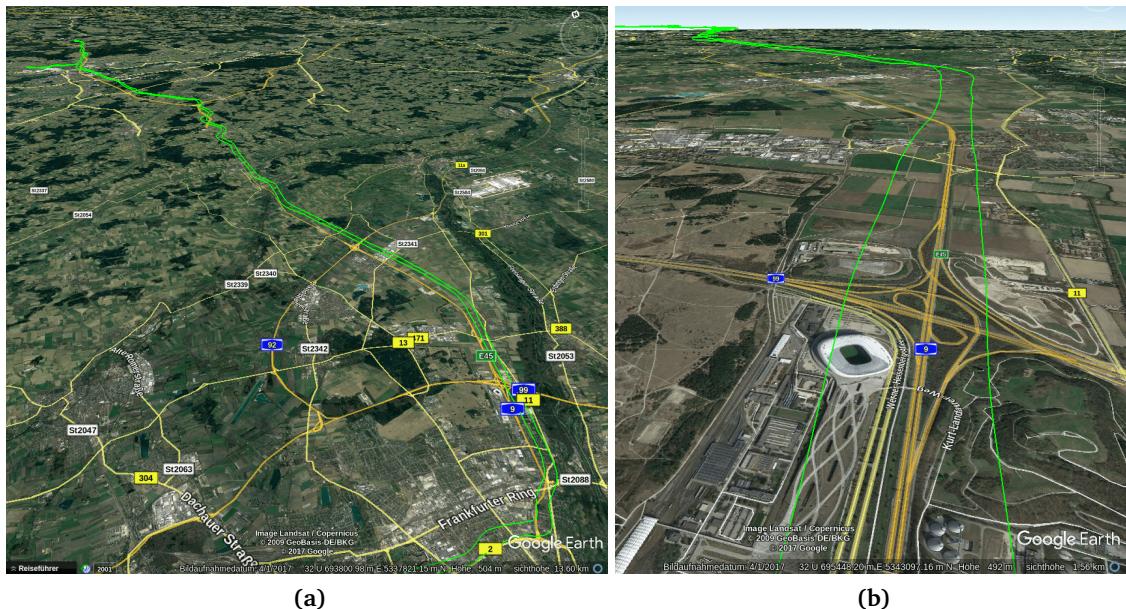
3.1 Materials

Canon EOS 1D-X	
Lenses	Zeiss Makro Planar f/2.0 50mm
Sensor / Pixel size	Full frame CMOS / $6.944 \mu\text{m}$
Image size	5184×3456 pixel, ratio 3:2 (17.9 MPix)
ISO	100–204800
max. frame rate / max. images	14 fps/ 180 images
Exposure time	30 s – 1/8000 s
Data interface	LAN (EDSDK software interface)

Table 3.1: Properties of the oblique camera

RGB, 50mm lens	
Oblique angle	$\pm 15^\circ$
FOV	$\pm 34^\circ$ across strip, $\pm 13^\circ$ along strip
Coverage @500m	780 m \times 230 m
GSD @500m	6.9 cm (nadir)

Table 3.2: Viewing geometry



3 Experimental Results and Evaluations

There are 15 images used in this work: 8 images in forward trip (first strip) and 7 in backward (second strip). The length of flight strips is around 890 meter.

Besides, the forward overlap is around 70%, and all the lane markings are covered by both strips, whereas the side overlap depends on the distance of flight strips, which is a result of the pilots navigation ability and other influences, like wind. Nevertheless, the motorway in its entire width was covered by the two flight strips. Altogether, this results in approximately 8-image coverage in road areas.

Exterior and Interior Orientations The aerial images are geo-referenced by Global Navigation Satellite System (GNSS)/Inertial system IGI AEROcontrol-IIId and further improved by Satellite Positioning Service of the German National Survey (SAPOS) correction. Additionally, a global terrain model (from Shuttle Radar Topography Mission (SRTM)) was introduced as pass information in the bundle adjustment, to improve the estimation of the focal length and boresight alignment. Self-calibrating bundle adjustment is applied to calibrate Interior Orientation (IO) parameters and to refine Exterior Orientation (EO) parameters. Table 3.3 and Table 3.4 show the precision of IO and EO parameters after self-calibrating bundle adjustment. To provide an overall quality on the interior orientations: from the calibration result of interior orientations (involving lens distortion), the residuals appear non-systematic and the biggest residual $r_{max,IO}$ is around 1 pixel.

position precisions	[meter]	attitude precisions	[degree]
σ_{north}	0.055	σ_{roll}	0.002
σ_{east}	0.035	σ_{pitch}	0.002
$\sigma_{altitude}$	0.069	σ_{yaw}	0.005

Table 3.3: Precisions of Exterior Orientations

Interior Orientations		precisions	unit
focal length c	0.051	σ_c	6.9e–7 [meter]
x coordinate of principal point pp_x	–42.259	σ_{pp_x}	0.167 [μm]
x coordinate of principal point pp_y	115.384	σ_{pp_y}	0.799 [μm]

Table 3.4: Interior Orientations and their precisions

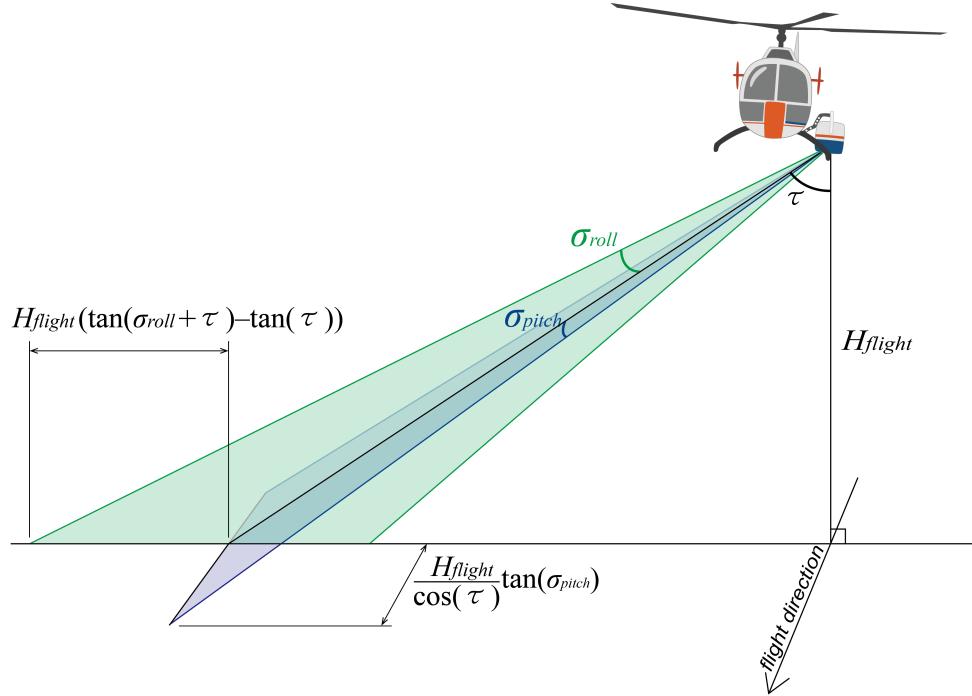


Figure 3.3: The impact of uncertain camera attitude in horizontal direction.

To evaluate the influence of the uncertainties in exterior and interior parameters on positioning precision in object space, the maximum values for each component based on the flight configuration was calculated. The quality of interior and exterior orientation parameters set would have a maximum impact in object space for around 16 [cm] in X,Y-direction:

- caused by inaccurate camera position:

$$\sqrt{\sigma_{north}^2 + \sigma_{east}^2} = \sqrt{0.055^2 + 0.035^2} \approx 0.065 \text{ [meter]}$$

- caused by inaccurate camera attitude, as illustrated in Figure 3.3:

$$\begin{aligned} & H_{flight} \times \sqrt{(\tan(\sigma_{roll} + \tau) - \tan \tau)^2 + \left(\frac{\tan \sigma_{pitch}}{\cos \tau}\right)^2} \\ &= 500 \times \sqrt{(\tan(0.002^\circ + 15^\circ) - \tan 15^\circ)^2 + \left(\frac{\tan 0.002^\circ}{\cos 15^\circ}\right)^2} \approx 0.026 \text{ [meter]} \end{aligned}$$

- caused by inaccurate Interior Orientations:

$$r_{max,IO} \times GSD = 1 \times 0.069 = 0.069 \text{ [meter]}$$

and around 6.9 [cm] in Z-direction:

- caused by inaccurate camera position:

$$\sigma_{altitude} \approx 0.069 \text{ [meter]}$$

3 Experimental Results and Evaluations

The above information tells the positioning precision in object space with measurements on a single image. With corresponding measurements from MVS, which allows the intersection of multiple projection rays from different directions, the positioning precision is expected to be improved.

DSM The DSM used in this experiment is generated by SGM based on MVS. As the asphalt road surfaces on which the lane markings are located are poorly textured, the SGM generated DSM is especially noisy in such poorly textured areas. However, such high-resolution DSM gives a good starting point for lane marking refinement. In other words, the DSM will be used only for setting up the initial values of the work flow, and will not influence the final results of 3D lane marking reconstruction.

The DSM has 20 cm grid spacing. Figure 3.4 shows a part of the DSM. Standard deviations of the height value in this part of the DSM is shown in Figure 3.5. The number of stereo image pairs used for each part on the DSM is shown in Figure 3.6.

Orthorectified Images The orthorectified images are processed using the DSM and the interior and exterior orientations derived from the bundle adjustment. One of the orthorectified images is shown in Figure 3.7. The orthorectified images are only used for setting up initial values and used as intermediary step for processing the road masks, but do not influence the results of 3D lane marking reconstruction.

Road Masks Road segments are masked out from original images based on OpenStreetMap (OSM) data: Firstly, the rasterized road segments from OSM data are written with 25 meter buffer width around road axes into orthorectified images. By back-projecting the mask from orthorectified image to original image using the 3D information from the DSM, it can then be used to mask out the road regions on the original images, as shown in Figure 3.8.

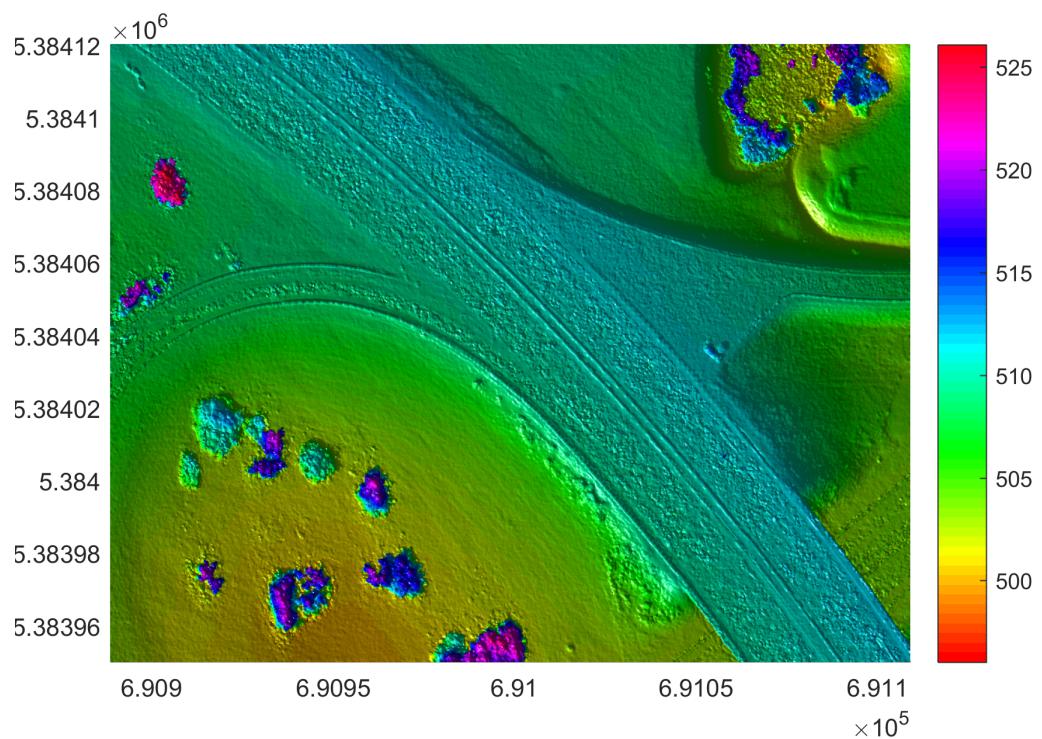


Figure 3.4: Part of the DSM in road area. It is noisy in the center of motorway.

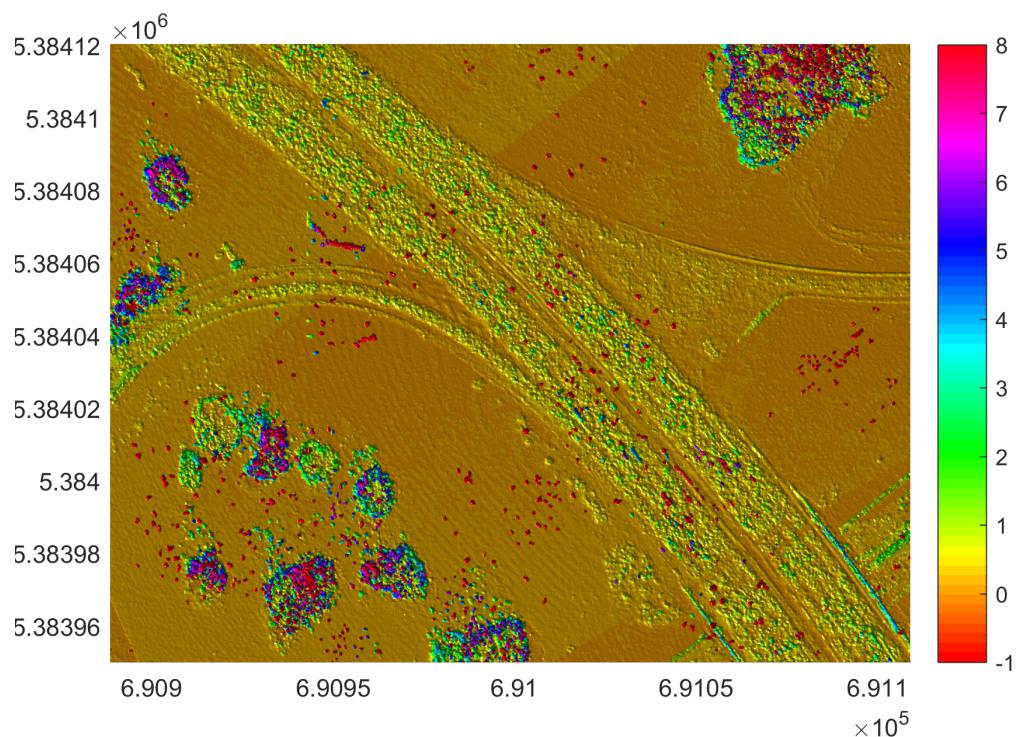


Figure 3.5: Standard deviations of the height value of the DSM in road area. It has higher values in the center of motorway.

3 Experimental Results and Evaluations

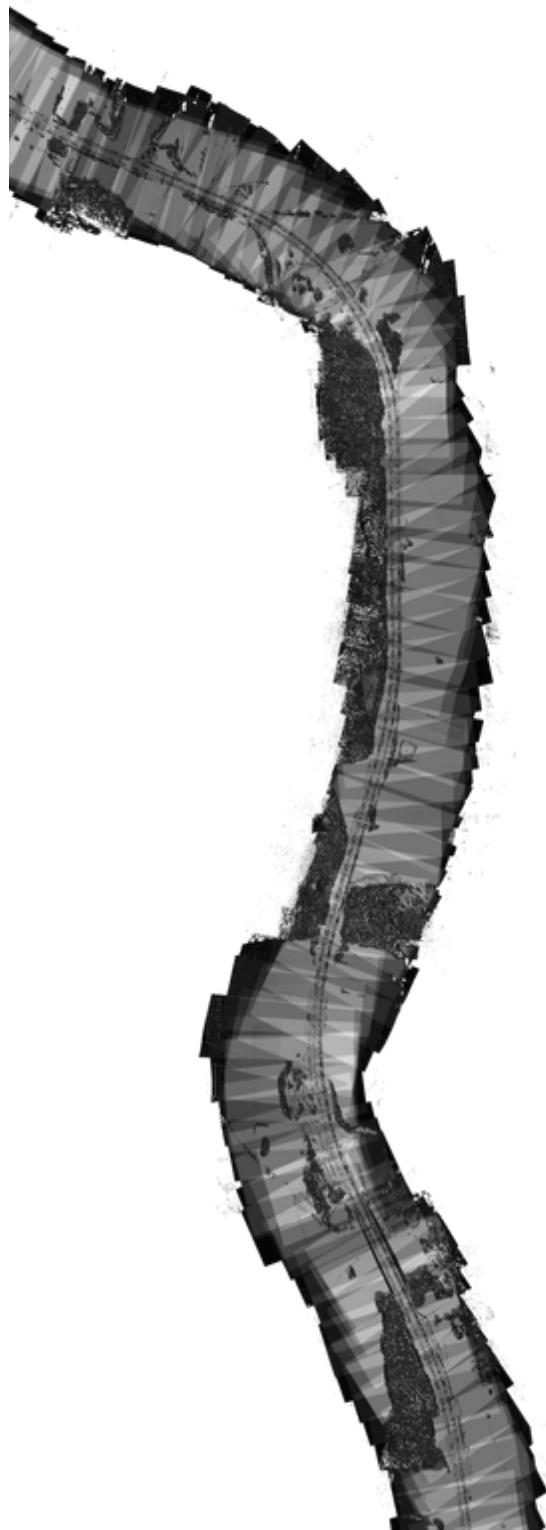


Figure 3.6: Distribution of stereo pairs used for DSM generation. Lighter color indicates more stereo pairs are used in that area. Maximum 27 stereo pairs are used for one pixel.

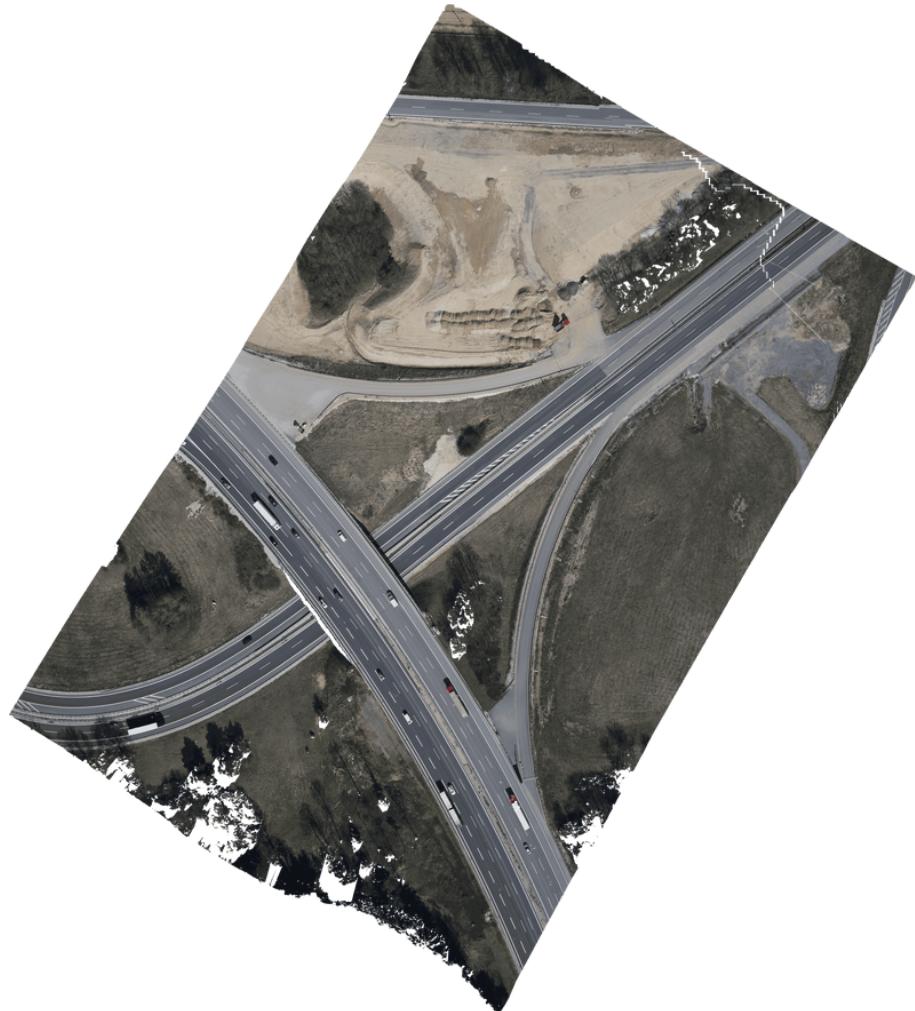


Figure 3.7: Orthorectified Image



Figure 3.8: Masked Image

3.2 Parameter Selection

In lane marking extraction process (see Section 2.1), the σ value for Gaussian smoothing is set to be 1.8 to slightly suppress the noise in images. A length threshold on the extracted lines is applied to reduce false positives —the detected line features which are indeed not lane markings. Regarding the fact that a dashed lane-line is around 6 meter long which corresponds to 62–87 pixels in image space (when parallel to one of the coordinate axes or in 45° direction), the extracted lines whose length is less than 65 pixels are rejected.

The convergence threshold for 2D image point projection on DSM (see Section 2.5) is selected as 0.5 m, as the height precision on road surface of the used DSM is expected to be around 50 cm.

The length of the sliding window should be decided base on the expected curvature of the targeted line and the robustness of the reconstruction model. In the cases of continuous lane markings on motorways, the sliding window length was fixed to 16 m, or for the last line segment it may be up to 24 m long, as a compromise between optimization robustness and the minimized systematic errors arisen from straight-approximated curvature. As to the cases of dashed lane-lines (6 m long), a sliding window has the same length as its targeted approximating line, i.e. it has 6 m length. The step size of sliding window was set to half of the sliding window length, i.e. 8 m.

The adapted lens distortion model is none shape-preserving, i.e. a 3D straight line is no more straight in image. However in every independent reconstruction process, only a short 16 m line segment is reconstructed, whose lens distortion correction along this line segment is no bigger than one pixel. Thus the bending of a line segment on an image, arose from lens distortion, is ignored in this work, but could be easily removed from the images by rectifying the images (i.e. calculating non distorted images).

3.3 Simulation

This section aims to verify the correctness of the derived LS model and to discover some characteristics of the reconstruction model using simulated data. The used materials are as described in Section 3.1. Only the measurements (the image coordinates of the extracted lines) as well as the true and approximate values of the unknowns (the object coordinates of a line segment) in the non-linear LS model are simulated, as described in Section 3.3.1.

Section 3.3.2 firstly evaluates whether the iteration scheme converges to the correct solution given imperfect initial values of the unknowns. The significant height differences between the approximate and the reconstructed line segments are also presented, indicating the refining ability of the proposed reconstruction approach.

The ability of the derived LS model on detecting the measurement errors is then evaluated. Furthermore, the increase of covering images would influence on the reconstruction result is elaborated.

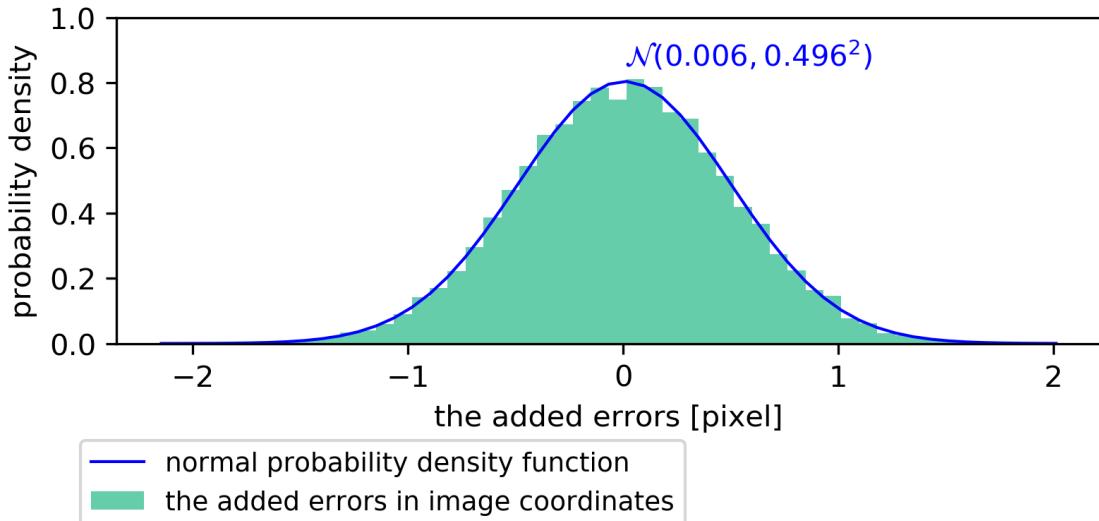


Figure 3.9: The added Gaussian random noise in the observations.

3.3.1 Simulated Data

The true line segment in object space Firstly, the object coordinates of the endpoints of a 3D line segment are defined, with 177.0 meters length, locating on the road surface in the test area (German highway A9) with 3 to 7 aerial images coverage. By linear interpolating several points with 0.2 meter spaces (considering DSM grid of 0.2 meter) between the two endpoints, a 3D line segment in the form of a set of 3D points is generated. This 3D line segment serves as the ground truth in the experiments in Section 3.3.

The observed line segments in image spaces The observations in the LS model are simulated by back-projecting the true line segment into the covering images. Gaussian random noise $e \sim \mathcal{N}(0, 0.5^2)$ is added in the observations for each LS adjustment, as line extraction process is of sub-pixel accuracy. The added noise is plotted in Figure 3.9

The approximate line segment in object space The initial estimates for non-linear LS adjustment is generated by projecting the observed line segments in image space onto the DSM.

3 Experimental Results and Evaluations

3.3.2 Simulation Result

Figure 3.10 shows the reconstructed and the true line segments in UTM 32N coordinate system. The distances from the reconstructed line nodes to the true line segment are computed and collected, resulting in sample size of 21. The sample mean is -0.023 [meter] and the sample variance is 0.106^2 [meter 2].

For such small sample size data, a two-tailed t-test is adopted to test if the population mean significantly not equals zero, i.e. if the reconstructed line segments are significantly far from the true line segments. Null hypothesis (H_0) and (two-tailed) alternative hypothesis (H_A) are stated as:

$$\begin{aligned} H_0 : \mu &= 0 \\ H_A : \mu &\neq 0 \end{aligned}$$

A significance level $\alpha = 0.05$ is selected, with degree of freedom being $21 - 1 = 20$, the two-tailed t-table value $T_{(0.975,20)}$ is

$$T_{(0.975,20)} = 2.086$$

which leads to the decision rule: if test statistic T_{obs} is less than $-T_{(0.975,20)} = -2.086$ or greater than $T_{(0.975,20)} = 2.086$, reject the null hypothesis.

With the sample mean $\bar{x} = -0.023$ [meter], the proposed population mean $\mu_0 = 0$ [meter], the sample standard deviation $\sigma = 0.106$ [meter], and sample size $n = 21$, the test statistic for One Sample T Test has the calculated value:

$$T_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{-0.023 - 0}{0.106/\sqrt{21}} \approx -0.994$$

which is neither less than $-T_{(0.975,20)} = -2.086$ nor greater than $T_{(0.975,20)} = 2.086$, i.e. not in the rejection region. As a result, we fail to reject the null hypothesis. In other words, we are not able to claim that the reconstructed line is significantly far away from the true line. This indicates that the derived non-linear LS adjustment model for 3D reconstruction is correct.

Figure 3.11 shows the reconstructed line segments and the DSM profile which serves as the initial approximation for non-linear LS adjustment, in UTM 32N coordinate system. The maximum distance between them is 1.97 meter, mainly in Z-direction. This tells that the reconstruction model is at least able to refine the initial approximation with 2 meters bias in Z-direction.

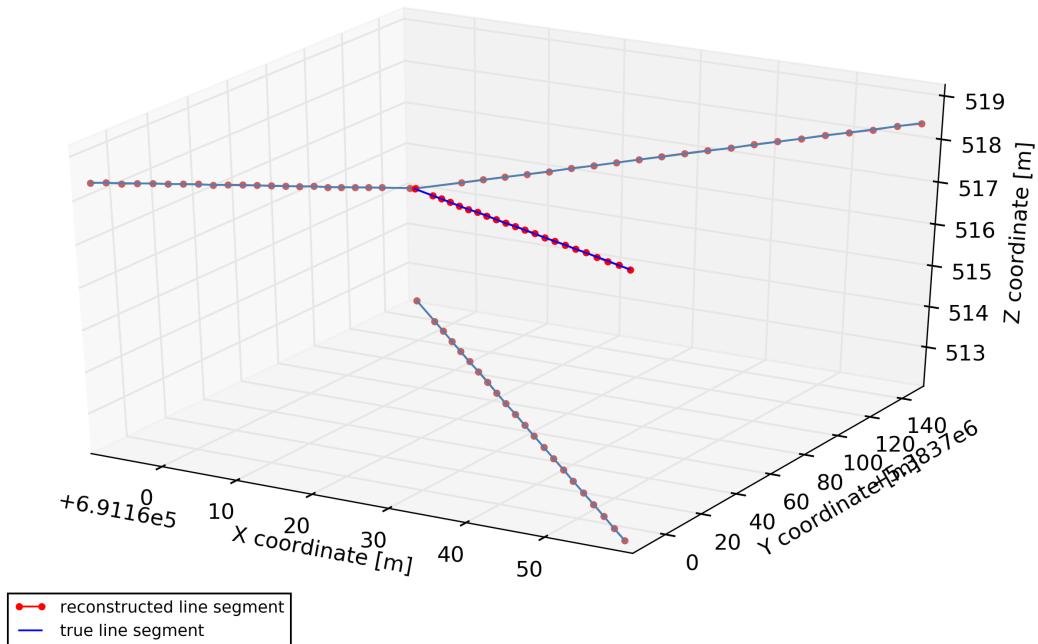


Figure 3.10: The reconstructed line segments and the true line segments in UTM 32N coordinate system.

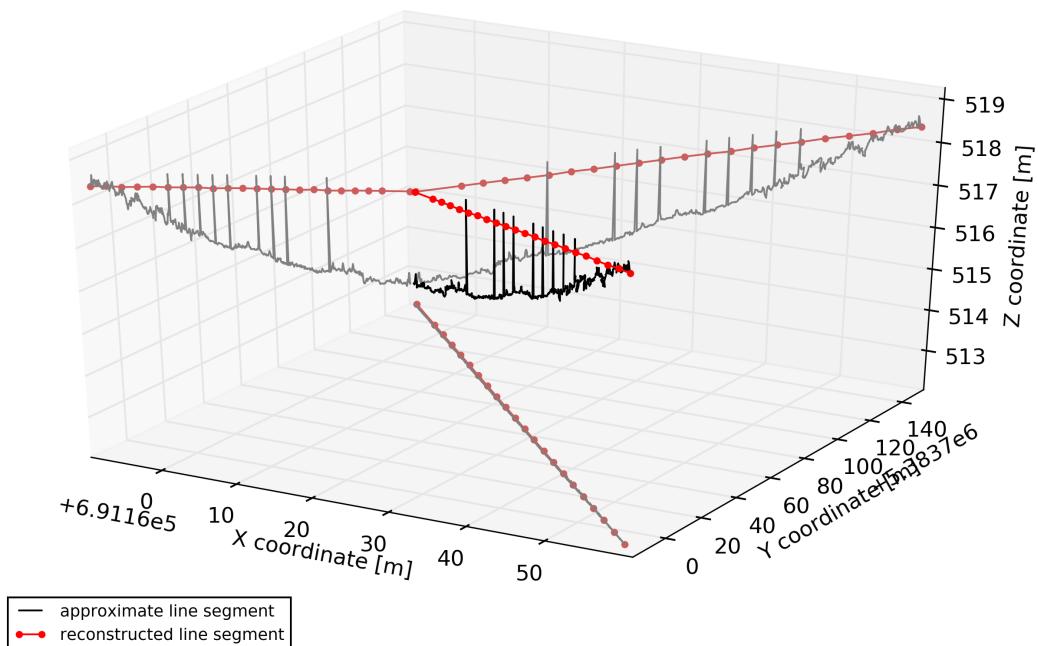


Figure 3.11: The reconstructed line segments and the unrefined DSM profile in UTM 32N coordinate system.

3 Experimental Results and Evaluations

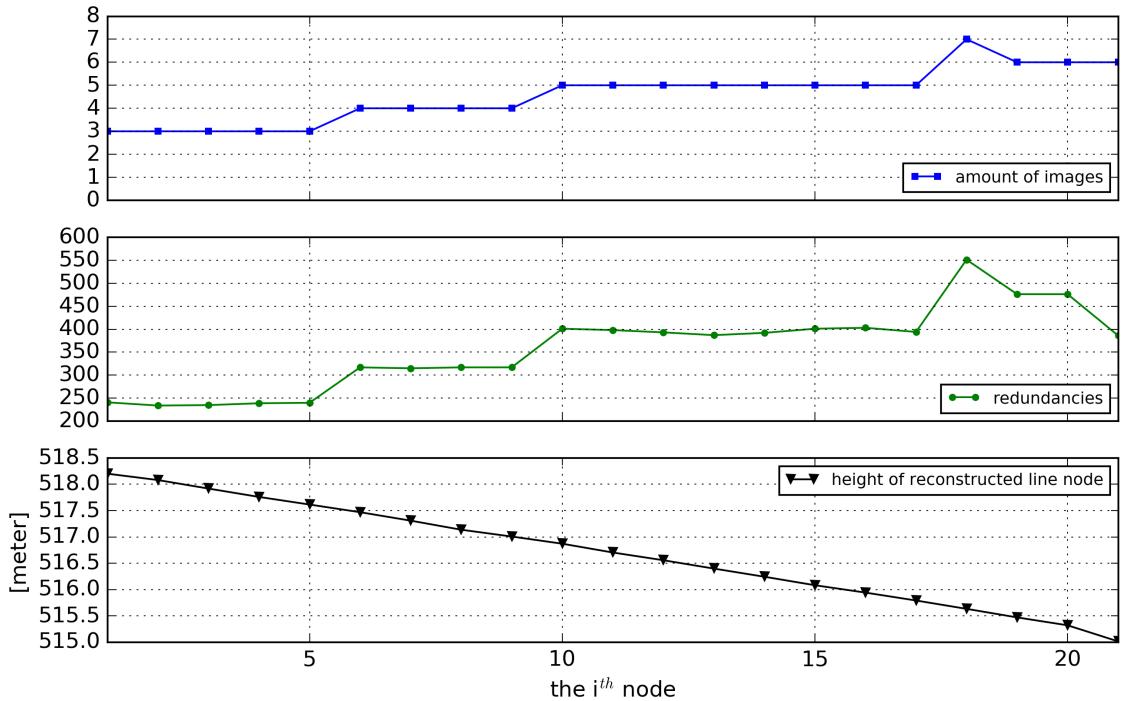


Figure 3.12: The amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment in simulation case.

The presented continuous lane line is 20 times segment-wise reconstructed, where each segment reconstruction is an independent LS adjustment process. Figure 3.12 gives the information on the amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment.

Figure 3.13 shows the mean and variance of the added random Gaussian noise and the adjusted residuals of each segment. This two samples are compared by applying a two-tailed two-sample T-test. Null hypothesis and alternative hypothesis (H_A) are stated as:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 - \mu_2 \neq 0$$

With significance level $\alpha = 0.05$ and degree of freedom ≈ 300 , the t-score is

$$T_{0.975,300} = 1.968$$

As shown in Figure 3.14, all the test statistics T_{obs} of each segment are not less than $-T_{0.975,300} = -1.968$ or greater than $T_{0.975,300} = 1.968$, i.e. not in the rejection region, the null hypothesis could not be rejected. In other words, we fail to claim that the adjusted residuals are statistically different from the added random noise.

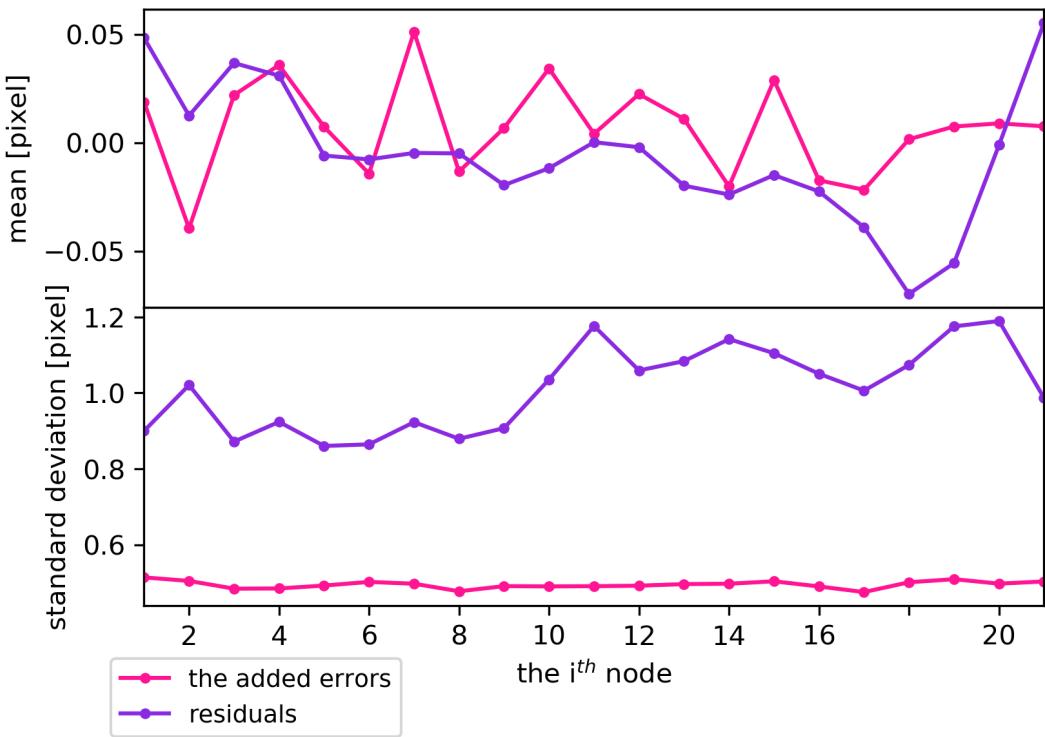


Figure 3.13: The relation between the added random Gaussian noise and the adjusted residuals.

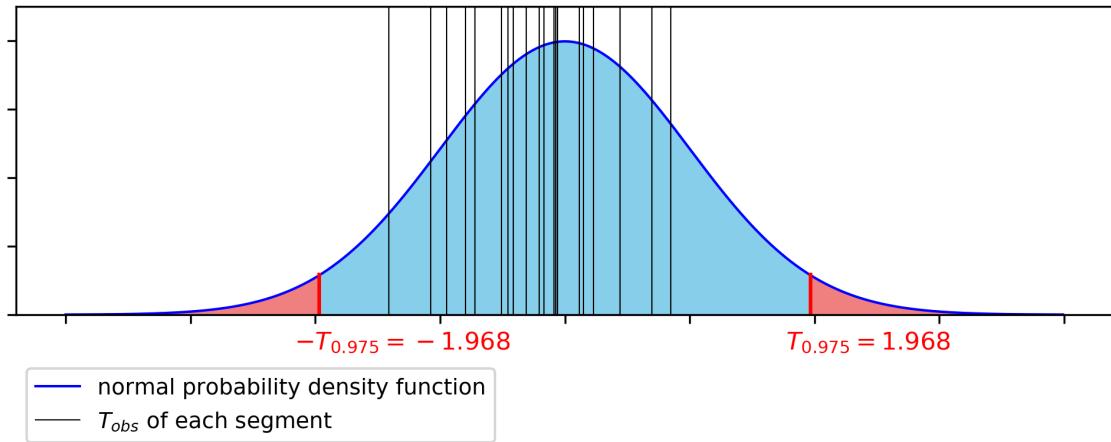


Figure 3.14: The red area under the probability density function is the rejection region, which is 5%. Since none of the T_{obs} falls in the rejection region, the null hypothesis could not be rejected.

3 Experimental Results and Evaluations

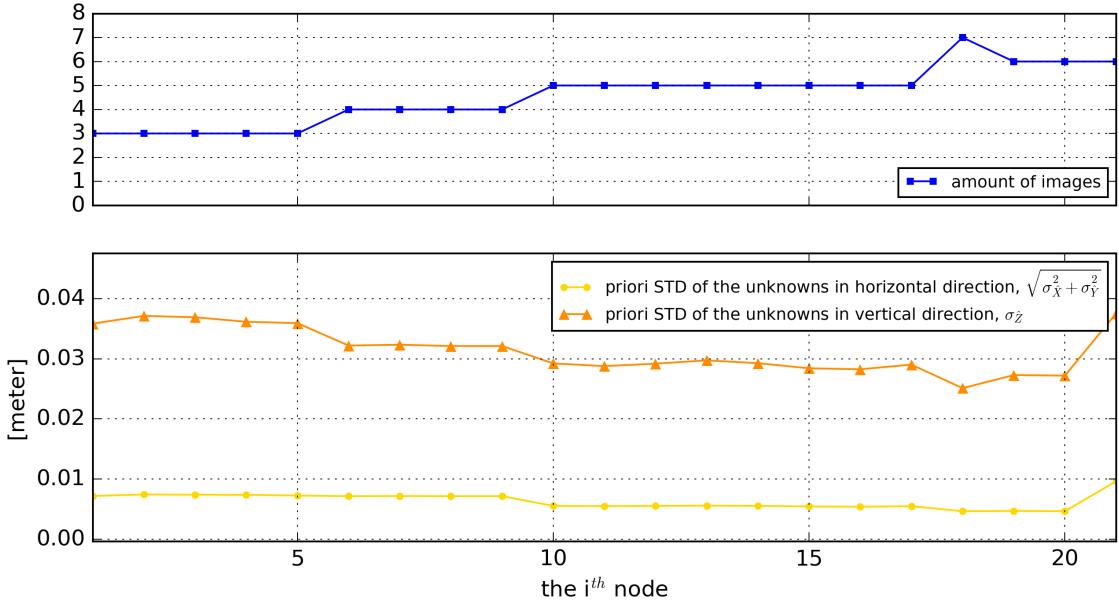


Figure 3.15: The amount of images, the posterior the standard deviation of the estimated parameters and .

By assuming a constant priori standard deviation of measurements to 1 [pixel] in all the LS adjustment processes, the priori variance-covariance matrix of the estimated parameters

$$\begin{aligned}\Sigma_{\hat{X}\hat{X}} &= \sigma_0^2((A^T W_{||} A)^{-1} - (A^T W_{||} A)^{-1} H (H^T (A^T W_{||} A)^{-1} H)^{-1} H^T (A^T W_{||} A)^{-1}) \\ &= \begin{bmatrix} \sigma_X^2 & \sigma_{X\hat{Y}} & \sigma_{X\hat{Z}} \\ \sigma_{\hat{Y}X} & \sigma_{\hat{Y}}^2 & \sigma_{\hat{Y}\hat{Z}} \\ \sigma_{\hat{Z}X} & \sigma_{\hat{Z}\hat{Y}} & \sigma_{\hat{Z}}^2 \end{bmatrix}\end{aligned}\quad (3.1)$$

reflects the quality of the design matrix in different LS adjustment processes. In other words, by assuming same measuring quality of 1 pixel level in each segment, $\Sigma_{\hat{X}\hat{X}}$ reflects the configuration strength of each line segments.

From Figure 3.15 it can be seen that the estimated parameters generally have smaller priori standard deviation with the increase of covering images. This tells that, the precision of the estimated parameters can be improved by increasing the configuration strength, i.e. increasing the amount of covering images with different orientations.

Besides, the priori standard deviation of the estimated parameters in horizontal direction $\sqrt{\sigma_X^2 + \hat{\sigma}_Y^2}$ is smaller than in vertical direction $\sigma_{\hat{Z}}$. This tells that the reconstructed nodes have higher precision in horizontal direction.

3.4 True Data

3.4.1 True Data Result

A continuous lane marking of 258.7 meters length is reconstructed. Figure 3.16 shows the reconstructed line segments and the DSM profile in UTM 32N coordinate system. The distances from the reconstructed line segment to the DSM profile are computed and plotted into histogram in Figure 3.17. They are collected along the reconstructed line segments with 0.2 meter spacing (considering the DSM grid of 0.2 meter), resulting in sample size of 1268. The sample mean is -0.162 [meter] and the sample standard deviation is 0.119 [meter].

Assuming DSM height profile being significantly lower than the reconstructed line segments for more than 15 centimeters, a lower-tailed Z-test is adopted. Null hypothesis (H_0) and (one-tailed) alternative hypothesis (H_A) are stated as:

$$\begin{aligned} H_0 : \mu &\geq -0.150 \\ H_A : \mu &< -0.150 \end{aligned}$$

A significance level $\alpha = 0.05$ is selected, i.e. the area in body is 0.950 out of 100%. The corresponding z-score is:

$$Z_{0.950} = 1.64$$

leads to the decision rule: if Z_{obs} is less than -1.64 , reject the null hypothesis.

With the sample mean $\bar{x} = -0.162$ [meter], the proposed population mean $\mu_0 = -0.150$ [meter], the sample standard deviation $\sigma = 0.119$ [meter], and sample size $n = 1268$, the test statistic for a One Sample Z Test has a calculated value:

$$Z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{-0.162 - (-0.150)}{0.119/\sqrt{1268}} \approx -3.59$$

As the test statistic $Z_{obs} \approx -3.59$ is less than $-Z_{0.95} = -1.64$, i.e. in the rejection region, the null hypothesis is rejected. In other words, with 95% confidence we can claim that the DSM profile is in average, statistically and significantly lower than the reconstructed line segments for at least 15 centimeters in this region.

3 Experimental Results and Evaluations

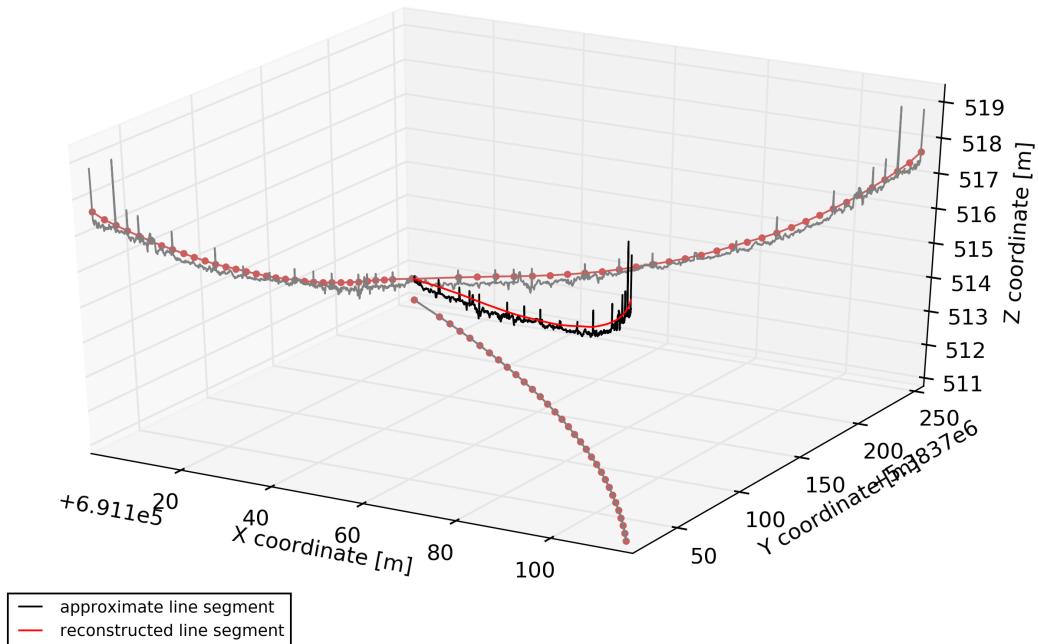


Figure 3.16: The reconstructed line segments and the unrefined DSM profile in UTM 32N coordinate system.

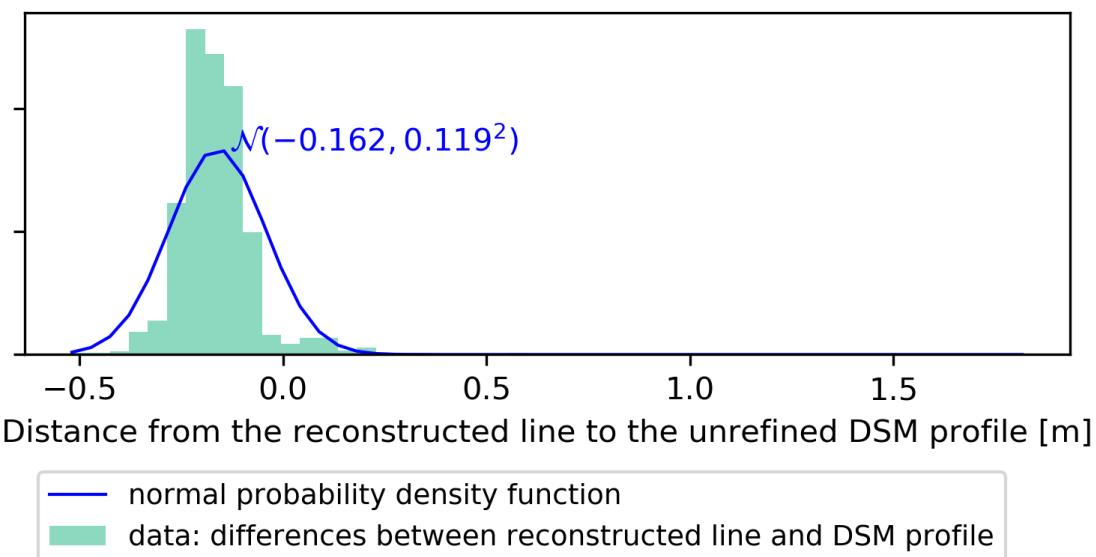


Figure 3.17: Histogram of the distances from the reconstructed line to the unrefined DSM profile.

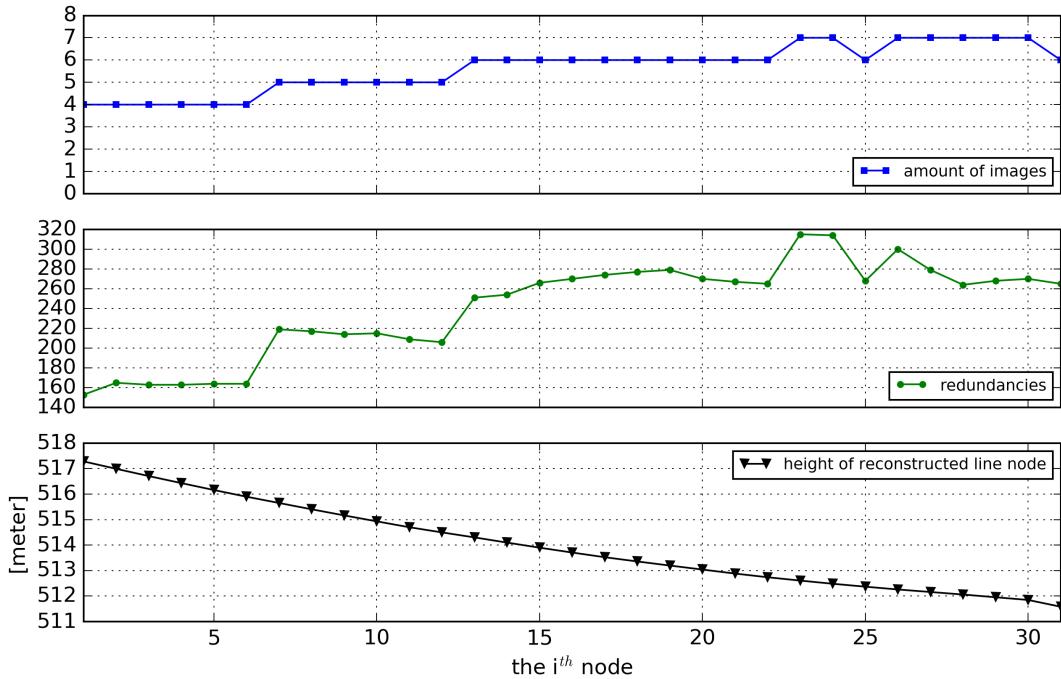


Figure 3.18: The amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment in the true data experiment.

Figure 3.18 gives the information on the amount of covering images, the redundancies and the height value of the reconstructed nodes, of each segment. Note that in each segment, LS adjustment is processed independently.

Compare to the simulation, where all line segments lie on a straight line and have same orientations, the true data (a continuous lane marking) is a curve line, whose line segments have different orientations from each other. Even some segments may have same image coverage configuration, their LS model may have different configuration due to their different orientations in 3D space. Nevertheless, the estimated parameters generally have smaller priori standard deviations with the increase of covering images, as shown in Figure 3.19. The priori standard deviations of the estimated parameters can be improved by increasing the configuration strength, i.e. the amount of images from different views.

Computed with Equation (2.39), Figure 3.20 shows: with sub-pixel posterior standard deviation of the measurements, the theoretical precision of the reconstructed line-nodes is within 2.5 cm in vertical direction and within 5 mm in horizontal direction.

The reconstructed lane-markings in the test area are shown in Figure 3.21 and Figure 3.22.

3 Experimental Results and Evaluations

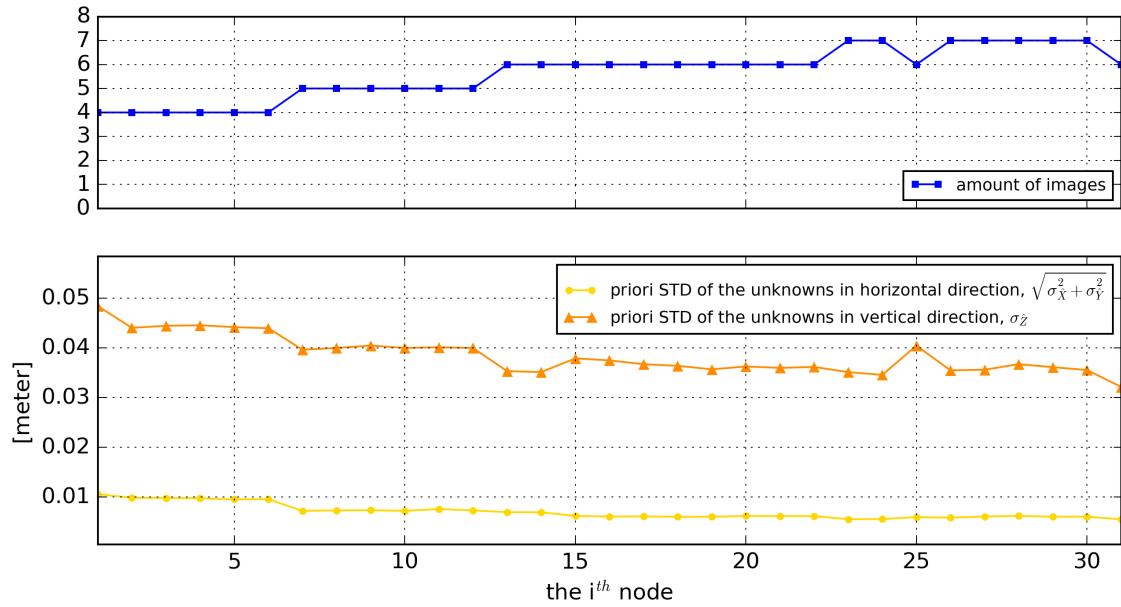


Figure 3.19: The variances of the estimated object coordinates, in horizontal and vertical directions.

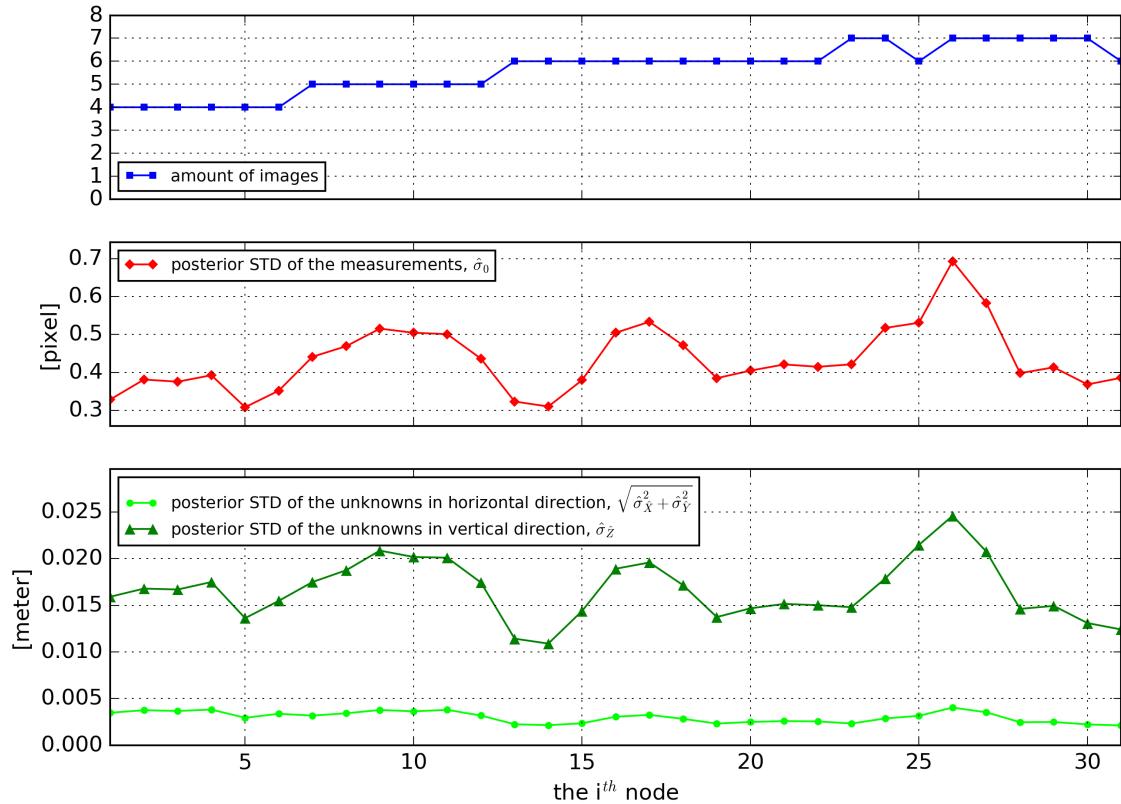
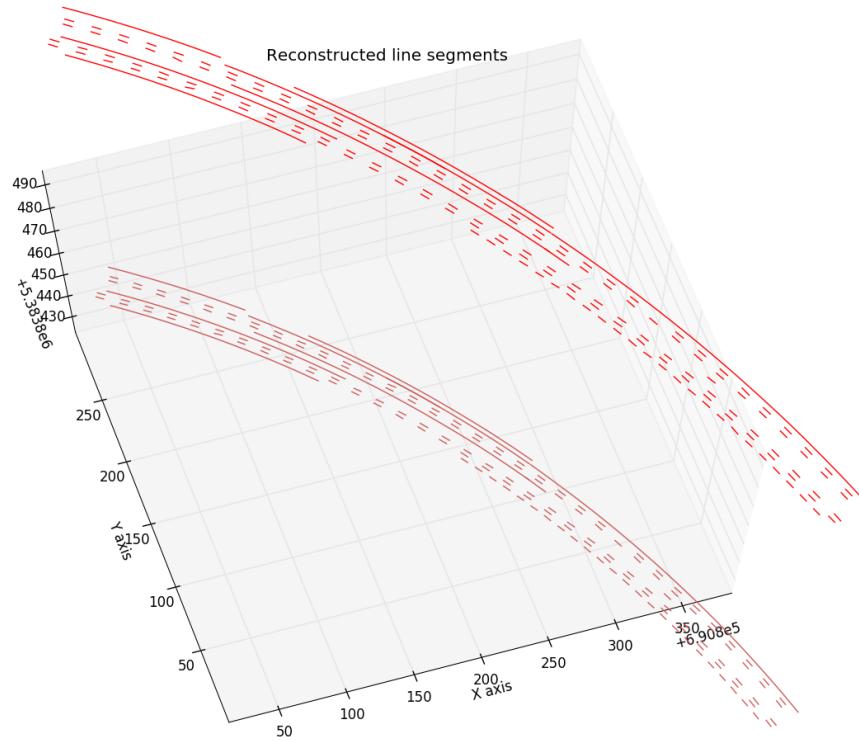
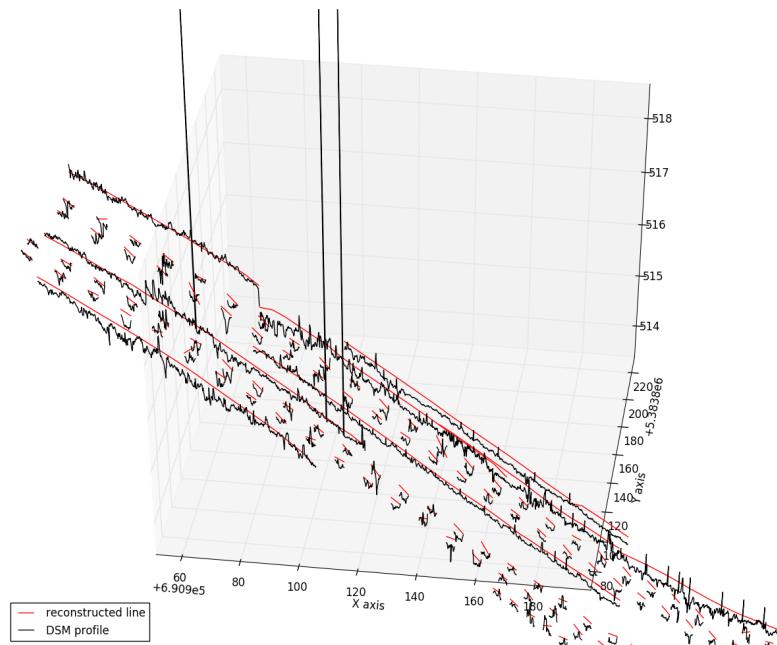


Figure 3.20: The estimated variances of the estimated object coordinates, in horizontal and vertical directions.



(a)



(b) Zoom in to show the difference between the DSM profiles and the reconstructed lines.

Figure 3.21: All the reconstructed lane-lines in the test area.

3 Experimental Results and Evaluations

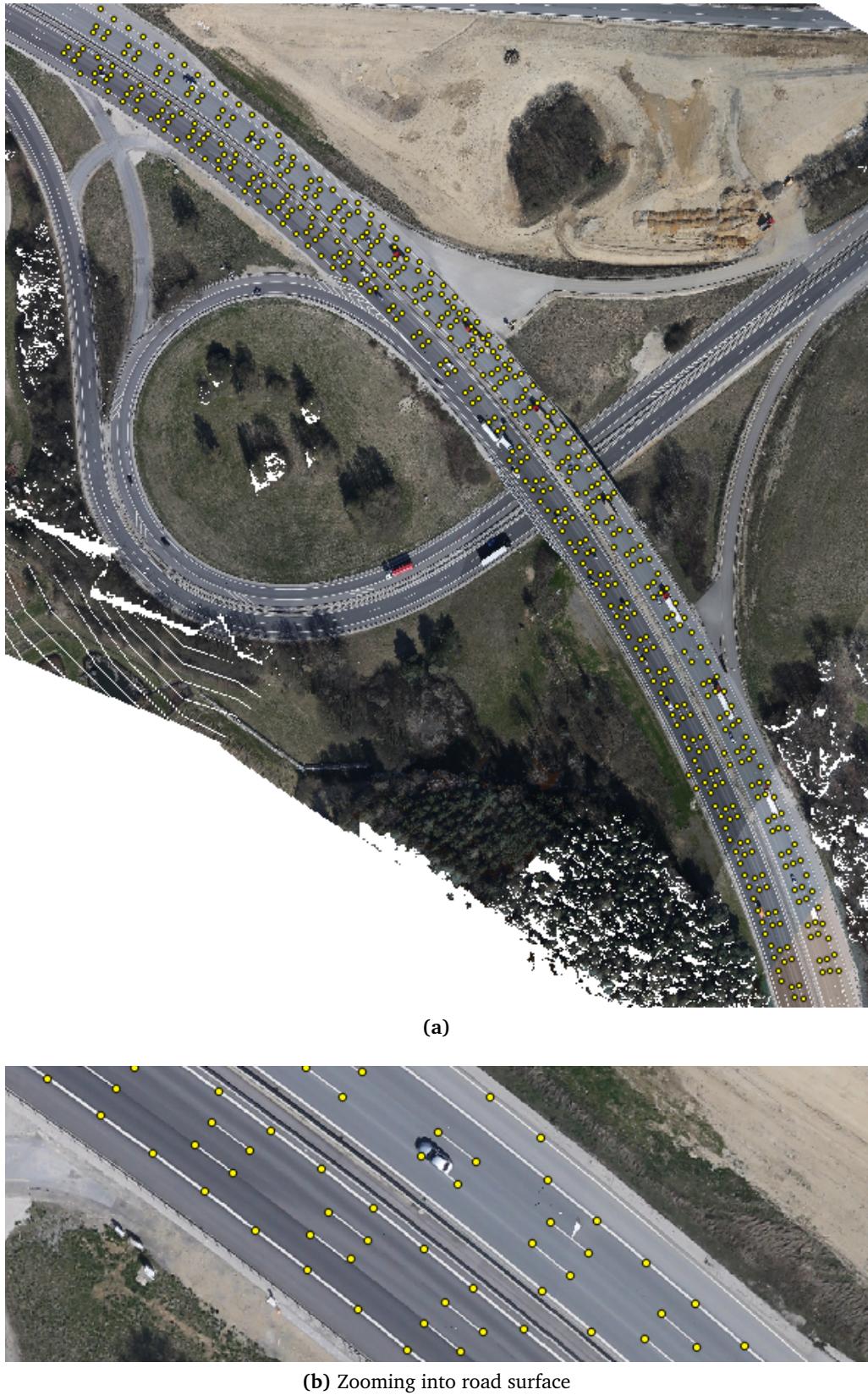


Figure 3.22: All the nodes of reconstructed lines on georeferenced aerial images, in UTM 32N coordinate system.

4 Conclusion and Future Work

This thesis proposed a framework for automatic 3D lane marking reconstruction using multi-view aerial imagery. Standard line detection algorithms are applied to extract lane markings on the aerial images. By exploiting the use of linear regression in image space and with the combination of collinearity condition, lane markings are reconstructed based on their detected positions in images and the viewing geometry. Without the utilization of the neighboring textures, the approach requires initial approximate 3D lines and is highly dependent on quality of pre-known image orientations. Nevertheless, it is robust to partial occlusions of the targeted lines on images and is applicable to the (quasi-)infinite line features as well as in cases of lowly textured neighboring and it improves the DSM at lowly textured road surfaces.

From simulated experiments in Section 3.3, some conclusions are given:

- Given approximations with 2 meters bias in Z-direction and sub-pixel random noises in the image coordinates of the detected/measured 2D lines, the proposed approach correctly refined the 3D positions of line segments.
- With the same line orientations in 3D space, the configuration strength increases with the increase of amount of covering images from different views.
- The reconstructed line segments have higher priori precision in horizontal direction than in vertical direction.

From experimental results of true data in Section 3.3, some conclusions are given:

- The DSM profile, i.e. the initial 3D line approximation, is significantly and systematically dozens of centimeters far away from the reconstructed line segments. This indicates the necessity of reconstruction based on detected 2D lines in image space and the viewing geometry, instead of simply reducing DSM noise by applying mean filter or such.
- The LS-estimated precision of the measurements, involving the lane marking extraction quality as well as the quality of image orientation parameters, is better than 1 pixel.
- The configuration defect barely happens: 1.when the linear regression functional model between image coordinates x and y are properly set up —according to the characteristics of the detected lane-lines in image space. 2.when the flight configuration guarantees stereo views whose base-lines are perpendicular to lane-lines orientations in 3D.

4 Conclusion and Future Work

- The theoretical precision of the reconstructed line-nodes is within 2.5 cm in vertical direction and within 5 mm in horizontal direction.

Some general conclusions are:

- Image configuration plays an important roll in 3D reconstruction. In the case of this work, being covered by more views whose base-lines are as perpendicular as possible to lane-marking directions, would improve the reconstruction result.
- Lane markings can be exploited to provide information on refining SGM-generated DSM.
- The proposed reconstruction workflow relies on initial approximation with enough accuracy, i.e. a global terrain model like SRTM would not be sufficient as starting height, as the error could be several meters in particular at roads.

Future Work

The proposed approach addresses 3D line features reconstruction problems without resorting to their appearances in images. Similar cases are railways. Applying this framework to such cases may be done in the future. Additionally, it worths trying the framework with images taken from drones instead of from helicopters.

Bibliography

- [ARB+15] M. Aeberhard, S. Rauch, M. Bahram, G. Tanzmeister, J. Thomas, Y. Pilat, F. Homm, W. Huber, N. Kaempchen. “Experience, Results and Lessons Learned from Automated Driving on Germany’s Highways.” In: *IEEE Intelligent Transportation Systems Magazine* 7.1 (Jan. 2015), pp. 42–57. ISSN: 1939-1390. DOI: [10.1109/MITTS.2014.2360306](https://doi.org/10.1109/MITTS.2014.2360306) (cit. on p. 11).
- [BFG05] H. Bay, V. Ferrari, L. V. Gool. “Wide-baseline stereo matching with line segments.” In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (June 2005), pp. 329–336. DOI: [10.1109/CVPR.2005.375](https://doi.org/10.1109/CVPR.2005.375) (cit. on pp. 14, 15).
- [Est] Estimation lemma. *3D reconstruction — Wikipedia, the Free Encyclopedia*. [Online; accessed 31-January-2018]. URL: https://en.wikipedia.org/wiki/3D_reconstruction (cit. on p. 12).
- [FPK+17] P. Fischer, B. Plaß, F. Kurz, T. Krauss, H. Runge. “Validation of HD maps for autonomous driving.” In: *International Conference on Intelligent Transport Systems in Theory and Practice, mobil.TUM* (July 2017) (cit. on p. 38).
- [Fra97] C. Fraser. “Digital camera self-calibration.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 52.4 (Aug. 1997), pp. 149–159. DOI: [10.1016/S0924-2716\(97\)00005-1](https://doi.org/10.1016/S0924-2716(97)00005-1) (cit. on p. 21).
- [HB08] H. Hirschmüller, T. Bucher. “Stereo Processing by Semiglobal Matching and Mutual Information.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (Feb. 2008), pp. 328–341. DOI: [10.1109/TPAMI.2007.1166](https://doi.org/10.1109/TPAMI.2007.1166) (cit. on p. 13).
- [HWB13] M. Hofer, A. Wendel, H. Bischof. “Line-based 3D Reconstruction of Wiry Objects.” In: *Proceedings of the 18th Computer Vision Winter Workshop* (Feb. 2013), pp. 78–85 (cit. on pp. 14, 15).
- [JKTS10] A. Jain, C. Kurz, T. Thormählen, H. Seidel. “Exploiting global connectivity constraints for reconstruction of 3D line segments from images.” In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2010), pp. 1586–1593. DOI: [10.1109/CVPR.2010.5539781](https://doi.org/10.1109/CVPR.2010.5539781) (cit. on p. 15).
- [Kep93] R. Keppler. *Richtlinien für die Markierung von Straßen (RMS) Teil 1 – Abmessungen und geometrische Anordnung von Markierungszeichen*. Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV), 1993 (cit. on p. 19).
- [Low99] D. Lowe. “Object recognition from local scale-invariant features.” In: *Proceedings of the Seventh IEEE International Conference on Computer Vision* 2 (Sept. 1999), pp. 1150–1157. DOI: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410) (cit. on p. 13).

- [SAE14] SAE International. *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. On-Road Automated Driving (Orad) Committee, 2014 (cit. on pp. 11, 12).
- [Ste98] C. Steger. “An unbiased detector of curvilinear structures.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.2 (Feb. 1998), pp. 113–125. ISSN: 0162-8828. DOI: [10.1109/34.659930](https://doi.org/10.1109/34.659930) (cit. on p. 20).
- [SZ97] C. Schmid, A. Zisserman. “Automatic Line Matching across Views.” In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (June 1997), pp. 666–671. DOI: [10.1109/CVPR.1997.609397](https://doi.org/10.1109/CVPR.1997.609397) (cit. on p. 14).
- [TA12] C. Topal, C. Akinlar. “Edge Drawing: A combined real-time edge and segment detector.” In: *Journal of Visual Communication and Image Representation* 23.6 (Aug. 2012), pp. 862–872. DOI: [10.1016/j.jvcir.2012.05.004](https://doi.org/10.1016/j.jvcir.2012.05.004) (cit. on p. 20).
- [Teu00] P. J. G. Teunissen. *Testing Theory: An Introduction*. Series on mathematical geodesy and positioning. Delft University Press, 2000. ISBN: 9789040719752 (cit. on p. 29).
- [TK95] C. Taylor, D. Kriegman. “Structure and motion from line segments in multiple images.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.11 (Nov. 1995), pp. 1021–1032. DOI: [10.1109/34.473228](https://doi.org/10.1109/34.473228) (cit. on pp. 15, 16).
- [WWH09] Z. Wang, F. Wu, Z. Hu. “MSLD: A robust descriptor for line matching.” In: *Pattern Recognition* 42.5 (May 2009), pp. 941–953. DOI: [10.1016/j.patcog.2008.08.035](https://doi.org/10.1016/j.patcog.2008.08.035) (cit. on pp. 14, 15).

All links were last followed on January 31, 2018.

Declaration

Hereby the following is confirmed – according to the Geomatics Engineering (GEOENGINE) examination regulations §22(7):

- I, the author, have prepared the thesis fully independently.
- I have not used any other than the specified sources. All statements that are literally taken from other sources are clearly marked. All these sources are explicitly listed in the thesis.
- The submitted thesis in whole or in substantial parts is not been involved in any other examination.
- This thesis has not been published previously in whole or in parts.
- The electronic version submitted fully complies with this thesis.

place, date, signature of thesis author