

**Integration of Geometric Computer Vision, Endoscopy and Computed  
Tomography for 3D Modeling of Gyroscopic Instruments**

Von der Fakultät für Luft- und Raumfahrttechnik und Geodäsie  
der Universität Stuttgart zur Erlangung der Würde eines  
Doktors der Ingenieurwissenschaften (Dr.-Ing.)  
genehmigte Abhandlung

vorgelegt von

**Kun Zhan**

geboren in  
Huanggang, China

Hauptberichter: Prof. Dr.-Ing. habil. Prof.h.c. Dieter Fritsch  
Mitberichter: Prof. Dr.-Ing. habil. Jörg F. Wagner  
Prof. Dr.-Ing. Sven Simon

Tag der mündlichen Prüfung: 30.05.2022

Institut für Photogrammetrie & Professur für Flugmesstechnik  
Universität Stuttgart  
2022

This thesis was published online on:  
<https://www.dgk.badw.de/publikationen/reihe-c-dissertationen.html>  
and <https://elib.uni-stuttgart.de>

# Contents

<b>Abstract</b>	<b>11</b>
<b>Kurzfassung</b>	<b>13</b>
<b>1 Introduction</b>	<b>15</b>
1.1 Background and Motivation . . . . .	15
1.1.1 Tech Heritage 3D digitization . . . . .	15
1.1.2 Importance of gyroscopes . . . . .	15
1.1.3 Integrated solutions . . . . .	16
1.2 Objectives . . . . .	17
1.2.1 Photogrammetric 3D reconstruction . . . . .	17
1.2.2 Endoscope calibration and 3D reconstruction . . . . .	17
1.2.3 CT 3D reconstruction . . . . .	18
1.2.4 Data integration . . . . .	18
1.2.5 CSG modeling of point clouds . . . . .	18
1.3 Contributions . . . . .	19
1.4 Outline . . . . .	19
<b>2 Related Work</b>	<b>21</b>
2.1 Cultural heritage preservation in 3D . . . . .	21
2.1.1 Review of photogrammetric 3D reconstruction . . . . .	21
2.1.2 Different sensors for 3D digital preservation . . . . .	22
2.2 Camera calibration . . . . .	22
2.2.1 Taxonomy of camera calibration solutions . . . . .	22
2.2.2 Stability analysis of camera calibration . . . . .	24
2.2.3 Endoscope calibration . . . . .	24
2.3 3D digitization . . . . .	25
2.3.1 Photogrammetry . . . . .	25
2.3.2 Endoscopy . . . . .	26
2.3.3 Computed Tomography . . . . .	28
2.4 3D model integration . . . . .	28
2.4.1 Gauss-Helmert Model based integration . . . . .	28
2.4.2 Automated data integration . . . . .	28
2.5 Deep learning and RANSAC-based CSG modeling . . . . .	29
2.5.1 RANSAC-based CSG modeling . . . . .	29
2.5.2 Deep learning-based CSG modeling . . . . .	29
<b>3 Camera Calibration</b>	<b>31</b>
3.1 Basic Concepts . . . . .	31
3.1.1 Collinearity equations . . . . .	31
3.1.2 Camera calibration models . . . . .	32
3.1.3 Experimental cameras . . . . .	33

3.2	Stability analysis of camera calibration . . . . .	35
3.2.1	Methods . . . . .	35
3.2.2	Experiments and results . . . . .	36
3.3	Endoscope calibration . . . . .	41
3.3.1	Methods . . . . .	41
3.3.2	Experiments and results . . . . .	43
3.4	Concluding remarks . . . . .	51
3.4.1	Stability of camera calibration . . . . .	51
3.4.2	Endoscope calibration . . . . .	51
<b>4</b>	<b>3D Digitization</b>	<b>53</b>
4.1	Imagery based 3D reconstruction . . . . .	53
4.1.1	Principle and workflow . . . . .	53
4.1.2	Photogrammetry . . . . .	55
4.1.3	Endoscopy . . . . .	62
4.2	Computed Tomography . . . . .	64
4.2.1	Basic principle and workflow . . . . .	64
4.2.2	CT 3D model pre-processing . . . . .	66
4.3	Concluding remarks . . . . .	69
<b>5</b>	<b>Data Integration</b>	<b>71</b>
5.1	Gauss-Helmert Model-based method . . . . .	71
5.1.1	Experiments and results . . . . .	71
5.2	CT surface coloring . . . . .	74
5.2.1	Related theory . . . . .	74
5.2.2	Experiments and results . . . . .	76
5.3	Virtual control points using geometric primitives . . . . .	78
5.3.1	Related theory . . . . .	78
5.3.2	Experiments and results . . . . .	80
5.4	Concluding remarks . . . . .	82
<b>6</b>	<b>CSG Modeling</b>	<b>85</b>
6.1	Manual and semi-automatic CSG modeling from point cloud . . . . .	85
6.1.1	Manual CSG modeling . . . . .	85
6.1.2	RANSAC-based CSG modeling . . . . .	86
6.2	Deep learning-based CSG modeling using point clouds . . . . .	87
6.3	Concluding remarks . . . . .	89
<b>7</b>	<b>Discussion and Conclusion</b>	<b>91</b>
7.1	Contributions . . . . .	91
7.2	Discussions . . . . .	91
7.3	Future work . . . . .	92
	<b>Bibliography</b>	<b>93</b>
	<b>Appendix</b>	<b>103</b>
	<b>Acknowledgements</b>	<b>118</b>
	<b>Relevant Publications</b>	<b>119</b>
	<b>Curriculum Vitae</b>	<b>120</b>

# List of Figures

Figure 1.1	Taxonomy for 3D tech heritage data collections [47]. . . . .	16
Figure 3.1	Camera systems used for calibration. From left to right: Sony $\alpha$ 7R III, Zeiss Loxia 25mm lens and Leica Q (Typ 116) (pictures are from respective official websites) [188]. . . . .	33
Figure 3.2	Camera calibration board (calib.io I/S). . . . .	34
Figure 3.3	Example image of the endoscope. . . . .	35
Figure 3.4	Self-designed endoscope imaging setup. . . . .	35
Figure 3.5	Workflow of camera calibration for stability analysis. . . . .	36
Figure 3.6	Gaussian fitting experiment results: the calibrated focal length (pixel) in x direction for (a) Sony $\alpha$ 7R III with Zeiss Loxia 25mm lens; (b) Leica Q (Typ 116) camera with Leica Summilux 28mm lens; (c) the Galaxy Note 8; (d)the iPhone 7 Plus. . . . .	37
Figure 3.7	Gaussian fitting standard deviations of the calibration for four types of cameras. . . . .	37
Figure 3.8	Gaussian fitting standard deviations (SD) of the calibrated focal length: (a) SD of the Sony camera; (b) the Leica camera; (c) Galaxy Note 8; (d) iPhone 7 plus. . . . .	38
Figure 3.9	Single factor analysis results of Gaussian fitting using different image compositions: (a) expected value for the calibrated focal length in X direction; (b) standard deviations of the calibrated focal length; (c) expected value for the principle point $x_0$ coordinate; (d) standard deviations of the principle point $x_0$ coordinate. . . . .	40
Figure 3.10	Temporal stability variation experiment results: (a) focal length of different calibrations with strong image block configuration; (b) focal length of different calibrations without strong image block configuration; (c) principle point of different calibrations with strong image block configuration; (d) principle point in X direction of different calibrations without strong image block configuration. . . . .	41
Figure 3.11	Oblique view indication. . . . .	42
Figure 3.12	Example image of endoscope calibration. . . . .	44
Figure 3.13	Example image of focal length Gaussian fitting from multiple endoscope calibration. . . . .	44
Figure 3.14	Expected values of the Gaussian fitting for endoscope calibrated focal length with different image number. . . . .	45
Figure 3.15	Standard deviation of Gaussian fitting variation with image numbers for the endoscope calibration. . . . .	45
Figure 3.16	Illustration of imaging distance and the focal length. . . . .	46
Figure 3.17	Undistorted image with different calibration parameter: (a) Model 1; (b) Model 2; (c) Model 3; (d) Model 7; (e) Model 8; (f) Model 12. . . . .	47

Figure 3.18	Straight line plot: (a) undistorted image with 1st order Fourier model; (b) zoomed view. The red line is the connection between the starting and ending feature of the row, the blue line is the connection between the extracted features of the undistorted image with an example plotted with green stars. . . . .	48
Figure 3.19	The sum of distance between undistorted feature and the straight line. . .	49
Figure 3.20	Calibrated endoscope focal length comparison with different models. . . .	49
Figure 4.1	Stereo vision illustration. . . . .	53
Figure 4.2	Topological structure of different SfM algorithms: (a) incremental SfM; (b) global SfM; (c) hierarchical SfM; (d) hybrid SfM. . . . .	54
Figure 4.3	Workflow for multiview imagery 3D reconstruction. . . . .	55
Figure 4.4	Photogrammetry turn-table setup. . . . .	55
Figure 4.5	Pose estimation results from VisualSfM: (a) KK12-09; (b) The Machine of Bohnenberger. . . . .	58
Figure 4.6	Manual registration using control points of KK17/03-09: (a) partial reconstruction with top side; (b) partial reconstruction with bottom side; (c) integrated reconstruction; (d) an original image of KK17-09. . . . .	59
Figure 4.7	Golden gnat registration using 3D features and ICP: (a) feature matching of point cloud pair; (b) initial pose of coarse registration; (c) refined registration using ICP algorithm; (d) integrated 3D model. . . . .	60
Figure 4.8	Spray operation for transparent or reflective surfaces: (a) original image; (b) sprayed image; (c) reconstructed mesh with sprayed image; (d) textured mesh with original image. . . . .	60
Figure 4.9	Image configurations and corresponding gyroscope 3D models by photogrammetry: (a)-(b) The Machine of Bohnenberger ; (c)-(d) KR05-17; (e)-(f) MFK01-09. . . . .	61
Figure 4.10	Undistortion of endoscope image: (a) original image; (b) undistorted image. . . . .	63
Figure 4.11	Specular highlight removal of endoscope image: (a) original image; (b) highlight removed image. . . . .	63
Figure 4.12	Brightness enhancement of endoscope image: (a) original image; (b) brightness enhanced image. . . . .	63
Figure 4.13	Drifted image alignment results with endoscopic images: (a) without sensor calibration; (b) with sensor calibration. . . . .	64
Figure 4.14	Drift-free Endoscopic SfM with manual control points . . . . .	64
Figure 4.15	Photogrammetry turn-table setup [48]. . . . .	65
Figure 4.16	X-ray slice of WK10-10 @ Gasim Mammadov. . . . .	65
Figure 4.17	CT histogram @ Gasim Mammadov. . . . .	66
Figure 4.18	Virtual dismantling of the stator of the rate gyro GG165A2 [48]. . . . .	66
Figure 4.19	CT point cloud generation from X-ray images. . . . .	67
Figure 4.20	De-noising process of WK10-10: (a) noise model of WK10-10; (b) de-noised model of WK10-10 @ Gasim Mammadov. . . . .	67
Figure 4.21	Two views of the iso-surface of G200 from the 3D volume data set reconstructed by the FBP [47]. . . . .	67
Figure 4.22	Two views of the iso-surface of the G200 calculated from the CT 3D volume data set reconstructed by the FBP from the denoised projections [47]. . . . .	68
Figure 4.23	Measurement setup inside the CT. . . . .	68
Figure 4.24	An example of surface extraction: (a) volumetric point cloud; (b) extracted surface data @ Gasim Mammadov. . . . .	68
Figure 5.1	Manual control points selection of Gyro G200. . . . .	72

Figure 5.2	Split views of integrated model of Gyro G200. . . . .	72
Figure 5.3	Split views of integrated model of direction gyro KK12-09. . . . .	73
Figure 5.4	Split views of integrated model of the Machine of Bohnenberger [47]. . . . .	73
Figure 5.5	Pipeline of CT surface point cloud colouring [189]. . . . .	75
Figure 5.6	Schematic diagram of ambiguous pixel on the depth image. . . . .	75
Figure 5.7	Photogrammetry 3D reconstruction of Golden Gnat: (a) example image of the Golden Gnat; (b) point cloud of the Golden Gnat. . . . .	76
Figure 5.8	Aligned photogrammetric and CT surface point clouds. . . . .	77
Figure 5.9	Depth image generation of Golden Gnat: (a) depth image of CT surface point cloud; (b) overlap display of the depth image and the corresponding photogrammetric image. . . . .	77
Figure 5.10	CT coloured point cloud of the Golden Gnat. . . . .	78
Figure 5.11	Result comparison of CT data coloring: (a) zoomed local display of CT coloured point cloud; (b) zoomed local display of photogrammetric point cloud. . . . .	78
Figure 5.12	Histogram of plane angles [187]. . . . .	80
Figure 5.13	Photogrammetric 3D surface model of KK10-09. . . . .	80
Figure 5.14	3D model of the object: (a) 3D volumetric model [47]; (b) point cloud @ Gasim Mammadov . . . . .	81
Figure 5.15	Extracted surface @ Gasim Mammadov. . . . .	81
Figure 5.16	Segmentation result. . . . .	81
Figure 5.17	Geometric primitive fitting of the object: (a) plane fitting; (b) cylinder fitting. . . . .	82
Figure 5.18	Virtual control points of photogrammetric model. . . . .	83
Figure 5.19	Cross sections of the integrated model. . . . .	83
Figure 6.1	3D modeling workflow. . . . .	86
Figure 6.2	An example (a) a cross section of the original model; (b) extrusion of the 2D cuts. . . . .	86
Figure 6.3	An example (a) 3D reconstructed model; (b) modeling of solid surfaces. . . . .	87
Figure 6.4	Virtual control points of photogrammetric model. . . . .	87
Figure 7.1	Illustration of collinearity equations. . . . .	104
Figure 7.2	The very first gyroscope of 1810—the Machine of Bohnenberger, Tuebingen, Germany: (a) original drawing; (b) photo @ J.F. Wagner ; and (c) 2.5D digital twin by CV/photogrammetry. . . . .	110
Figure 7.3	Golden Gnat: (a) general view; (b) front view of the Golden Gnat turning top . . . . .	111
Figure 7.4	The Gyro G200 of the LN3 Inertial Platform: (a) Gyro G200; (b) G200 transportation box. . . . .	112
Figure 7.5	Accelerometer BM06-10: (a) general view; (b) side view of the accelerometer. . . . .	112
Figure 7.6	Directional gyro KK13-09: (a) general view; (b) side view without the caging device. . . . .	113
Figure 7.7	Directional gyro KK01-09: (a) general view; (b) back view. . . . .	113
Figure 7.8	Pneumatically driven direction gyro: (a) general view; (b) side view of the course gyro. . . . .	114
Figure 7.9	Directional gyro KK12-09: (a) general view; (b) back view. . . . .	115
Figure 7.10	Gyroscopic rotor KR04-17: (a) general view; (b) side view. . . . .	115
Figure 7.11	Magnetic and long-range compasses MFK01-09: (a) general view; (b) side view. . . . .	116
Figure 7.12	Operator unit KH09-09: (a) general view; (b) side view. . . . .	116

# List of Tables

3.1	Specifications of camera systems used for calibration . . . . .	33
3.2	Features of the rigid endoscope . . . . .	34
3.3	Pixel size of the four different cameras . . . . .	37
3.4	Calibration parameters for Sony $\alpha$ 7R III, Leica Q, Galaxy Note 8 and iPhone 7 Plus . . . . .	39
3.5	Parameters for Sony $\alpha$ 7R III and the endoscope . . . . .	46
3.6	Comparison results of calibrated focal length and principle point coordinates for the endoscope. . . . .	50
4.1	Summary of employed datasets . . . . .	57
4.2	Summary of employed datasets . . . . .	57
5.1	Precision estimation of photogrammetry/CV and CT point cloud estimation by the Gauss-Helmert model . . . . .	74
5.2	Coefficients of 3D primitives . . . . .	79
6.1	Summary of employed datasets . . . . .	88
7.1	Dimensional information for the Machine of Bohnenberger . . . . .	110
7.2	Dimensional information for the Golden Gnat . . . . .	111
7.3	Dimensional information for the Gyro G200 of the LN3 Inertial Platform . . . . .	111
7.4	Dimensional information for the accelerometer BM06-10 . . . . .	112
7.5	Dimensional information for the directional gyro KK13-09 . . . . .	113
7.6	Dimensional information for the directional gyro KK01-09 . . . . .	114
7.7	Dimensional information for course gyro KK10-09 . . . . .	114
7.8	Dimensional information for the directional gyro KK12-09 . . . . .	115
7.9	Dimensional information for the gyroscopic rotor KR04-17 . . . . .	116
7.10	Dimensional information for the magnetic and long-range compasses MFK01-09 . . . . .	116
7.11	Dimensional information for the Operator unit KH09-09 . . . . .	117



# Acronyms

**3D** Three-dimensional array—3D points meshed by 3D topology. 15, 21

**AR** Augmented Reality. 16, 25, 27, 102

**BMBF** German Ministry of Education and Research. 21

**CAD** Computer-aided Design and Drafting. 99

**CLT** Central Limit Theorem. 43, 45

**CMVS** Clustering Multi-view Stereo. 32

**CNN** Convolutional Neural Network. 76

**CSG** Constructive Solid Geometry. 11, 12, 16, 25–27, 36, 95–100

**CT** Computed Tomography. 15, 16, 91, 102

**DIM** Dense Image Matching. 32, 63–67

**DLT** Direct Linear Transformation. 37, 61

**DT** Digital Twin. 25

**FBP** Filtered Back Projection. 4, 74, 77, 82

**FPFH** Fast Point Feature Histogram. 35

**GSD** Ground Sampling Distance. 82

**ICP** Iterative Closest Point (algorithm). 16, 35, 91

**LiDAR** Light Detection and Ranging. 27

**MEMS** Micro-Electro-Mechanical Systems. 22, 123

**MLE** Maximum Likelihood Estimation. 42, 43

**MS** Agisoft Metashape. 65–67

**NARF** Normal Aligned Radial Feature. 35

**OP** Orthogonal Polynomial. 117, 118

**PCG** Preconditioned Conjugate Gradient. 34

- PCL** Point Cloud Library. 85
- PFH** Point Feature Histogram. 88
- PMVS** Patch-based Multi-view Stereo. 32
- PPT** Python Photogrammetry Toolbox. 32
- RANSAC** Random Sample Consensus. 16, 36, 95, 96, 100, 101
- RBT** Rigid Body Transform. 79
- ReCap** RealityCapture. 63, 65–67
- SfM** Structure-from-Motion. 16, 24, 26, 31–34, 50
- SFS** Shape-from-Shading. 33
- SGM** Semi-global Matching. 32
- TH** Tech Heritage. 15, 21
- tSGM** Tube-based Semi-global Matching. 32
- VR** Virtual Reality. 16, 25, 27, 95, 96, 102

# Abstract

3D digitization is of vital importance for cultural heritage assets for modern civilizations regarding safekeeping and promotion. Generally, cultural heritage indicates old buildings, ancient status or unearthed relics for the public. However, the objectives to be digitized also include tools and instruments that have been widely applied in the past decades, even though they have been replaced with more advanced technologies. We call these technical instruments and artifacts Tech Heritage (TH). Gyroscopes are one group of such fascinating instruments with a history dating back to 200 years. The main characteristics of gyroscopes regarding 3D digitization are (1) having highly complex structure; (2) consisting of different materials; (3) not only the surfaces but also the internal structures are important. All these features decide that no single methodology could meet the demand for their 3D digitization.

To fulfill the requirements of gyroscopes in our research, photogrammetry, endoscopy and Computed Tomography (CT) are introduced for complete 3D digitization. With colored point clouds or textured meshes as result, photogrammetry is mainly for the global surface reconstruction of the object. For some cavities, holes or other parts that the regular camera hardly has access to, endoscopy is applied for a local 3D reconstruction, as supplement. As internal structures are also important, X-Ray computed tomography is utilized for volumetric 3D digitization. These three 3D sensor data should then be integrated for a complete 3D model. Additionally, the registration method should be adaptive to the data characteristics such as the geometry, point cloud density, etc. In this thesis, 3D reconstructions with each method as well as the data fusion are investigated.

1. Firstly, we study the stability and reliability of camera calibration before 3D reconstruction with photogrammetry and endoscopy. As the standard pre-calibration solution, Zhang's method suffers from the instability due to the correlations between the calibration parameters. To reduce this effect, the image configuration should be well considered with adequate oblique angles, distance difference as well as roll angles for a convergent image block. In our research, a quantitative analysis is implemented by a statistical approach using large bundles of images and get calibrations from randomly chosen image subsets. In addition, the recovered expected values of parameters are utilized as ground truth to scrutinize the single influencing factors of the imaging configuration.
2. Secondly, the 3D reconstruction processes are investigated with practical implementations. For the endoscope 3D reconstruction, the data acquisition process is the first challenge resulting from the image blur which may caused by the hand shaking as well as the small overlap. The imaging assistant setup and a mixture of image and video strategy are the methods adopted in our research as the solution. With the accurate calibration information and the improved image quality and configuration, we optimize the entire process through optimization of the Structure-from-Motion (SfM) method. As for CT 3D reconstruction, a stack of X-ray images, carrying the information of attenuation, is to be collected from different perspectives of the object. All reconstructed slices are integrated into an uniform 3D coordinate system to construct the complete 3D volumetric representation.
3. Thirdly, data registration methods are proposed regarding different data features. To register these two 3D data with few overlaps such as photogrammetry and endoscopic

point clouds, a Gauss-Helmert model with manually picked control points is applied for transformation estimation with precision assessments. To take advantage of the pairwise point cloud registration research, CT point cloud conversion and surface extraction are implemented from the volumetric CT data. As for the CT and photogrammetry data registration, it could be divided into two cases regarding the completeness of the CT surface representation. If the surface material is completely indicated in the CT data, we could directly project the color information from photogrammetric images to the CT surface after both datasets are transformed into the same coordinate system. In this way, we combine the high precision of CT data with the rich texture information. While low density surface material causes an incomplete representation of the CT surface, the transformation is estimated via the primitive based virtual control points from both surface data. With the determined transformation, the photogrammetric model could then be integrated with the CT model for a complete 3D representation.

4. Finally, in terms of 3D model expression, point clouds are of too big data volume if precision is required and have limited interaction possibilities. Therefore, the point clouds need to be vectorized into Constructive Solid Geometry (CSG) models to enable easier human-computer interaction. This process could be precisely done via manual work with sufficient caution via a Random Sample Consensus (RANSAC)-based geometric fitting process or even with a deep learning strategy via an end-to-end trained framework. The vectorized 3D model could be applied in AR/VR related applications to make full use of the work of 3D digitization.

For the first time, three totally different sensors are studied for a fused 3D reconstruction in this research. Among the workflow, the practical application of endoscopy is fully investigated. The integration methods are adaptively designed according to the characteristics of each sensor as well as of the reconstructed object. It provides more possibilities and ideas for the digital tasks of different types of cultural heritage.

# Kurzfassung

Die 3D-Digitalisierung ist von entscheidender Bedeutung für die Aufbewahrung und Förderung des Kulturerbes moderner Zivilisationen. Im Allgemeinen bezeichnet Kulturerbe alte Gebäude, antike Statuen oder ausgegrabene Relikte. Zu den zu digitalisierenden Objekten gehören jedoch auch Werkzeuge und Instrumente, die in den letzten Jahrzehnten weit verbreitet waren, auch wenn sie zum jetzigen Zeitpunkt durch fortschrittlichere Technologien ersetzt wurden. Wir nennen diese technischen Instrumente und Artefakte Tech Heritage (TH). Gyroskope sind eine Gruppe dieser faszinierenden Instrumente mit einer 200-jährigen Geschichte. Die Hauptmerkmale von Gyroskopen sind (1) hochkomplexe Struktur; (2) Herstellung aus verschiedenen Materialien; (3) nicht nur die Oberfläche, sondern auch die innere Struktur ist wichtig.

Um die Digitalisierung all dieser Merkmale zu realisieren, benötigt man eine umfangreiche Methodik. Um den Besonderheiten eines Gyroskops gerecht zu werden, werden in unserer Forschung Photogrammetrie, Endoskopie und CT für eine vollständige 3D-Digitalisierung eingeführt. Für die globale Oberflächenrekonstruktion des Objekts eignen sich insbesondere farbige Punktwolken und texturierte Netze der Photogrammetrie. Bei einigen Kavitäten, Löchern oder anderen Stellen, zu denen die normale Kamera kaum Zugang hat, wird das Endoskop für eine lokale 3D-Rekonstruktion ergänzend eingesetzt. Da für die 3D-Digitalisierungsaufgabe auch die interne Struktur wichtig ist, werden CT-Scans für eine volumetrische 3D-Digitalisierung verwendet. Diese drei Sensorarten sollen dann für ein vollständiges 3D-Modell integriert werden. Außerdem sollte die Registrierungsmethode an die Dateneigenschaften wie Geometrie, Punktwolkendichte usw. adaptierbar sein. In dieser Arbeit werden 3D-Rekonstruktionen mit jeder Methode sowie die Datenfusion untersucht.

1. Zunächst untersuchen wir die Kamerakorrektur vor der 3D-Rekonstruktion. Als Standard-Vorkalibrierungslösung leidet die Methode von Zhang unter der Instabilität aufgrund der Korrelationen zwischen den Kalibrierungsparametern. Um einen solchen Effekt zu reduzieren, sollte die Bildkonfiguration mit ausreichendem Schrägwinkel, Entfernungswinkel sowie Rollwinkeln berücksichtigt werden. In unserer Forschung wird eine quantitative Analyse durch einen statistischen Ansatz implementiert, der große Bildbündel verwendet und Kalibrierungen aus zufällig ausgewählten Bilduntermengen enthält. Darüber hinaus werden die gewonnenen Erwartungswerte von Parametern als Ground Truth verwendet, um die einzelnen Einflussfaktoren der Bildgebungskonfiguration zu untersuchen.
2. Als nächsten Schritt werden die 3D-Rekonstruktionsprozesse untersucht. Für die Endoskop-3D-Rekonstruktion ist der Datenerfassungsprozess der erste zu lösende Schritt aufgrund der Bildunschärfe, die durch das Handzittern verursacht werden kann, sowie der kleinen Überlappung, die sich aus der geringen Bildabdeckung ergibt. Der Aufbau eines Bildgebungsassistenten und eine Mischung aus Bild- und Videostrategie sind die Methoden, die in unserer Forschung als Lösung verwendet werden. Mit den genauen Kalibrierinformationen und der verbesserten Bildqualität und Konfiguration haben wir den gesamten Prozess durch Bildvorverarbeitung und Optimierung des Structure-from-Motion (SfM)-Verfahrens optimiert. Was die CT-3D-Rekonstruktion betrifft, wird aus verschiedenen Perspektiven des Objekts ein Stapel von Röntgenbildern gesammelt, die Informationen über die Ab-

- schwächung enthalten. Alle rekonstruierten Schichten werden in ein einheitliches 3D-Koordinatensystem integriert, um die volumetrische 3D-Darstellung zu konstruieren.
3. Im dritten Schritt werden Datenregistrierungsverfahren bezüglich verschiedener Datenmerkmale vorgeschlagen. Um die 3D-Daten mit wenigen Überschneidungen photogrammetrische und endoskopische Punktwolken zu registrieren, werden Gauss-Helmert-Modelle mit manuell ausgewählten Kontrollpunkten verwendet. Um die paarweise Punktwolkenregistrierung zu nutzen, werden CT-Punktwolkenkonvertierung und Oberflächenextraktion aus den volumetrischen CT-Daten implementiert. Die CT Daten können hinsichtlich der Vollständigkeit der CT-Oberflächendarstellung in zwei Fälle unterteilt werden. Wenn das Oberflächenmaterial in den CT-Daten vollständig enthalten ist, können wir die Farbinformationen aus photogrammetrischen Bildern direkt auf die CT-Oberfläche projizieren. Auf diese Weise kombinieren wir die hohe Präzision der CT-Daten mit den reichhaltigen Texturinformationen. Wenn dagegen das Oberflächenmaterial aufgrund geringer Dichte eine unvollständige Darstellung der CT-Oberfläche verursacht, wird die Transformation über die primitiven virtuellen Kontrollpunkte aus beiden Oberflächendaten geschätzt. Mit der ermittelten Transformation kann dann das photogrammetrische Modell mit dem CT-Modell für eine vollständige 3D-Darstellung integriert werden.
  4. Schließlich verursachen Punktwolken in Bezug auf die 3D-Modellgenerierung ein zu großes Datenvolumen, wenn Präzision erforderlich ist, und haben begrenzte Interaktionsmöglichkeiten. Daher muss die Punktwolke in ein CSG Modell vektorisiert werden, um eine einfachere interaktive Verwendung und Visualisierung mit entsprechender Datengröße zu ermöglichen. Dieser Prozess kann manuell mit ausreichender Sorgfalt durchgeführt werden, mit einem RANSAC-basierten geometrischen Anpassungsprozess oder sogar mit einer Deep-Learning-Strategie via ein durchgängig trainiertes Netz. Das vektorisierte 3D-Modell kann in AR/VR Applikation angewendet werden, um die 3D-Digitalisierung voll auszunutzen.

In dieser Arbeit werden erstmals drei vollkommen unterschiedliche Sensoren für eine fusionierte 3D-Rekonstruktion untersucht. Neben dem Arbeitsablauf wird die praktische Anwendung des Endoskops umfassend untersucht. Die Integrationsverfahren werden entsprechend den Eigenschaften jedes Sensors sowie des rekonstruierten Objekts adaptiv gestaltet. Es bietet daher mehr Möglichkeiten und Ideen für die digitalen Aufbereitung verschiedener Arten von Kulturerbe, vor allen Dingen im technischen Bereich.

# Chapter 1

## Introduction

### 1.1 Background and Motivation

#### 1.1.1 Tech Heritage 3D digitization

The definition of 3D digitization is the content that provides a faithful, often photorealistic representation of real-world objects. 3D digitization for cultural heritages is of vital importance regarding digital preservation as well as dissemination purposes. The digitized 3D model could be shared widely and freely without shipping them around the world for exhibitions in scene. With the rapid development of 3D digitization related sensors, a big community is working on applying them in virtual museum related projects. Most of them concentrate on cultural heritage objects in a common sense, such as old buildings, ancient status or unearthed relics. However, 3D digitization of Tech Heritage (TH), that is old technical instruments and artifacts, is leading to a new trend, though at the moment draws limited attention. These objects allow insights into developments that fundamentally shaped today's civilization. Without professional processing, however, these objects remain silent; especially when they are technically complex and heavily encapsulated. Gyroscopes are one group of such fascinating instruments with a history dating back to 200 years. At the University of Stuttgart, Germany, an unique collection of over 220 exhibits, including most of the known types of gyroscopes, as well as accelerometers and complete initial platforms, is available and in transition towards a sustainable future. 3D digitizations of these objects, not only for the surface, but also for the internal structure, are very important for research and didactic purposes as done within the project Gyrolog, which was supported by German Ministry of Education and Research (BMBF). The 3D digitizations of gyroscopes can be made accessible to draw more attention instead of being in the black box only serving for teaching and research. Among the whole collection, only part of them are chosen as experimental objects to be presented in this work for demonstrating the advantages as well as the problems of our methods. In the Appendix B, images of the example objects together with the introduction of their working principle, manufacture as well as other characteristics are presented.

#### 1.1.2 Importance of gyroscopes

The gyroscope is the invention that has revolutionized routing and navigation for more than a century with lasting impact on society. The main function of the gyroscope is the measurement of precise angles and rotations, which are frequently needed in many applications in our daily life such as Micro-Electro-Mechanical Systems (MEMS) in smartphones. In addition to its importance for modern everyday technologies, the gyroscope has always been of eminent military relevance and as "dual use" object, it is paradigmatic for technical developments of the very

latest decades. Due to its historical and cultural significance for over 200 years, gyroscopes are classic examples within the context of Digital Technological Cultural Heritage.

Except the uniqueness of the gyroscopes itself, the organization of the Gyrolog project also represents a close cooperation between several institutes at the University of Stuttgart, including the Chair of Adaptive Structures in Aerospace Engineering (now the Chair of Flight Measuring Technology), the Historical Institute, Department of the History of Natural Sciences and Technology, the Institute for Photogrammetry as well as the Institute for Parallel and Distributed Systems, which enables the overall goal of the 3D Virtual Reality modeling. In addition to the main collection in Stuttgart, the branches at the Technical University of Munich, Germany, and Johannes Kepler University of Linz, Austria, make also contribution to the digitization objectives.

### 1.1.3 Integrated solutions

The goal of 3D digitization is challenging due to the complex structure of the gyroscope and the special characteristics of necessary sensors. Gyroscopes have multiple materials and highly complex structures as well as different object sizes, which enable the angular velocity or orientation measuring abilities of different principles. To realize the goal of 3D digitization for gyroscopes, there are lots of possibilities that could be applied according to specific requirements. A technology taxonomy for 3D tech heritage data collections is shown in Figure 1.1. Though photogrammetry is normally regarded as the technology that could achieve a photo-realistic 3D reconstruction of the object with a mature workflow, the complete perspective of data acquisition, special material surface processing as well as special structures are still problems, that may limit its' ideal performance. These problems are discussed in this work with sufficient experiment validation. As for some cavities or hollow structures, it is difficult for a regular camera to take images from appropriate perspectives. Therefore, the close-range imaging devices that are frequently used in medical field, such as an endoscope, could be taken advantage of. Furthermore, the internal information of the object is impossible to be realized by 3D reconstruction with passive sensors. In this case, active image-based sensors could be applied. In the medical field, CT reconstruction related applications are either for local reconstruction or for localization assistance purposes instead of visualization applications with a complete 3D model. The 3D reconstruction applications with an endoscope and CT are innovative attempts, with different challenges accordingly to be solved.

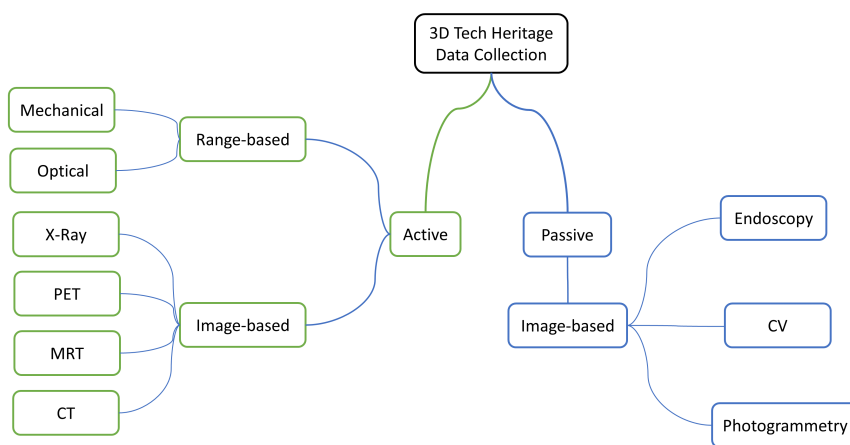


Figure 1.1: Taxonomy for 3D tech heritage data collections [47].



## 1.2 Objectives

Our research content is mainly for solving stated problems for a complete 3D model, including surface and internal representations of gyroscopes with complex structures. Generally, we innovatively introduce photogrammetry, endoscopy and CT technologies for an integrated solution.

### 1.2.1 Photogrammetric 3D reconstruction

In practice, photogrammetry is a mature topic and has been widely applied in many scenarios. Through reliable camera calibration and an appropriate image capture process, global colored point clouds or textured meshes could be obtained. Thanks to the long-term algorithm development and wide application exploration by a big community, photogrammetry has advantage over the other two sensors regarding applicability, data format as well as data registration methods. All these factors lead to a decision that photogrammetry data will be the reference for the transformation estimation of other sensor data. According to the characteristics of photogrammetry, its' main function in this work is the global reconstruction of gyroscopes. The general rule of image acquisition from different perspectives is realized in practice by the application of a turn table, a height-adjustable tripod together with a high resolution and wide-angle mirrorless camera. On the one hand, the overlap between images is required to be aligned together in the SfM process. Big overlap means small movement of the neighboring views which makes it easier for the correspondence estimation of the extracted feature points. On the other hand, too dense distribution of the images causes challenges for image collection as well as the computation of the 3D reconstruction, more specifically in the dense image matching process. Therefore, a compromise should be made when designing the image configuration for the 3D reconstruction considering both completeness of the views and computational efforts. In addition, part of gyroscopes from the collection contain homogeneous, reflective or even transparent surfaces such as a dash board of the instruments. These situations are challenging to apply the conventional imaging strategy due to ambiguities caused by the correspondences estimation from image pairs, which are influenced by the above factors. The solution could be either the processing on the object before image collection or processing of the image correspondingly, which are both investigated within this work here.

### 1.2.2 Endoscope calibration and 3D reconstruction

Endoscope 3D reconstruction is a big challenge due to the facts that (1) special optic design with a small scope cylinder which limits the f-number of the optical system and thus the image brightness; (2) oblique view design with a prism attached to the distal tip for bigger field-of-view which leads to the difficulty of image configuration design; (3) different magnification level in the center and in the periphery, which leads to a severe barrel distortion. For regular camera imagery 3D reconstruction, the task of calibrating the camera could be avoided by auto-calibration using directly multiple uncalibrated images for metric reconstructions. However, a pre-calibration is a prerequisite for the endoscope due to the low image quality and the difficulty of its image acquisition. The standard calibration method suffers from stability as well as the incompatibility when applied to endoscope calibration. Thus, different calibration models of diverse principles should be investigated using endoscopic images with appropriate configuration and quality. Due to the complex structure and severe distortion, calibration models with more modeling power, such as calibration models based on mathematical principles, are necessary to precisely recover the intrinsic parameters. Moreover, by applying a normal sequential SfM, it is very likely to cause severe drift of the endoscope trajectory resulting from the iterative execution of the bundle adjustment and inaccuracies due to the limited image quality. In comparison, the global SfM, which implements the bundle adjustment only, is free from the accumulation of errors while outliers could affect the final results. Considering the pros and cons of the sequential and

global SfM, alternatives such as distributed SfM or hybrid SfM could be alternatives for the pose estimation with endoscopic images. In addition, for the improvement of the endoscopic 3D reconstruction, more efforts could be made regarding the selection of the endoscope diameter, imaging assistant with better control system as well as the image pre-processing steps.

### 1.2.3 CT 3D reconstruction

CT is a non-destructive imaging technique used to reconstruct internal structures of an object that is not visible from outside in three dimensions. This ability is the ideal choice for the internal digitization of valuable historical collections without destroying their structure. It is the method that takes a large number of radio-graphic measurements from different angles, in order to generate slices of the object. All reconstructed slices are then integrated into a uniform 3D coordinate system to construct the complete 3D volumetric representation. Each voxel of the 3D model carries the attenuation values of the corresponding material when X-rays pass through. A first problem of CT reconstruction is the data completeness, which mainly relies on the material penetration characteristic. If the material is of low density, the corresponding information might be eliminated during the threshold selection process of the CT 3D reconstruction. Another problem is that the X-ray images have information not only about the accumulated attenuation on the detector but also the noise, which can come from the nature of monochromatic X-ray source causing beam hardening artefacts and scattering. Therefore, data completeness and noise are the main problems that need to be dealt with for the application in our research.

### 1.2.4 Data integration

After obtaining the 3D models from three sensors individually, data registration is to be implemented. First of all, for low overlapping ratios between the point clouds such as endoscopic local 3D reconstruction and the photogrammetric global 3D reconstruction, the integration could be solved by manually picked control points and a Gauss-Helmert model estimation process. Though manual effort is necessary in picking up the control points, this method gives precision assessments of the estimation process and has few limitations on the initial data in comparison to the conventional ICP method. Basically, it could be applied to any situations where at least four identical points could be extracted from the point cloud pair to be registered. For another, CT data has different attributions as photogrammetry, such as the internal and surface reconstruction, intensity and texture representation. Besides this manual registration method, more efforts are to be made for realizing an automated process according to the data integrity of the CT surface. On condition that the CT surface data shows good completeness, the color information from the photogrammetric images could be projected onto the corresponding CT points which have been transformed into the same coordinate system. Thus, the combination of color from photogrammetric data and high precision of the CT data could be realized. Otherwise, the application of the overall feature rather than the local ones will be used by fitting geometric primitives to the CT surface point cloud and photogrammetric point cloud. As a following step, the control points, which are generated from the primitives directly or by an intersection are utilized to estimate the transformation between point clouds from different coordinate systems. This method has advantage over the manually picked control points regarding the accuracy as well as other 3D feature point extraction algorithms concerning the data requirements and robustness.

### 1.2.5 CSG modeling of point clouds

After having the 3D Digital Twin (DT) of the gyroscopes in the format of a point cloud or mesh, there is still an obstacle in applying them in scenarios such as web visualization or VR/AR presentations mainly due to their large data volume. Therefore, the procedure of CSG modeling

based on the original dense point cloud is implemented for the vectorization of the data by replacing them with combinations of the regular geometry shapes. The manual modeling requires demanding work, especially for the objects with rich details, though it works as the most practical way for the visualization task at the moment. Another option is by taking advantage of deep learning principles. Because of the special characteristics, it is much more difficult to apply deep learning methods for the point cloud than other data types such as images or even audios and videos. In our research, with the training of the large amount of point clouds with different geometric shapes, density, completeness and other attributions, the primitives are expected to be extracted from the complex point clouds, which could largely accelerate the CSG modeling procedure. It has to be noted, that the application of deep learning in automatic point cloud modeling is still a challenging task. However, it will be the trend in near future, also for the wider application of 3D data in different scenarios.

### 1.3 Contributions

The main contribution of this thesis is dealing with the following aspects:

- **Camera calibration:** The stability and reliability of the camera calibration using the conventional Brown model is investigated by a statistical analysis with a large number of image subsets calibration results. The influence of different imaging configurations are quantitatively analyzed. By implementing the stated process, the problem caused by the parameter correlations and the image configuration requirements could be avoided in a degree by using qualified calibration images.
- **Endoscope calibration and reconstruction:** Except the reliability analysis, different calibration models based on different principles are compared for a satisfying calibration of the endoscope. In addition, the SfM process is also optimized according to the characteristics of endoscopic imagery. With well calibrated parameters, pre-processed images and also the appropriate SfM strategy, the drift of the pose estimation is compensated.
- **Integration of data from different sensors:** The endoscope reconstruction result is integrated with the photogrammetry data via the Gauss-Helmert model estimation using manually picked control points. As for the CT 3D model, if the surface representation is complete regarding different materials, the color information of photogrammetric images from the aligned photogrammetric 3D model could be directly projected onto the CT surface data. While in case of incomplete CT surface, a geometric primitive based virtual control points registration could be applied to be integrated with the photogrammetric model.
- **CSG modeling of the point cloud:** To vectorize the 3D point cloud or a mesh, a CSG modeling process is implemented via manual operation and validated for the purpose of VR/AR application. Moreover, the deep learning is also reviewed for the process with certain degree of automation.

### 1.4 Outline

This thesis will be divided into four sections as follows. Firstly, a review of the related work of our research, which includes all subtopics, is presented in chapter 2. In chapter 3, camera calibration regarding the stability and reliability analysis as well as different models for the endoscope calibration is introduced. After the camera calibration work, three 3D digitization sensors are discussed individually regarding the detailed principles, problems as well as corresponding solutions within chapter 4. Chapter 5 and 6 deal with the integration of separately reconstructed

3D models and the CSG modeling of those data respectively. In the last chapter 7, the discussion and the conclusion are presented regarding the work summary, disadvantages analysis as well as an outlook for future work.

## Chapter 2

# Related Work

This chapter will give a comprehensive review of recent development methods, which are needed to realize an integrated 3D reconstruction with photogrammetry, endoscopy and CT. Among many subtopics, camera calibration, endoscope reconstruction, 3D point cloud registration and point cloud CSG modeling will be particularly studied. With the state-of-art of all the related methodologies, we are making efforts to improve the 3D reconstruction process of each sensor, to develop an adaptive 3D point cloud registration strategy as well as to explore the way to apply the 3D model into wider use by CSG modeling.

### 2.1 Cultural heritage preservation in 3D

Cultural heritage digital preservation has been put forward for decades to avoid the potential damage that might be caused by wars, natural disasters or human negligence. Moreover, different digitization technologies could also be helpful to attract audiences' attention and ease their understanding of the extend and depth of the culture [93] with the digitized materials. The 3D digitization technologies, which are frequently used for heritage sites and objects, could be classified into passive and active sensors represented by image data and range data respectively [16, 126]. In addition, conventional survey methods could also be an option to enrich the possibilities.

#### 2.1.1 Review of photogrammetric 3D reconstruction

Among others, photogrammetric 3D reconstruction, which could recover the 3D shape of the objects in reality, has drawn the most interest. The very first project that applied photogrammetry was conducted for the documentation of buildings by Albrecht Meydenbauer, a young civil engineer, in 1858 as introduced in [3]. A recent survey, which has been carried out by the biggest 3D model and AR/VR sharing platform Sketchfab, has shown that over 60 % of the architecture and archaeological sites digitization tasks applied photogrammetry. The other popular techniques such as Light Detection and Ranging (LiDAR), laser scanning take no more than 20 % proportion. The disadvantages of LiDAR or laser scanning methods could be the incapability of capture of texture, highly reflective surfaces and the sharp-edge objects despite of the depth perception efficiency and portability characteristics. As an example, the workflow of European project "Four Dimensional Cultural Heritage World" [8] consists of data acquisition using terrestrial laser scanning, airborne and close-range cameras, 3D reconstruction, data fusion, 3D modeling as well as the VR/AR app development. As the result, the testbed Calw, hometown of Nobel Prize Winner in Literature Hermann Hesse, has been fully digitized and rendered in the "3D Calw VR" app and the app "Tracing Hermann Hesse in Calw", which has realized the goal of virtual traveling [8]. Other pilot works include the Digital Michelangelo

Project introduced in [89], the Great Buddha Project introduced in [75], the Beauvais Cathedral Project in [4] and many others. Except the conventional cultural heritage sites such as old buildings, ancient status or unearthed relics, nowadays, there is a trend of digitizing important objects from different disciplines, such as insects 3D digitization in the work [154], specimen in the work [81], historical violin digitization in the work [36] and many more.

### 2.1.2 Different sensors for 3D digital preservation

With the increase of the diversity of the digitization objectives, the choice of appropriate technology is becoming an important issue. There are various factors that could have an effect on this decision, such as the objective of the inventory, the portability, the size, complexity of the object, the accessibility of the object, inside or outside of the object where digitization occurs and the accuracy requirement [93,164]. According to these factors, appropriate methods are utilized with set technical specifications to meet the demand of the final output. Budak et al. have developed a database, which could help the users to make a decision based on all related parameters for the cultural heritage preservation including even price and customers' requirements [26]. In reality, each object is different, and each project has a different scope. Thus, naturally there is no standard 3D digitization solution that can be universally applied. The selection of the technologies is crucial for the success of the project. Generally, there are light-dependent systems and light-independent systems, which are also called active and passive sensors, for sensing the placement of the object with the light or the internal structure with the geometric principles [32]. Along with the development of the digitization sensors as well as the diverse requirements of the digitization tasks, there is a scope for additional research to enhance the methods and complete solutions for the topic of cultural heritage 3D preservation. Especially interdisciplinary applications and the data fusion are important topics to add more options.

## 2.2 Camera calibration

### 2.2.1 Taxonomy of camera calibration solutions

#### 2.2.1.1 Camera calibration model

Camera calibration has been researched for decades with various calibration models, calibration test fields as well as algorithms being put forward. In the year 1971, Brown has put forward the first analytical calibration model in his work [37], which has long-lasting impact also on the digital camera calibration research. His model with interior parameters, containing three radial distortion parameters as well as two decentering parameters has laid foundation for the following investigations, such as additional in-plane distortion by Fraser [45], zooming effect by Fraser and Al-Ajlouni in [46] and the industrial application issues [99]. These are representing works of physical camera calibration models with parameters indicating specific physical phenomenon of the lens in reality. Though the physical camera model is capable of solving many problems in practice, many other models based on a mathematical principle are also proposed with the increasing demand of more compatible modeling power for wide-angle lens, fish-eye lens or other highly distorted lenses. Representative works are models in [38, 60, 159]. Ebner and Grün proposed polynomial models by using orthogonal polynomials of second and fourth order respectively. The most interesting work with a basic math foundation is given by Tang [159]. He demonstrated that the additional parameters for self-calibration are based on the Weierstrass foundation of function approximation. Finally the Ebner and Gruen models are subsets and have got by Tang's thorough and comprehensive research a math foundation. In addition, Tang proposed Legendre and Fourier polynomials to be used also in this thesis. For a simpler though less accurate calibration, the division model [88] has been also proposed to fulfill the efficiency

requirement of some practical application. Afterwards, the model has been further improved in the work [43] and has drawn a wider attention. The investigation of the work could be found in [27, 153, 170, 180]. Besides those works, that are attempts to model the distortion explicitly or implicitly using physical or mathematical calibration models, there are also works [10, 65] considering parameter-free calibration and the former one also estimates the radial distortion center.

### 2.2.1.2 Camera calibration pattern

Camera calibration could also be classified into target or target-free calibration. The target-free camera calibration indicates the calibration that only applies feature points on the image instead of extracted target points. Generally, the self-calibration using image point correspondences rather than feature points from the calibration object is much easier regarding the implementation to mention here [42, 119, 151]. There are references that present a critical view [19] and investigate both strategies regarding the stability analysis [67]. The stability as well as the precision of the target-free calibration are considered not as reliable as the test field calibration. There are circular targets [78, 104], grid pattern or other pre-coded targets [34, 52, 53]. In the early days of camera calibration, many researchers pay a lot of attention to the calibration efficiency regarding the feature extraction process with even manual effort. To realize accurate pattern localization, different feature extraction algorithms are put forward. These works include the early Forstner operator [44], Harris operator [64], and also the more efficient SIFT algorithm [94, 98]. With the development of the related algorithms, high accuracy and fully automatic feature extraction from the calibration images is becoming possible with Zhang's method [191] using a chess board as the most popular pattern. In addition, there are also well-designed 3D test field with depth information, such as [68, 127, 155, 186]. However, the manufacturing expense of different principled 3D test fields is normally too expensive for many applications in practice.

### 2.2.1.3 Camera calibration algorithm

With an appropriate camera calibration model as well as a test field, the calibration procedure and the algorithm are important issues for a reliable calibration. Early works such as [156] has evaluated three popular calibration methods regarding the accuracy. Among many others, there are some classical works that lead the trend of the research such as [166, 191]. With collected calibration images from different perspectives, a coarse-to-fine estimation is implemented after the precise and automatic feature extraction process from the images. Initially, a pinhole model is adopted to estimate the interior and exterior parameters. Afterwards, additional distortion parameters are embedded in the model for a global optimization for the final calibration results. There is following research working on the improvement of Zhang's method from different aspects, especially the discussion of the calibration image configuration. Ricolfe-Viala et al. have conducted rigorous mathematical calculation for optimal poses for the images [132], which, however, has not been adopted in practice widely yet. Furthermore, the [131] from Richardson has introduced the idea of using optimal combination from a pose library which has been pre-defined. Peng [118] made the procedure more flexible by taking few images as initial poses and then calculate the next best poses taking the uncertainty of the extracted feature points on the image into consideration. There are also works [100, 127] which have analyzed the influence of the image configurations on the final calibration results.

In summary, though camera calibration is more or less a mature technology, there is still scope for additional research for various different types of lens distortions. With the requirement of a high accurate calibration, all aspects of the workflow including hardware and the algorithms should be fully considered.

### 2.2.2 Stability analysis of camera calibration

With the current standard method of camera calibration which has been introduced above, some casually taken images are still being used for calibration, though the disadvantage of parameter correlation exists. Normally, the calibration could still deliver results with small re-projection errors, while the stability and the reliability deserve further and deeper research. [127] provides an overall analysis about such phenomenon, including the necessity of large oblique images, roll-angle images, different distance images etc. In [66, 100, 155], further analyses and investigation have been given regarding the influence of the weak configuration image sets for the calibration. With regards to this situation, the correlation between the calibration parameters has also been discussed in [159] and [101]. In the root of Brown's work, this correlation has been mathematically analyzed with the suggested solution to decrease the influence of it. Following efforts have been made in [141]. Even in Zhang's work [191], practical advises are also provided to achieve a satisfying calibration. However, in some available open source software and work packages, guidelines are normally less serious about such influences. Instead, the re-projection error and precision assessment have attached more importance for the users, which can be misleading for high-accuracy demanding projects. The work such as [118, 131] explored the idea of next best pose estimation problem with rigorous mathematical analysis. Though theoretically images taken under such real time estimation have compensated the problem of parameter correlations, their work lacks for real images for experimental investigations. The camera calibration is a process involving many steps, which could all lead to unsatisfying results. Therefore, more practical and detailed suggestions are important for the users of such standard camera calibration implementation.

### 2.2.3 Endoscope calibration

The state-of-art of endoscope calibration research is mostly focused on the medical field. Early works such as [69, 92, 173] attached much importance on designing appropriate calibration boards normally with circular markers. In these works, one of the key process is automatic feature extraction. With the development during recent years, there are a few tools available based on Zhang's method [191] which could realize a fully automatic and accurate feature extraction such as Matlab Calibration toolbox [161] and OpenCV calibration [113]. Endoscope calibration for the operating room has been conducted in many researches as given by [9]. These works concern more about the efficiency rather than accuracy, which might turn to a least squares based closed-form solution instead of an iterative optimization process. The works in [177, 182] have taken the freedom of scope cylinder rotation into account in the calibration model. However, the implementation is mostly depending on an extra optical tracker system, which is also appeared in [86, 142]. Those works mostly focus on some specialty of the endoscope or some application limitations. The widely used Zhang's method for regular cameras has not been frequently discussed regarding endoscope calibration due to following concerns: (1) whether the model is appropriate to correct the distortion of the endoscope; (2) whether the convergent image block of endoscopic image is applicable to compensate the correlation between different correction parameters for a reliable and realistic result. Regarding the image configuration for the camera calibration, [100, 141, 188] have emphasized the importance of convergent image blocks to compensate the correlation between different correction parameters for a reliable and realistic result. In the work [120], a study on the impact of camera calibration conditions on mean reprojection error as well as the quality of following 3D reconstruction is accomplished. Yet, few works have studied the same issue for an endoscope calibration.



## 2.3 3D digitization

### 2.3.1 Photogrammetry

#### 2.3.1.1 History of photogrammetric 3D reconstruction

Photogrammetry is the technique that uses photos for mapping and surveying, which has been put forward over 150 years. The development of photogrammetry has gone through great improvement of the theory as well as the devices for complex calculations. The very first stereoscope by Wheatstone of England in 1838 and Stereocomparator by Carl Pulfrich in 1901 are representatives of the analog photogrammetry era. In the following analytical photogrammetry times, the autcartograph by Reinhard Hegershoff in 1921, modified plotter prototypes by Heinrich Wild in 1926 as well as many of optomechanical autographs, comparators and analytical plotters developed and manufactured based on Heinrich's idea have made great contribution to the development of photogrammetry [58]. However, due to the incomplete workflows and insufficient computation power, the photogrammetric 3D reconstruction of the cultural heritages has not been widely applied until the beginning of 21 century. With the development of hardware, software as well as complete concepts, the application of photogrammetry and laser scanning technology made the 3D digitization of cultural heritages possible. Nowadays, high-performance computers and digital cameras with high resolution are available applied to multi-view images to reconstruct the surface by digital photogrammetry.

#### 2.3.1.2 Procedures of multi-imagery 3D reconstruction

With regard to image pose estimation in the SfM process, it could be classified into calibrated image and uncalibrated image category for the position and orientation estimation as well as for the reconstruction of the initial sparse point cloud. In the work [6], a detailed mathematical theory has been illustrated. In practice, SfM with uncalibrated images provides easier and more convenient implementation at the cost of reduced accuracy and reliability in comparison with calibrated image SfM. Works that apply incremental SfM strategy for processing large bundle of images are introduced in [148,178]. The research [140] has reviewed the state-of-art and proposed an optimized SfM method as an open source implementation COLMAP. The sparse point cloud from the SfM, which is also the result of the pose estimation process, can not sufficiently represent the shape of the object to be reconstructed. Therefore, a dense image matching (DIM) with various algorithms [129] is necessary to be implemented for the calculation of a dense point cloud of the object.

Dense image matching is the most computation intensive process, which calculate the 3D point for every corresponding image point pair based on the estimated image poses. The algorithms for Dense Image Matching (DIM) could be divided into global [108] and local [190] based matching. As the initial step, pixelwise matching cost calculation could be calculated using different techniques [63], such as absolute differences in [110,111,171], square differences in [105,183], features based methods in [145,184], normalized cross correlation in [137,174]. After the first step, global matching methods implement a disparity refinement following the disparity calculation while the local matching method conducts a matching cost aggregation before the determination of the disparity. Moreover, the global matching algorithm has the advantage of high accuracy over local matching while the computation is very time consuming. The Semi-global Matching (SGM) algorithm of [70] has demonstrated a good balance between the global and local strategy. To further improve the performance of stereo matching process, many variants such as the Tube-based Semi-global Matching (tSGM) [133] has been put forward by applying the disparity priors to limit the search range. [39] propose a More-Global-Matching algorithm which uses a different strategy of cost aggregation for a more consistent disparities measure. In addition, nowadays, there are more and more methods based on deep learning principles being

put forward. [192] provides a comprehensive review of such stereo matching methods based on deep learning.

In the final step, the point cloud could then be converted into a triangular mesh, which could be further optimized and textured for a better visualization. These complex processes have been integrated and optimized in many open source or commercial software packages. VisualSfM [179] by Wu Changchang et al. and Bundler [147] are representatives for open source tools. Furthermore, Python Photogrammetry Toolbox (PPT) introduced in the work [14] has integrated Bundler, Clustering Multi-view Stereo (CMVS) [49] and Patch-based Multi-view Stereo (PMVS) [50] modules together for a complete pipeline of 3D reconstruction. CMVS and PMVS take the images with pose from SfM process as input and produce dense and accurate 3D colored point clouds or textured meshes. While the widely used commercial software RealityCapture [29] and Agisoft Metashape (before it was Agisoft Photoscan) have better completeness regarding the whole workflow and also easier user interaction. In addition, those commercial softwares pay also attention to the computation efficiency within the black box due to the high requirement on the computer configuration. [76] has compared the performance of the commercial software Agisoft PhotoScan Pro and an open source toolbox IGN MicMac with UAV images under sub-optimal conditions. In the work [57], four software including Erdas-LPS, EyeDEA (Univ. of Parma), Agisoft Photoscan, Pix4UAV are compared regarding the accuracy of the pose estimation and final model. More researches such as [107, 125, 128], give comparisons and assessments of those open source as well as commercial software regarding the functions, efficiency as well as the accuracy.

### 2.3.1.3 Challenges and solutions

Though the workflow has been developed into a complete one, the complete 3D reconstruction is still not a fully automatic process. As for application in practice, the 3D digitization of medium size and portable objects could be normally realized by a turn-table or a hanging imaging method. In addition, there are also projects using transparent turn-table for the image acquisition from the bottom view. However, to ensure a complete imaging perspective, the movement of the camera station or the object itself could bring challenges to the image alignment process. Because of many uncertainties, such as the object size, complexity, texture etc., there are difficulties in developing a uniform distance or the interval between images to guarantee the connection between images. Cultlab3D from Fraunhofer society has led the research on the automation of the imaging process by using the robotic arm with next best pose estimation [160]. However, the production and promotion are still on the way for a wider application with reasonable prices.

Furthermore, the 3D reconstructions of some special material surfaces is still under development due to that the situation does not go with the assumption of photogrammetry co-linearity basis. There are refraction on the transparent surface as well as non-diffuse reflection on the homogeneous surface which will lead to confusion in the image matching process. Normally spray could be used to change the attribution of the object surface without damaging it. Spray image and unsprayed image could be applied for geometry reconstruction and texturing respectively. Other possibilities are using a combination of polarized light and polarization filter to reduce the effect of the surface reflection such as indicated in work [106].

## 2.3.2 Endoscopy

### 2.3.2.1 The development of the endoscope

Endoscopy was invented by German mechanics Philipp Bozzini in the year 1806 in Mainz with the name "Lichtleiter". The application of endoscopes is mainly in the medical field for assisting the surgery implementation. And in the mid-1800s, several other scientists attempted to construct the instruments such as the American physician Fisher and the French scientist Segales [150].

The illumination of the endoscope has gone through the phase of using reflection of candlelight, burner made of alcohol and turpentine until the recent fiber technology. The endoscope design has been explored regarding the inclined view, smaller diameter while high ratio of lens and illumination fiber, higher resolution, binocular lens, as well as the flexible types [87]. Nowadays, with the development of optic and precision manufacturing technology, the endoscope plays an important role in minimum invasive surgery for its' ability in providing internal picture of the human body which is hard to observe with human eyes. In addition, there are more and more endoscope applications in industry for the purposes such as fault inspecting and diagnosing.

### 2.3.2.2 The application of the endoscope

Though endoscope is becoming more and more important in the field where is not accessible for human eyes, the demand for depth perception with the endoscopic image is urgent. The works [11, 55] have presented a overall review of the 3D techniques using endoscopes. In principle, the theory of stereo vision in photogrammetry could be directly applied to endoscope 3D reconstruction. However, similar to the situation for its' calibration, the low image quality, difficulty of taking images are the factors that limited the depth estimation using endoscopic imagery. According to the problems, different methods have been proposed, such as the stereo endoscope in the work [115]. However, the manufacture challenge as well as the much bigger entry holes have limited its wider application. Instead, there are also mono endoscope 3D reconstruction research, such as AR Widya et al. have mentioned in the work [175] using a dye on the surface of the object to be reconstructed to enhance the image information. Tokgozoglu's work [165] uses Shape-from-Shading (SFS) method, which is less influenced by the low texture information than the SfM process. Relatively there is less research concentrating on the pure endoscope SfM process. Thormählen et al. has presented the SfM implementation with an efficient outlier detector in the research [163]. Besides, Chinese scholar Jianxun Zhang has introduced in his work a weighted guide filter for endoscopic image processing method to enhance the image feature for easier feature matching process and 3D reconstruction result. Other works pay also attention to the endoscopic image pre-process such as the specular reflection removal in [109, 136]. As for the challenge of endoscopic image collection, many works such as [40, 149, 185] apply videos instead of images as input for 3D reconstruction.

### 2.3.2.3 SfM using endoscopic imagery

The classical SfM algorithm could be divided into incremental and global strategy according the initial camera pose estimation method. The incremental SfM, as has been stated in the work [71], it is effected by the long-rang drift though has advantage in robustness and accuracy. In the work [15], a review has been made concerning the evaluation of incremental SfM pipelines with ground truth dataset. Wu has introduced in his work [178] an acceleration on incremental SfM with the help of Preconditioned Conjugate Gradient (PCG). Another classical principle is global SfM, which estimates simultaneous poses for all cameras, is sensitive to the outliers. To combine the advantages and avoid the weakness of both methods, different algorithms are put forward. Cui has proposed a hybrid method in [33], which applies the global SfM for rotation estimation and estimate the camera center in a incremental way. The hybrid SfM in Li et al.'s work [91] applies a robust graph based on global SfM as initial input with a incremental SfM for remaining images. Due to the small coverage of the endoscopic images with short imaging distance than the images from a regular camera, the endoscopic 3D reconstruction of a certain surface area would require a larger number of images. Many research are focused on solving the large dataset SfM problem such as [158, 194, 195], they are mostly based on a distributed manner. For many commercial 3D reconstruction software, SfM process is integrated in the black box without necessary explanation of the implementation details. Therefore, in practice,

partial image alignment result from the SfM is common. The works such as [41, 95] could solve the problem using intermediate results for a complete reconstruction.

### 2.3.3 Computed Tomography

CT is also the non-destructive imaging technology that is widely used in medical field for obtaining both external and internal features of the human body. In comparison to the other method, CT scan can achieve small voxel size in the single digit micrometer range. Similar to the endoscope, except the application in medical field, CT has been more and more widely used in industry. The X-ray image obtained by CT indicates the information of the certain cross section of the object. With many of such X-ray images of small rotation angle, 3D model could be reconstructed. The application of CT 3D reconstruction has been discussed in the work [81], especially the limitations and the problems to be solved. First of all, as stated in the work [17], noises existed in the X-ray image, which are caused by the statistical error of low photon counts and are presented as scattered bright and dark streaks in the slices, are also effecting the CT 3D reconstruction result. The noise of the CT data can be reduced via denosing algorithms in the sinogram or the 3D CT data domain, see [74, 157].

Another problem is the threshold selection for the segmentation of different materials. When various materials exist in the object being scanned, distinguishing of them depends on the peak distribution of the histogram. However, if the difference between materials is too small or could not be extracted from the noise or the air, it is very likely to lead to the loss of the material or inadequate segmentation. Classical work in [114] has introduced peak selection method based on discriminant criterion, which was initially for 2D image segmentation task. Other works for threshold selection on histogram of CT volume values are presented in the works [176, 193].

## 2.4 3D model integration

In the work [32], the author reviewed all related sensors for 3D digitization including the practical application status. Although the possibility and the ability of 3D digitization technology is increasing, there is no single sensor which could combine all the advantages and avoid shortcomings. Therefore, sensor integration is necessary to expand the potential of 3D digitization applications.

### 2.4.1 Gauss-Helmert Model based integration

The most typical 3D sensor fusion is for photogrammetry and laser scanner with point cloud as objective data format. In practice, artificial markers, which are also called control points or targets, are applied for a least-squares based transformation estimation. The early representative work in photogrammetry field to solve the problem is presented in [61], which focused on image matching. [62] extended the solution to 3D surfaces. Later work such as [2] improved the algorithm regarding the efficiency or the registration accuracy. It should be noted that the above three references applied Gauss-markov model for the least-squares estimation, which takes only the stochastic error from target into account. In the work of [54], the authors considered the errors from both source and target point cloud separately, which could better reflect the reality.

### 2.4.2 Automated data integration

Concerning a more automated registration method, among many other algorithms, ICP [13] is the most widely used method, which calculate the transformation information by iteratively minimizing the distance between the closet points of source and target point clouds. There are many variants based on the classical ICP method such as [59, 196]. Though being the most popular algorithm, ICP and its' variants need a adequate initial registration as the basis

for the precise transformation estimation. Therefore, there are also other methods that are based on automatic 3D features extraction such as the Fast Point Feature Histogram (FPFH) [134] or Normal Aligned Radial Feature (NARF) [135] etc. The principle is to extract the key points according to the local or global characteristics of the point cloud. The correspondences are estimated according to the key points with their descriptors for the transformation matrix determination. The effectiveness of such methods rely on the point cloud quality as well as the similarity of both point clouds to be registered. Due to the reason that the 3D data of artificial objects contains rich regular geometric primitives, there are also lots of works taken the geometric primitive information within the point cloud into account. [5] utilizes lines rather than point for an iterative estimation. In the work [152], the authors calculates the intersection lines of neighboring planes and estimate the transformation matrix with at least two corresponding line pairs. [162] put forward a terrestrial point clouds registration method by using virtual tie points from the intersection of planar surfaces. One drawback could be that the descriptor of the control point generated by the primitives is difficult to design without neighboring points.

## 2.5 Deep learning and RANSAC-based CSG modeling

In recent years, 3D reconstruction using multi-view imagery, 3D scanning or other principles has been widely used in practice for the acquisition of point clouds of the reality, which are then used for reconstructing the surface [12] by algorithms such as poisson reconstruction [80] and many other possibilities. However, the large data volume, the noise as well as the gross error of point clouds or meshes are crucial factors that may limit the high level manipulation or user interaction without structural information. Therefore, it is necessary to converting the point clouds or meshes to other appropriate format such as low-poly models [123] or CSG model, which could largely reduce the data size as well as eliminate the influence of noise or data incompleteness [138, 146].

### 2.5.1 RANSAC-based CSG modeling

The manual process of converting the point cloud into CSG model needs a lots of extensive skills and intervention, which improves the application cost as well as reduces the accuracy when trying to fit the point cloud with too many details. Two widely used principles are RANSAC and hough transformation [73]. With regard to RANSAC-based principle, it has been the gold standard for a long time since the early work introduced in [85, 102]. In recent and more robust work such as [139], basic geometric primitives such as planes, cylinders, toris, spheres etc. are extracted precisely from the input point cloud by an efficient means of sampling the points for fitting and evaluating scores based on locality sensitive methods. It has also been introduced in the work [84] the process of creating CSG models from point cloud of large-scale city, which concentrated more on the urban elements, especially the reconstruction of the building. As the commonly used 3D representations applied in BIM (Building Information System), the work [117] provides an overview of the process focusing on geometric modeling from point cloud.

### 2.5.2 Deep learning-based CSG modeling

Though being regarded as the standard solution, RANSAC-based methods have disadvantages such as the large space of parameter as well as the processing of large data-sets with diverse geometric primitives. Therefore, researchers have been exploring deep learning based methods for solving the problems. As the pioneer work, [121] has applied permutation invariant feature learning and multi-scale feature aggregation for the purpose of 3D classification and segmentation. And [172] is designed for the same purpose via a neural network acting on graphs dynamically computed in each layer of the network. To adopt the idea of RANSAC with deep

learning, [23] proposed a pipeline of solving the non-differentiable properties by learning an average hypotheses. In [122], the deep networks are applied to estimate the weight of homogeneous least-squares problems for geometric model fitting. Preliminary work such as [31, 35] generate compact polygon meshes which could be further processed for the generation of CSG model. [22] introduced a method for point cloud labeling, which puts more focus on the understanding of the 3D scenes. Some works aim for CSG program using deep learning with voxels [143, 168] or range image [197], which are both effective while can only serve as a rough abstraction of the input data. [90] is a representative work applying deep learning algorithm for fitting primitive to point cloud by an initial prediction of per-point properties following with a differential model estimation module for the related parameters.

## Chapter 3

# Camera Calibration

As for every camera sensor, it is essential to implement the calibration before applying to close-range photogrammetric measurements. The assumption of using the images for metric measurements is that the camera model is fully known and normally based on the ideal design, i.e. the pinhole projection or other standard models. However, the ideal design is impractical due to the manufacturing limitations of the optics in reality. Therefore, accurate intrinsic camera parameters through the calibration process are of vital importance when utilizing multi-view imagery for applications such as 3D reconstructions for the purpose of cultural heritage preservation. For normal scenarios, there are even open source or free software packages for the camera calibration, which are mostly based on linear approaches with simplified camera models such as Tsai's method [167] and Zhang's method [191]. However, due to the convenience and accessibility of such calibration solutions, the stability and the reliability of camera calibration are largely ignored. In Section 3.2, we will analyze the stability of the standard camera calibration method through statistical behavior of multiple selected image subsets calibrations from large number of images. Section 3.3 deals with the calibrations of an endoscope. Different calibration models are compared through the distances between the extracted feature points on the undistorted image and the straight line connecting the starting and ending point of each line.

### 3.1 Basic Concepts

#### 3.1.1 Collinearity equations

For the relationships between the 2D and 3D world in Euclidean geometry, different coordinate systems are defined (see Appendix A2). Initially, the camera model is assumed as an ideal mathematical model of the central perspective without distortion, which complies with a Direct Linear Transformation (DLT) as shown in (3.1).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dX & 0 & x_0 \\ 0 & 1/dY & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.1)$$

where

- $X, Y, Z$  are the coordinates of an object point in the world coordinates;
- $\mathbf{R}$  and  $\mathbf{t} = [X_0, Y_0, Z_0]^T$  are the rotation and translation respectively between the world coordinate system and the camera coordinate system;
- $dX$  and  $dY$  are the scale factors in the image frame;
- $x$  and  $y$  are the coordinates of an image point in the camera coordinate system;
- $x_0$  and  $y_0$  are the coordinates of the principle point and  $f$  is the focal length.

When extending the ideal model with distortion representing the reality of the camera, it could be written into the collinearity equations (see Appendix A2) as (3.2 and 3.3).

$$x = x_0 - f \cdot \frac{r_{11} \cdot (X - X_0) + r_{21} \cdot (Y - Y_0) + r_{31} \cdot (Y - Y_0)}{r_{13} \cdot (X - X_0) + r_{23} \cdot (Y - Y_0) + r_{33} \cdot (Y - Y_0)} + x_{distorted} \quad (3.2)$$

$$y = y_0 - f \cdot \frac{r_{12} \cdot (X - X_0) + r_{22} \cdot (Y - Y_0) + r_{32} \cdot (Y - Y_0)}{r_{13} \cdot (X - X_0) + r_{23} \cdot (Y - Y_0) + r_{33} \cdot (Y - Y_0)} + y_{distorted} \quad (3.3)$$

Here  $x_{distorted}$  and  $y_{distorted}$  are the correction terms for the image coordinates,  $r_{ij}$  are the components of the rotation matrix  $\mathbf{R}$ .

### 3.1.2 Camera calibration models

As for lenses with different kinds of distortion, additional distortion terms  $x_{distorted}$  and  $y_{distorted}$ , represented by parametric models according to the characteristics of different lenses, could be added. In this section, the division model, Brown model as well as examples of calibration models that are based on the mathematical principle are introduced.

**The Standard Brown Model** For single camera calibration, the toolbox of Matlab [161] is applied widely for its convenience and friendly evaluation features. Here the distortion modeling of the camera is mainly based on Brown's [25] method with 3 radial distortion parameters and 2 tangential parameters (optional) as indicated in (3.4) .

$$\begin{aligned} x_{distorted} &= x \cdot (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + [2 \cdot p_1 \cdot x \cdot y + p_2 \cdot (r^2 + 2 \cdot x^2)] \\ y_{distorted} &= y \cdot (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + [2 \cdot p_1 \cdot (r^2 + 2 \cdot y^2) + 2 \cdot p_2 \cdot x \cdot y] \end{aligned} \quad (3.4)$$

With  $k_1, k_2, k_3$  being the radial distortion parameters,  $p_1, p_2$  are the tangential parameters, and  $r^2 = x^2 + y^2$ . In addition, the skew could also be estimated during the calibration process.

Despite the fact that Zhang's method [191] has a mature algorithm and practical application foundation, its parameter correlations for the calibration model need to be compensated by a well convergent image block. Generally, for most lenses applied in industrial photogrammetry or close-range photogrammetry, the requirements could be fulfilled with adequate caution. However, due to the ongoing emergence of new camera systems nowadays, it brings more challenges for the practical implementation. In addition, except the systematic errors covered in this standard model, there are also other types mentioned in [100] such as chromatic aberration, image-variant changes in the principle point and sensor array unflatness to be compensated, where extended models may be necessary.

**The Division Model** is another classical calibration model, which utilizes less parameters than the Brown model to compensate a large distortion as indicated in (3.5). In many cases, only one parameter is applied for modeling the camera distortion with few calibration images.

$$\begin{aligned} x_{distorted} &= \frac{x}{1 + \lambda_1 \cdot r^2 + \lambda_2 \cdot r^4 + \dots} \\ y_{distorted} &= \frac{y}{1 + \lambda_1 \cdot r^2 + \lambda_2 \cdot r^4 + \dots} \end{aligned} \quad (3.5)$$

Here  $\lambda_1, \lambda_2$  are the distortion parameters. Considering the pros and cons of the division model, it would be preferable in case of efficiency demands while considering less accuracy requirements.

**The Mathematical self-calibration models** Mathematical self-calibration models are based on the principle of function approximation with additional parameters that have no specific interpretation regarding the distortion behavior. General introduction as well as two specific camera calibration models, which are based on polynomial approximations and the Fourier theorem [159] are given in Appendix A3.

The  $n$ -th order Fourier model is defined as  $M = N = n$ . Generally, the more complicated the distortion is, the higher orders are needed to be applied for the approximation. However,



there are also over-parameterization and under-parameterization issues to be considered when choosing the order of the model for different level of distortions. Over-parameterization should be taken into account when high orders are used, while under-parameterization implies that the additional parameters are not sufficient to compensate the distortion. In practice, this mathematical model has only been verified in the aerial photogrammetry field. Therefore, the application of this model for other camera systems should be further investigated.

The Legendre self-calibration model is another model developed in [159] based on the algebraic polynomials using two variables multiplied by coefficients. Due to the parameter correlations, some terms of the original complete polynomials must be eliminated. Ebner [38] and Gruen [60] models are earlier good examples of classical mathematical calibration models, but with the similar issue as the Legendre self-calibration model regarding the correlations. Theoretically, the Fourier self-calibration has advantage over other mathematical calibration models. The reason is that the  $\Delta x$  and  $\Delta y$  are independent and there is no need for eliminating any APs (additional parameters). On the one hand, it is more rigorous regarding the completeness of the model, on the other hand there are less strict requirements for calibration of any image configuration. Therefore, among many mathematical self-calibration models, only the Fourier self-calibration of different orders are experimented for comparison.

### 3.1.3 Experimental cameras

We conduct the experiment of camera calibration stability analysis with two mirrorless cameras, a Sony  $\alpha$  7R III with Zeiss Loxia 25mm lens, and a Leica Q (Typ 116) camera with Leica Summilux 28mm lens (see Figure 3.1). Besides, the potential of two smartphone cameras are investigated: a Galaxy Note 8 from Samsung and an iPhone 7 Plus from Apple. The related specifications for the cameras are listed in Table 3.1. In this work, all experimental results are based on a special manufactured  $18 \times 29$  chess-board by the company calib.io I/S (see Figure 3.2).



Figure 3.1: Camera systems used for calibration. From left to right: Sony  $\alpha$  7R III, Zeiss Loxia 25mm lens and Leica Q (Typ 116) (pictures are from respective official websites) [188].

Table 3.1: Specifications of camera systems used for calibration

	Camera type	Focal length(mm)	Sensor type
	Sony $\alpha$ 7R III	25 (engraved value)	35mm (Full frame)
	Leica Q	28 (engraved value)	35mm (Full frame)
Galaxy Note 8	Wide-angle lens	-	1/2.55"
	Extra telephoto lens	-	1/3.6"
iPhone 7 Plus	Wide-angle lens	-	1/3"
	Extra telephoto lens	-	1/3.6"

The experiment of reliability analysis has also been tested with the images captured with an endoscope model from the Endo Industrial company with specifications indicated in Table 3.2 and an image in Figure 3.3. Among other parameters, the diameter of the endoscope makes a big difference for the imaging ability. According to our research, most of the work applied using the endoscopes is based on a diameter around or bigger than 4mm, such as 3.8mm and 10mm diameter laparoscopes in [173], 4mm diameter arthroscope in [103]. According to Leiner, Dennis C in [87], the diameter of the endoscope has great influence on the image quality as well as the manufacturing process. The small diameter will lead to less space for the illumination fibers and limit the f-number of the optical system, which will directly influence the image brightness. Moreover, the oblique view characteristic will further restrict the space design.

Table 3.2: Features of the rigid endoscope

	Rigid Endoscope
Sensor size	1/3"
Image resolution	752 * 576
Pixel size	8.5 $\mu$ m
Diameter	2.7mm
Distance from the object	<2cm
Direction of view	30 °
Opening angle	From 50 °to 105 °
Length	140mm

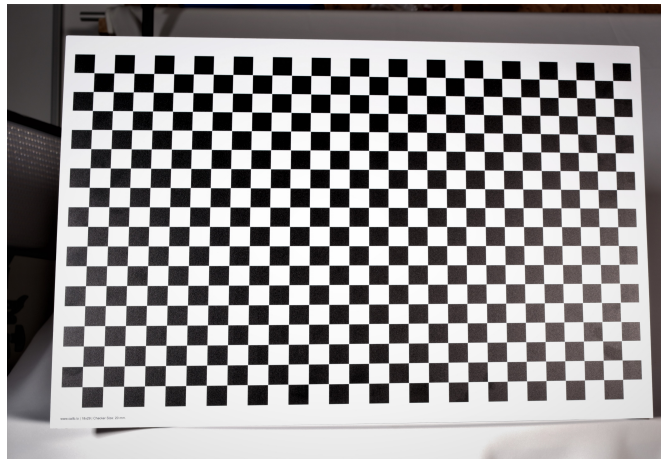


Figure 3.2: Camera calibration board (calib.io I/S).

Since the image capturing process with an endoscope is easily influenced by the hand tremor of the user, a self-designed setup using three orthogonal cross slides and an additional head ball as shown in Figure 3.4 is applied to guarantee the imaging stability. By twisting the knob on each slide, the endoscope could move accordingly with small steps ensuring adequate overlap between neighboring images.

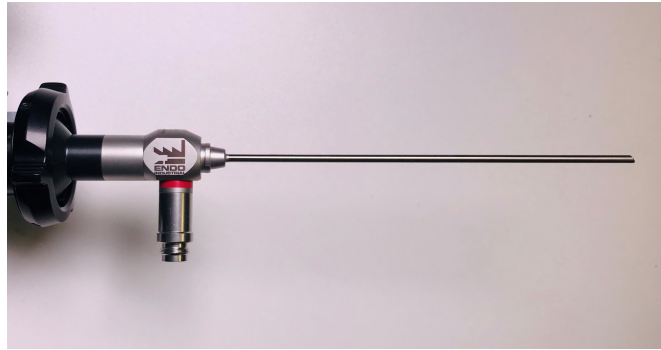


Figure 3.3: Example image of the endoscope.



Figure 3.4: Self-designed endoscope imaging setup.

## 3.2 Stability analysis of camera calibration

### 3.2.1 Methods

#### 3.2.1.1 Workflow

As the first step, ground truth estimation of the intrinsic parameters are to be estimated by considering the statistical behavior of the calibrated parameters with regard to random errors. According to the implementation instruction of OpenCV or Matlab Calibration Toolbox [20], 10 to 20 images from different perspectives are collected for a calibration. However, in reality, images taken without enough considerations of the configuration will lead to unstable intrinsic parameters of the same camera. If some images are ill-conditioned, they may severely influence the final calibration results. Therefore, in our experiment, far more than the required number of images are taken, with different orientations including slightly tilted or highly tilted images, rotated images, and a mix of images with different distances. In case of ground truth estimation of intrinsic parameters, the ill-conditioned images most probably will have an impact on the confidence interval of the Gaussian fitted functions. We use the Maximum Likelihood Estimation (MLE) method to calculate the expected values of the calibration parameters. The workflow for the calibration is given in Figure 3.5.

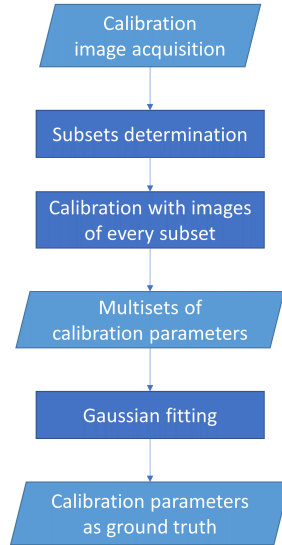


Figure 3.5: Workflow of camera calibration for stability analysis.

### 3.2.1.2 Central Limit Theorem

The general idea of the Central Limit Theorem (CLT) [24] is, that regardless of the population distribution model, if the sample size increases, the sample mean tends to be normally distributed around the population mean, and its standard deviation shrinks as the number of samples  $n$  increases. The theorem holds on condition that 1) the sample size must be independent and 2) the sample size must be “big enough”. The CLT is stated in (3.6).

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1) \quad (3.6)$$

where  $\bar{X}$  stands for the sample mean,  $\mu$  and  $\sigma$  are the nature mean and standard deviation of the normal distribution.

Since the extraction uncertainty of each image feature is independent, different image orientations are considered to be independent as well. Therefore, the final calibration result should fulfill the CLT, which means the parameters from succeeding calibrations should fulfill a normal distribution. More specifically, when taking a big number of images following the guidance to assume that all the images provide useful information for the calibration process, a sufficient number of random subsets can be chosen. It is assumed that also the unstable calibration parameters should present a normal random error distribution behavior.

## 3.2.2 Experiments and results

### 3.2.2.1 Gaussian fitting for multisets of calibrations

In the experiments, different numbers of images for the subsets are tested. For a certain subset image number, 1000 subsets are chosen and thus 1000 calibrations are implemented. Therefore, every calibration parameter has 1000 results from different subsets. As stated in 3.2.1.1, a MLE is applied to obtain the expected values (in unit pixels). The respective pixel size of each camera is listed in Table 3.3, which can be used to convert the given results to metric units. Additionally, for non-full-frame sensors, the crop factor related to the sensor size is needed for the conversion to a 35mm equivalent value. The results of the focal length in X direction are displayed in Figures 3.6.

Table 3.3: Pixel size of the four different cameras

Camera Type	Sony $\alpha$ 7R III	Leica Q	Galaxy Note 8	iPhone 7 Plus
Pixel size ( $\mu\text{m}$ )	4.5	5.97	1.4	1.2
Crop factor	1	1	6	7.2

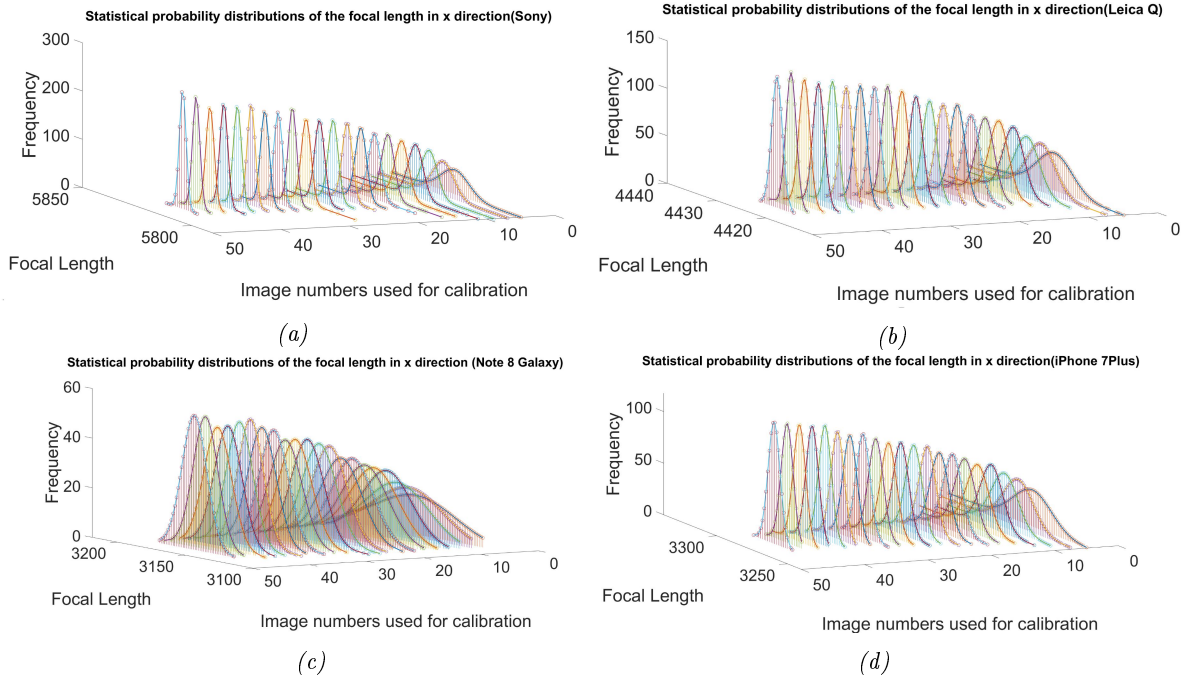


Figure 3.6: Gaussian fitting experiment results: the calibrated focal length (pixel) in x direction for (a) Sony  $\alpha$  7R III with Zeiss Loxia 25mm lens; (b) Leica Q (Typ 116) camera with Leica Summilux 28mm lens; (c) the Galaxy Note 8; (d) the iPhone 7 Plus.

Figure 3.6 demonstrates that with different image subsets, the natural mean of the focal length from all the calibrations will converge. Moreover, the Gaussian fitting results indicate that more images for a subset will make a more stable calibration, though more than 20 images do not show a significant improvement. From the comparison of the four curves of Figure 3.7, the standard deviations of all camera calibrations converge to small values, while camera Sony  $\alpha$  7R III and Leica Q (Typ 116) cameras show a much better stability than the smartphone cameras of the Galaxy Note 8 and iPhone 7 Plus.

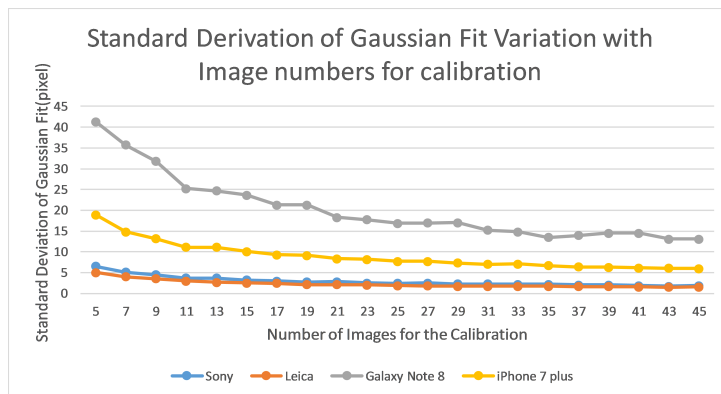


Figure 3.7: Gaussian fitting standard deviations of the calibration for four types of cameras.

From the above processing, for different image numbers of subsets we obtain different Gaussian fitting curves with different expected values and standard deviations. Examples of the calibrated focal length in x direction and y direction for the four different cameras' calibration are displayed in Figure 3.8.

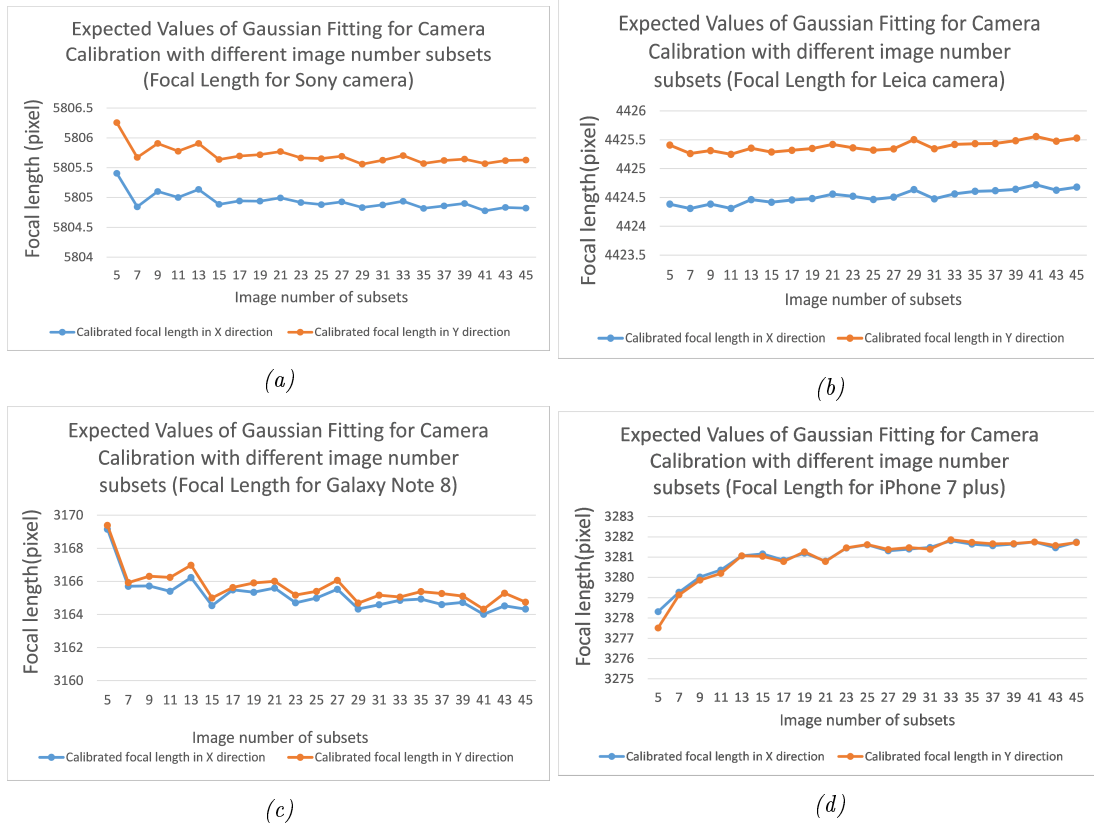


Figure 3.8: Gaussian fitting standard deviations (SD) of the calibrated focal length: (a) SD of the Sony camera; (b) the Leica camera; (c) Galaxy Note 8; (d) iPhone 7 plus.

The outcome of the experiments above is that the calibrated parameters are stable and within small range for the calibrated focal length and the other calibration parameters. This means that the influence of the image configuration is greatly reduced by using many subsets for the calibration from a huge image bundle and then estimate the natural mean according to the CLT. A weighted arithmetic mean for the expected value is determined as the final calibration parameter in (3.7).

$$X = \frac{\sum_{i=1}^n x_i / \sigma_i^2}{\sum_{i=1}^n 1 / \sigma_i^2} \quad (3.7)$$

with  $X$  as ground truth of the calibration parameter,  $x$  is the expected value of the  $i^{th}$  Gaussian fitting curve, and  $\sigma_i$  is the corresponding standard deviation of the normal distribution. Table 3.4 lists the results of the individual calibration parameters.

The values of the calibration parameters from Table 3.4 may differ from the specifications of the vendor when transferred to mm. A reason can be the difference between the design value and the production tolerance of the product. However, the concept introduced in this thesis is not suggesting readers to implement this workflow for every camera calibration, instead, the method is applied here for the estimation of a ground truth information in order to analyse the single factor influence in Section 3.2.2.2, so that practical suggestions can be given for a stable calibration with around 20 images.

Table 3.4: Calibration parameters for Sony  $\alpha$  7R III, Leica Q, Galaxy Note 8 and iPhone 7 Plus

Calibration parameters	Estimated Values			
	Sony $\alpha$ 7R III	Leica Q	Galaxy Note 8	iPhone 7 Plus
Focal length $f_x$ (pixels)	5819.6096	4424.5633	3164.8576	3281.3755
Focal length $f_y$ (pixels)	5819.4368	4425.4199	3165.3254	3281.3854
Principal point $x_0$ (pixels)	-32.4949	3.2504	27.8190	-24.6529
Principal point $y_0$ (pixels)	-1.4404	2.5041	13.6578	22.9677
Radial distortion $K_1$	-0.0566	0.0479	0.1566	0.1951

### 3.2.2.2 Image block analysis based on ground truth estimation

After having estimated the intrinsic parameters through the Gaussian Fitting method with the whole image dataset, the composition of the images used for calibration is changed with regard to different factors for further comparison. Oblique imaging angles and roll angles (rotations along the optical axis) are important to make a strong image block for the camera calibration. Among the whole image dataset, a group of photos is taken with relatively small oblique angles, which follows the typical description of camera calibration in most open-source calibration toolboxes. Additional images with bigger oblique angles and roll angles are taken, which compose a more convergent image block to optimize the calibration configurations. Thus, a control experiment with and without big oblique view images and roll movement images is designed to validate the suggestion for taking images. In this section, only the result of the Sony camera experiments is presented in Figure 3.9. From Figure 3.9, the following conclusions can be made:

- big oblique view images are important for a reliable and stable camera calibration;
- weak image blocks without big oblique image and roll angle images can lead to unstable calibration results and even deviate from estimated ground truth with more images;
- roll angle images can contribute to more stable calibration results;
- roll angle images affects more the focal length estimation than the coordinates of the principle point.

Therefore, for non-expert users, it is of vital importance to pay more attention to the image configuration for camera calibration regarding oblique view angles as well as roll angles particularly.

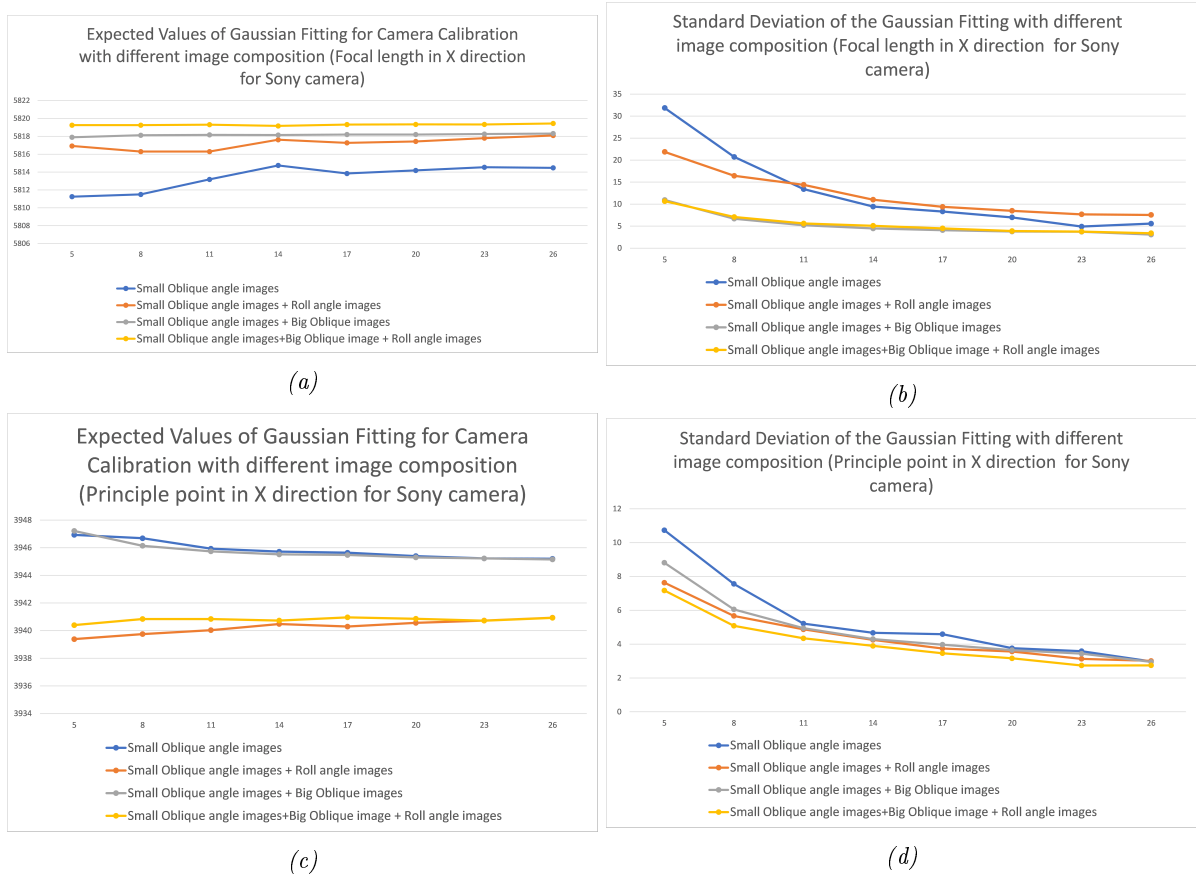


Figure 3.9: Single factor analysis results of Gaussian fitting using different image compositions: (a) expected value for the calibrated focal length in X direction; (b) standard deviations of the calibrated focal length; (c) expected value for the principle point  $x_0$  coordinate; (d) standard deviations of the principle point  $x_0$  coordinate.

### 3.2.2.3 Temporal stability variation

We have learned that the above factors have a strong influence on the calibration. Therefore, we need to pay attention to the obtained results by taking approximate images. For validation, the experiments with the Sony camera are shown as an example. Twenty independent sets of images, for which each consists of 15-25 images, are collected within one week at different epochs for camera calibration. Ten sets of images follow the typical description of camera calibration in most open-source calibration toolboxes, while another ten sets of images with the big oblique angle and roll angle images supplement the validation, so that the temporal stability of the proposed method is finally proven. Part of the resulting calibration parameters including the calibrated focal length and the principal point coordinates is shown in Figure 3.10.



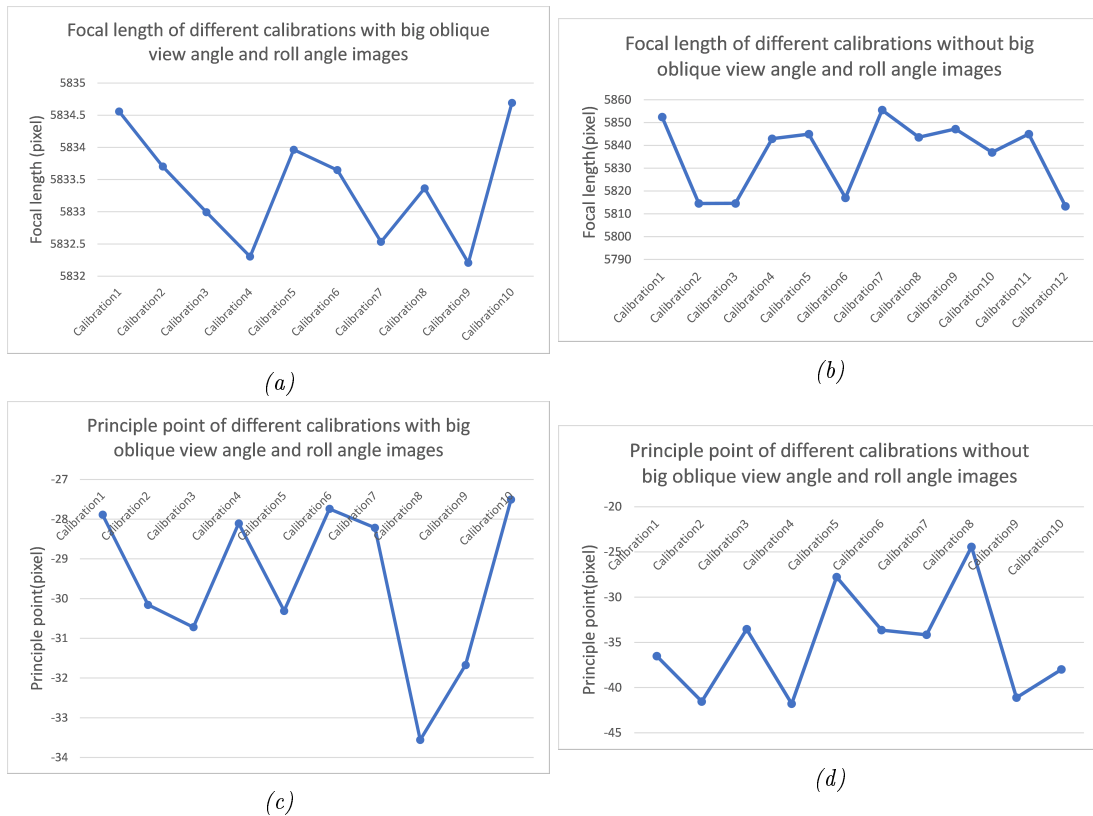


Figure 3.10: Temporal stability variation experiment results: (a) focal length of different calibrations with strong image block configuration; (b) focal length of different calibrations without strong image block configuration; (c) principle point of different calibrations with strong image block configuration; (d) principle point in X direction of different calibrations without strong image block configuration.

According to the figures, it can be seen that for the image datasets with big oblique angle images and roll angle images, the differences of the calibrated focal length or the principal point calibration of different calibrations are within 6 pixels. However, image datasets without strong image block configurations indicate high instability with calibrated parameter deviations of up to 40 pixels. In addition, it must be noted that the experimental conditions in this section are deviating from Section 3.2.2.1 regarding focus setting and other changes, which might cause the difference of the calibrated parameters between section 3.2.2.3 and previous experiments.

### 3.3 Endoscope calibration

In Section 3.2, the calibration experiment is using the Brown model integrated in Zhang's method, which is under an assumption that the Brown model is capable to model the lens distortion behaviors. However, regarding special lenses such as wide-angle lenses, long focal length lenses, more adaptive calibration models should be adopted for the reliability of the calibration. Within this section, one of the special optics endoscope is under investigation.

#### 3.3.1 Methods

##### 3.3.1.1 Endoscope structure

Endoscopes, as special optical structure, are normally divided into flexible endoscopes and rigid endoscopes. In our work, the latter one is applied. A new degree of freedom, the rotation between the scope cylinder and the camera head is introduced to ensure a wider viewing field. However, we keep the unchanged structure between the scope cylinder and the camera head for

a more stable intrinsic information for the purpose of metric information obtainment. Another design for wider viewing field is the oblique scope in the tip as shown in Figure 3.11.

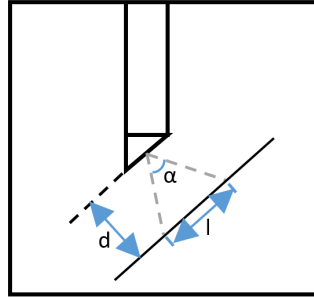


Figure 3.11: Oblique view indication.

Suppose the imaging distance is  $d$ , the angle of view is  $\alpha$ , the coverage of the object space can be calculated via (3.8). It could be easily concluded that a short imaging distance leads to a small viewing coverage in object space.

$$l = 2 \cdot d \cdot \tan(\alpha/2) \quad (3.8)$$

Normally, the diameter of the scope cylinder is designed in the range of millimeter level. The small aperture though enables the access to small cavities, it also introduce challenges to the imaging transmission system as well as lighting system construction. Furthermore, the periphery of the endoscopic image has normally severe distortions. In summary, an endoscope has several disadvantages over a regular camera regarding 3D reconstructions:

- severe lens distortions;
- low image quality;
- unstability of the optics;
- unstability of the image collection process.

Therefore, though the camera calibration process can rely on mature algorithms and toolboxes, it is still an challenging task to accomplish an endoscope calibration.

### 3.3.1.2 Necessity of pre-calibration using different models

Due to the compound lenses and the imperfection of the manufacture, there is a discrepancy between the real imaging process and the pinhole imaging in case of ideal lenses. Therefore, various calibration models based on different principles are put forward to model the distortion behavior of the lens. The common distortion type as described in Section 3.1.2 are radial distortion and tangential distortion, which could be well modelled by the Brown model. The mature algorithm, including an initial closed-form solution under the assumption of an ideal lens and global optimization process, has been widely applied. However, with the development of the optics, many different lenses for specific purposes are designed with distortions that are beyond the ability of common calibration algorithms. The main problems are in two aspects:

- the calibration model is not powerful enough to model the lens distortion, such as wide-angle lens or fish-eye lens, therefore more powerful calibration models or specific models should be developed for this task;
- the high correlation between parameters of the standard calibration algorithm will together with inappropriate image configurations have an influence on the mathematical stability.

In addition, the difference between pre-calibration and the calibration in the SfM process are also worth consideration. In the SfM process, camera calibration is calculated as part of the projective matrix of each single image. Even though normally more feature points are applied than in a pre-calibration, the SfM based camera calibration is neither metrically precise, scene independent nor image invariant. Therefore, the analysis of the results would be difficult for the optimization of the process.

### 3.3.1.3 Camera calibration for 3D reconstruction

As one of the steps for 3D reconstruction, camera calibration could be determined by auto-calibration [6] directly from multiple uncalibrated photogrammetric images or implemented beforehand. In case of auto-calibration, with the well overlapped images for 3D reconstruction, matched feature points lead to a projective reconstruction, for which several methods are available to obtain the intrinsic parameters. The projective reconstruction could be decomposed as  $P^i = K^i [R^i | \mathbf{t}^i]$  with  $K^i, R^i, \mathbf{t}^i$  representing the camera calibration matrix, rotation and translation, respectively. In the case of pre-calibration, the calibration parameters could be applied to obtain the normalized image coordinates of matched features for 3D reconstruction. For a 3D reconstruction, the fundamental matrix as shown in 3.9 is then degraded to the essential matrix as shown in (3.10) for the determination of image poses.

$$\mathbf{x}^T \mathbf{F} \mathbf{x}' = 0 \quad (3.9)$$

$$\hat{\mathbf{x}}'^T \mathbf{E} \hat{\mathbf{x}} = 0 \quad (3.10)$$

where  $\mathbf{x}$  and  $\mathbf{x}'$  are the corresponding image coordinates,  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{x}}'$  are the corresponding normalized image coordinates,  $\mathbf{F}$  and  $\mathbf{E}$  represent the fundamental and essential matrix, respectively.  $\mathbf{F}$  and  $\mathbf{E}$  are connected via a camera matrix.

Though the auto-calibration could offer a complete solution to metric reconstruction under appropriate circumstances, it will fail according to [6] if the motion of the camera is too restricted, which is very likely to happen in case of endoscope imaging. Thus, for endoscope metric reconstruction, a pre-calibration has definite advantages over relying on auto-calibration during the 3D reconstruction.

## 3.3.2 Experiments and results

With the introduction of endoscopy above, an example of an endoscope calibration image is shown in Figure 3.12. As can be seen from the image, the periphery shows severe distortion in comparison to the central part. In addition, the brightness and the resolution is lower than that of a regular camera. These are all the challenges to be dealt with in the following section.

### 3.3.2.1 Stability experiment of the endoscope

According to the implementation of stability analysis introduced in section 3.2, an example of Gaussian fitting for focal length in X direction is shown in Figure 3.13. Over 800 endoscopic images were collected, the number of subsets were defined from 5-50 images and the random subset selection was implemented 1000 times. Here the times of subset selection is decided taking consideration of the computation time and the significance of the statistical distribution.

A more intuitive view of the expected values of the focal length and the standard derivation of the Gaussian fitting process are shown in Figure 3.14 and Figure 3.15, respectively. For a comparison, an example of Gaussian fitting result for the regular camera Sony  $\alpha$  7R III can be found in Section 3.2.2.1. From these results, it indicates that the convergence of the intrinsic parameters shows more stability than the regular camera in [188], numerically in units of pixel. However, due to the imaging distance and configuration as well as the pixel size, the derivation

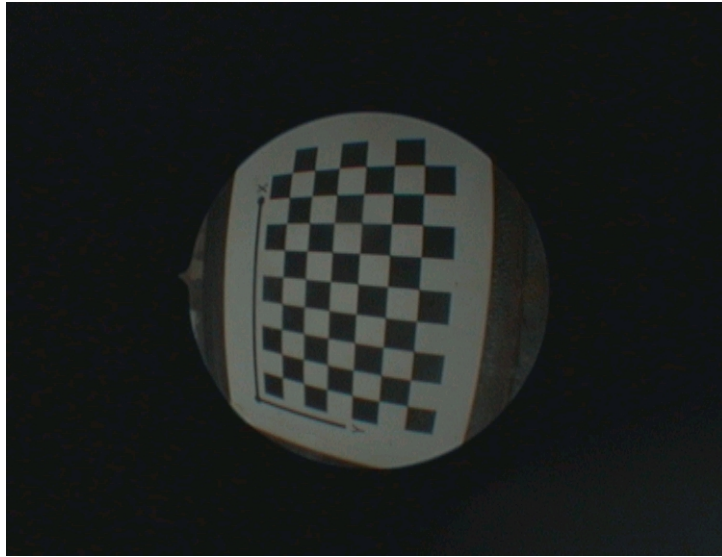


Figure 3.12: Example image of endoscope calibration.

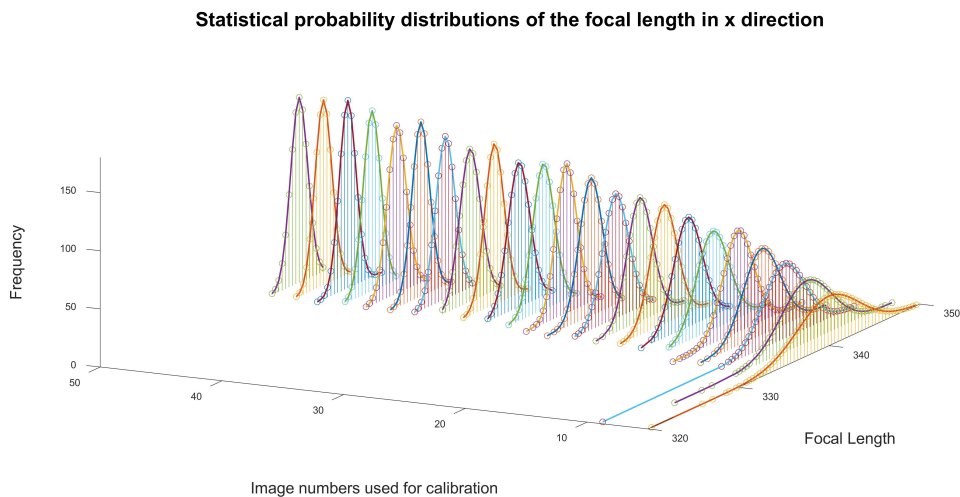


Figure 3.13: Example image of focal length Gaussian fitting from multiple endoscope calibration.

from endoscope calibration contributes a much bigger influence than the regular camera. Take the Sony  $\alpha$  7R III and the endoscope as example, the corresponding parameters are shown in Table 3.5.

According to the above values, it could be seen that the standard deviations in metric space are nearly the same for both cameras. However, the ratio of the imaging distance to the focal length of the regular camera is much bigger than that of the endoscope as illustrated in Figure 3.16. Even though the standard deviation of the endoscopic calibration by a Gaussian fitting process is smaller than that of the Sony  $\alpha$  7R III, the instability of the endoscope has much bigger effect on the application obtaining metric information. Therefore, the result of endoscope calibration stability analysis and the comparison with Sony  $\alpha$  7R III camera indicate that it is of vital importance for the endoscope to achieve accurate calibration for metric information determination.

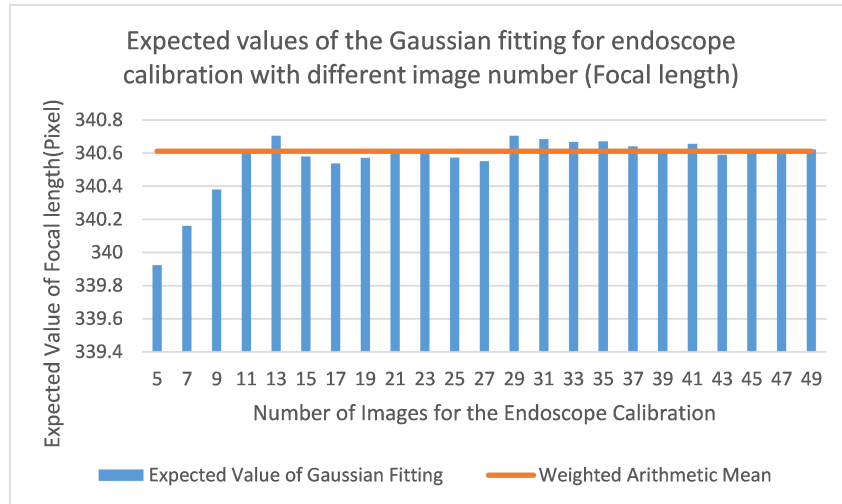


Figure 3.14: Expected values of the Gaussian fitting for endoscope calibrated focal length with different image number.

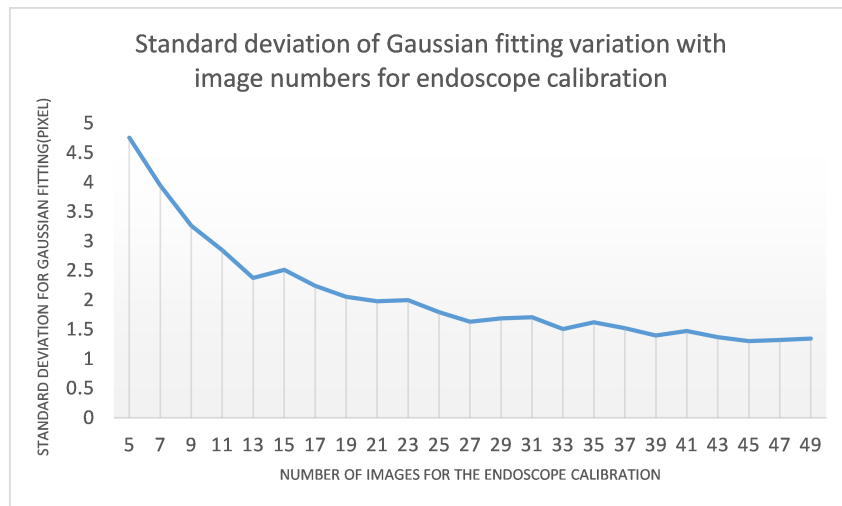


Figure 3.15: Standard deviation of Gaussian fitting variation with image numbers for the endoscope calibration.

### 3.3.2.2 Endoscope calibration model comparison

With the stability and reliability analysis from section 3.2, we captured a set of well-configured images with regard to the reference from the Gaussian fitting process for the purpose of a comparison with different camera models. Among them, introduction of Fourier mode could be found in Appendix A3. The applied camera models are:

- Model 1: Division calibration model with single parameter
- Model 2: Classical Brown model with 3 radial distortion parameters and 2 tangential parameters
- Model 3: Fourier model of first order with 16 distortion parameters
- Model 4: Fourier model between first and second order with 32 distortion parameters
- Model 5: Fourier model of second order with 48 distortion parameters
- Model 6: Fourier model of third order with 96 distortion parameters
- Model 7: Fourier model of fourth order with 160 distortion parameters
- Model 8: Combination of Brown model and Fourier model of first order

Table 3.5: Parameters for Sony  $\alpha$  7R III and the endoscope

	Sony $\alpha$ 7R III	Endoscope
Resolution	7952 * 5762	752 * 576
Sensor size	36mm * 24mm	6.4mm * 4.8mm
Pixel size	4.5 $\mu\text{m}$	8.5 $\mu\text{m}$
Standard deviation	4 pixels	2 pixels
Imaging distance	>50cm	<1cm

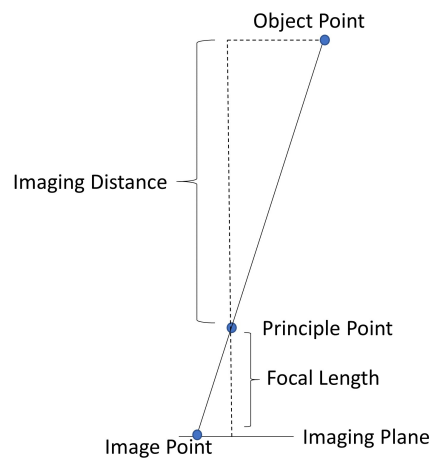


Figure 3.16: Illustration of imaging distance and the focal length.

- Model 9: Combination of Brown model and Fourier model between first and second order
- Model 10: Combination of Brown model and Fourier model of second order
- Model 11: Combination of Brown model and Fourier model of third order
- Model 12: Combination of Brown model and Fourier model of fourth order

It should be noted that the Fourier self-calibration model with 32 parameters is the intermediate model between the first and the second model, which has been explicitly indicated in Tang's work [159]. The mathematical foundations of the applied models could be found in section 3.1.2 and in Appendix A3. The realization of calibration with different models is based on [21] as well as the tutorial from [28]. An example of undistorted images with the results of the Brown model calibration can be seen in Figure 3.17. The peripheral of (b) and (d) from Figure 3.17 is due to the multiple solution of undistorted coordinates from the Brown model as well as the relationship between the sensor area and the imaging area.

In addition, to evaluate the effectiveness of radial distortion estimation from different models, the comparison between the straight line and the line on the image is plotted with a zooming view on the undistorted image as Figure 3.18.

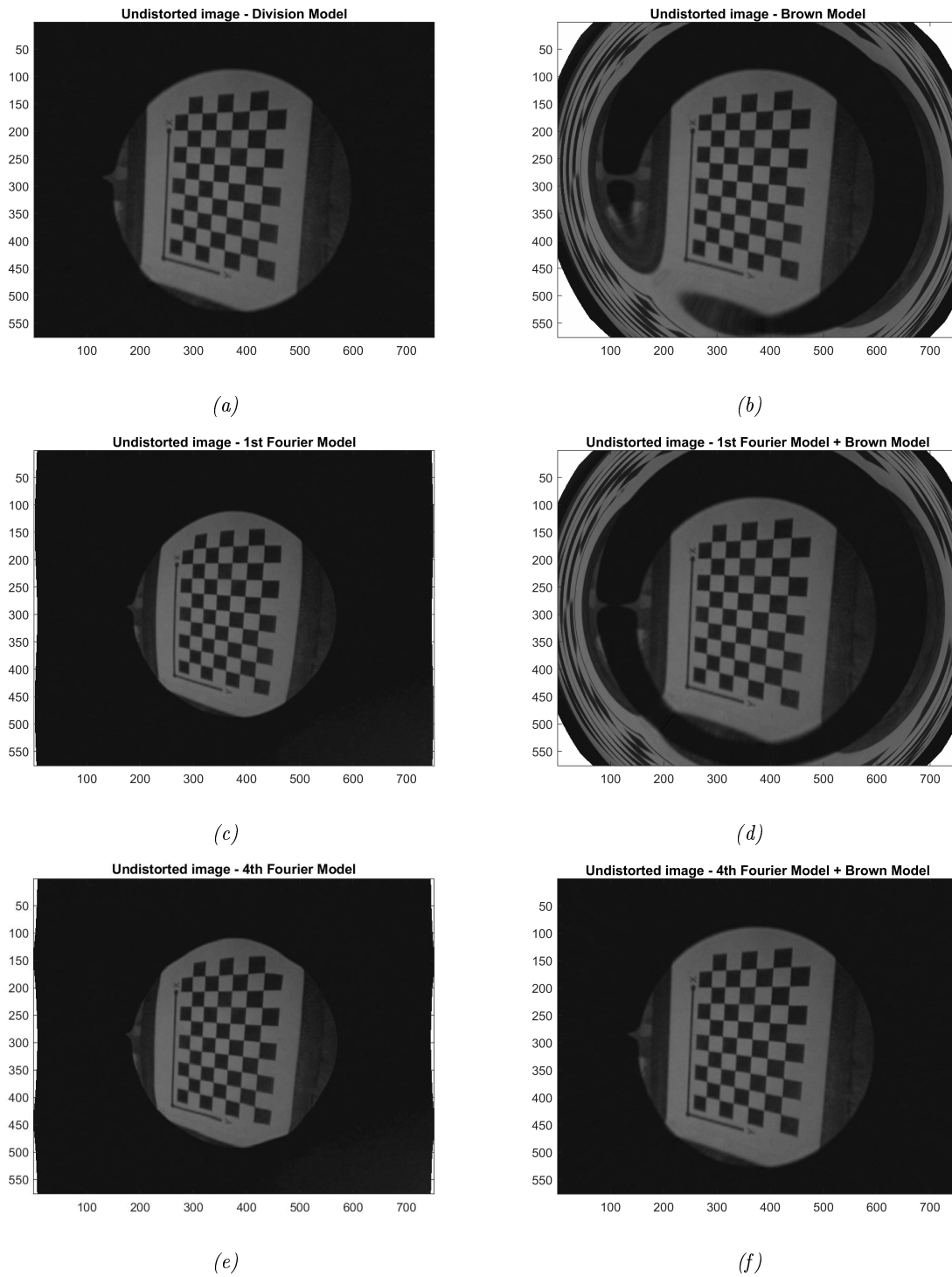
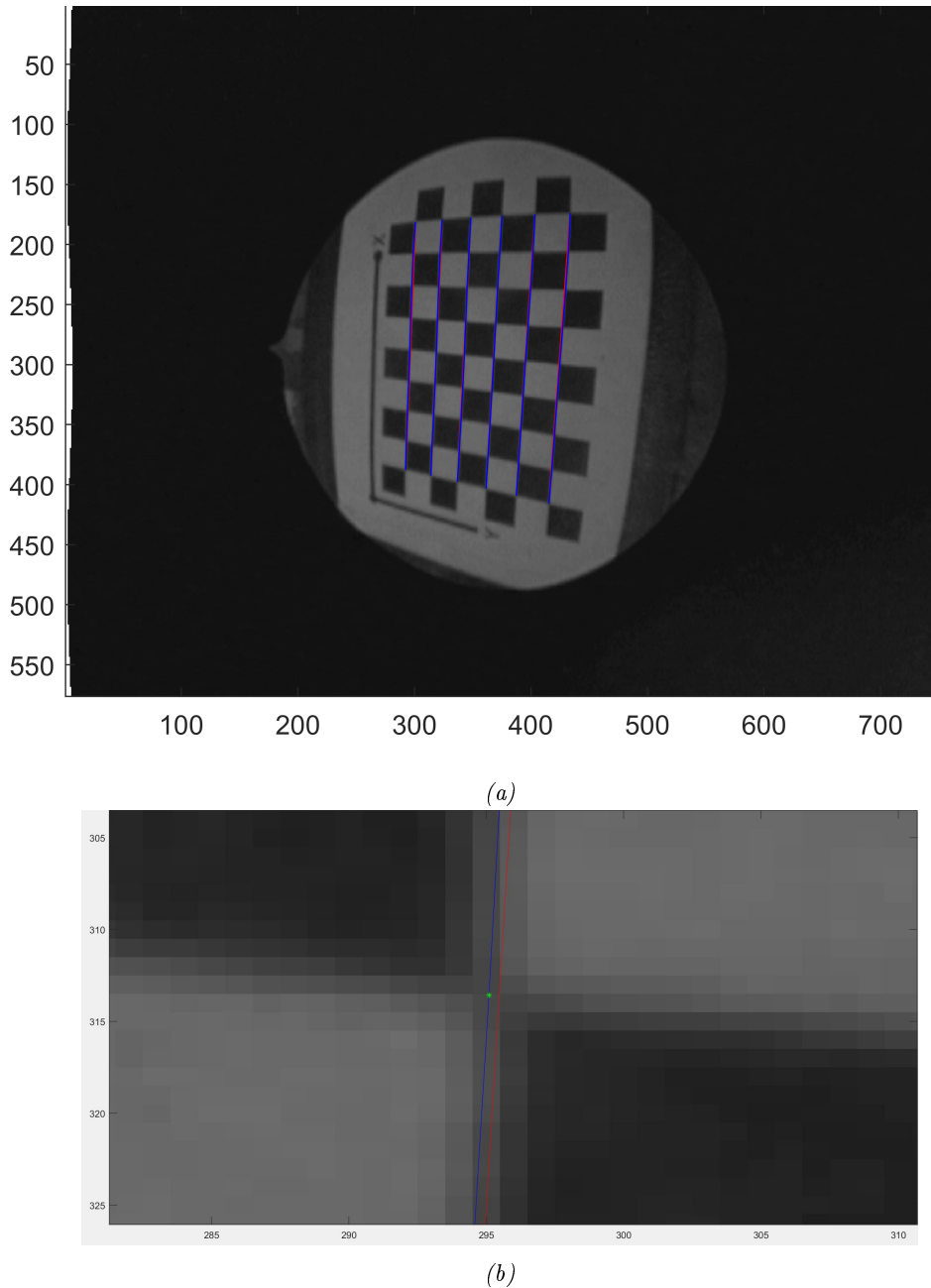


Figure 3.17: Undistorted image with different calibration parameter: (a) Model 1; (b) Model 2; (c) Model 3; (d) Model 7; (e) Model 8; (f) Model 12.



*Figure 3.18:* Straight line plot: (a) undistorted image with 1st order Fourier model; (b) zoomed view. The red line is the connection between the starting and ending feature of the row, the blue line is the connection between the extracted features of the undistorted image with an example plotted with green stars.

To present a concise view of the comparison between different models, the sum of distances between the extracted features and the line connecting the starting and ending point in each row is calculated and shown in the Figure 3.19. Regarding the undistortion effectiveness of the radial components with different calibration models, it could be concluded that:

- The Brown model and division model could both well estimate the radial distortion behaviour of the endoscope with limited number of parameters;
- Though low order Fourier self-calibration models presents bigger values, the distance is decreasing dramatically with the increase of the parameters employed, and the fourth order of the Fourier model could show comparable behavior with the Brown model;



- The combination of the Fourier model and the Brown model shows further improvements in compensating the radial distortion.

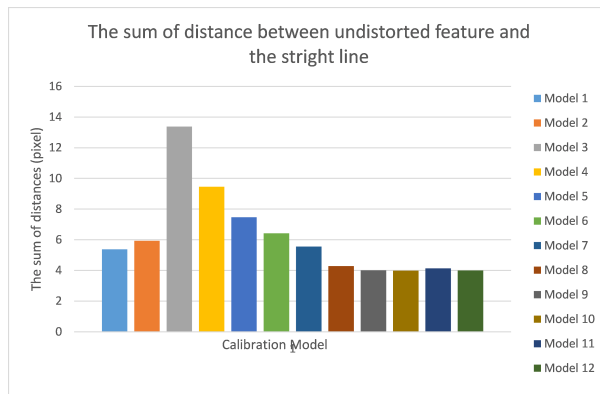


Figure 3.19: The sum of distance between undistorted feature and the straight line.

Though it is commonly recognized that the radial component is the elementary distortion of endoscopic images, other results should also be analyzed for a more convincing comparison. Take the focal length in X direction as an example, the calibration results from different models are shown in Figure 3.20. The complete comparison results of calibrated focal length and principle point coordinates are shown in Table 3.6, where  $f_x$  and  $f_y$  stand for focal length in x and y directions,  $c_x$  and  $c_y$  stand for principle point coordinates.

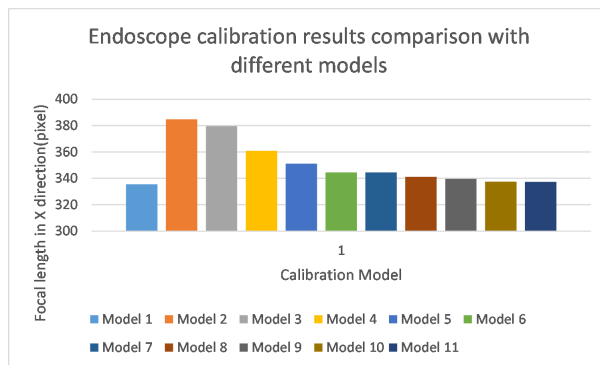


Figure 3.20: Calibrated endoscope focal length comparison with different models.

Table 3.6: Comparison results of calibrated focal length and principle point coordinates for the endoscope.

Calibration Model	$f_x$	$f_y$	$c_x$	$c_y$
Model 1	327.29	324.39	382.55	289.66
Model 2	335.45	332.42	382.57	287.23
Model 3	384.71	386.24	345.222	305.31
Model 4	379.56	380.08	340.01	299.71
Model 5	360.94	361.83	354.92	299.77
Model 6	351.12	350.85	359.91	296.05
Model 7	344.45	343.53	360.63	291.20
Model 8	344.51	342.01	383.25	283.97
Model 9	341.18	338.67	378.59	286.29
Model 10	339.57	336.97	378.62	285.06
Model 11	337.51	334.65	379.28	286.43
Model 12	337.37	334.47	379.22	285.59

With the endoscope calibration experiments with the conventional Brown model, Fourier self-calibration model from first order to fourth order as well as their combinations, the results indicate the following points:

- The division calibration is efficient with a single parameter while it is not an appropriate choice for high accuracy camera calibration. While compared to the low order Fourier self-calibration model, the division model gives closer calibration results to the estimated values;
- The performance of a Fourier self-calibration model largely depends on the number of parameters applied. From the first order with 16 parameters to the fourth order with 160 parameters, the calibration results show a dramatic change up to over 40 pixels with also an approaching tendency to the expected value estimated in Section 3.3.2.1. The phenomenon indicates that it is necessary to apply high orders of the Fourier self-calibration model to compensate severe radial distortions. However, matrix singularities might happen in the calculation process for high order Fourier self-calibration polynomials. The experiment results of endoscope calibration are consistent with theory in Tang's work, which has only been validated for airborne cameras;
- The combination of the Brown model with the Fourier self-calibration model has largely improved the stability of the calibration results. On the one hand, it indicates that the radial distortion plays a dominant role for the endoscope calibration, which could be well modeled by the radial distortion components of the Brown model. On the other hand, the difference between the combination model calibration and the Brown model calibration reveals that except distortions that are modeled by the Brown model, there are other types of distortion that could be well compensated by the supplementary Fourier self-calibration model.

In summary, it could be concluded that the Fourier self-calibration model is not the appropriate choice for modeling the radial distortion alone. While it could be combined with models such as conventional Brown model as a supplement to compensate any other unknown distortion types.

## 3.4 Concluding remarks

### 3.4.1 Stability of camera calibration

The main work of this section is the analysis of relationships between stability of camera calibration using Zhang's method and the configuration of images for calibration. Through multiple subset random selections from the collection of a big number of images, the camera calibration results from all the image subsets could be applied for a Gaussian-fitting process. A comparison experiment validated the importance of different factors, including oblique angles and 90-degree roll angles along the camera axis. In addition, the temporal stability has been verified by 20 independent calibration image groups with and without strong image block respectively.

### 3.4.2 Endoscope calibration

In our research, endoscope calibration is first of all analyzed regarding the stability of the results through its statistical behavior. The experiment shows that the stability of the endoscope calibration is worse than that of regular cameras even though the numerical pixel indexes indicate the opposite conclusion. Besides the standard Brown model, we experimented with the classical division model and Fourier self-calibration model, which is based on a mathematical principle. The experiment results indicate that the radial distortion as the main distortion is necessary to be compensated with a radial distortion parameter, while the mathematical calibration model is an appropriate model for other unknown distortion types that may exist in the endoscopic images.



## Chapter 4

# 3D Digitization

With related technology development in recent years, various sensors based on different principles could be applied, which could be generally classified into passive and active sensors. As our objects of 3D digitization - gyroscopes - are important tech heritage with complex surface and internal structures, our solution for a complete digitization is the integration of photogrammetry, endoscopy as well as CT scanning technology. In Section 4.1, photogrammetry and endoscopy, which are basically of the same principle, are introduced with reference to the workflow as well as the challenges. In Section 4.2, readers could get a view of the CT 3D reconstruction, especially the theorem and the image processing.

### 4.1 Imagery based 3D reconstruction

#### 4.1.1 Principle and workflow

Photogrammetry and endoscopy 3D reconstructions are both procedures using multiview imagery following the same principle. In photogrammetry, spatial resection determines the poses of images. A common used method is the DLT, which is frequently applied when no approximate values or interior parameters are available. After the camera poses are estimated, spatial intersection is applied to find out the 3D coordinates of object points as shown in Figure 4.1. The solution comes from the collinearity equations (see Appendix A2).

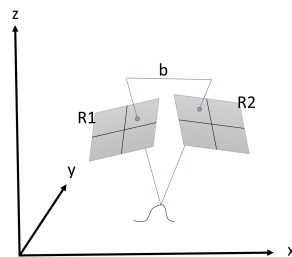


Figure 4.1: Stereo vision illustration.

With a well-calibrated camera, sufficiently overlapped images are collected for a SfM process with accurate image poses and a sparse point cloud as results. In practice, to obtain a satisfying image pose estimation, the overall overlap of 60 % should be reached. Additionally, the settings of the camera such as the resolution, focal length, imaging distance etc., should also be considered for pictures with appropriate exposure and sharpness.

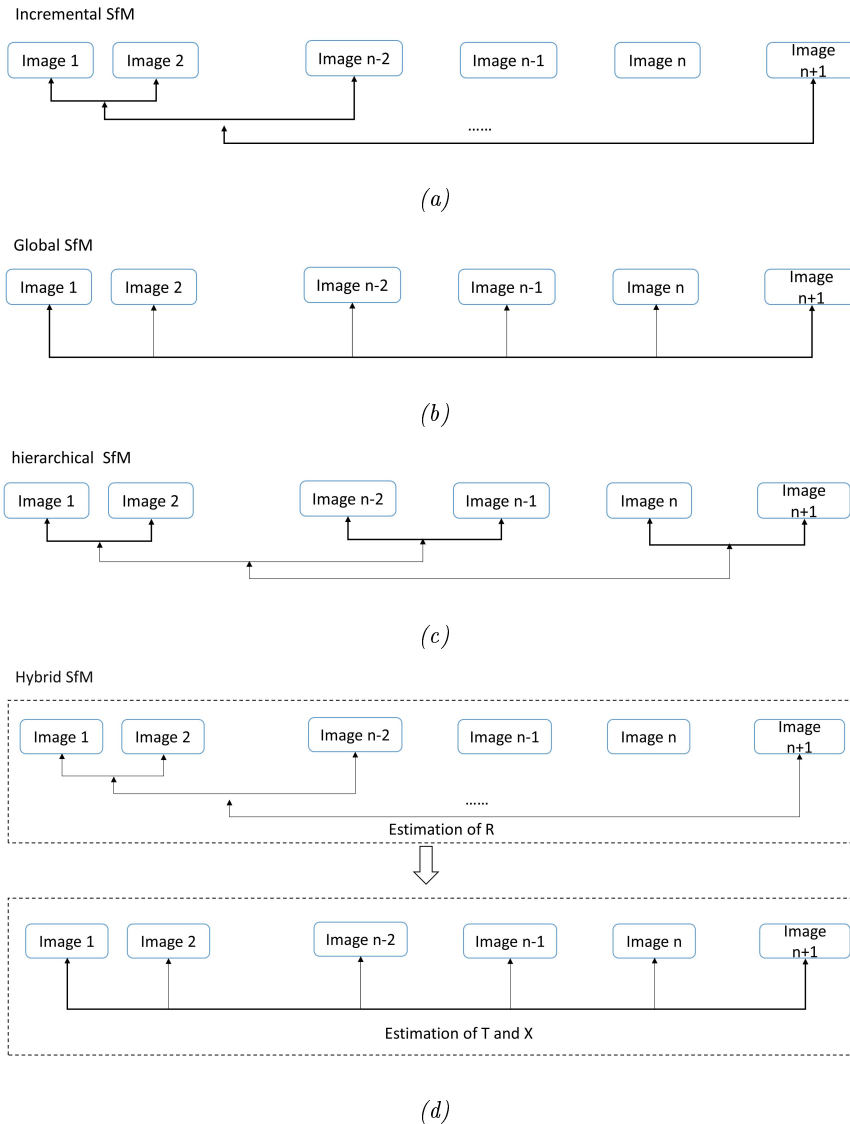


Figure 4.2: Topological structure of different SfM algorithms: (a) incremental SfM; (b) global SfM; (c) hierarchical SfM; (d) hybrid SfM.

There are different SfM strategies proposed for an accurate and efficient pose estimation as well as sparse point cloud generation. SfM algorithms could be generally classified into incremental SfM, global SfM, hierarchical SfM and hybrid SfM. The topological structure of these four classes could be seen in Figure 4.2. The widely applied incremental SfM has the advantage of robust and high accuracy due to its iterative execution of bundle adjustment over the global SfM while it will suffer from the drift as well as the computation efficiency. Relatively, the global SfM implements the bundle adjustment only once but the outliers will influence the final results especially the translation between images. Hybrid and hierarchical SfM combine the advantages of these two methods regarding the computation effort and accuracy.

After having recovered the image poses, a Dense Image Matching (DIM) process (see Appendix A4) will generate a dense point cloud, which could also be used for generating water-tight 3D surface model. As the last step, color informations from images are projected onto the 3D surface model for the final result as summarized in Figure 4.3.

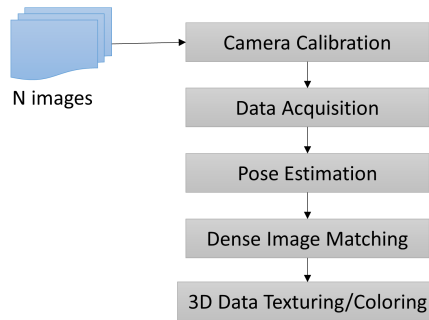


Figure 4.3: Workflow for multiview imagery 3D reconstruction.

## 4.1.2 Photogrammetry

### 4.1.2.1 Practical procedure

According to the requirements of calculating 3D points in Section 4.1.1, a partial 3D reconstruction of the object with non-reflective and non-transparent surface is normally achievable without too much effort. However, to obtain a complete 3D model in an efficient way for movable objects, it is not advisable to use hand-held camera images with unpredictable camera movements. Instead, assisting setup facilities such as a turn-table in our research as Figure 4.4, is necessary for an easier well-overlapped image acquisition. Approaches such as predefined planar or cylindrical trajectories of the camera has been summarized in [72]. Other options are multi-camera systems or more automatic setups using robotic arms with next best pose estimation at first and movement of the camera such as the work presented in [160]. As those options are more expensive, this could largely limit a wider application in reality.



Figure 4.4: Photogrammetry turn-table setup.

Generally, with a turn-table and a stationary camera, overlapping image configurations could be obtained through the appropriate interval rotation movement in horizontal direction. In principle, high image overlaps under appropriate relative orientation could contribute to the stability. In reality, to increase the calculation efficiency with the ReCap application, less neighboring images are considered in SfM processes with a high image overlap setting. In contrast, low image

overlap will take more images into the calculation process and consequently lead to a longer computation time. Therefore, a compromise should be made between the image alignment results and the time cost. Another factor that may influence the SfM process is the feature detection and matching process. Theoretically, in the process of interior and exterior parameter estimation from fundamental matrix via SfM, at least 8 corresponding points [97] could lead to a unique solution while more than that will turn to a least-square estimation process. The redundancy of the measurements could not only be helpful to eliminate gross errors, but also could strengthen the stability of the computation. During the calculation of black-box software, feature detection could be set according to different sensitivity. Too many detected features under a low sensitivity setting may contain even noise or other invalid feature points due to the situations such as specular highlights or transparent materials. It should be noted that in different software packages for 3D reconstructions, the related settings are different and should be set accordingly. Because of the different characteristics of the objects to be reconstructed in reality, it is impossible to set a common parameter for all projects. The solution is to visualize the results of each step for an adaptive adjustment.

Nevertheless, camera movements in vertical direction as well as the change of object placement are necessary to obtain the whole perspective images, especially the contact surface between the object and the placing platform. In practice, if the views have been changed dramatically when the position of the camera or the object changes, the SfM process may fail to align the whole perspective images into a uniform coordinate system. Instead, several image groups would be determined into different local coordinate system. Thus, DIM could not be implemented for a complete 3D model. The problem could be solved via a manual control points picking registration, which, however, could be labor consuming and reduce the accuracy. To avoid such situations, the following rules should be obeyed in the image acquisition phase:

- the illustration condition of the images from different stationary camera should be consistent for the image matching result;
- the background of the image should be as clean as possible to avoid the effect of noise on image matching process;
- the overlap between views before and after the movement of the object should be guaranteed;
- the whole perspective images should be collected to avoid missing information on the 3D model;
- the number of images should be appropriate, though basically two overlapped images are enough to recover the 3D information, too many images will lead to expensive computation cost while too few images may lead to inaccurate 3D models [1].

#### 4.1.2.2 Evaluation of the software for 3D reconstruction

After the image acquisition, there are several issues to be taken into account in the procedure of 3D reconstruction. Even though commercial software as a black-box is able to implement the calculation in an automatic way, the setting of parameters related to the corresponding theory will influence the performance of integrated functionalities.

The software for 3D reconstructions are classified into two categories, i.e open source software and commercial software as introduced in section 2.3.1.2. The adopted software in the research are

- Agisoft Metashape 1.6.2.10247 (MS, previous name was Agisoft Photoscan).
- Capturing Reality 1.2.0.16813 RealityCapture (ReCap)
- VisualSfM
- PMVS



The commercial software has the advantage regarding the completeness of the 3D reconstruction workflow, while the open source software packages provide algorithm details regarding the intermediate results. In our work, those software are compared regarding the alignment performance, dense image matching quality as well as the computation efficiency with three datasets from the collection of the Gyrolog project, which is indicated in Table 4.1. For a rigorous comparison, all implementations are conducted with the same computer, which uses a i7-8700 CPU processor and a NVIDIA GeForce GTX 1080 graphic card. Besides, the feature detection is performed with the same numbers and same resolution images are utilized in all related procedures such as image correspondence extraction and DIM steps [128].

Image number	Image overlap	Object Dimension
Object KK12-09		
46	High overlap	$15.8 * 13.5 * 21.5cm^3$
Object KH09-09		
225	High overlap	$4.4 * 14.5 * 14.8cm^3$
Object Bohnenberger Maschine		
86	High overlap	$15.5 * 8.6 * 8.6cm^3$

Table 4.1: Summary of employed datasets

#### 4.1.2.3 Results and analysis

The results of software evaluation with 3 datasets are reported in the Table 4.2

	MS	ReCap	VisualSFM	PMVS
Dataset1 - KK12-09				
Aligned Image Number	46/46	44/46	44/46	
Alignment Calculation Time	91.6s	47s	133s	
Number of 3D points	15,430	15,016	15,481	
Mean error (pixels)	1.8315	0.494947		
DIM Calculation Time	9mins	17mins	-	20mins
Number of dense point	1,868,390	1,420,204		770,071
Dataset2 - KH09-09				
Aligned Image	225/225	225/225	225/225	
Calculation Time	5mins(error)	4mins	15mins	
Number of 3D points	44,790	44,719	48,339	
Mean error (pixels)	0.763384	0.425272		
DIM Calculation Time	165mins	75mins		101mins
Number of dense point	6,402,978	2,739,365		1,384,752
Dataset3 - Bohnenberger Maschine				
Aligned Image	86/86	86/86	86/86	
Calculation Time	70.145s	86.867s	241s	
Number of 3D points	31,990	28,451	29,289	
Mean error (pixels)	0.605	0.336	-	
DIM Calculation Time	26mins	27mins		21mins
Number of dense point	8,617,456	1,623,642		165,261

Table 4.2: Summary of employed datasets

- (1) Dataset1 applies the directional gyro KK12-09 as introduced in Section Appendix B. The experiment is designed using an easy image alignment configuration, where the camera is

stationary and the object is placed on a turn-table without flip or move. Obviously, all software packages are capable to align the images together into an uniform coordinate system and show similar results of the sparse point cloud. However, ReCap shows an evident advantage over others regarding the computation efficiency in the image alignment phase. Of particular note is that the reconstruction of Agisoft Metashape includes the background information which lead to a much bigger 3D point number of the reconstruction.

- (2) Dataset2 uses the operator unit platform KH09-09 also introduced in Section Appendix B. The experimental data consists of images from the object that has been moved and flipped its position, which might cause uncertainty for the image alignment process. Among all, Recap and VisualSFM show best results with the most images oriented. The MS failed to align all the images together. For comparison, three chunks are processed separately and merged for the later DIM comparison. Within this dataset with over 200 images, the computation efficiency shows even a bigger difference, while the final numbers of 3D points from DIM processing are close.
- (3) The object for dataset3 is one of the most important objects - the Bohnenberger Maschine. The reason of the selection is due to the homogeneous surface of the object, which could reveal the feature detection and matching ability of the software. From the results, it could be seen all softwares are possible to align the images together.

In brief, the commercial softwares MS and ReCap provide a more user friendly and compact operation procedure. Relatively, ReCap has better performance as for the computation efficiency, accuracy as well as the alignment stability. With regard to DIM, MS could generated more 3D points while it could not exclude the background, i.e the platform effectively. However, MS has several advanced functions, such as the application of the image mask, camera calibration function, and could work as a supplementary tool when dealing with complex objects using ReCap. In addition, though the open source toolbox VisualSFM could achieve comparable results as MS and ReCap, its results of pose estimation indicates obvious errors as shown in Figure 4.5 from the actual circular trajectories.

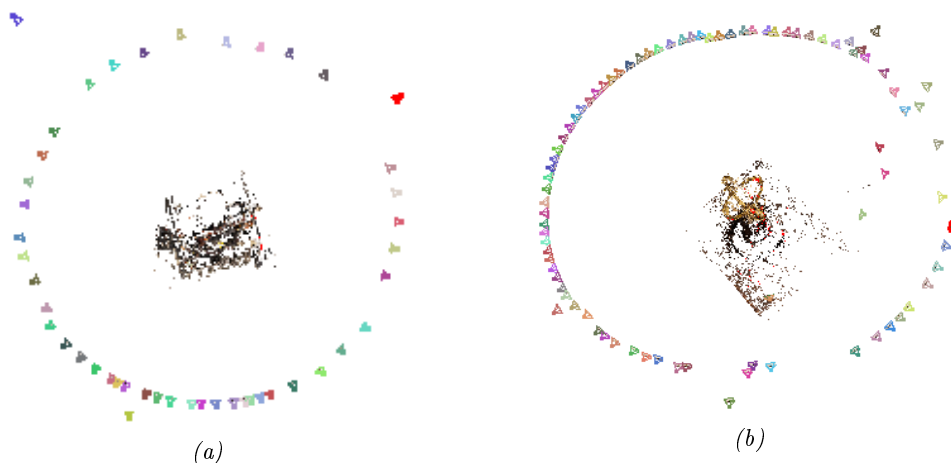
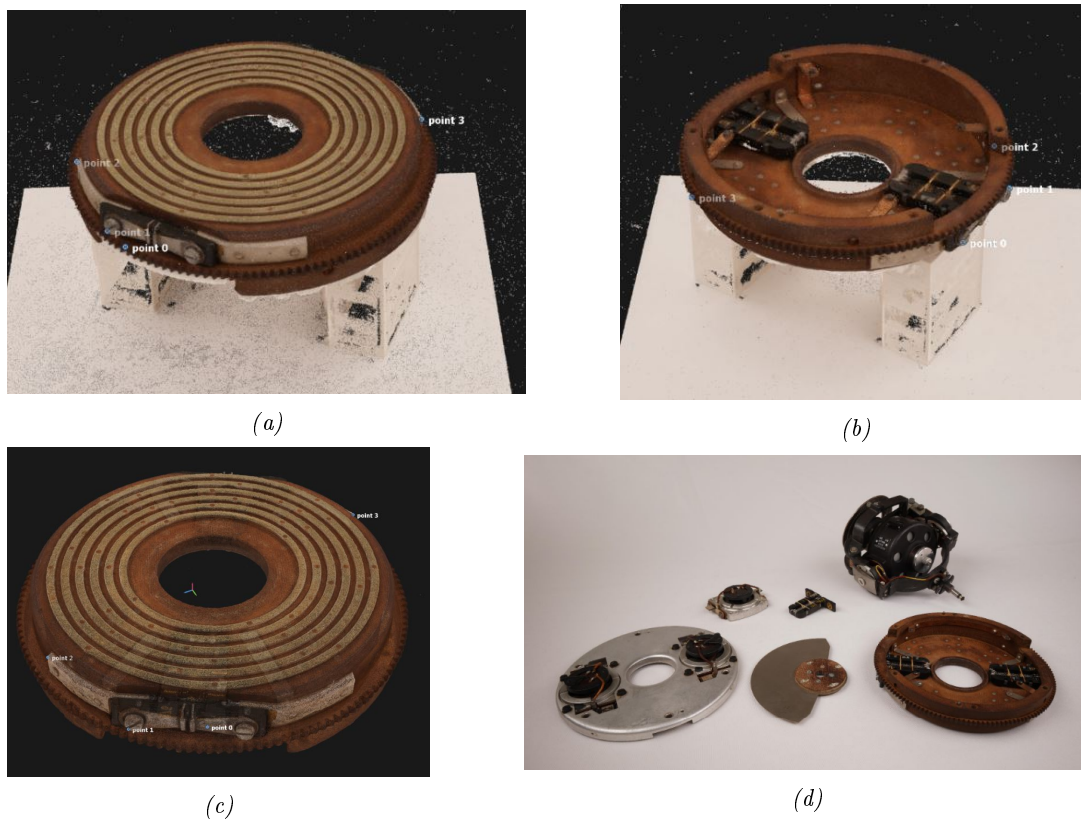


Figure 4.5: Pose estimation results from VisualSfM: (a) KK12-09; (b) The Machine of Bohnenberger.

For situations, where the images obtained before and after the object movement are difficult to be aligned together, two possibilities are available to solve the problem. We could manually choose control points on the images of different components for an implementation of the image alignment process. Alternatively, components containing partial view images could also be reconstructed separately to generate several dense partial point clouds of the object. In the next step, those point clouds are to be registered according to the overlap information. As can be

seen from the Figure 4.6 with the example of a component from KK17-09, two initial image groups created with all the input images have been registered successfully. It should be noted, that the control points on the image should be picked evenly in corresponding 3D space and also precisely. In Figure 4.7, two partial point clouds as well as the registered results are shown of the object Golden Gnat (Section 7.3). This is normally realized by a coarse-to-fine process, which applies 3D feature points for the initial registration and an ICP like algorithm for refinement. For the convenience of implementation, the upper and lower sections are truncated from the corresponding partial reconstructed point clouds.



*Figure 4.6:* Manual registration using control points of KK17/03-09: (a) partial reconstruction with top side; (b) partial reconstruction with bottom side; (c) integrated reconstruction; (d) an original image of KK17-09.

A special case of Gyroscope 3D reconstruction is the reflective or transparent material processing. The principle of photogrammetry introduced in Section 4.1.1 is under the assumption of a diffuse reflection on the surface of the object to be reconstructed. However, transparent surfaces will introduce refractions and a very shiny surface will lead to confusion for the 3D reconstruction program. In this case, one of the solutions is using spray on the corresponding surface. After having aligned both sprayed and non-sprayed images together in an uniform coordinate system, only sprayed images are applied for the dense image matching process and non-sprayed images for texturing. An example is given in Figure 4.8.

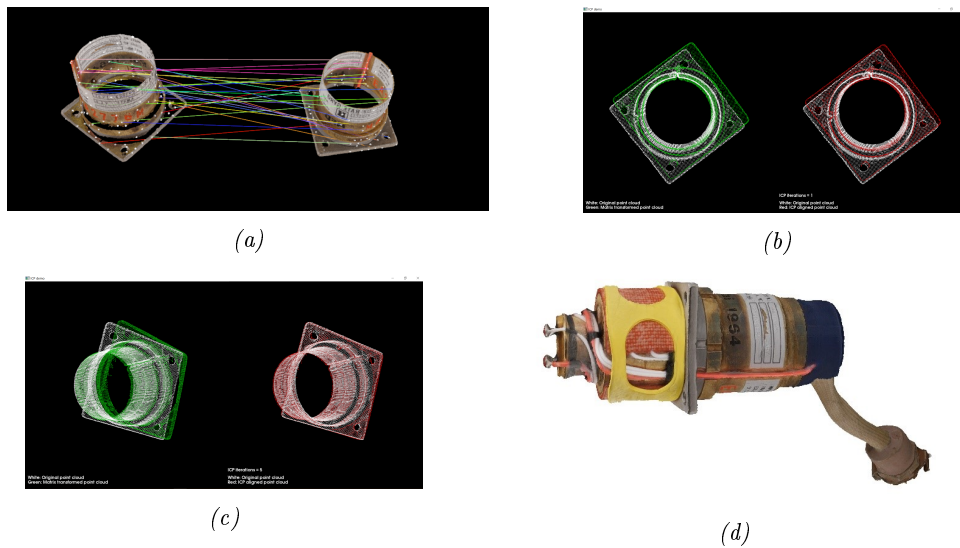


Figure 4.7: Golden gnat registration using 3D features and ICP: (a) feature matching of point cloud pair; (b) initial pose of coarse registration; (c) refined registration using ICP algorithm; (d) integrated 3D model.

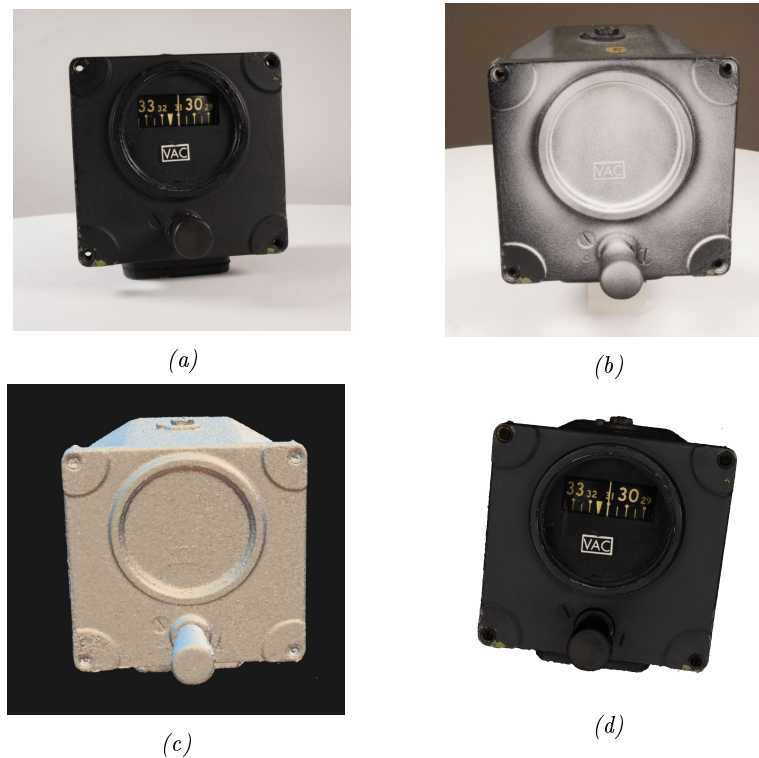
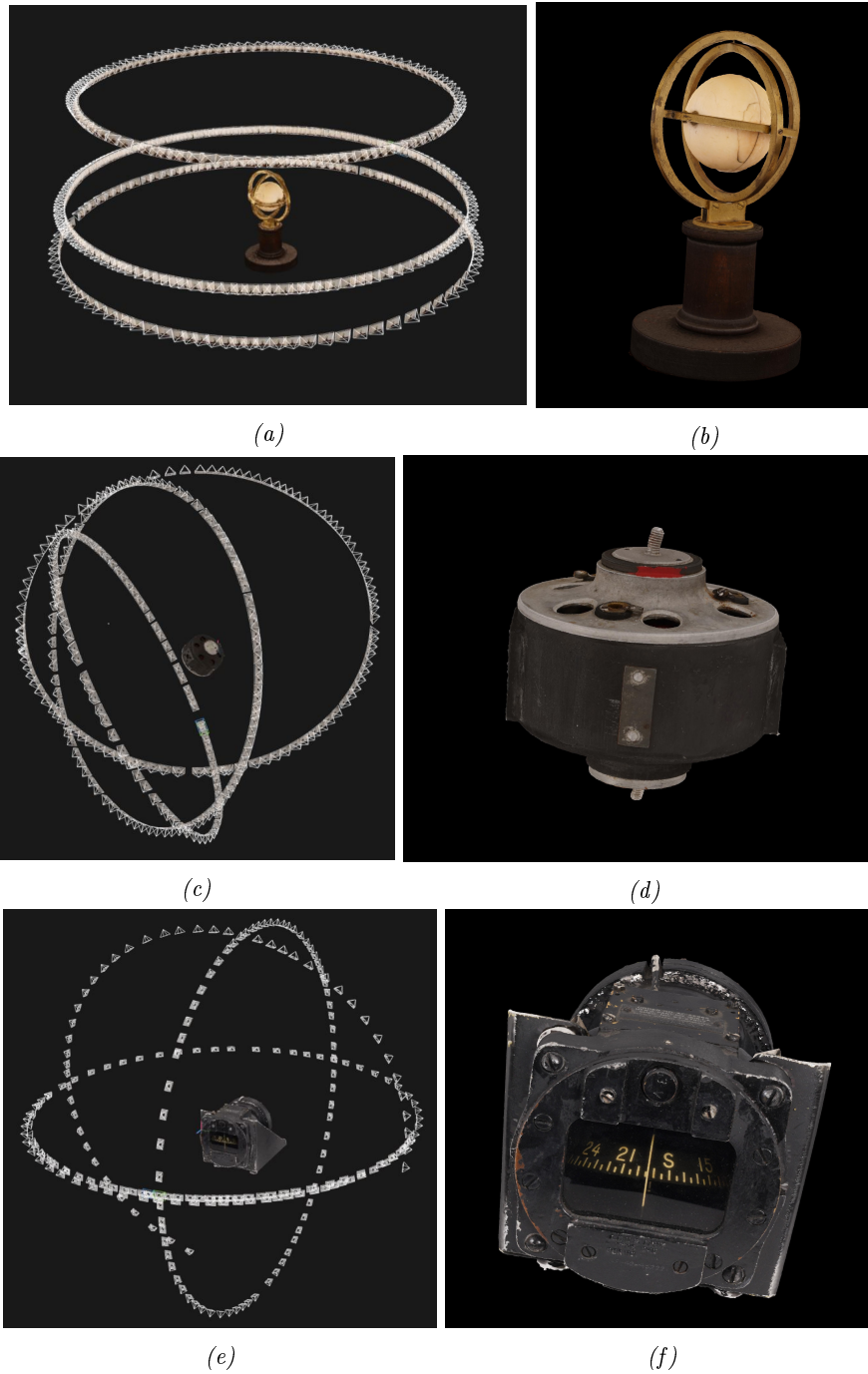


Figure 4.8: Spray operation for transparent or reflective surfaces: (a) original image; (b) sprayed image; (c) reconstructed mesh with sprayed image; (d) textured mesh with original image.

According to the evaluation results of the software, ReCap is used as the main tool due to its calculation efficiency with MS as the supplementary solution for its various advanced functions. Successful examples of gyroscope 3D reconstructions are given in Figure 4.9. From the pictures, it could be seen that the overlap between images from different camera stations are well aligned and the surfaces have been well reconstructed.



*Figure 4.9:* Image configurations and corresponding gyroscope 3D models by photogrammetry: (a)-(b) The Machine of Bohnenberger ; (c)-(d) KR05-17; (e)-(f) MFK01-09.

### 4.1.3 Endoscopy

#### 4.1.3.1 Image acquisition

As introduced in Section 4.1.1, as a special camera system, the endoscope 3D reconstruction follows the same principle as conventional photogrammetry. However, the special characteristics of the endoscope differs from the regular camera regarding the 3D reconstruction in following aspects:

- the difficulty of obtaining appropriate image configurations due to its special structure as shown in Figure 3.3;
- the limited image quality with reference to resolution, brightness, specular highlight and severe distortions;
- short imaging distance and small field of view lead to larger number of images, which is likely to cause drift in the result of trajectory estimation.

Due to the fact that the above differences bring difficulties for endoscopic image collections, it is necessary to apply a fixture to avoid the image blur caused by the shake of hand-held endoscopic images. In addition, the trajectory of the endoscope should be precisely controlled to ensure the overlap between neighboring images. Therefore the setup as shown in Figure 3.4 in Section 3.1.3 has been applied. Furthermore, to increase the efficiency of the image acquisition process, the strategy of combination of images and videos is adopted. We capture videos instead of images for those less important areas which are only used for bridging reasons.

#### 4.1.3.2 Image preprocessing

According to the characteristics of endoscopic images, the necessary pre-processing steps include:

- Boundary detection. Due to the bigger sensor area than the imaging area, only the central area could collect effective information with the useless ambient area. Therefore, the circular boundary of the imaging area could be extracted to avoid the noise influence of the ambient area.
- Image undistortion. The severe endoscopic image distortion should be corrected according to the calibration results obtained in Section 3.3. A pair of images in Figure 4.10 is presented to show the endoscope images before and after the image undistortion.
- Specular highlight removal. This is also an issue due to the metal surface of the Gyroscope as well as the near-field lighting. The specular highlight removal is implemented according to the method introduced in [7]. The highlight removal could avoid the feature matching errors as well as texture ambiguities because of the view changes of the highlight area. A pair of examples is given in Figure 4.11.
- Image brightness enhancement. It is caused by the limited illumination space of the small diameter scope cylinder, which has great influence for the SfM process. The image sequence is processed via a histogram normalization to ensure the uniform and enhanced brightness of the images. Original and enhanced images are shown in Figure 4.12.
- The noise coming along with the low illumination could be handled with different filters. The method introduced in [77] could improve the feature detection as well as the correspondence matching.

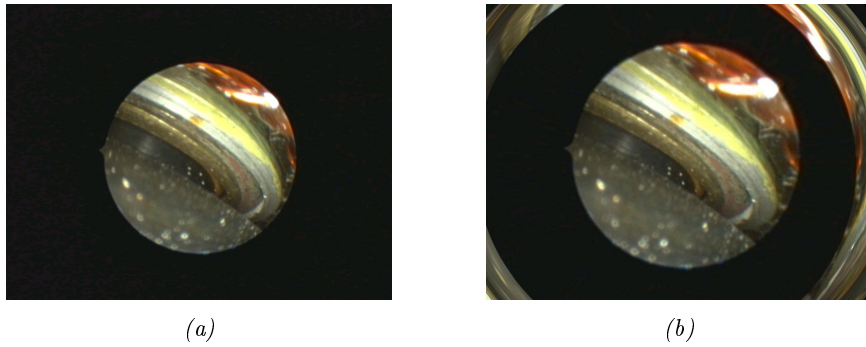


Figure 4.10: Undistortion of endoscope image: (a) original image; (b) undistorted image.

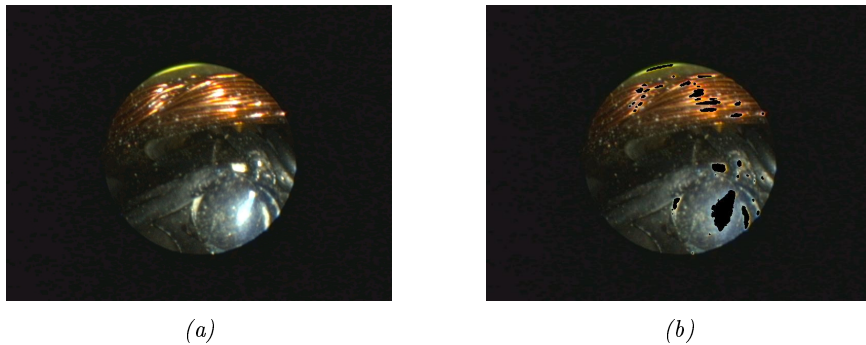


Figure 4.11: Specular highlight removal of endoscope image: (a) original image; (b) highlight removed image.

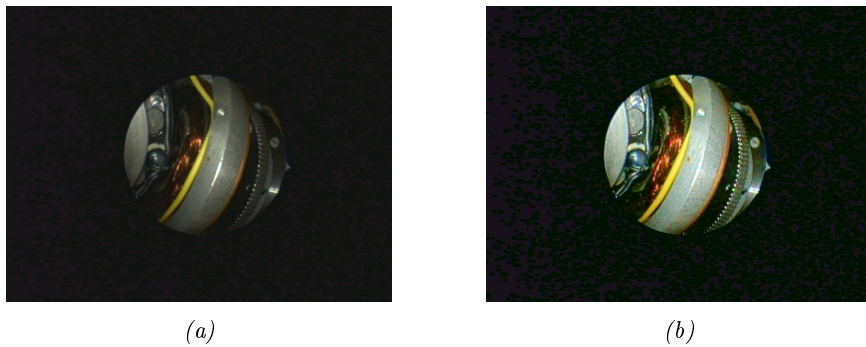
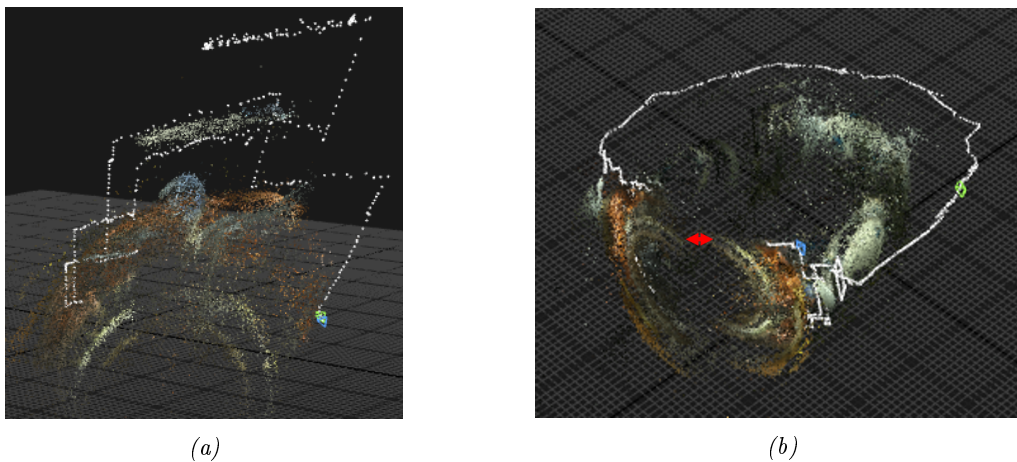


Figure 4.12: Brightness enhancement of endoscope image: (a) original image; (b) brightness enhanced image.

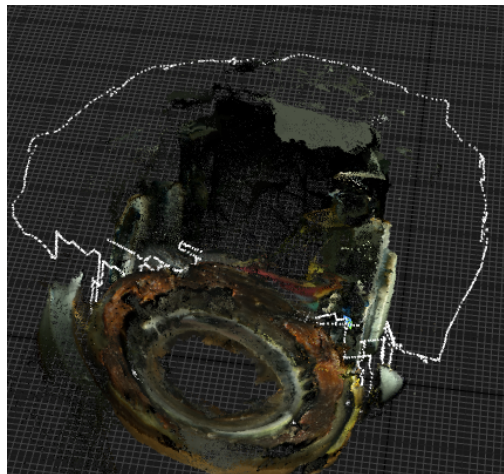
#### 4.1.3.3 SfM procedure

With the collected endoscopic images, image alignment is to be implemented according to the workflow for estimating the accurate image poses. Although spatial resections are using well matched correspondences, the low quality of endoscopic images brings challenges as introduced in Section 4.1.3.1, which stresses the necessity of image pre-processing before the image alignment procedure starts. Additionally, due to the small coverage of a single endoscope image, a larger number of endoscopic images is needed in comparison with a normal camera. However, such long trajectory may lead to drift problems in the pose estimation results. Examples can be seen in Figure 4.13, which are obtained by RealityCapture image alignment process with and without image undistortion respectively. The left image show much bigger distortion than the right image due to the lack of calibrated intrinsic information. Though the implementation details

are embedded in the black-box, the estimated camera poses with an obvious drift phenomenon indicates that the SfM is an incremental like strategy. By using additional manual control points, drift-free SfM results by RealityCapture as comparison is shown in Figure [?].



*Figure 4.13:* Drifted image alignment results with endoscopic images: (a) without sensor calibration; (b) with sensor calibration.



*Figure 4.14:* Drift-free Endoscopic SfM with manual control points .

## 4.2 Computed Tomography

### 4.2.1 Basic principle and workflow

CT 3D reconstruction is initially realized by the acquisition of a stack of radio-graphic projections (2D images) of the object around a rotation axis, which are then processed for a volumetric representation. The object to be scanned is placed between the source and detector of the X-ray as shown in Figure 4.15. In comparison to the medical CT scanners, the industrial CT scanners are more powerful to penetrate through metals, which are essential components of gyroscopes. Thus, industrial CT is an ideal solution for non-destructive preservation of gyroscopes regarding the internal structure.

As the basis of the 3D reconstruction process in computed tomography, the 2D image acquisition process as well as its quality is highly depending on the power of the X-rays, which is



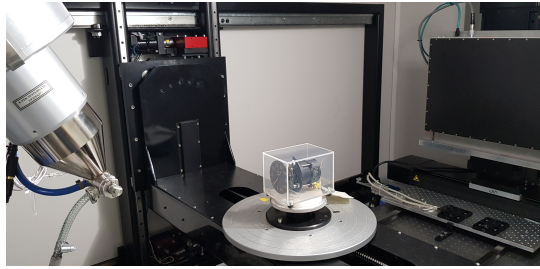


Figure 4.15: Photogrammetry turn-table setup [48].

generated by electron beams. The X-ray source contains an X-ray generator (a vacuum tube) in which electrons are released from a filament (the cathode) and are highly accelerated by an electric potential difference. Then, the electronic beam is focused on a metal target (the anode) and produces X-rays according to two different processes: The electrons are decelerated by the atomic nucleus of the target and part of their kinetic energy is converted into an emitted X-ray photon. This phenomenon is called ‘braking radiation’ or ‘Bremsstrahlung’. The level of the Bremsstrahlung depends essentially on the selected acceleration voltage (anode voltage). The output spectrum consists of a continuous spectrum of X-ray energies ranging from 0 to the electronvolt of the X-ray source. As a result, a projection is generated on the X-ray detector with an example as given by Figure 4.16. The penetration depth of the X-ray radiation and the associated image quality depends on various factors: selected acceleration voltage in kV, material properties and thus absorption properties of the object to be scanned.

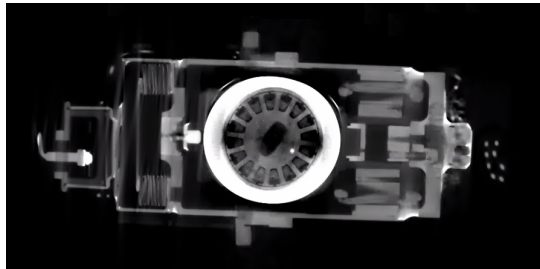


Figure 4.16: X-ray slice of WK10-10 @ Gasim Mammadov.

In general one pixel on the projected image is represented by 14 bits which later extends to 32 bits single precision float type. The pixel value changes after post processing such as shading correction and bad pixel estimation. All the projections are influenced to the poisson noise, beam hardening and metal artifacts. After obtaining the set of projections, they are stored in hdf5 file which is a data model which stores large datasets and large variety of metadata. The projections further are fitted to the reconstruction algorithm such as Filtered Back Projection (FBP) and the output is a volume with the 3D pixel called voxel. The volume is also stored in hdf5format, which contains the certain metadata as grid origin, grid space and volume dimensions. To extract the layers of different materials, the histogram could be applied for the whole volume data as shown in Figure 4.17 for the purpose of thresholds selection. Figure 4.18 presents an example that the CT images represent a possibility of virtually dismantling the reconstructed 3D models in order to examine their individual parts more closely. For this purpose, the partially disassembled Honeywell GG165A2 rate gyro, a variant of the Golden Gnat gyro, is shown on the left, which is an object of the gyro collection. In the middle is the inner part of the electric rotor drive, the so-called stator. The picture on the right shows a 3D reconstruction of the stator, both as a complete component under positions  $i$  and  $j$  and broken down into its individual parts. The parts, generated purely by computed tomography, are artificially colored to distinguish them. The visual creation was done interactively with the software Scan-Mess 3D.

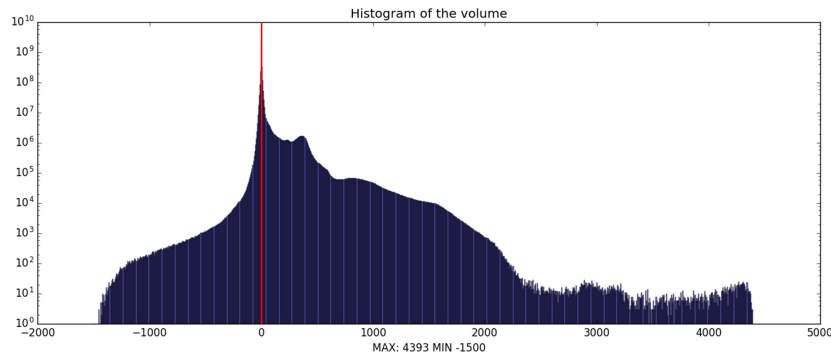


Figure 4.17: CT histogram @ Gasim Mammadov.

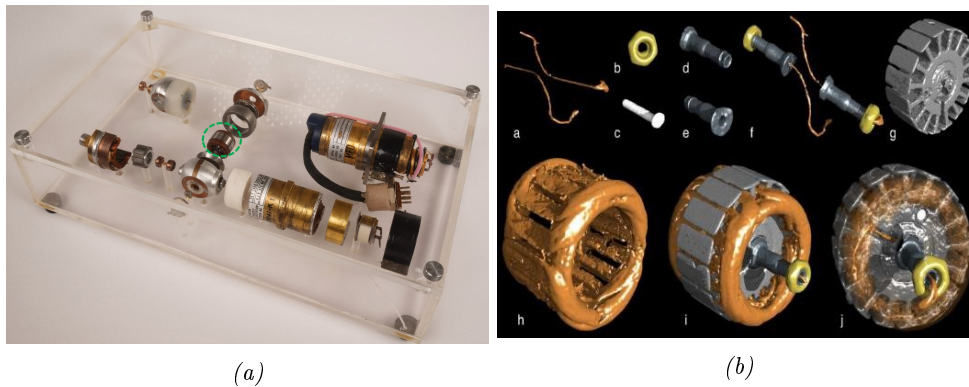


Figure 4.18: Virtual dismantling of the stator of the rate gyro GG165A2 [48].

For the computed tomographic examination in our research, a Nikon XT H 225 X-ray Computed Tomography (CT) system (see Figure 4.15) has been used to generate 3D models of the gyroscopic objects. The micro-focus X-ray source of the CT system has a maximum electron acceleration voltage of 225kV with a maximum power of 225W. The X-ray source has a minimum focal spot of  $3\mu\text{m}$  at 7W using a reflection tungsten target. The detector of the Nikon CT is Varian 4030 model with dimensions of  $2300 * 3200$  active pixels and the pixel size of  $127\mu\text{m}$ . The rotation and the further manipulations of the object in CT is performed by the manipulator XT H 225. It has 5 axes which are X,Y,Z, Tilt  $\pm 30^\circ$  and Rotate ( $n * 360^\circ$ ), where n is step size. The maximum sample weight is 15 kg.

## 4.2.2 CT 3D model pre-processing

The volume achieved from the CT 3D reconstruction is then used to generate the point cloud. The point cloud is the collection of the voxel positions, which are calculated using the grid origin and grid space. In addition to the voxel location described by the three coordinates  $(x, y, z)$ , the color value is also specified by using the RGB color model. The workflow could be illustrated as Figure 4.19.

### 4.2.2.1 De-noising of CT 3D model

In general, all voxels can be transferred to the points, however, the interesting volume region should be part of an object and not the background. Thus, the thresholding is set to distinguish the object from the noise and background. Despite the case that the range of voxel values starts from negative values, we take only the positive range which starts from zero. The negative values, as shown in Figure 4.17 on the left side of the red line, is background noise which does

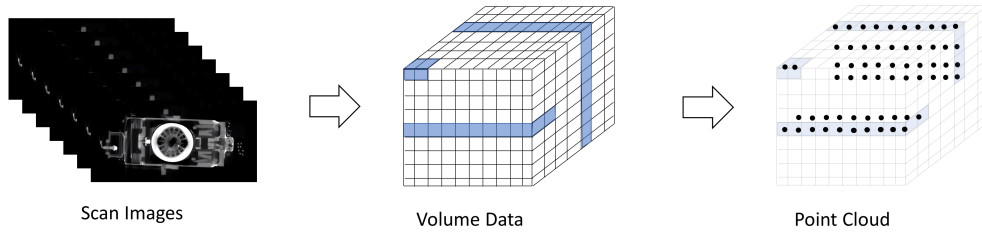


Figure 4.19: CT point cloud generation from X-ray images.

not carry any information. An example of de-noising with the object Golden Gnat is shown in Figure 4.20.

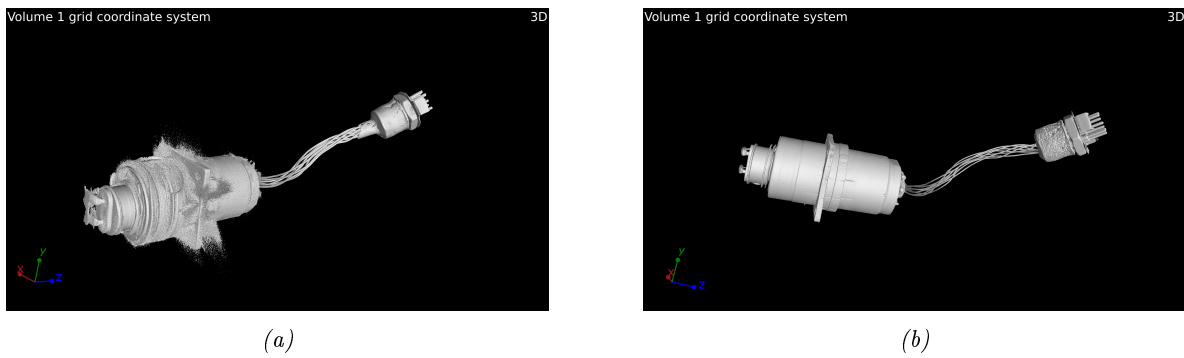


Figure 4.20: De-nosing process of WK10-10: (a) noise model of WK10-10; (b) de-noised model of WK10-10 @ Gasim Mammadov.

Besides the noise of the background, the complex components of the object could also lead to a imperfect reconstruction. As can be seen from Figure 7.3, it has several cables and pins which might lead to streak artefacts. Besides, it consists of different metal alloys such as iron, copper, and aluminum. Therefore, high energy is necessary to be applied for reducing the beam hardening artefacts. Two iso-surfaces are shown in Figure 4.21. With the method based on Convolutional Neural Network (CNN) proposed in the work [124], the de-noised projections are obtained for the reconstruction of a volumetric model as shown in Figure 4.22.

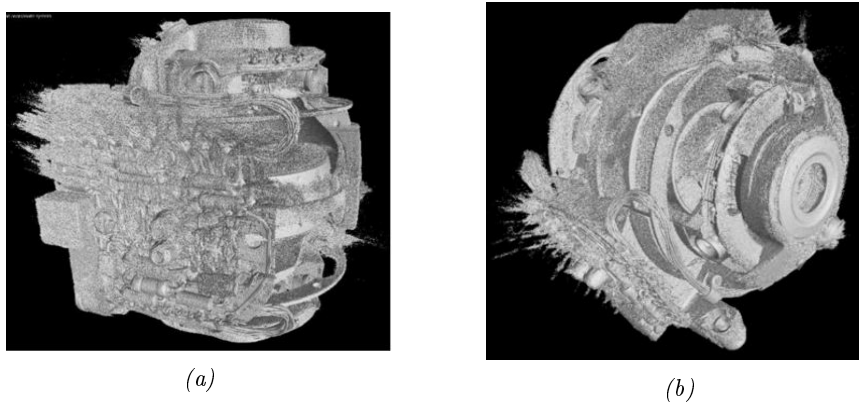


Figure 4.21: Two views of the iso-surface of G200 from the 3D volume data set reconstructed by the FBP [47].

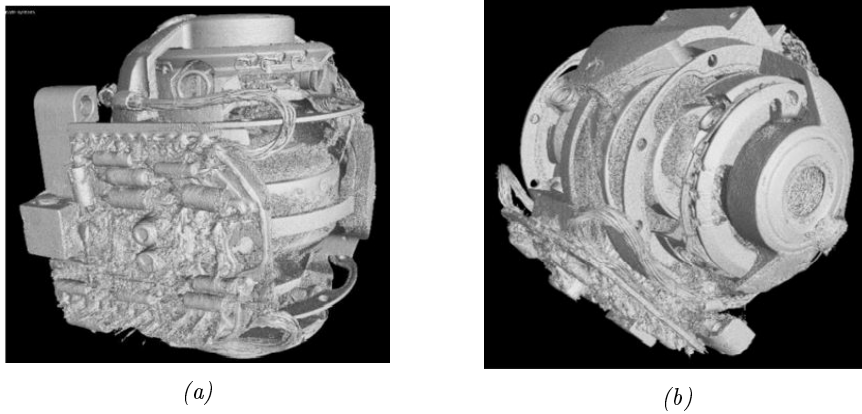


Figure 4.22: Two views of the iso-surface of the G200 calculated from the CT 3D volume data set reconstructed by the FBP from the denoised projections [47].

#### 4.2.2.2 Volumetric point cloud surface extraction

With the point cloud transformed from the volumetric data, the further surface extraction can be performed, for the convenience of data registration with the photogrammetric point cloud. The surface extraction process can be characterized as the problem of finding the convex hull, which is the set of points which are enclosed in a convex polygon. The solution is to project outer voxels to the surface of the object. When the line of each voxel projection touches the object it is counted as a surface. However, the noise can still exist in volume and thus the next couple of voxels are also checked to guarantee that it is solid object not a noise. The complete workflow of the surface extraction can be summarized in Figure 4.23. In addition, an example of the original volumetric point cloud and the extracted surface data are presented in Figure 4.24.

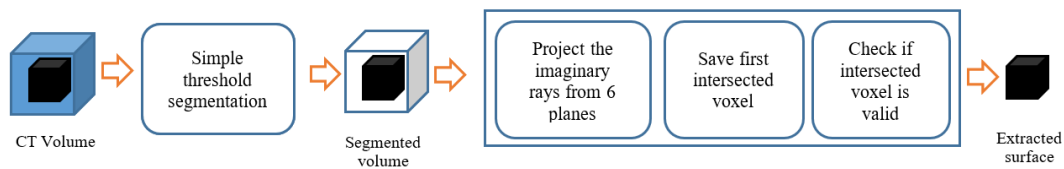


Figure 4.23: Measurement setup inside the CT.

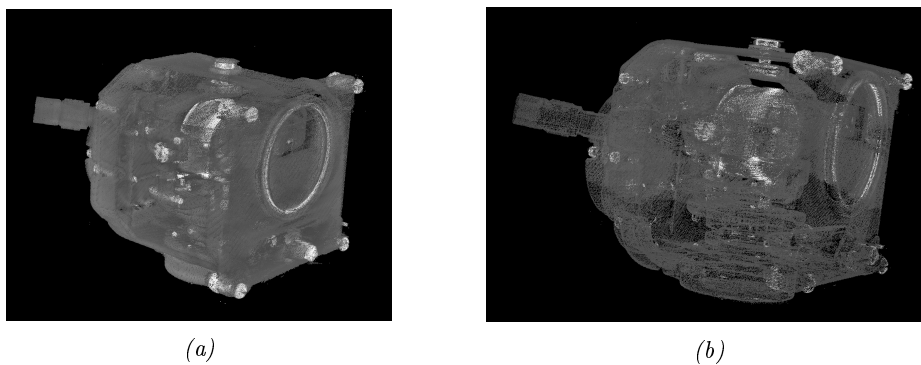


Figure 4.24: An example of surface extraction: (a) volumetric point cloud; (b) extracted surface data @ Gasim Mammadov.

### 4.3 Concluding remarks

With three sensors of different characteristics, photogrammetry could realize a global surface 3D reconstruction with a regular camera, an endoscope aims for local reconstruction for hollow or cavity structures on the surface, while CT supplements with 3D internal information. The integration of 3D models from the three sensors above could provide a complete 3D digitization with both surface and internal information. Among three methods, photogrammetry has the best maturity for non-transparent, non-reflective surfaces with appropriate image configurations. As for application of endoscopes for heritage preservation, it is a totally new trial and the image acquisition, image pre-processing and 3D reconstruction workflow are necessary to be improved and defined. The CT mainly suffers from the thresholding selection for 3D reconstruction and the noise problems. Excluding the cost factor, it is promising to have a wider application of CT in other 3D reconstruction projects.



## Chapter 5

# Data Integration

With the recent development of the technology, the integration of various sensors - also called sensor fusion - based on different principles is the big trend to enrich 3D digitization applications. However, there is no uniform algorithm to realize the 3D data integration under the situation that measurements are from different sensors having different standards, formats and other characteristics. Within this chapter, Section 5.1 presents a common solution based on artificial markers using the Gauss-Helmert model for transformation parameter estimations, which has the least limitations among many other registration algorithms but requires manual work. In Section 5.2, under the assumption that the CT data has a complete representation, the color information from the photogrammetric images could be directly projected onto the CT data surface which has already been transformed into the same coordinate frame as the photogrammetric point cloud. Section 5.3 deals with the situation where the assumption under Section 5.2 is not fulfilled. According to the structure characteristics of the gyroscopes, the virtual intersection points, which are obtained by extracting geometric primitives from the photogrammetry and CT point clouds, respectively, could be determined to act as control points for the transformation matrix estimation.

### 5.1 Gauss-Helmert Model-based method

The task of integrating several point clouds from different sources is also called registration, or the problem of a Rigid Body Transform (RBT). A common solution is the Iterative Closest Point (ICP) algorithm. However, there is a restriction on the overlap and the necessity of a coarse registration. Alternatively, we propose a comprehensive accuracy analysis for point cloud registrations with corresponding common points. These points are called “Control Points” or simply “Targets”. With manually selected control points, we formulate an overall 7 parameters transformation embedded in a Gauss-Helmert model. More details of the mathematical derivation for the process could be found in Appendix A5.

#### 5.1.1 Experiments and results

To implement the transformation parameter estimation using the Gauss-Helmert model, the corresponding control points are to be manually selected. The requirements for this process are: (1) high precision, which needs cautious operation due to the reason that the cursor might fall on the point with wrong depth information; (2) even spatial distribution, which is of vital importance for the error propagation during the estimation process. In particular, co-planar control points should be avoided. For the CT and photogrammetry point cloud registration, only geometric feature point clouds could be applied because of the lack of texture information on the CT data. This could be challenging for the objects containing few corner points such as a ring,

sphere or cylinder shape. Thus, the details should be carefully observed for this control points picking process.

According to the last section of theory introduction, we have chosen the G200 from Figure 7.4 for investigation. Figure 5.1 presents the views of 10 chosen corresponding control points which are evenly distributed on the object surfaces.

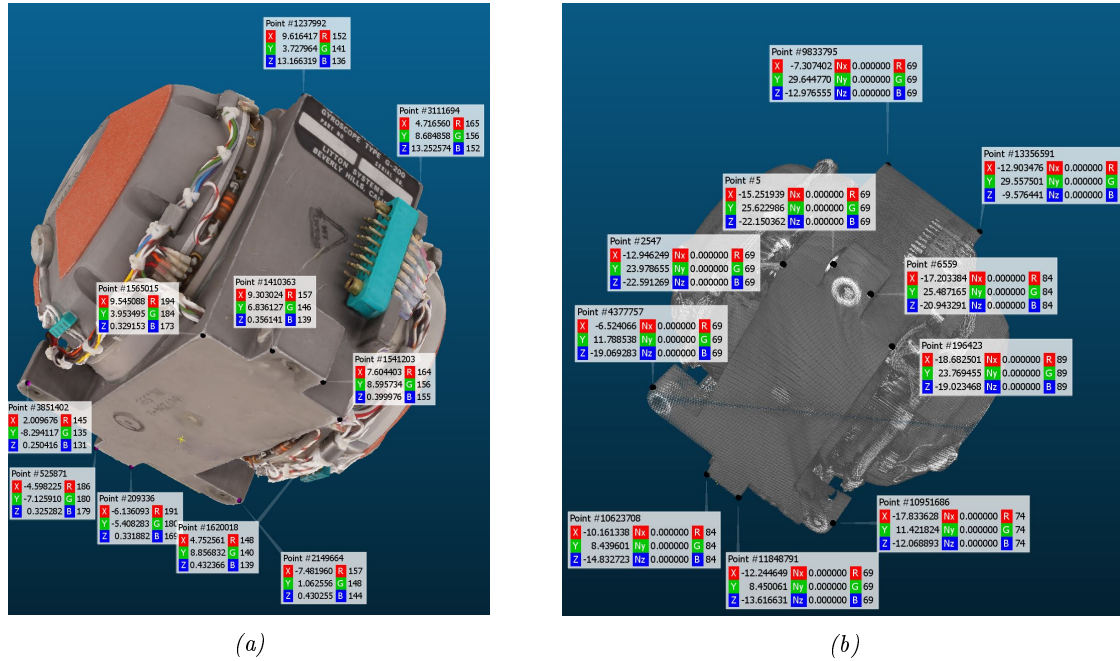


Figure 5.1: Manual control points selection of Gyro G200.

After the estimation with the Gauss-Helmert model using 7 parameters, direct split views of the integrated model is shown in Figure 5.2. For comparison, an estimation with 4 corresponding control points has also been implemented with precision assessment presented in Table 5.1. In addition, two more objects the Bohnenberger Maschine and the Ternstedt gyro are also included in the experiments. The cross section of the integrated model and precision assessments are shown in Figure 5.3, Figure 5.4 and Table 5.1, respectively. It should be noted, that the integrated models are obtained from the transformation matrix with 10 control points and the CT data are down-sampled to increase the efficiency of visualization. Experiments with 4 control points could also realize the transformation estimation, while the standard deviation of the unit weight  $\sigma_0$  deviates up to 38% from the nominal value, which is 1.0.

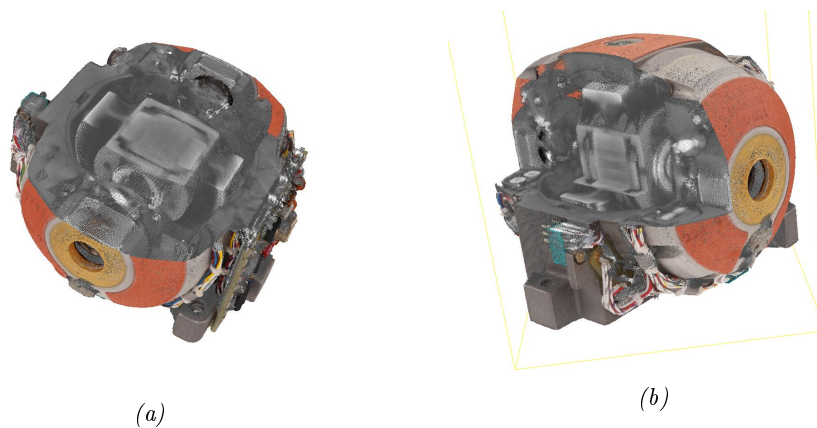


Figure 5.2: Split views of integrated model of Gyro G200.



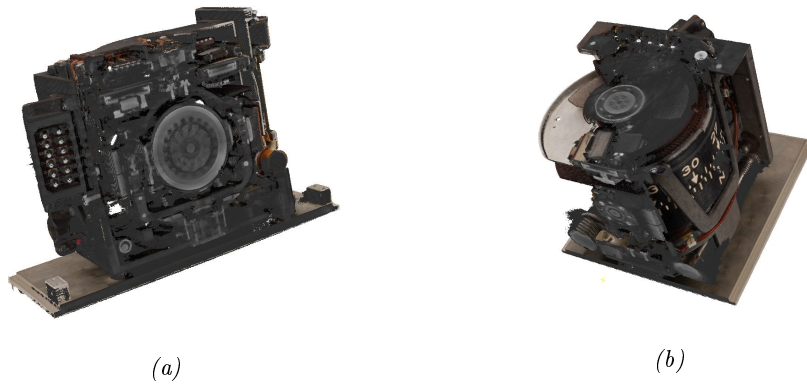


Figure 5.3: Split views of integrated model of direction gyro KK12-09.

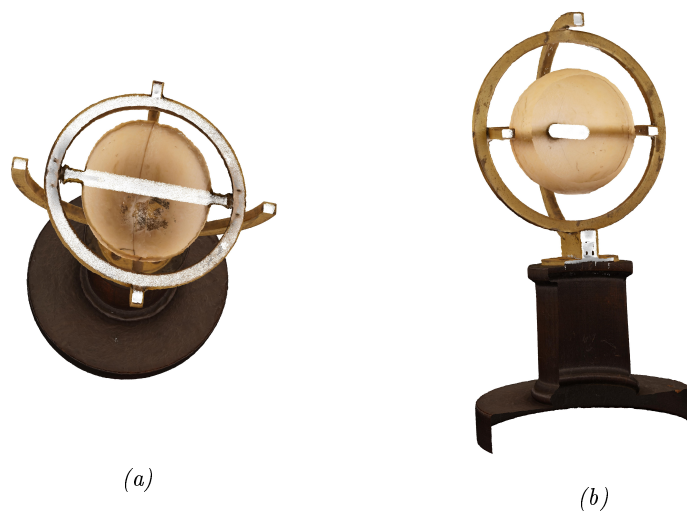


Figure 5.4: Split views of integrated model of the Machine of Bohnenberger [47].

Table 5.1: Precision estimation of photogrammetry/CV and CT point cloud estimation by the Gauss-Helmert model

Objects	CPs	$\sigma_0$	$\sigma_{CT}$	$\sigma_{CV}$	CT GSD	CV GSD	File Size	Integrated Size
G200	4	1.38	0.10mm	0.08mm	0.22mm	0.05mm	440MB	1.05GB
	10	1.07	0.08	0.06				
BM	4	1.36	0.125mm	0.149mm	0.035mm	0.029mm	171MB	378MB
	10	1.02	0.094	0.113				
Ternstedt	4	1.24	0.115mm	0.140mm	0.22mm	0.05mm	165MB	1.46GB
	10	1.09	0.101	0.124				

## 5.2 CT surface coloring

Different from the last Section 5.1 using manual control points based on a Gauss-Helmert parameter estimation for point clouds registration, this section will mainly focus on combining individual advantages of both photogrammetric and CT data. Under the assumption that CT data has a complete representation on the surface regarding different materials, the overall concept is using a coarse-to-fine registration procedure to determine the spatial relationship between photogrammetric images and CT data. Afterwards the color information from the optimal images is projected onto the CT surface point cloud. Thus the advantage of color information from photogrammetric data and the high geometry precision from the CT data are combined together.

### 5.2.1 Related theory

#### 5.2.1.1 Workflow

The whole workflow for a CT surface data coloring process is summarized in Figure 5.5. The first step is the generation of CT surface point clouds and photogrammetric point clouds of the object's surface. Secondly, we take the surface of CT data for accomplishing the alignment with the photogrammetric 2.5D surface model. Thirdly, the typical procedure for alignment between CT surface data and photogrammetric data is to select control points from both data sets to obtain a coarse-registered model and then apply an ICP (Iterative Closest Point) algorithm for refinement.

Fourthly, after the alignment, the transformations between CT data and images for the 3D reconstruction are determined. According to each image pose, the CT surface data can generate a depth image (see Section 5.2.1.3), which corresponds with the original photogrammetric image. Fifthly, the exclusion of the ambiguous pixels on the depth image due to the projection of 3D points should be implemented by making use of range values. And finally, all points from CT surface data are processed in the same way in order to obtain a full coloured CT surface representation.

#### 5.2.1.2 Histogram based threshold determination

The reconstructed CT volume represents the object with attenuation values carried by each voxel. The overall histogram of the final volumetric data illustrates the different peaks with respect to the various materials used in the object, since each material attenuates in a different way when the X-rays pass through. Theoretically, the overall histogram of the final CT volume calculated with the FBP algorithm carries multiple peaks representing different materials (see Figure 4.17). However, the X-ray images have information not only about the accumulated attenuation on the detector, but also the noise which can come from the detector and the source as well as other artifact types, see [83]. Based on the severity of the noise which can distort the

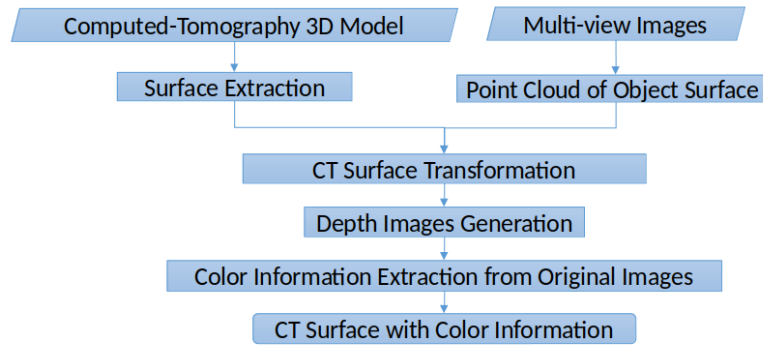


Figure 5.5: Pipeline of CT surface point cloud colouring [189].

shape of each peak one can separate materials with the simple thresholding [114]. Additionally, some parts of a scanned object can be simply not visible during the visualization as a result of high X-ray energy spectrum: materials such as paper or plastic will have low attenuation in comparison with the metal and dense materials with high atomic numbers.

### 5.2.1.3 Depth image generation and application

A depth image is an image channel containing the distance information from a viewpoint to the scene object. It can be obtained from multi-view images for the purpose of 3D point calculation. However, a depth image could also be generated from a 3D point cloud with adequate camera parameters for other applications. The process is a projection from the 3D space to the 2D image frame, where pixels do not carry colour but range values to the corresponding 3D point. If the camera parameters from the pose estimation and calibration are adopted, the depth image should also be overlapping with the original photogrammetric images in the projection areas. To get the conceived result, the intrinsic parameters, including the calibrated focal length and the principle point coordinates, and extrinsic parameters are to be provided to define the camera. In addition, the width and height of the depth image are needed to be defined.

### 5.2.1.4 Colour information assignment

Because the range image generation process projects all points onto the image plane with the camera intrinsic parameters, some points on the back side and out of the field of view, which should not be projected according to the law of light propagation, will fall on the image plane. This will cause ambiguities when finding the corresponding pixel between the depth image and the undistorted image. The ambiguity may happen as indicted in Figure 5.6.

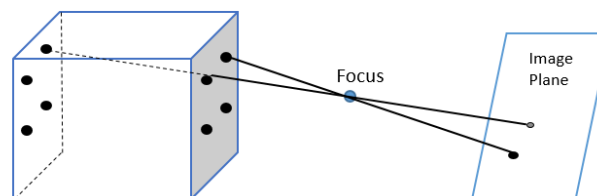


Figure 5.6: Schematic diagram of ambiguous pixel on the depth image.

Under normal conditions, the neighbouring pixels should have similar range values, if they are neighbouring points in 3D space. Another possibility for neighbouring pixels can also be

that the light ray travels through two surfaces and have two intersection points. Therefore, the way to solve the problem is by taking advantages of the range values of neighbouring pixels. The procedure to eliminate the wrong pixels is as follows:

- create the depth image according to the camera pose;
- list the pixel coordinates and colour information according to the depth image and corresponding photogrammetric image for each 3D point;
- take a certain size window of the pixel, and compare the range values between the target pixel and its neighbouring pixels;
- the pixel which holds much bigger value ranges will be excluded
- the qualified pixel will get the colour information from the corresponding pixel of the undistorted photogrammetric images. from the colouring process.

By applying this processing pipeline, the corresponding pixels of the back side of the object are eliminated. However, a single point is still visible from multiple images, which means one point from CT surface might get more than one colour values. Therefore, a further range comparison is necessary to determine the final colour value.

## 5.2.2 Experiments and results

### 5.2.2.1 Data acquisition process of the Golden Gnat

With the equipment and configurations introduced in Section 4.1.2, the 3D reconstruction using photogrammetry and CT have been realized respectively. The Golden Gnat is imaged with Lego bricks fixing its base. One of the original images is shown on the left image of Figure 5.7. The photogrammetric 3D reconstruction is implemented via RealityCapture software. To better focus on the object part for the experiment, the point cloud is filtered to be displayed in the right image of Figure 5.7. A CT reconstructed model is given in Figure 4.20 and was derived from a large number of X-ray images with an example of Figure 4.16. Additionally, both datasets have been pre-processed regarding noise and other imperfections. It should be noted, that only the upper part above the lego base of the point cloud has been used in this experiment.



*Figure 5.7:* Photogrammetry 3D reconstruction of Golden Gnat: (a) example image of the Golden Gnat; (b) point cloud of the Golden Gnat.

### 5.2.2.2 Point Cloud Registration of the Golden Gnat

The registration of CT and photogrammetric data of the Golden Gnat is divided into two steps. First of all, control points from both datasets are manually selected as stated in Section

5.1. Secondly, with an initial registration, an ICP algorithm is applied with Point Cloud Library (PCL) under C++, to calculate the closest distance between the overlap of the two point clouds. The aligned CT point clouds and the photogrammetric source point cloud are displayed in Figure 5.8. It has to be noted that the color of CT data has been darkened and the photogrammetric point cloud has been sampled to enable the visualization of both surface and internal information.



Figure 5.8: Aligned photogrammetric and CT surface point clouds.

After the application of the ICP algorithm, the aligned CT point cloud is considered in the same position with the photogrammetric point cloud. Therefore, the pose relationship between images and photogrammetric point clouds is valid for the transformed CT point cloud.

### 5.2.2.3 Depth image generation of the Golden Gnat

After the transformed CT surface data is acquired, the depth image of the CT point cloud needs to be calculated. The depth image generation process is realized under PCL from aligned CT point clouds. In this research, the needed camera intrinsic and extrinsic information are exported from the Reality Capture software. However, the definition of the parameters, especially the camera poses, which have multiple representation possibilities, are crucial for the implementation of the procedure. The depth image corresponds to the pose of the original image, which is shown on the left side of Figure 5.7, and is displayed on the left image of Figure 5.9. Furthermore, an overlap display of the original image and the depth image is presented on the right image of Figure 5.9.

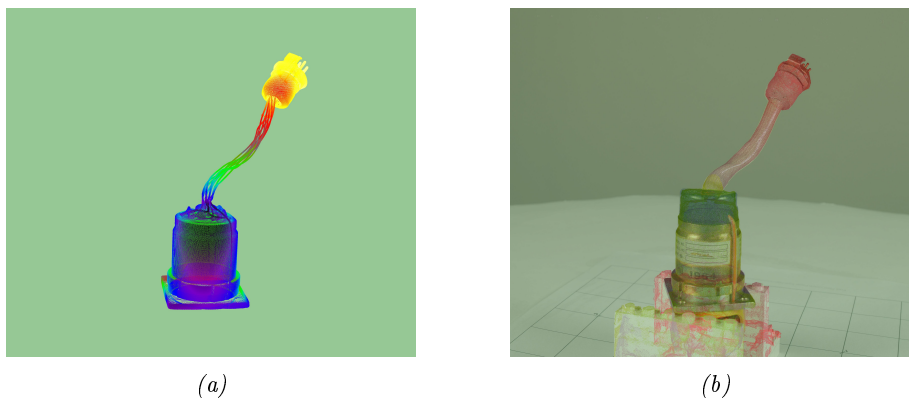


Figure 5.9: Depth image generation of Golden Gnat: (a) depth image of CT surface point cloud; (b) overlap display of the depth image and the corresponding photogrammetric image.

It should be noted here, that the result does not correspond to the original photogrammetric image but to the undistorted image, which needs to be calculated through the original image

and applying camera distortion parameters. The image on the right hand side of Figure 5.9 indicates the overlap between the depth image and the undistorted photogrammetric image.

#### 5.2.2.4 Colour information assignment for the Golden Gnat CT surface point cloud

By applying the method stated in Section 5.2.1.4, a window size of 5\*5pixels is chosen to exclude the ambiguous pixels for the colour assignment process. Figure 5.10 displays the final results of the coloured CT point clouds. It can be seen that except the wired part coloured with the information of its wrapping plastics, the CT point cloud is coloured with satisfying quality from photogrammetric images.



Figure 5.10: CT coloured point cloud of the Golden Gnat.

A comparison of a zoomed local display of both the original photogrammetric point cloud and the coloured CT point cloud is shown in Figure 5.11. According to the results above, it could be concluded that the high precision geometry of the CT data and texture advantage of photogrammetry data are integrated in the coloured CT point cloud.



Figure 5.11: Result comparison of CT data coloring: (a) zoomed local display of CT coloured point cloud; (b) zoomed local display of photogrammetric point cloud.

## 5.3 Virtual control points using geometric primitives

### 5.3.1 Related theory

#### 5.3.1.1 Primitive fitting

The preliminary step of primitive fitting is a point cloud segmentation, which has not been mentioned in other point cloud registration methods using geometric primitives. Furthermore, for a watertight 3D model of a complex structured gyroscope object, the fitting process will be influenced by the spatial distribution of the points. Therefore, it is of vital importance

to implement the point cloud segmentation into different clusters before fitting the primitives. Region growing is a conventional segmentation method, which could apply various information such as normals or colors for the algorithm. The CT point cloud has high geometric accuracy regarding the voxel positions as well as the normals. Therefore, a region growing algorithm using the normals and curvatures works well in CT surface point cloud segmentation. As for photogrammetry, the primitives generally differ from each other in material and color. Hence the color-based region growing segmentation is applied.

For each segmented cluster, parameterized geometric primitives are used to fit to the discrete point cloud data. Like other fitting problems, various principles are available but mostly originate from RANSAC (see Appendix A1). Due to the diversity of the point clouds, a simple threshold definition is hardly to meet the requirements for a best primitive fitting problem. Maximum Likelihood Consensus (MLESC), which improves the RANSAC by optimizing the inlier scoring rather than simply counting the inlier numbers, is adopted. The coefficients of the most frequent 3D primitives are listed in Table 5.2.

Table 5.2: Coefficients of 3D primitives

3D Primitives	Point	Vector	Constant	Angle
Cone	Apex	Axis		Open angle
Plane		Normal	Distance	
Sphere	Center		Radius	
Cylinder	Point on axis	Axis	Radius	
3D Circle	Center	Normal	Radius	

### 5.3.1.2 Correspondence estimation

With good fitting results, these parameterized primitives can be used to directly or indirectly obtain virtual control points. Because the fitting process is based on the RANSAC principle, the virtual control points are the representations of the overall characteristics of the point cloud clusters. Theoretically, they are more reliable than the manually selected control points or the feature points calculated using the neighboring information from the incomplete point cloud. After obtaining the virtual control points, their corresponding relationship needs to be solved. In general, after obtaining the key points, the corresponding descriptor will be calculated for the process of correspondence estimation. Global and local descriptors use global and neighborhood point information respectively, but these are not applicable to calculate virtual points, due to the fact that virtual points are mostly outside the point cloud. The information of geometric primitives, which is used for virtual control point calculation, could be encoded. The virtual control points could be obtained from the center of a sphere, the intersection of the cylinder axis and the plane, or the intersection of three planes. [162] applied mainly the plane angle for the condition of three planes intersection. If  $(a_1, b_1, c_1)$  and  $(a_2, b_2, c_2)$  stand for normal vectors of two intersecting planes, the angle between can be expressed as (5.1),

$$\alpha = \arccos \frac{a_1 \cdot a_2 + b_1 \cdot b_2 + c_1 \cdot c_2}{\sqrt{a_1^2 + b_1^2 + c_1^2} \cdot \sqrt{a_2^2 + b_2^2 + c_2^2}} \quad (5.1)$$

We denote planes that are related to the intersection point as target planes, which can be ordered by the angles, and the rest non-target planes. Except the angles between target planes, the angles between each target plane and all non-target planes could also be calculated. All angle values are binned into a histogram. Similar to the idea of a Point Feature Histogram

(PFH), the binning process divides the angles into certain subdivisions, and counts the number of occurrences in each subinterval. The advantage of the virtual control point descriptor is that when the number of control points is relatively large, the computational efficiency can be significantly improved. A similar process could also be applied to the cylinder axis and plane intersection situation by replacing the plane angle to line-plane angle. An example is shown in Figure 5.12.

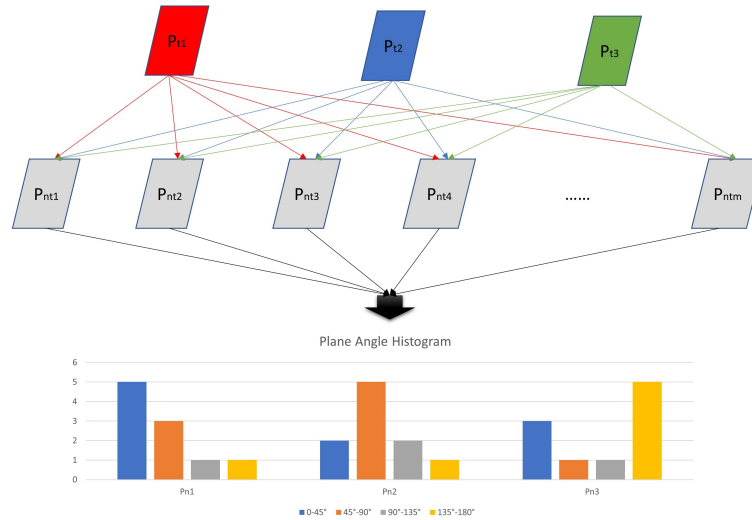


Figure 5.12: Histogram of plane angles [187].

### 5.3.2 Experiments and results

The main experimental object for this study is a pneumatically driven directional gyro, manufactured by GM Corp in US, as shown in Figure 7.8. More information could be obtained from Appendix B. The object is chosen due to its rich geometric primitives design while few evident corner points within the rounded edges could be manually picked for CT and photogrammetry point cloud registration. As for photogrammetric 3D reconstructions, Figure 5.13 shows the camera poses represented by circular white rectangles via SfM and the watertight 3D surface model from DIM and texturing procedure. CT is used to reconstruct internal structures of an

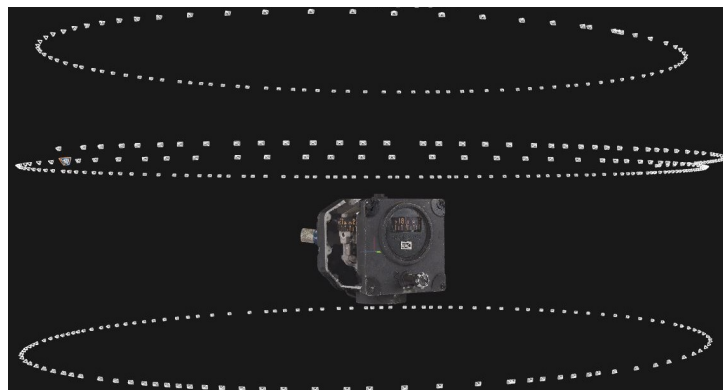


Figure 5.13: Photogrammetric 3D surface model of KK10-09.

object that is not visible from outside. Left image of Figure 5.14 shows the complete CT recon-



structured 3D model visualized in VGStudio [169]. The right image of Figure 5.14 is the point cloud of CT data by representing each voxel by the center point with its intensity value.

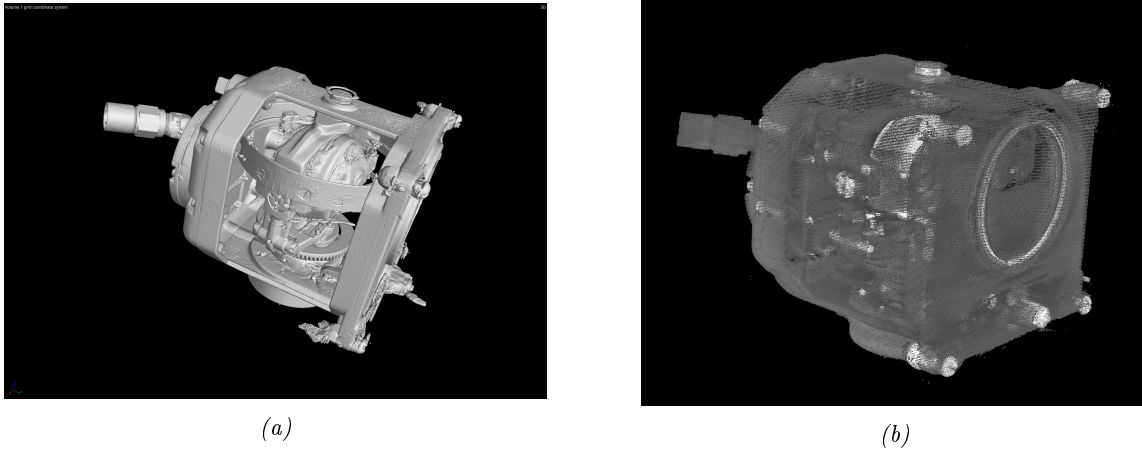


Figure 5.14: 3D model of the object: (a) 3D volumetric model [47]; (b) point cloud @ Gasim Mammadov

### 5.3.2.1 Point clouds pre-processing

The transformation from CT volume data to point clouds is implemented according to the introduction in section 4.2.2. Thereafter a CT surface extraction is necessary to provide the similar information as given by the photogrammetric point cloud for registration. We apply a ball query method for each point to extract the points with less neighbouring points within a threshold range, which are interpreted as the points on the surface of the 3D CT model. An alternative method finding the convex hull could be found in 4.2.2.2. Figure 5.15 shows the surface extraction result. In addition, by removing this shell of certain thickness will also avoid the visualization ambiguity of the integrated model on the surface, which could be seen on the integrated model of Golden Gnat from Figure 5.8.

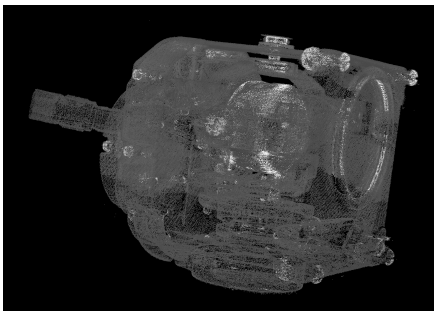


Figure 5.15: Extracted surface @ Gasim Mammadov.

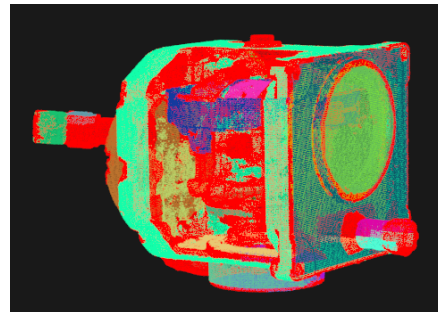


Figure 5.16: Segmentation result.

### 5.3.2.2 Point cloud segmentation and primitive fitting

After obtaining a CT surface and photogrammetric point cloud, the implementation of a direct geometry primitive fitting may result in unsatisfying output, due to the highly complex structure of gyroscopes. Therefore, a point cloud segmentation as shown in Figure 5.16 using region growing methods based on different principles is necessary to be conducted beforehand, considering the differences of the sensors as well as the object structure characteristics.

A well-segmented point cloud contributes to an easier and more precise best-fit of geometric

primitives based on RANSAC. The plane, as the most frequent geometric primitive within gyroscopes, is determined by the inliers, i.e. extracted co-planar points with a predefined threshold. Figure 5.17 displays examples of plane and cylinder fitting. A control point could be determined by the intersection of every three non-parallel parameterized planes. In addition, other best-fit primitives such as spheres or 3D circles could provide their center points directly as control points.

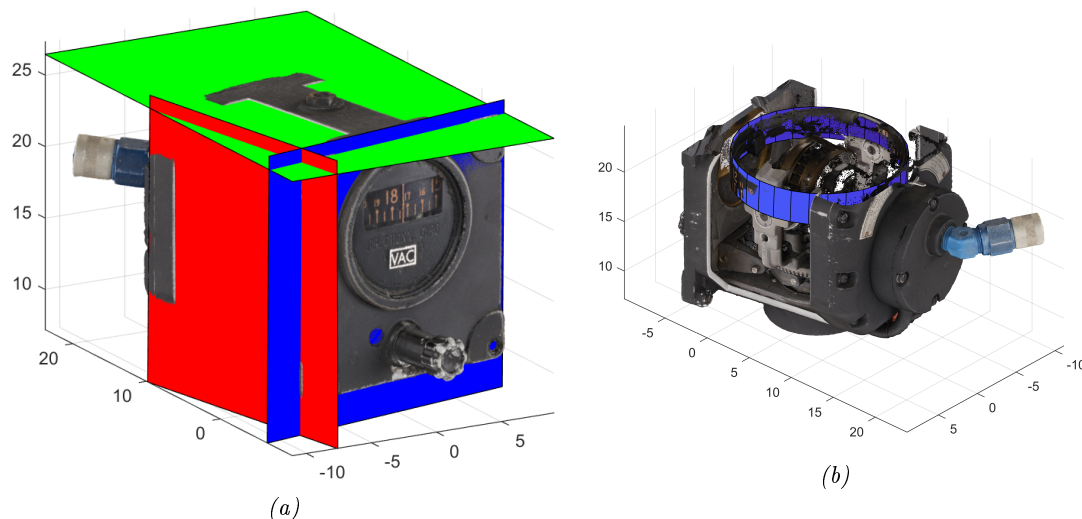


Figure 5.17: Geometric primitive fitting of the object: (a) plane fitting; (b) cylinder fitting.

### 5.3.2.3 Point cloud registration

In the next step, two sets of points are obtained as control points for the estimation of the transformation parameters as shown in Figure 5.18. For the differences between the two data-sets as well as the primitive fitting procedures, the control points may be disordered and different in quantity. To determine the transformation, the control points together with their descriptors as introduced in Section 5.3.1.2 are to be estimated for corresponding point pairs. The obtained result is used for the transformation matrix calculation as initial registration for the later refinement by ICP and/or the Gauss-Helmert Model. With two transformation matrices, the photogrammetry point cloud and the CT point cloud could be registered together. In our experiment both refinement methods delivered good registration results without too much differences. Cross sections of the photogrammetry and CT models are displayed in Figure 5.19.

## 5.4 Concluding remarks

Though the method in Section 5.1 is easy to implement and has precision estimation, the manual process of control points picking is much more complex than in 2D images, due to the lack of color information for CT data and the influence of the unsharpness of the photogrammetry data. In this case, increasing the number of the control points and considering the spatial distribution of them could improve the precision of this method.

When CT surface representation shows good completeness and good overlap with the photogrammetric point cloud, the registration could be further developed into a CT surface coloring process with corresponding photogrammetric images. The main point of the method is that the advantages of CT data's high geometric precision and photogrammetry data's color information are combined together. Nevertheless, the experiment is implemented with some assumptions, which can also be regarded as limitations of the methods such as the robustness of the CT data threshold extraction .



Figure 5.18: Virtual control points of photogrammetric model.

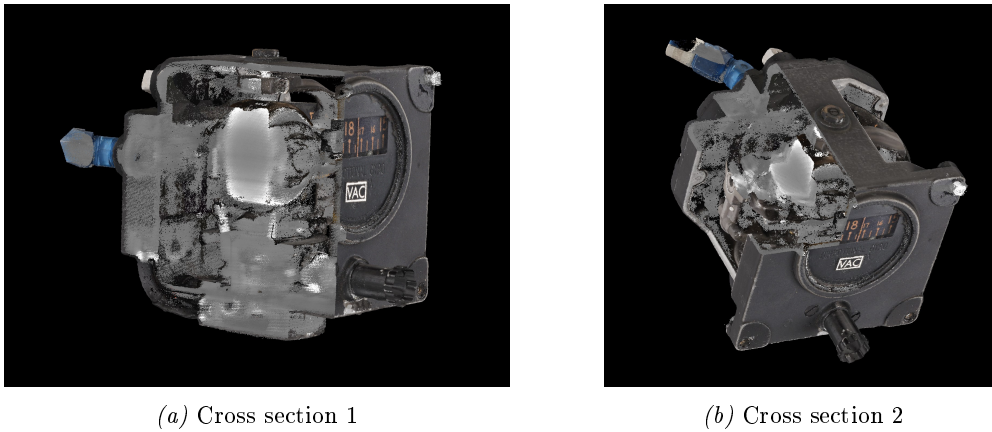


Figure 5.19: Cross sections of the integrated model.

To further increase the automation level of the data integration process, our method using virtual control points which is independent from the accuracy of the local control points shows good performance in practice. Specifically, the proposed workflow has advantages in the following aspects

- robustness against the noise of the point cloud;
- robustness against incomplete CT data due the low penetration of some materials;
- high efficiency on condition of good primitive fitting results.

The proposed workflow involves several steps to reach the final registration. However, each single step could be improved, such as well-extracted CT surfaces, more robust point cloud segmentation, more adaptive primitive fittings regarding different objects as well as finding a more efficient descriptors for the virtual control points.



## Chapter 6

# CSG Modeling

Among all the 3D sensors applied in our research, the results of photogrammetry and endoscopy are point clouds or meshes, the CT outcome is a volumetric model. Although those dense discrete representation could well recover the geometric information of the objects, all the data has too big data volume, which lead to obvious disadvantages such as the long processing time, efficiency in real-time visualization for web-based applications or VR environment renderings. In addition, the unstructured characteristic makes it difficult for user interactions. Among many solutions proposed, one common way in practice is to vectorize the data into CSG (constructive solid geometry) model for a more efficient handling of the visualization. In section 6.1, the manual CSG modeling implementation and the semi-automatic modeling methods are introduced based on point clouds via the RANSAC principle. Section 6.2 gives an overview of the application with deep learning.

### 6.1 Manual and semi-automatic CSG modeling from point cloud

#### 6.1.1 Manual CSG modeling

##### 6.1.1.1 Theory

Point clouds or structured models obtained through computer vision and photogrammetry or other related methods, are used as references in the 3D modeling process using software such as Autodesk 3ds Max, Maya etc,. The general workflow is shown in Figure 6.1.

##### 6.1.1.2 Experiments and results

The object BM0610, an accelerometer, serves as an example (see introduction in Appendix B), which has advantages in processing artificial objects for CSG modeling due to the regular shape design in many cases. Thus, default geometry primitives could be directly fitted into the point clouds or through boolean operation between diverse primitives. However, the fitting process is challenging due to the fact that it involves an alignment of 3 axes between the reference point cloud and the geometric primitives. An effective solution would be conducting cross-sections of a well-positioned reference model for the 2D outlines from different sections, which could be seen from the left image of Figure 6.2. Subsequently, the 2D cuts were extruded into 3D shapes according to their height as shown on the right image of Figure 6.2.

Different components are then assembled with boolean operations within the program to model the real geometric relationship of the object. After the CSG generation process, the texture will be baked from the reference data for a photo-realistic model of the original object. The assembled model and textured CSG model are shown in Figure 6.3

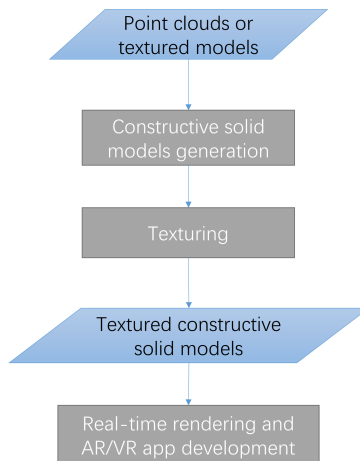


Figure 6.1: 3D modeling workflow.

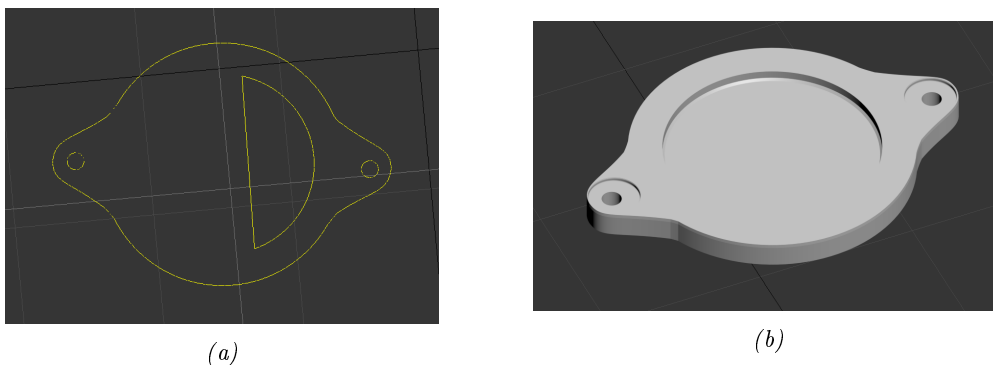


Figure 6.2: An example (a) a cross section of the original model; (b) extrusion of the 2D cuts.

As the results, the animation or rendering in a VR environment could be implemented with the generated 3D CSG model. An example of the cross section design is developed and on display in Figure 6.4.

### 6.1.2 RANSAC-based CSG modeling

To realize the automatic CSG modeling from the 3D data, RANSAC (see Appendix A1) has been the most widely used method for decades. In our case with the input of point clouds containing rich geometric primitives, the minimal set is randomly selected for an initial estimation of the parameters for the shape. All other points are tested with the estimated shape to find inliers, the maximal of which will be the best fitted primitive. To achieve a satisfying results, the related parameters should be fine-tuned. Some examples could be found in Section 5.3.2.2. For large datasets with various geometric primitives, the above process will be run iteratively. After successful detection of one shape, the remaining points will repeat the step until no more shape could be fitted. However, the uniform parameter setting is difficult in the iterative process for all shapes. In the work [139], the number of minimal datasets as well as the fitting quality are solved. The minimal set is estimated adaptively according to the outlier ratio of the point cloud and the probability of the fitting. In addition, the fitting quality is measured using a score function based on the point number, the normal as well as the connectivity. Another widely applied method is based on Hough transformation for both 2D and 3D applications. The main idea is to transform the input from the Cartesian space to the Hough space, which is denoted

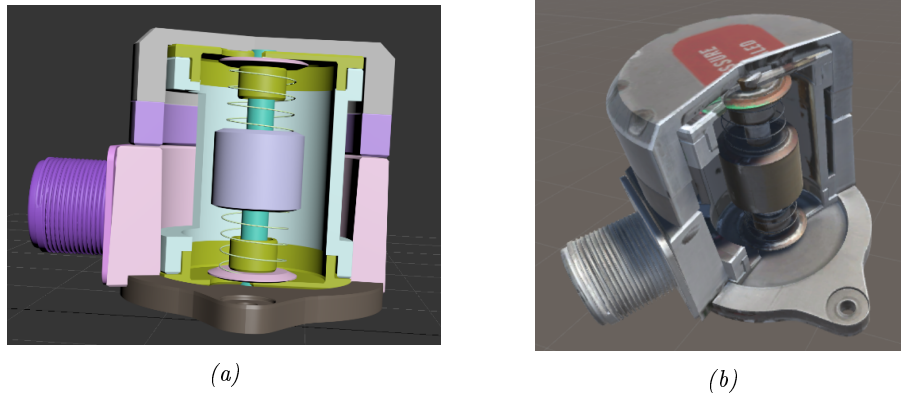


Figure 6.3: An example (a) 3D reconstructed model; (b) modeling of solid surfaces.

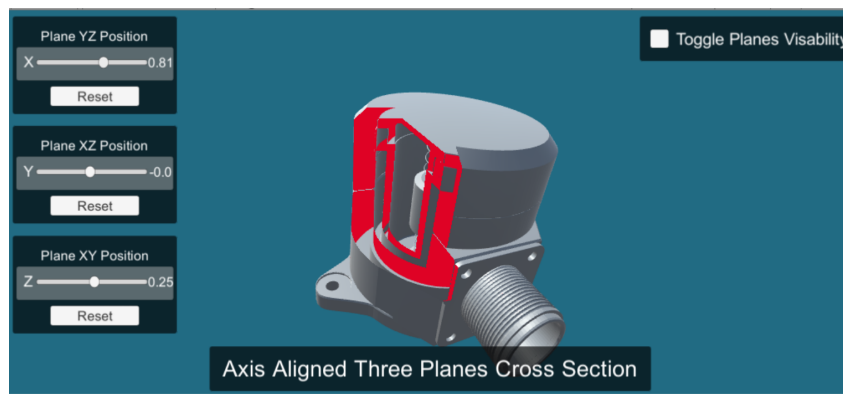


Figure 6.4: Virtual control points of photogrammetric model.

using polar coordinates. According to the transformed results, each cell will be voted for the selection of the best fitted shape with the most votes.

## 6.2 Deep learning-based CSG modeling using point clouds

Deep learning has been becoming more and more popular in the application solving many 2D vision problems. With reference to 3D representations such as voxels, patches, implicit representations, boxes or gaussians, the deep learning could also be used for different purposes. However, for the widely used 3D data types in reality, such as point clouds, the unstructured nature as well as the few data-sets have significantly limited the application of deep learning. Some works transform the point cloud data into multiview images or volume data to avoid the disadvantages, while this will also introduce new challenges. Nevertheless, related references, especially concerning the CSG modeling are still boosting with an increasing number of methods being proposed. For different deep learning methods to solve the CSG modeling from 3D data, table 6.1 summarizes regarding the important factors.

With a comprehensive review of the related works, it could be summarized as follows:

1. **Training data source** The training data are limited and some popular datasets frequently used in the point cloud deep learning research such as ABC [82] and ShapeNet [30] are mostly simple shapes such as chairs, tables and airplanes. In addition, manual created CAD models rather than real scanning data are provided. Though the training of those data could achieve promising results under certain conditions, it is impractical to apply those deep learning methods with complex 3D models.
2. **Input data for deep learning** The input data for the deep learning methods contains

Table 6.1: Summary of employed datasets

	Input	Supervision	Data characteristics	Primitive type	Prediction	Strengthens
[90]	Point cloud	Supervised	ANSI Mechanical Component dataset	plane, sphere, cylinder, cone	CSG Model	(1)direct application from point cloud to CSG model; (2) test with real scan data
[143]	Depth image	unsupervised	furniture	sphere, cylinder, cube	CSG Model	(1) the input could be 2D or 3D for generating the shape;(2)recurrent neural network parsing the shape in a top-down manner
[168]	volumetric grid	unsupervised	chair, airplane from ShapeNet [30]	cuboids	volumetric primitives	data-driven 3D understanding with cuboidal primitives;
[130]	Point cloud	unsupervised	toy dataset generated from OpenSCAD	box, sphere, cylinder, cone	CSG stump	(1) interpretable, compact and editable nature of CSG-Stump, but has no complex tree structure; (2)simple and regular structure, and constant dimensionality for deep learning
[31]	64 <sup>3</sup> volume	unsupervised	airplane, car,chair,lamp,table	convex sets	polygonal meshes	BSP-Net uses few primitives for reconstruction; (2)no convex shape decompositions are needed for training
[197]	depth image	supervised	chair, table, nightstand from ShapeNet [30]	cuboid	shape primitive	(1)models complex shapes with a low parametric model; (2)Evaluations on synthetic and real depth map
[112]	RGB image	unsupervised	chair, table, airplane from ShapeNet [30]	cuboid	3D shape structure	detailed recovery of diverse 3D part structures from single-view 2Dimages
[143]	64 <sup>3</sup> voxel grid	supervised	Synthetic 3D shape	sphere, cylinder, cube	CSG shape	(1)recurrent neural network parsing the shape in a top-down manner;(2)effective shape detection
[51]	partial point cloud	supervised	shape collection with self-defined template	box	structural template and semantic labels	automatic inferring of CSG programs from 2D and 3D shapes
[56]	water tight mesh	self-supervised	airplanes, cars, and chairs from ShapeNet [30]	implicit shape	parameters for shape template	constructing a general shape parametric template that fits any shape
[121]	point cloud	supervised	40 man made object including tables, chairs, airplanes etc. from ShapeNet [30]	-	segmentation	consuming point cloud with permutation invariance for effective classification and segmentation
[35]	single image	self supervised	man made objects	convex polytopes	polygonal mesh	(1) outperformance regarding reconstruction; (2)direct downstream application of the result
[116]	single image	unsupervised	humans and man made objects	cuboids and superquadric surfaces	primitive and latent hierarchical structure	(1) structure prediction without supervision; (2)variety of experimental data
[79]	CAD model	unsupervised	chair, desk, and lamps	box and sphere	CSG parse trees	Boolean operation of CSG primitives for meshes
[181]	point cloud	supervised	unkown	CAD model	CAD command sequence	unknown
[144]	point cloud	supervised	CAD models of mechanical objects	parametric surface patches	(1)parameters for the primitive type	fitted parametric surfaces including B-spline patches and geometric primitives; (2) rich class of primitives
[96]	point cloud	supervised	ABC and PartNet dataset	-	part boundaries	accurate localization boundaries of semantic parts or geometric primitives



single images, multi-view images, voxels, meshes, CAD models as well as point clouds. The works are dealing with single RGB images, range images or partial point clouds focus more on the scene understanding rather than the precise 3D modeling due to the limited information from the input. The methods using voxel based datasets have limitations on both the training accuracy as well as consistency of the conventional data collection. Though there are some references applying point clouds, they are mostly converted from the CAD models instead of coming from the real scans, which indicates the gap between the experimental data and the reality.

3. **Tasks of deep learning** Most of the references concentrate on the tasks such as classification, segmentation, parsing or a rough fitting, which attach more importance to the 3D scene understanding instead of precise reconstruction. On the one hand, it is of vital importance to understand the 3D scenes more effectively rather than the complete visualization. On the other hand, limited data sources for training and the complex characteristics have limited more researches on the 3D data modeling using real scan point clouds.
4. **Prediction results of deep learning** The prediction primitive results of reviewed studies contain boxes, cuboids, gaussians, convex, implicit structures, meshes as well as the CSG models. Besides, labels for the point cloud segment, boundaries of the point cloud part also draw researchers' attention. In brief, different 3D representations have different advantages for applications in reality.

Overall, the possibilities of collecting enough 3D data for training and the increase of the computation power have enabled the deep learning algorithm development for different 3D tasks. However, limited works have been focusing on the CSG modeling from the point cloud data. It is necessary to establish such benchmarks as well as to trigger more research enthusiasm of converting real 3D data into CSG models for wider application in the metaverse.

### 6.3 Concluding remarks

For the purpose of CSG modeling from point cloud data, manual and semi-automatic method based on RANSAC have been conducted and experimented with the gyroscopes. For manual CSG modeling, various software could be applied for the primitive fitting process and for the texture rendering. As for the missing information from the original input point cloud, symmetric characteristics or other empirical knowledge could be applied for a complete modeling. In addition, physical attributions and animations are possible to be added to the final CSG model. However, the process is expertise demanding and could be very time consuming by fully manual operations. Therefore, the parameterized primitives could be defined and fitted to the point cloud data with the distance as the cost function, which is minimized in the process for a best fit. To achieve a satisfying fitting result, this semi-automatic method needs the input of appropriate parameters regarding the primitives. The manual and semi-automatic method could be applied in practice for a precise CSG modeling with the 3D data of the reality.

There is an evident increase of the interests in point cloud oriented deep learning methods for the purpose of segmentation, classification and abstraction etc,. For the understanding of the 3D data, many works have already shown great potential in different application scenarios. However, for a precise CSG modeling of the objects in reality with certain complexity, deep learning based algorithm could play an assisting role for the acceleration of the process instead of completely substituting the workflow introduced in Section 6.1.



## Chapter 7

# Discussion and Conclusion

### 7.1 Contributions

The thesis discusses the possibility of applying three different methods, photogrammetry, endoscopy, and CT for the 3D digitization of gyroscopes, which is quite different from the previous virtual museum-related research projects using mainly photogrammetry and laser scanning and concentrating on cultural relics in the common sense. In addition, the main goal of reconstruction was the geometric restoration of the outer surfaces of an object. Our research has made breakthroughs in the above three aspects and is an innovative representation of this type of research. Regarding the objects and technical methods involved in our application we could provide the following contributions: (1) It is a pioneering work on the three-dimensional digitization of industrial cultural relics, which enriches the context. (2) Pre-calibration of the sensors for 3D digitization using different calibration models for comparison as well as analyzing the stability and reliability quantitatively. (3) Industrial endoscopes were introduced, and low-quality photo pre-processing and 3D reconstructions obtained by special sensors were accomplished. (4) We introduced CT as a sensor for internal information acquisition, and analyzed the advantages and characteristics of CT data to realize the integration with other data. (5) We explored the possibility of different point cloud vectorization, practiced manual and RANSAC based methods as well as provided a comprehensive review of the application of deep learning.

### 7.2 Discussions

Concerning the contributions of this work, the points that are necessary to be discussed are listed below:

- The research of camera calibration quantitatively analyzes the stability and reliability of commonly used calibration methods, and provides good guidance for practice, but lacks automation and a more user-friendly design. In the research of endoscopic correction we studied, how to obtain high-quality and good-configuration photos, so that the model comparison could finally be applied;
- In terms of the 3D reconstruction using an endoscope, due to the small diameter, it has brought great difficulties to the relevant reconstruction work. An endoscope with a more suitable diameter should be used, considering the size of the accessible cavity and the endoscope. At the same time, more reasonable photographing schemes should be explored to better integrate with other data;
- In terms of 3D data fusion, the coloring process of the CT surface point cloud combines the advantages of photogrammetry and CT data to overcome missing material information. The problem has not yet been fully explored by our research. The method of virtual

control points has a certain dependence on the quality of geometric feature extraction. This process has not yet been fully automated to obtain high-precision results, and the descriptors of control points should be further studied to achieve a more efficient matching process.

### **7.3 Future work**

Based on this research, future work should have the following work directions: (1) Improving the process of endoscopy three-dimensional reconstruction, and trying to apply the corresponding photogrammetry or CT data assistance. (2) In addition to data fusion, further explorations can be made. The characteristics of the object itself are rich geometric features. A more robust and automated, high-precision fusion scheme is carried out through the application of features and the analysis of the characteristics of different sensor data. Therefore we should search for a more automated integrated data fusion scheme. (3) In the data application room, different principles can be compared. RANSAC and deep learning methods should be tested on our data, while considering the characteristics of the three-dimensional gyroscope data. On the basis of extraction, explore the automatic extraction of combined shapes, apply deep learning methods for CSG modeling, and expand the application of related data in the VR/AR field.

# Bibliography

- [1] A. H. Ahmadabadian, S. Robson, J. Boehm, and M. Shortis. Stereo-imaging network design for precise and dense 3d reconstruction. *The Photogrammetric Record*, 29(147):317–336, 2014.
- [2] D. Akca. Co-registration of surfaces by 3d least squares matching. *Photogrammetric Engineering & Remote Sensing*, 76(3):307–318, 2010.
- [3] J. Albrecht. Albrecht meymenbauer-pioneer of photogrammetric documentation of the cultural heritage. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(5/C7):19–25, 2002.
- [4] P. K. Allen, A. Troccoli, B. Smith, I. Stamos, and S. Murray. The beauvais cathedral project. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, volume 1, pages 10–10. IEEE, 2003.
- [5] M. Alshawa. lcl: Iterative closest line a novel point cloud registration algorithm based on linear features. *Ekscentar*, (10):53–59, 2007.
- [6] A. M. Andrew. Multiple view geometry in computer vision. *Kybernetes*, 2001.
- [7] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *EURASIP Journal on Image and Video Processing*, 2010:1–12, 2010.
- [8] J. Balsa-Barreiro and D. Fritsch. Generation of 3d/4d photorealistic building models. the testbed area for 4d cultural heritage world project: The historical center of calw (germany). In *International Symposium on Visual Computing*, pages 361–372. Springer, 2015.
- [9] J. Barreto, J. Roquette, P. Sturm, and F. Fonseca. Automatic camera calibration applied to medical endoscopy. In *BMVC 2009-20th British Machine Vision Conference*, pages 1–10. The British Machine Vision Association (BMVA), 2009.
- [10] F. Bergamasco, L. Cosmo, A. Gasparetto, A. Albarelli, and A. Torsello. Parameter-free lens distortion calibration of central cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3847–3855, 2017.
- [11] T. Bergen and T. Wittenberg. Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods. *IEEE journal of biomedical and health informatics*, 20(1):304–321, 2014.
- [12] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, G. Guennebaud, J. A. Levine, A. Sharf, and C. T. Silva. A survey of surface reconstruction from point clouds. In *Computer Graphics Forum*, volume 36, pages 301–329. Wiley Online Library, 2017.
- [13] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.
- [14] A. Bezzi, L. Bezzi, R. Gietl, and N. Pisu. Archeos and uavp, a free and open source platform for remote sensing: the case study of monte s. martino ai campi di riva del garda (italy). *Archaeology in the Digital Era*, pages 792–799, 2013.
- [15] S. Bianco, G. Ciocca, and D. Marelli. Evaluating the performance of structure from motion pipelines. *Journal of Imaging*, 4(8):98, 2018.

- [16] F. Blais. Review of 20 years of range sensor development. *Journal of electronic imaging*, 13(1):231–243, 2004.
- [17] F. E. Boas, D. Fleischmann, et al. Ct artifacts: causes and reduction techniques. *Imaging Med*, 4(2):229–240, 2012.
- [18] J. G. F. Bohnenberger. Beschreibung einer Maschine zur Erläuterung der Geseze der Umdrehung der Erde um ihre Axe, und der Veränderung der Lage der letzteren, 1817.
- [19] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pages 790–796. IEEE, 1998.
- [20] J.-Y. Bouguet. Camera calibration toolbox for matlab. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html), 2004.
- [21] J.-Y. Bouguet. Camera calibration toolbox for matlab (2008). URL [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc), 1080, 2008.
- [22] A. Boulch, B. Le Saux, and N. Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. *3DOR@ Eurographics*, 3, 2017.
- [23] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother. Dsac-differentiable ransac for camera localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6684–6692, 2017.
- [24] P. Breuer and P. Major. Central limit theorems for non-linear functionals of gaussian fields. *Journal of Multivariate Analysis*, 13(3):425–441, 1983.
- [25] D. C. Brown. An advanced reduction and calibration for photogrammetric cameras. Technical report, Instrument Corp of Florida Melbourne, 1964.
- [26] I. Budak, Z. Santosi, V. Stojakovic, D. Korolija Crkvenjakov, R. Obradovic, M. Milosevic, and M. Sokac. Development of expert system for the selection of 3d digitization method in tangible cultural heritage. *Tehnički vjesnik*, 26(3):837–844, 2019.
- [27] F. Bukhari and M. N. Dailey. Automatic radial distortion estimation from a single image. *Journal of mathematical imaging and vision*, 45(1):31–45, 2013.
- [28] W. Burger. Zhang’s camera calibration algorithm: In- depth tutorial and implementation. *HGB16-05*, pages 1–6, 2016.
- [29] Capture Reality, Slovakia. *RealityCapture*, 2021.
- [30] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.
- [31] Z. Chen, A. Tagliasacchi, and H. Zhang. Bsp-net: Generating compact meshes via binary space partitioning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 45–54, 2020.
- [32] E. Cieslik. 3d digitization in cultural heritage institutions guidebook.
- [33] H. Cui, X. Gao, S. Shen, and Z. Hu. Hsfm: Hybrid structure-from-motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1212–1221, 2017.
- [34] S. Daftry, M. Maurer, A. Wendel, and H. Bischof. Flexible and user-centric camera calibration using planar fiducial markers. In *BMVC*. Citeseer, 2013.
- [35] B. Deng, K. Genova, S. Yazdani, S. Bouaziz, G. Hinton, and A. Tagliasacchi. Cvxnet: Learnable convex decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 31–44, 2020.
- [36] P. Dondi, L. Lombardi, M. Malagodi, and M. Licchelli. 3d modelling and measurements of historical violins. *Acta IMEKO*, 6(3):29–34, 2017.
- [37] C. B. Duane. Close-range camera calibration. *Photogramm. Eng*, 37(8):855–866, 1971.
- [38] H. Ebner. Self calibrating block adjustment. *Bildmessung und Luftbildwissen*, 44:128–139, 1976.

- [39] G. Facciolo, C. De Franchis, and E. Meinhardt. Mgm: A significantly more global matching for stereovision. In *BMVC 2015*, 2015.
- [40] Y. Fan, M. Q.-H. Meng, and B. Li. 3d reconstruction of wireless capsule endoscopy images. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 5149–5152. IEEE, 2010.
- [41] M. Fang, T. Pollok, and C. Qu. Merge-sfm: Merging partial reconstructions. In *BMVC*, page 29, 2019.
- [42] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *European conference on computer vision*, pages 321–334. Springer, 1992.
- [43] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.
- [44] W. Förstner. A feature based correspondence algorithm for image matching. *ISPRS ComIII, Rovaniemi*, pages 150–166, 1986.
- [45] C. S. Fraser. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote sensing*, 52(4):149–159, 1997.
- [46] C. S. Fraser and S. Al-Ajlouni. Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 72(9):1017–1026, 2006.
- [47] D. Fritsch, J. F. Wagner, B. Ceranski, S. Simon, M. Niklaus, K. Zhan, and G. Mammadov. Making historical gyroscopes alive—2d and 3d preservations by sensor fusion and open data access. *Sensors*, 21(3):957, 2021.
- [48] D. Fritsch, J. F. Wagner, S. Simon, B. Ceranski, M. Niklaus, K. Zhan, T. Schweizer, and Z. Wang. Gyrolog—towards vr preservations of gyro instruments for historical and didactical research. In *2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, pages 1–7. IEEE, 2018.
- [49] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1434–1441. IEEE, 2010.
- [50] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis (pmvs). In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 2007.
- [51] V. Ganapathi-Subramanian, O. Diamanti, S. Pirk, C. Tang, M. Niessner, and L. Guibas. Parsing geometry using structure-aware shape templates. In *2018 International Conference on 3D Vision (3DV)*, pages 672–681. IEEE, 2018.
- [52] Z. Gao, M. Zhu, and J. Yu. A self-identifying checkerboard-like pattern for camera calibration. In *2020 39th Chinese Control Conference (CCC)*, pages 6351–6356. IEEE, 2020.
- [53] Z. Gao, M. Zhu, and J. Yu. A novel camera calibration pattern robust to incomplete pattern projection. *IEEE Sensors Journal*, 21(8):10051–10060, 2021.
- [54] X. Ge and T. Wunderlich. Surface-based matching of 3d point clouds with variable coordinates in source and target system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 111:1–12, 2016.
- [55] J. Geng and J. Xie. Review of 3-d endoscopic surface imaging techniques. *IEEE Sensors Journal*, 14(4):945–960, 2013.
- [56] K. Genova, F. Cole, D. Vlasic, A. Sarna, W. T. Freeman, and T. Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7154–7164, 2019.
- [57] R. Gini, D. Pagliari, D. Passoni, L. Pinto, G. Sona, and P. Dosso. Uav photogrammetry: Block triangulation comparisons. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.*, 1:W2, 2013.
- [58] S. Gosh. History of photogrammetry. *Laval University, Canada*, 1981.

- [59] D. Grant, J. Bethel, and M. Crawford. Point-to-plane registration of terrestrial laser scans. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72:16–26, 2012.
- [60] A. Gruen. Progress in photogrammetric point determination by compensation of systematic errors and detection of gross errors. *Inform. Relative to Cartography and Geodesy. Ser. 2: Transl. p 113-140(SEE N 80-12455 03-43)*, 1978.
- [61] A. Gruen. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187, 1985.
- [62] A. Gruen and D. Akca. Least squares 3d surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(3):151–174, 2005.
- [63] R. A. Hamzah and H. Ibrahim. Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, 2016, 2016.
- [64] C. Harris, M. Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.
- [65] R. Hartley and S. B. Kang. Parameter-free radial distortion correction with center of distortion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1309–1321, 2007.
- [66] H. Hastedt and T. Luhmann. Investigations on the quality of the interior orientation and its impact in object space for uav photogrammetry. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 40, 2015.
- [67] F. He and A. Habib. Target-based and feature-based calibration of low-cost digital cameras with large field-of-view. In *Proceedings of the ASPRS 2015 Annual Conference, Tampa, FL, USA*, pages 4–8, 2015.
- [68] J. Heikkila and O. Silvén. A four-step camera calibration procedure with implicit image correction. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pages 1106–1112. IEEE, 1997.
- [69] J. P. Helferty, C. Zhang, G. McLennan, and W. E. Higgins. Videoendoscopic distortion correction and its application to virtual guidance of endoscopy. *IEEE transactions on medical imaging*, 20(7):605–617, 2001.
- [70] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341, 2007.
- [71] A. Holynski, D. Geraghty, J.-M. Frahm, C. Sweeney, and R. Szeliski. Reducing drift in structure from motion using extended features. In *2020 International Conference on 3D Vision (3DV)*, pages 51–60. IEEE, 2020.
- [72] A. Hosseininaveh, B. Sargeant, T. Erfani, S. Robson, M. Shortis, M. Hess, and J. Boehm. Towards fully automatic reliable 3d acquisition: From designing imaging network to a complete and accurate point cloud. *Robotics and Autonomous Systems*, 62(8):1197–1207, 2014.
- [73] P. V. Hough. Method and means for recognizing complex patterns. *US patent*, 3(6), 1962.
- [74] J. Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*, volume 114. SPIE press, 2003.
- [75] K. Ikeuchi, T. Oishi, J. Takamatsu, R. Sagawa, A. Nakazawa, R. Kurazume, K. Nishino, M. Kamakura, and Y. Okamoto. The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *International Journal of Computer Vision*, 75(1):189–208, 2007.
- [76] M. Jaud, S. Passot, R. Le Bivic, C. Delacourt, P. Grandjean, and N. Le Dantec. Assessing the accuracy of high resolution digital surface models computed by photoscan® and micmac® in sub-optimal survey conditions. *Remote Sensing*, 8(6):465, 2016.
- [77] Z. Jian-Xun, H. Ming-Hui, and D. Yu. Three-dimensional porous structure reconstruction for low-resolution monocular endoscopic images.
- [78] G. Jiang and L. Quan. Detection of concentric circles for camera calibration. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 333–340. IEEE, 2005.



- [79] K. Kania, M. Zięba, and T. Kajdanowicz. Ucs-g-net—unsupervised discovering of constructive solid geometry tree. *arXiv preprint arXiv:2006.09102*, 2020.
- [80] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006.
- [81] K. Keklikoglou, S. Faulwetter, E. Chatzinikolaou, P. Wils, J. Brecko, J. Kvaček, B. Metscher, and C. Arvanitidis. Micro-computed tomography for natural history specimens: a handbook of best practice protocols. *European Journal of Taxonomy*, (522), 2019.
- [82] S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin, and D. Panozzo. Abc: A big cad model dataset for geometric deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9601–9611, 2019.
- [83] M. Krumm, S. Kasperl, and M. Franz. Reducing non-linear artifacts of multi-material objects in industrial 3d computed tomography. *Ndt & E International*, 41(4):242–251, 2008.
- [84] F. Lafarge and C. Mallet. Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. *International journal of computer vision*, 99(1):69–85, 2012.
- [85] D. H. Laidlaw, W. B. Trumbore, and J. F. Hughes. Constructive solid geometry for polyhedral objects. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 161–170, 1986.
- [86] S. Lee, H. Lee, H. Choi, S. Jeon, and J. Hong. Effective calibration of an endoscope to an optical tracking system for medical augmented reality. *Cogent Engineering*, 4(1):1359955, 2017.
- [87] D. C. Leiner. *Digital endoscope design*. SPIE Press, 2016.
- [88] R. Lenz. Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3d—Messungen in Echtzeit. In *Mustererkennung 1987*, pages 212–216. Springer, 1987.
- [89] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144, 2000.
- [90] L. Li, M. Sung, A. Dubrovina, L. Yi, and L. J. Guibas. Supervised fitting of geometric primitives to 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2652–2660, 2019.
- [91] R. Li, D. Gong, J. Sun, Y. Zhu, Z. Wei, and Y. Zhang. Robust and accurate hybrid structure-from-moti. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 494–498. IEEE, 2019.
- [92] W. Li, S. Nie, M. Soto-Thompson, C.-I. Chen, and Y. I. A-Rahim. Robust distortion correction of endoscope. In *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, volume 6918, page 691812. International Society for Optics and Photonics, 2008.
- [93] Z. Li, J. Chen, and E. Baltsavias. *Advances in photogrammetry, remote sensing and spatial information sciences: 2008 ISPRS congress book*, volume 7. CRC Press, 2008.
- [94] A. Lingua, D. Marenchino, and F. Nex. Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors*, 9(5):3745–3766, 2009.
- [95] A. Locher, M. Havlena, and L. Van Gool. Progressive structure from motion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 20–35, 2018.
- [96] M. Loizou, M. Averkiou, and E. Kalogerakis. Learning part boundaries from 3d point clouds. In *Computer Graphics Forum*, volume 39, pages 183–195. Wiley Online Library, 2020.
- [97] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133–135, 1981.
- [98] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [99] T. Luhmann. Close range photogrammetry for industrial applications. *ISPRS journal of photogrammetry and remote sensing*, 65(6):558–569, 2010.

- [100] T. Luhmann, C. Fraser, and H.-G. Maas. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:37–46, 2016.
- [101] T. Luhmann, S. Robson, S. Kyle, and J. Boehm. *Close-range photogrammetry and 3D imaging*. de Gruyter, 2019.
- [102] D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 200(1140):269–294, 1978.
- [103] R. Melo, J. P. Barreto, and G. Falcao. A new solution for camera calibration and real-time image distortion correction in medical endoscopy—initial technical evaluation. *IEEE Transactions on Biomedical Engineering*, 59(3):634–644, 2011.
- [104] X. Meng and Z. Hu. A new easy camera calibration technique based on circular points. *Pattern Recognition*, 36(5):1155–1164, 2003.
- [105] A. Miron, S. Ainouz, A. Rogozan, and A. Benschair. A robust cost function for stereo matching of road scenes. *Pattern Recognition Letters*, 38:70–77, 2014.
- [106] O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In *Machine Vision Applications in Industrial Inspection XIII*, volume 5679, pages 178–186. International Society for Optics and Photonics, 2005.
- [107] V. Mousavi, M. Khosravi, M. Ahmadi, N. Noori, S. Haghshenas, A. Hosseinaveh, and M. Varshosaz. The performance evaluation of multi-image 3d reconstruction software with different sensors. *Measurement*, 120:1–10, 2018.
- [108] M. G. Mozerov and J. Van De Weijer. Accurate stereo matching by two-step energy minimization. *IEEE Transactions on Image Processing*, 24(3):1153–1163, 2015.
- [109] B. Münzer, K. Schoeffmann, and L. Böszörményi. Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications*, 77(1):1323–1362, 2018.
- [110] L. Nalpantidis and A. Gasteratos. Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence. *Robotics and Autonomous Systems*, 58(5):457–464, 2010.
- [111] L. Nalpantidis and A. Gasteratos. Stereo vision for robotic applications in the presence of non-ideal lighting conditions. *Image and Vision Computing*, 28(6):940–951, 2010.
- [112] C. Niu, J. Li, and K. Xu. Im2struct: Recovering 3d shape structure from a single rgb image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4521–4529, 2018.
- [113] OpenCV. Open source computer vision library, 2015.
- [114] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [115] M. Parchami, J. A. Cadeddu, and G.-L. Mariottini. Endoscopic stereo reconstruction: A comparative study. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2440–2443. IEEE, 2014.
- [116] D. Paschalidou, L. V. Gool, and A. Geiger. Learning unsupervised hierarchical part decomposition of 3d objects from a single rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1060–1070, 2020.
- [117] V. Pătrăucean, I. Armeni, M. Nahangi, J. Yeung, I. Brilakis, and C. Haas. State of research in automatic as-built modelling. *Advanced Engineering Informatics*, 29(2):162–171, 2015.
- [118] S. Peng and P. Sturm. Calibration wizard: A guidance system for camera calibration based on modelling geometric and corner uncertainty. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1497–1505, 2019.
- [119] M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, PhD thesis, ESAT-PSI, KU Leuven, 1999.

- [120] A.-S. Poulin-Girard, S. Thibault, and D. Laurendeau. Influence of camera calibration conditions on the accuracy of 3d reconstruction. *Optics express*, 24(3):2678–2686, 2016.
- [121] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.
- [122] R. Ranftl and V. Koltun. Deep fundamental matrix estimation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 284–299, 2018.
- [123] I. Reisner-Kollmann. *Reconstruction of 3D Models from Images and Point Clouds with Shape Primitives*. PhD thesis, Reisner-Kollmann, 2013.
- [124] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein. Deep convolutional denoising of low-light images. *arXiv preprint arXiv:1701.01687*, 2017.
- [125] F. Remondino, S. Del Pizzo, T. P. Kersten, and S. Troisi. Low-cost and open-source solutions for automated image orientation—a critical overview. In *Euro-Mediterranean Conference*, pages 40–54. Springer, 2012.
- [126] F. Remondino and S. El-Hakim. Image-based 3d modelling: a review. *The Photogrammetric Record*, 21(115):269–291, 2006.
- [127] F. Remondino and C. Fraser. Digital camera calibration methods: considerations and comparisons. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5):266–272, 2006.
- [128] F. Remondino, E. Nocerino, I. Toschi, and F. Menna. A critical review of automated photogrammetric processing of large datasets. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42, 2017.
- [129] F. Remondino, M. G. Spera, E. Nocerino, F. Menna, and F. Nex. State of the art in high density image matching. *The Photogrammetric Record*, 29(146):144–166, 2014.
- [130] D. Ren, J. Zheng, J. Cai, J. Li, H. Jiang, Z. Cai, J. Zhang, L. Pan, M. Zhang, H. Zhao, et al. Csg-stump: A learning friendly csg-like representation for interpretable shape parsing. *arXiv preprint arXiv:2108.11305*, 2021.
- [131] A. Richardson, J. Strom, and E. Olson. Aprilcal: Assisted and repeatable camera calibration. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1814–1821. IEEE, 2013.
- [132] C. Ricolfe-Viala and A.-J. Sanchez-Salmeron. Camera calibration under optimal conditions. *Optics Express*, 19(11):10769–10775, 2011.
- [133] M. Rothmel, K. Wenzel, D. Fritsch, and N. Haala. Sure: Photogrammetric surface reconstruction from imagery. In *Proceedings LC3D Workshop, Berlin*, volume 8, 2012.
- [134] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pages 3212–3217. IEEE, 2009.
- [135] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *2011 IEEE international conference on robotics and automation*, pages 1–4. IEEE, 2011.
- [136] C.-A. Saint-Pierre, J. Boisvert, G. Grimard, and F. Cheriet. Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications*, 22(1):171–180, 2011.
- [137] S. Satoh. Simple low-dimensional features approximating ncc-based image matching. *Pattern Recognition Letters*, 32(14):1902–1911, 2011.
- [138] R. Schnabel, P. Degener, and R. Klein. Completion and reconstruction with primitive shapes. In *Computer Graphics Forum*, volume 28, pages 503–512. Wiley Online Library, 2009.
- [139] R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. In *Computer graphics forum*, volume 26, pages 214–226. Wiley Online Library, 2007.
- [140] J. L. Schonberger and J.-M. Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.

- [141] O. Semeniuta. Analysis of camera calibration with respect to measurement accuracy. *Procedia CIRP*, 41:765–770, 2016.
- [142] R. Shahidi, M. R. Bax, C. R. Maurer, J. A. Johnson, E. P. Wilkinson, B. Wang, J. B. West, M. J. Citardi, K. H. Manwaring, and R. Khadem. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Transactions on Medical imaging*, 21(12):1524–1535, 2002.
- [143] G. Sharma, R. Goyal, D. Liu, E. Kalogerakis, and S. Maji. Csgnet: Neural shape parser for constructive solid geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5515–5523, 2018.
- [144] G. Sharma, D. Liu, S. Maji, E. Kalogerakis, S. Chaudhuri, and R. Měch. Parsenet: A parametric surface fitting network for 3d point clouds. In *European Conference on Computer Vision*, pages 261–276. Springer, 2020.
- [145] K. Sharma, K.-y. Jeong, and S.-G. Kim. Vision based autonomous vehicle navigation with self-organizing map feature matching technique. In *2011 11th International Conference on Control, Automation and Systems*, pages 946–949. IEEE, 2011.
- [146] C.-H. Shen, H. Fu, K. Chen, and S.-M. Hu. Structure recovery by part assembly. *ACM Transactions on Graphics (TOG)*, 31(6):1–11, 2012.
- [147] N. Snavely. Bundler: Structure from motion (sfm) for unordered image collections. <http://phototour.cs.washington.edu/bundler/>, 2008.
- [148] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM siggraph 2006 papers*, pages 835–846. 2006.
- [149] T. D. Soper, M. P. Porter, and E. J. Seibel. Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance. *IEEE Transactions on Biomedical Engineering*, 59(6):1670–1680, 2012.
- [150] S. J. Spaner and G. L. Warnock. A brief history of endoscopy, laparoscopy, and laparoscopic surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 7(6):369–373, 1997.
- [151] C. Stamatopoulos and C. Fraser. Automated target-free network orientation and camera calibration. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2(5), 2014.
- [152] I. Stamos and M. Leordeanu. Automated feature-based range registration of urban scenes of large scale. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–ii. IEEE, 2003.
- [153] R. Strand and E. Hayman. Correcting radial distortion by circle fitting. In *BMVC*, 2005.
- [154] B. Ströbel, S. Schmelzle, N. Blüthgen, and M. Heethoff. An automated device for the digitization and 3d modelling of insects, combining extended-depth-of-field and all-side multi-view imaging. *ZooKeys*, (759):1, 2018.
- [155] P. Sun, N.-G. Lu, M.-L. Dong, B.-X. Yan, and J. Wang. Simultaneous all-parameters calibration and assessment of a stereo camera pair using a scale bar. *Sensors*, 18(11):3964, 2018.
- [156] W. Sun and J. R. Cooperstock. An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Machine Vision and Applications*, 17(1):51–67, 2006.
- [157] M. Sutton, I. Rahman, and R. Garwood. *Techniques for virtual palaeontology*. John Wiley & Sons, 2014.
- [158] C. Sweeney, V. Fragoso, T. Höllerer, and M. Turk. Large scale sfm with the distributed camera model. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 230–238. IEEE, 2016.
- [159] R. Tang. *Mathematical methods for camera self-calibration in photogrammetry and computer vision*. 2013.
- [160] R. Tausch, M. Domajnko, M. Ritz, M. Knuth, P. Santos, and D. Fellner. Towards 3d digitization in the glam (galleries, libraries, archives, and museums) sector: Lessons learned and future outlook. *IPSI BgD Transactions on Internet Research (TIR)*, 16(1):1–9, 2020.

- [161] I. The MathWorks. *Computer Vision Toolbox*. Natick, Massachusetts, United State, 2018.
- [162] P. Theiler, K. Schindler, et al. Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:173–178, 2012.
- [163] T. Thormahlen, H. Broszio, and P. N. Meier. Three-dimensional endoscopy. In *Falk Symposium*, pages 199–214. Kluwer Academic Publishers, 2002.
- [164] A. Tobiasz, J. Markiewicz, S. Łapiński, J. Nikel, P. Kot, and M. Muradov. Review of methods for documentation, management, and sustainability of cultural heritage. case study: Museum of king jan iii’s palace at wilanów. *Sustainability*, 11(24):7046, 2019.
- [165] H. N. Tokgozoglul, E. M. Meisner, M. Kazhdan, and G. D. Hager. Color-based hybrid reconstruction for endoscopy. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 8–15. IEEE, 2012.
- [166] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.
- [167] R. Y. Tsai. An efficient and accurate camera calibration technique fro 3d machine vision. In *CVPR’86*, 1986.
- [168] S. Tulsiani, H. Su, L. J. Guibas, A. A. Efros, and J. Malik. Learning shape abstractions by assembling volumetric primitives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2635–2643, 2017.
- [169] M. VgStudio. 3.3.6, Volume Graphics GmbH, D - 69123 Heidelberg, Germany, 2020.
- [170] A. Wang, T. Qiu, and L. Shao. A simple method of radial distortion correction with centre of distortion estimation. *Journal of Mathematical Imaging and Vision*, 35(3):165–172, 2009.
- [171] L. Wang, M. Gong, M. Gong, and R. Yang. How far can we go with local optimization in real-time stereo matching. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT’06)*, pages 129–136. IEEE, 2006.
- [172] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.
- [173] C. Wengert, M. Reeff, P. C. Cattin, and G. Székely. Fully automatic endoscope calibration for intraoperative use. In *Bildverarbeitung für die Medizin 2006*, pages 419–423. Springer, 2006.
- [174] M. Werner, B. Stabernack, and C. Riechert. Hardware implementation of a full hd real-time disparity estimation algorithm. *IEEE transactions on consumer electronics*, 60(1):66–73, 2014.
- [175] A. R. Widya, Y. Monno, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki. Whole stomach 3d reconstruction and frame localization from monocular endoscope video. *IEEE journal of translational engineering in health and medicine*, 7:1–10, 2019.
- [176] R. Wiemker and A. Zwartkruis. Optimal thresholding for 3d segmentation of pulmonary nodules in high resolution ct. In *International Congress Series*, volume 1230, pages 653–658. Elsevier, 2001.
- [177] C. Wu. *3D reconstruction of anatomical structures from endoscopic images*. PhD thesis, Citeseer, 2010.
- [178] C. Wu. Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision-3DV 2013*, pages 127–134. IEEE, 2013.
- [179] C. Wu et al. Visualsfm: A visual structure from motion system. 2011.
- [180] F. Wu, H. Wei, and X. Wang. Correction of image radial distortion based on division model. *Optical Engineering*, 56(1):013108, 2017.
- [181] R. Wu, C. Xiao, and C. Zheng. Deepcad: A deep generative network for computer-aided design models. *arXiv preprint arXiv:2105.09492*, 2021.

- [182] T. Yamaguchi, M. Nakamoto, Y. Sato, Y. Nakajima, K. Konishi, M. Hashizume, T. Nishii, N. Sugano, H. Yoshikawa, K. Yonenobu, et al. Camera model and calibration procedure for oblique-viewing endoscope. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 373–381. Springer, 2003.
- [183] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [184] S.-J. Yao, L.-H. Wang, C.-L. Lin, and M. Zhang. Real-time stereo to multi-view conversion system based on adaptive meshing. *Journal of Real-Time Image Processing*, 14(2):481–499, 2018.
- [185] Y. Yuan, Y. Ding, L. Zhao, and L. Lv. An improved method of 3d scene reconstruction based on sfm. In *2018 3rd International Conference on Robotics and Automation Engineering (ICRAE)*, pages 228–232. IEEE, 2018.
- [186] N. Zeller, C. Noury, F. Quint, C. Teulière, U. Stilla, and M. Dhome. Metric calibration of a focused plenoptic camera based on a 3d calibration target. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:449–456, 2016.
- [187] K. Zhan, D. Fritsch, and J. Wagner. Photogrammetry and computed tomography point cloud registration using virtual control points. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:265–270, 2021.
- [188] K. Zhan, D. Fritsch, and J. F. Wagner. Stability analysis of intrinsic camera calibration using probability distributions. In *IOP Conference Series: Materials Science and Engineering*, volume 1048, page 012010. IOP Publishing, 2021.
- [189] K. Zhan, Y. Song, D. Fritsch, G. Mammadov, and J. Wagner. Computed tomography data colouring based on photogrammetric images. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43(B2):361–368, 2020.
- [190] Y. Zhan, Y. Gu, K. Huang, C. Zhang, and K. Hu. Accurate image-guided stereo matching with efficient matching cost and disparity refinement. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(9):1632–1645, 2015.
- [191] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.
- [192] K. Zhou, X. Meng, and B. Cheng. Review of stereo matching algorithms based on deep learning. *Computational intelligence and neuroscience*, 2020, 2020.
- [193] S. Zhou, Y. Cheng, Y. Wang, K. Dong, C. Guo, J. Bai, and S. Tamura. Segmentation of the hip joint in ct volumes using adaptive thresholding classification and normal direction correction. *Journal of the Chinese Institute of Engineers*, 36(8):1059–1072, 2013.
- [194] S. Zhu, T. Shen, L. Zhou, R. Zhang, J. Wang, T. Fang, and L. Quan. Parallel structure from motion from local increment to global averaging. *arXiv preprint arXiv:1702.08601*, 2017.
- [195] S. Zhu, R. Zhang, L. Zhou, T. Shen, T. Fang, P. Tan, and L. Quan. Very large-scale global sfm by distributed motion averaging. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4568–4577, 2018.
- [196] T. Zinßer, J. Schmidt, and H. Niemann. A refined icp algorithm for robust 3-d correspondence estimation. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 2, pages II–695. IEEE, 2003.
- [197] C. Zou, E. Yumer, J. Yang, D. Ceylan, and D. Hoiem. 3d-prnn: Generating shape primitives with recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 900–909, 2017.

# Appendix

## Appendix A: Theory for 3D Reconstruction and Data Registration

### A1: Random sample consensus

The Random sample consensus (RANSAC) is an iterative process estimating the mathematical model from the given data which contains outliers. The algorithm is working under the assumption, that the data consist of inliers, which matches well with the mathematical model, and outliers. If the outliers are too few, the problem is degraded to a least-squares problem. The process of RANSAC could be summarized below,

1. A sample subset consists of minimum data from the given dataset is randomly selected;
2. Parameters of the model fitted by the subset are calculated;
3. Inliers of the fitted model among the entire dataset are checked according to a specific loss function;
4. The estimated model with sufficient inliers is classified into consensus set;
5. The final model is obtained by optimization with all consensus sets.

For the implementation in practice, parameters such as the number of iterations, the threshold for the inlier as well as the number of inliers to define a consensus set are necessary and have to be defined beforehand. Suppose the probability of the inlier within the entire dataset is  $w$  to be determined by the number of inliers  $n_{inlier}$  and outliers  $n_{outlier}$  as (7.1), and  $k$  is the number of random selection, minimum points for fitting the model is  $n$ , then the desired probability of RANSAC could be denoted as (7.2).

$$w = \frac{n_{inlier}}{n_{inlier} + n_{outlier}} \quad (7.1)$$

$$p = 1 - (1 - w^n)^k \quad (7.2)$$

This could lead to the the determination of  $k$  (7.3).

$$k = \frac{\log(1 - p)}{\log(1 - w^n)} \quad (7.3)$$

The RANSAC algorithm has the advantage of robustness in estimation the parameters for the model even if a certain number of outliers exists. However, if there is limitation on the iterations or too small inlier ratios, the algorithm is likely to provide unsatisfying results.

## A2: Collinearity equation

The collinearity equation is the formulation relating 2D points on the image plane to 3D points in the space via extrinsic and intrinsic parameters in the photogrammetry and computer vision field. As shown in Figure 7.1,  $XYZ$  and  $xyz$  defines the object and camera coordinate system respectively,  $uvw$  is a transition coordinate system between  $XYZ$  and  $xyz$  and has the same origin as  $xyz$  and same orientation as  $XYZ$ , point  $A(X_a, Y_a, Z_a)$  is in object coordinate system,  $a(x, y)$  is the associated image point,  $C(X_c, Y_c, Z_c)$  is the optic center in object coordinate system.

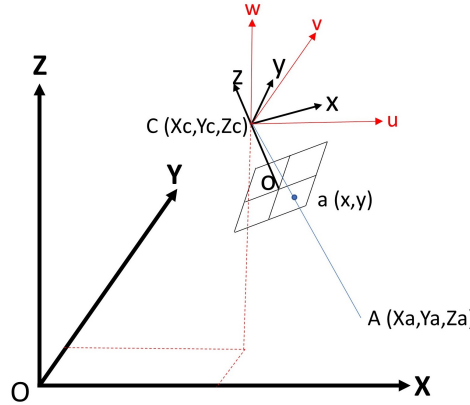


Figure 7.1: Illustration of collinearity equations.

According to the rule of central projection, point  $C(X_c, Y_c, Z_c)$ , point  $a(x, y)$  and point  $A(X_a, Y_a, Z_a)$  should be collinear, which leads to (7.4), where  $m$  is the scale factor.

$$\vec{CA} = m * \vec{CO} \quad (7.4)$$

Point  $a(x, y)$  could be transformed into the camera coordinate system as  $a^2(x-x_0, y-y_0, -f)$ , where  $x_0, y_0$  are coordinates of the principle point and  $f$  is the focal length. For the further formulation, an image point is necessary to be transformed from the camera coordinate system to the transition coordinate system  $uvw$  via a rotation matrix  $R$  consisting of  $r_{ij}$  ( $i$  and  $j$  from 1 to 3). In addition, the collinear relation could be written as (7.5) .

$$\begin{pmatrix} X_a \\ Y_a \\ Z_a \end{pmatrix} = \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} + m \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \\ -f \end{pmatrix} \quad (7.5)$$

The equation could be further written into (7.6) by inverting the rotation matrix and separating the known image coordinates.

$$\begin{pmatrix} x - x_0 \\ y - y_0 \\ -f \end{pmatrix} = \frac{1}{m} \begin{pmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{pmatrix} \begin{pmatrix} X_a - X_c \\ Y_a - Y_c \\ Z_a - Z_c \end{pmatrix} \quad (7.6)$$

By dividing the first and second row by the third row of the above equation, the scale factor  $m$  could be eliminated and the collinearity equation could be derived as (7.7).



$$\begin{aligned}
x &= x_0 - f \cdot \frac{r_{11} \cdot (X_a - X_c) + r_{21} \cdot (Y_a - Y_c) + r_{31} \cdot (Z_a - Z_c)}{r_{13} \cdot (X_a - X_c) + r_{23} \cdot (Y_a - Y_c) + r_{33} \cdot (Z_a - Z_c)} \\
y &= y_0 - f \cdot \frac{r_{12} \cdot (X_a - X_c) + r_{22} \cdot (Y_a - Y_c) + r_{32} \cdot (Z_a - Z_c)}{r_{13} \cdot (X_a - X_c) + r_{23} \cdot (Y_a - Y_c) + r_{33} \cdot (Z_a - Z_c)}
\end{aligned} \tag{7.7}$$

### A3: Mathematical self-calibration models

Except the classical physical camera calibration model, early works such as [38, 60] proposed mathematical models using orthogonal polynomials of second and fourth order. However, these references have insufficient consideration on the aspects of overparameterization, correlations and math foundations. In the work [159], the rationality of using algebraic polynomials for modeling the camera distortion has been rigorously proven by the Weierstrass theory and the Fourier theorem. In addition, two groups of mathematical self-calibration models have been built based on the function approximation problem.

#### Legendre self-calibration model

As the optimal approximation in the least-squares sense, the Legendre polynomial could be used for the camera distortion modeling. The polynomials are defined as (7.8):

$$\begin{aligned}
|L_m(x)| &\leq 1, -1 \leq x \leq 1 \\
\int_{-1}^1 L_m(x)L_n(x)dx &= \begin{cases} 0, m \neq n. \\ 1, m = n \end{cases}
\end{aligned} \tag{7.8}$$

When apply the Legendre polynomial for the image coordinates, if  $2b_x$  and  $2b_y$  denote the width and length of the image frame, the polynomial could be scaled by (7.9). The first few terms are in (7.10).

$$l_m(\bar{x}, b_x) = L_m(\bar{x}/b_x), l_n(\bar{y}, b_y) = L_n(\bar{y}/b_y) \tag{7.9}$$

$$\begin{aligned}
l_0 &= 1 \\
l_1 &= \bar{x}/b_x \\
l_2 &= [(\bar{x}/b_x)^2 - 1]/2 \\
l_3 &= [5(\bar{x}/b_x)^3 - 3(\bar{x}/b_x)]/2 \\
l_4 &= [35(\bar{x}/b_x)^4 - 30(\bar{x}/b_x)^2 + 3]/8 \\
l_5 &= [63(\bar{x}/b_x)^5 - 70(\bar{x}/b_x)^3 + 15(\bar{x}/b_x)]/8 \\
l_6 &= [231(\bar{x}/b_x)^6 - 315(\bar{x}/b_x)^4 + 105(\bar{x}/b_x)^2 - 5]/16
\end{aligned} \tag{7.10}$$

where  $\bar{x}$  and  $\bar{y}$  are the corrected coordinates of image measurements and  $L_m$  and  $L_n$  are univariate Legendre polynomials.

In 2D image frame, the bivariate Othogonal Polynomial (OP) could be derived as (7.11)

$$p_{m,n} \triangleq 10^{-6} f_{m,n}(\bar{x}, \bar{y}; b_x, b_y) = l_m(\bar{x}, b_x)l_n(\bar{y}, b_y) \tag{7.11}$$

The foundation of applying bivariate OP for camera distortion modeling is indicated in 7.12, which indicates that  $p_{m,n}$  is orthogonal over the image frame.

$$\int_{-b_y}^{b_y} \int_{-b_x}^{b_x} p_{m,n}(x, y)p_{i,j}(x, y)dxdy = 0, \text{ if } m \neq i \text{ or } n \neq j. \tag{7.12}$$

The distortion could be approximated by the continuous bivariate OPs with certain degrees. Additionally, some highly correlated terms are to be eliminated to ensure the calibration quality using a Legendre self-calibration model.

### Fourier self-calibration model

According to the Fourier theorem, another optimal function approximation is the Fourier self-calibration model with two variables defined in  $[-\pi, \pi] \times [-\pi, \pi]$  as indicated in (7.13).

$$\begin{aligned} u &= \bar{x}\pi/b_x, v = \bar{y}\pi/b_y, u, v \in [-\pi, \pi] \\ c_{m,n} &\triangleq 10^{-6} \cos(mu + nv), s_{m,n} \triangleq 10^{-6} \sin(mu + nv) \end{aligned} \quad (7.13)$$

where  $\bar{x}$  and  $\bar{y}$  are image pixel coordinates,  $b_x$  and  $b_y$  are width and height of the image frame,  $10^{-6}$  is added for numerical stability.

The general form is given as (7.14).

$$\begin{aligned} \Delta x &= \sum_{m=1}^M \sum_{n=-N}^N (a_{m,n}c_{m,n} + b_{m,n}s_{m,n}) + \sum_{n=1}^N N(a_{0,n}c_{0,n} + b_{0,n}s_{0,n}) \\ \Delta y &= \sum_{m=1}^M \sum_{n=-N}^N (a'_{m,n}c_{m,n} + b'_{m,n}s_{m,n}) + \sum_{n=1}^N N(a'_{0,n}c_{0,n} + b'_{0,n}s_{0,n}) \end{aligned} \quad (7.14)$$

The coefficients  $a_{m,n}$ ,  $b_{m,n}$ ,  $a'_{m,n}$ ,  $b'_{m,n}$  are to be determined during the camera calibration process. According to the distortion of the lens, different orders with different numbers of additional parameters are applied for appropriate modeling. If the degrees are M and N respectively, the number of additional parameters for the Fourier self-calibration  $n_{AP}$  is (7.15).

$$n_{AP} = 4(2MN + M + N) \quad (7.15)$$

### A4: Dense image matching

For dense image matching, the strategies are classified into global and local methods. The widely used semi-global matching algorithm combines the advantage of both ideas in the following four main steps.

#### Pixelwise matching cost calculation

The process searches for the identical point on two overlapped images with known epipolar geometry. The core idea of the method is minimizing a smart approximation of a global cost function, which represents the goodness of the alignment of all pixels across two views. As for global cost function, in comparison with the matching process using absolute differences, normalized cross-correlation etc. the mutual information has the advantage concerning the robustness of illumination and noise. When using mutual information for pixelwise matching cost calculation, the entropy of the image is to be defined as (7.16).

$$MI_{I_1, I_2} = H_{I_1} + H_{I_2} - H_{I_1, I_2} \quad (7.16)$$

where the entropies are calculated from the probability distributions of intensities of the associated images as (7.17).

$$\begin{aligned} H_{I_1} &= - \int_0^1 P_I(i) \log P_I(i) di \\ H_{I_1, I_2} &= - \int_0^1 \int_0^1 P_{I_1, I_2} \log P_{I_1, I_2}(i_1, i_2) di_1 di_2 \end{aligned} \quad (7.17)$$

The joint entropy in (7.17) working for the whole image could be transformed into a sum over pixels using a Taylor expansion in (7.18).

$$H_{I_1, I_2} = \frac{1}{n} \sum_p h_{I_1, I_2}(I_{1p}, I_{2p}) \quad (7.18)$$

To take the pixel occlusions into consideration, the entropy could also be calculated analog to the joint entropy in (7.19).

$$H_I = \sum_{\mathbf{p}} h_I(I_{\mathbf{p}}) \quad (7.19)$$

$$h_I(i) = -\frac{1}{n} \log(P_i \otimes g_i) \otimes g_i$$

Then the mutual information cost is defined in (7.20).

$$MI_{I_1, I_2} = mi_{I_1, I_2}(I_{1\mathbf{p}}, I_{2\mathbf{p}}) \quad (7.20)$$

$$mi_{I_1, I_2}(i, k) = h_{I_1}(i) + h_{I_2}(k) - h_{I_1, I_2}(i, k)$$

For the pixel  $\mathbf{p}$ , if the disparity is  $d$ , the corresponding cost is defined by (7.21).

$$C_{MI}(\mathbf{p}, d) = -mi_{I_b, f_D(I_m)}(I_{b\mathbf{p}}, I_{m\mathbf{q}}) \quad (7.21)$$

$$\mathbf{q} = e_{bm}(\mathbf{p}, d)$$

For the goal of calculating the disparity image, it is needed as input for the mutual information determination. The problem could be solved by an iterative process or a hierarchical calculation. The output of this step is a disparity space image which stores the matching cost considering the local information.

### Aggregation of costs

The second step optimizes the disparity image globally to improve the instability of the pixel-wise matching using additional constraints supporting the smoothness of it. The energy with smoothness constraints to be minimized is defined in (7.22).

$$E(D) = \sum_{\mathbf{p}} (C(\mathbf{p}, D_{\mathbf{p}})) + \sum_{\mathbf{q} \in N_{\mathbf{q}}} P_1 T[|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] + \sum_{\mathbf{q} \in N_{\mathbf{q}}} P_2 T[|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] \quad (7.22)$$

where the first term is the sum of MI cost, the second and third terms add constraints for pixels for which the disparity changes with different penalty.

Due to the difficulty of solving such a NP-complete problem, the minimization along individual image rows in 1D could be performed. Moreover, to avoid big differences of constraints between different directions, the matching costs in 1D from all directions are aggregated. The cost  $L_{\mathbf{r}}$  is derived in (7.23) with the consideration of avoiding too big cost values.

$$L_{\mathbf{r}}(\mathbf{p}, d) = C(\mathbf{p}, d) + \min(L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d - 1) + P_1, L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \min_i L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) - \min_k L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, k) \quad (7.23)$$

For the final cost for the pixel  $\mathbf{p}$ , the cost that choose the disparity  $d$  is indicated in 7.24.

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d) \quad (7.24)$$

### Disparity calculation

After the aggregation of the costs, the optimal disparity of each pixel is determined. Normally, the smallest cost value is regarded to be optimal. Since there are still occlusions and false matches in the disparity image  $D_b$  and  $D_m$  corresponding to the base image and the matched image respectively, a consistency check is necessary to be implemented for a reliable disparity computation.

## Disparity refinement

Phenomenons such as low texture, reflections, noise etc. may bring outliers in the results of calculated disparities. In the last step of stereo matching, normally a disparity refinement is implemented by intensity consistent disparity checks, discontinuity preserving interpolating etc.

## A5: Gauss-Helmert Model

For 3D registrations, a 7 parameter transformation is formulated as 7.25.

$$\mathbf{X} = \mathbf{X}_0 + \mu \mathbf{R} \mathbf{x} \quad (7.25)$$

where  $\mathbf{X}$  the  $(3*1)_u$  vector of world coordinates of  $u$  control points,  $\mathbf{X}_0$  the  $(3*1)_u$  vector of the 3 translation parameters  $(X_0, Y_0, Z_0)$ ,  $\mu$  is the scale,  $\mathbf{R}$  the  $(3*3)_u$  rotation matrix depending on the unknown rotation angles  $\alpha, \beta, \gamma$  and  $\mathbf{x}$  the  $(3*1)_u$  vector of the local  $u$  control point coordinates. This non-linear transformation is linearized considering only differential changes in the three translations, the rotations and the scale, and therefore replacing 7.25 by 7.26 as follows.

$$d\mathbf{x} = \mathbf{S} d\mathbf{t} \quad (7.26)$$

$\mathbf{S}$  is the  $(3*7)_u$  similarity matrix resulting from the linearization process of 7.25, given as 7.27.

$$\mathbf{S} = \left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -z & y & x \\ 0 & 1 & 0 & z & 0 & -x & y \\ 0 & 0 & 1 & -y & x & 0 & z \end{array} \right] \quad (7.27)$$

where  $d\mathbf{t}' = [dx, dy, dz, d\alpha, d\beta, d\gamma, d\mu]$  representing the seven unknown registration parameters. If no scale adjustment is necessary, we set  $\mu = 1$ . In order to estimate also the precision of the datum transform, the transformation is embedded in a least-squares Gauss-Helmert model for  $u \geq 3$  and  $\mathbf{B} := \mathbf{S}$  leading to 7.28.

$$1^{st} \text{ order} : \mathbf{A} \mathbf{v} + \mathbf{B} \mathbf{x} + \mathbf{w} = 0; 2^{nd} \text{ order} : D(\mathbf{v}) = D(\mathbf{I}) = \sigma^2 \mathbf{P}^{-1} \quad (7.28)$$

Solving  $\|\mathbf{v}\|^2 = \min$  subject to  $\mathbf{A} \mathbf{v} + \mathbf{B} \mathbf{x} + \mathbf{w} = 0$  the corresponding normal equation system is obtained as 7.29.

$$\begin{bmatrix} \mathbf{A} \mathbf{P}^{-1} \mathbf{A}' \\ \mathbf{B}' \end{bmatrix} \hat{\boldsymbol{\lambda}} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \hat{\mathbf{x}} + \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} = \mathbf{0} \quad (7.29)$$

Resolving 7.29 wrt  $\hat{\boldsymbol{\lambda}}$  we obtain 7.31.

$$\hat{\boldsymbol{\lambda}} = -(\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} (\mathbf{B} \hat{\mathbf{x}} + \mathbf{w}) \quad (7.30)$$

and resubstitute it again we obtain 7.31

$$\begin{aligned} -\mathbf{B}' (\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} (\mathbf{B} \hat{\mathbf{x}} + \mathbf{w}) &= \mathbf{0} \\ -\mathbf{B}' (\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} \mathbf{B} \hat{\mathbf{x}} - \mathbf{B}' (\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} \mathbf{w} &= \mathbf{0} \\ -\mathbf{B}' (\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} \mathbf{B} \hat{\mathbf{x}} &= \mathbf{B}' (\mathbf{A} \mathbf{P}^{-1} \mathbf{A}')^{-1} \mathbf{w} \\ \hat{\mathbf{x}} &= -(\mathbf{B}' \tilde{\mathbf{P}}^{-1} \mathbf{B})^{-1} \mathbf{B}' \tilde{\mathbf{P}}^{-1} \mathbf{w} \end{aligned} \quad (7.31)$$

mit  $\tilde{\mathbf{P}} = \mathbf{A} \mathbf{P}^{-1} \mathbf{A}'$  The error propagation problem of any precision parameters is solved by the well-known law of error propagation (of statistical inference). For the  $(3*1)_u$  vector of residuals as given in 7.32.

$$\hat{\mathbf{v}} = \mathbf{P}^{-1} \mathbf{A}' \hat{\boldsymbol{\lambda}} \rightarrow D(\hat{\mathbf{v}}) = \mathbf{P}^{-1} \mathbf{A}' D(\hat{\boldsymbol{\lambda}}) \mathbf{A} \mathbf{P}^{-1} \quad (7.32)$$

and for the vector of the estimated unknown datum parameters 7.33.

$$\hat{\mathbf{x}} = -(\mathbf{B}'\tilde{\mathbf{P}}^{-1}\mathbf{B})^{-1}\mathbf{B}'\tilde{\mathbf{P}}^{-1}\mathbf{w} \rightarrow D(\hat{\mathbf{x}}) = (\mathbf{B}'\tilde{\mathbf{P}}^{-1}\mathbf{B})^{-1}\mathbf{B}'\tilde{\mathbf{P}}^{-1}D(\mathbf{w})\tilde{\mathbf{P}}^{-1}\mathbf{B}(\mathbf{B}'\tilde{\mathbf{P}}^{-1}\mathbf{B})^{-1} \quad (7.33)$$

with  $D(\mathbf{w}) = \sigma^2\mathbf{A}\mathbf{P}^{-1}\mathbf{A}'$  and substituted in 7.33 the precision of the datum transformation parameter we obtain 7.34.

$$D(\hat{\mathbf{x}}) = \sigma^2[\mathbf{B}'(\mathbf{A}\mathbf{P}^{-1}\mathbf{A}')^{-1}\mathbf{B}]^{-1} \quad (7.34)$$

### A6: Gauss-Markov Model

The relationship between observations and true values in a statistical sense is shown in (7.35).

$$\mathbf{E}(\hat{\mathbf{x}}) = \mathbf{x} \quad (7.35)$$

This proof is given by a simple definition of the Gauss-Markov model for estimating unknown parameters in a least-squares parameter estimation process. Let be given the linear model in (7.36).

$$1st\ order : \mathbf{E}(\mathbf{l}) = \mathbf{A}\mathbf{x}; \quad 2nd\ order : \mathbf{D}(\mathbf{l}) = \sigma^2\mathbf{I} \quad (7.36)$$

$$\mathbf{l} + \mathbf{v} = \mathbf{A}\mathbf{x}; \quad \mathbf{D}(\mathbf{l}) = \sigma^2\mathbf{I} \quad (7.37)$$

which could be derived from 7.38.

$$\mathbf{E}(\mathbf{l} + \mathbf{v}) = \mathbf{E}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathbf{E}(\mathbf{x}) = \mathbf{A}\mathbf{x} \quad (7.38)$$

with  $\mathbf{E}(\mathbf{l}) + \mathbf{E}(\mathbf{v}) = \mathbf{A}\mathbf{x}$  if and only if  $\mathbf{E}(\mathbf{v}) = 0$ . The parameter estimation of  $\|\mathbf{v}^2\| = \min$  finally leads to 7.39 and 7.40.

$$\hat{\mathbf{x}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{l} \quad (7.39)$$

$$\mathbf{E}(\hat{\mathbf{x}}) = \mathbf{E}((\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{l}) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{E}(\mathbf{l}) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{x} \quad (7.40)$$

## Appendix B: 3D Digital Twins of Gyros

### The Machine of Bohnenberger

The very first gyroscope, called the Machine of Bohnenberger, was invented by the astronomer, mathematician and physicist Johann Gottlieb Friedrich Bohnenberger (1765-1831) at the University of Tuebingen in 1810. It was initially introduced as a tool assisting the teaching during the lectures in astronomy. A systematic and complete description and the use of the instrument could be found in [18]. According to the content in this work, in 2004 and 2010, original copies were discovered respectively. The year of construction is between 1810 and 1836 with brass, ivory and wood as the main materials. The related pictures are shown in Figure 7.2. And Table 7.1 gives additional data regarding the outlook. This very initial work has inspired and contributed greatly to the development of the gyroscope technology such as laser gyros, fiber-optical gyros as well as the widely applied MEMS gyroscopes of today.

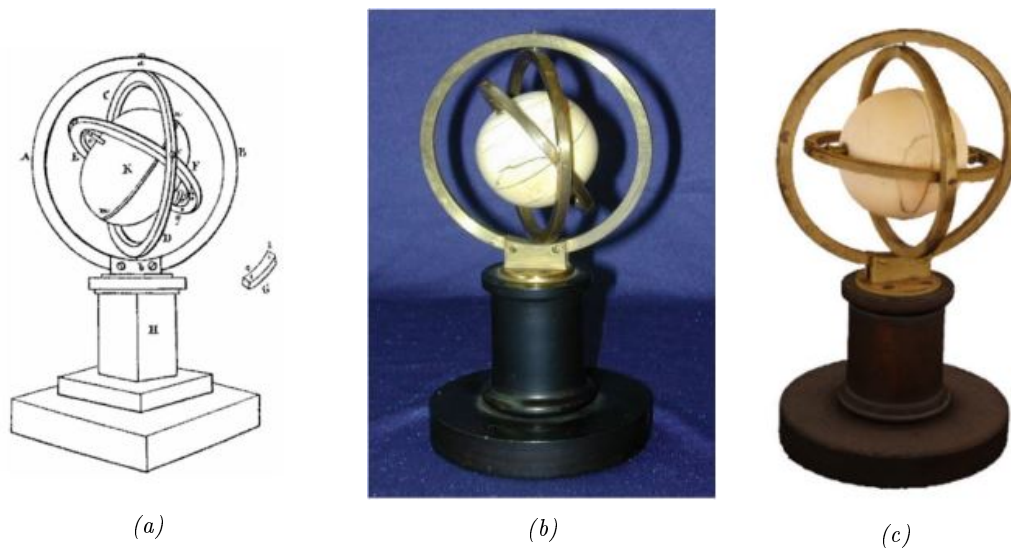


Figure 7.2: The very first gyroscope of 1810—the Machine of Bohnenberger, Tuebingen, Germany: (a) original drawing; (b) photo @ J.F. Wagner; and (c) 2.5D digital twin by CV/photogrammetry.

Table 7.1: Dimensional information for the Machine of Bohnenberger

Parameter	Value
Height (mm)	155
Width (mm)	86
Length (mm)	86
Sphere Diameter (mm)	44
Weight (kg)	0.289

### Golden Gnat WK10-10

Golden Gnat is a miniature rate gyro and mostly part of guidance and control systems, installed in INSs, missiles and radar systems during the cold-war era. Currently the object with the inventory number WK10-10 is possessed by the Chair of Flight Measuring Technology, University of Stuttgart. It is a floated, single-axis gyroscope, which is driven by a hysteresis motor that turns the rotor (the gyro part of the gyroscope) with 24,000 rpm (rotations per minute) and

senses angular velocities from around  $0.01^\circ/\text{sec.}$  to  $50^\circ/\text{sec.}$  This device was manufactured by the American company Honeywell International Incorporation in November 1962. More dimensional information could be found in Table 7.2.

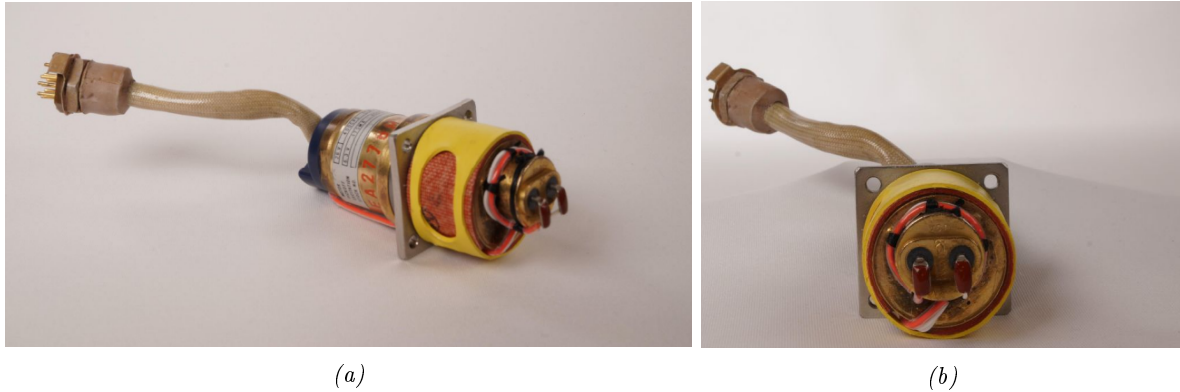


Figure 7.3: Golden Gnat: (a) general view; (b) front view of the Golden Gnat turning top .

Table 7.2: Dimensional information for the Golden Gnat

Parameter	Value
Height (mm)	31
Width (mm)	31
Length (mm)	143
Weight (kg)	0.15

### LN3-G200

The object named LN3-G200 gyro as left image of the Figure 7.4 in the Gyrolog collection was part of an inertial platform, the so-called LN3, manufactured by Litton Technische Werke, Freiburg (Litef). The manufacture year of the object is deduced before September of 1968. The German part of the American company was founded for the purpose of producing the LN3 platform for the Lockheed F104G Starfighter, purchased by the German government during the Cold War period. The object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. Other similar objects within the collection are LK06-17 and DM-LN3. The right image of the Figure 7.4 is the special transportation box for LN3-G200. More dimensional informations are presented in Table 7.3.

Table 7.3: Dimensional information for the Gyro G200 of the LN3 Inertial Platform

Parameter	Value
Height (mm)	80
Width (mm)	99
Length (mm)	100
Weight (kg)	0.95



Figure 7.4: The Gyro G200 of the LN3 Inertial Platform: (a) Gyro G200; (b) G200 transportation box.

### BM06-10

BM06-10 is an accelerometer manufactured by Honeywell International Inc in the year 1965. It is identical to the accelerometers BM01-09 and BM02-09 except that a sectional model was made from it for the purpose of illustration. The object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. For the purpose of 3D reconstruction, the transparent bounding box was taken off for image acquisition. Images and more dimensional informations could be found in Figure 7.5 and Table 7.4.

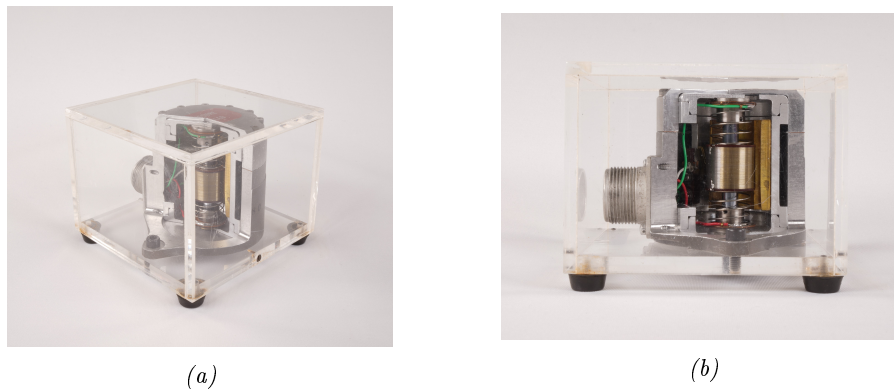


Figure 7.5: Accelerometer BM06-10: (a) general view; (b) side view of the accelerometer.

Table 7.4: Dimensional information for the accelerometer BM06-10

Parameter	Value
Height (mm)	58
Width (mm)	82
Length (mm)	249
Weight (kg)	0.40



**KK13-09**

This is a directional gyroscope S.F.I.M by BEZU, France. The object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. Within the collection, KK03-09 is of the same type as KK13-09.



Figure 7.6: Directional gyro KK13-09: (a) general view; (b) side view without the caging device.

Table 7.5: Dimensional information for the directional gyro KK13-09

Parameter	Value
Height (mm)	114
Width (mm)	120
Length (mm)	213
Weight (kg)	2.15

**KK01-09**

This object is a directional gyro AN5731-1 with the inventory number KK01-09, which was manufactured during the Second World War. Currently, the object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. Within the collection, other similar gyroscopes of the same type are KK06-09, KK07-09, KK08-09, KK10-09, KK11-09, KK26-10 .



Figure 7.7: Directional gyro KK01-09: (a) general view; (b) back view.

Table 7.6: Dimensional information for the directional gyro KK01-09

Parameter	Value
Height (mm)	126
Width (mm)	115
Length (mm)	180
Weight (kg)	1.70

### KK10-09

This directional gyro is a device that was produced by Ternstedt Manufacturing Div, GM Corp, Detroit, USA, and was installed in a large number of American aircraft during the Second World War. It was found in many aircraft models such as P40, P51, B17, B24, B25, A26. Compared with this gyro, the other gyros that have the number AN5731-1, have been produced from other companies, but they are all identical because they were built under license from Sperry. This same design is due to the Second World War, in which the device was installed as standard in a large number of aircraft, - this device is operated pneumatically. Therefore it could not be checked for its functionality. However, it should be functional as soon as compressed air is connected. The object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. Pictures and more dimensional information could be referred to Figure 7.8 and Table 7.7 respectively.

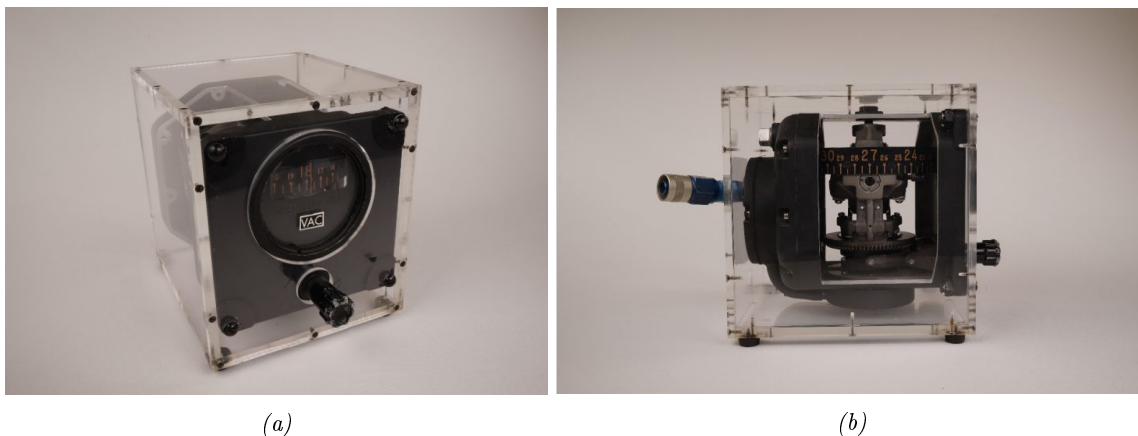


Figure 7.8: Pneumatically driven direction gyro: (a) general view; (b) side view of the course gyro.

Table 7.7: Dimensional information for course gyro KK10-09

Parameter	Value
Height (mm)	150
Width (mm)	129
Length (mm)	230
Weight (kg)	2.25

### KK12-09

The object is an electrical directional gyro, which was produced by Siemens-LGW, Berlin, Germany. Currently, the object is owned by the Chair of Flight Measuring Technology, University

of Stuttgart. Within the collection, other similar gyroscopes of the same type are KK02-09, KK09-09, KK22-09.

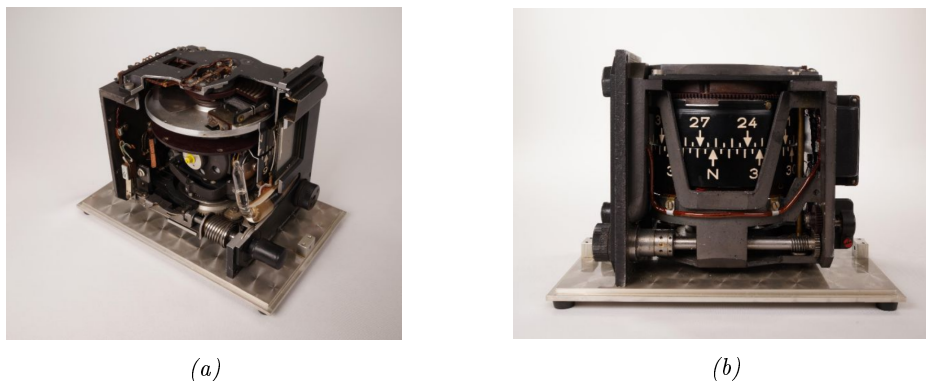


Figure 7.9: Directional gyro KK12-09: (a) general view; (b) back view.

Table 7.8: Dimensional information for the directional gyro KK12-09

Parameter	Value
Height (mm)	158
Width (mm)	135
Length (mm)	215
Weight (kg)	2.80

### KR04-17

This object is a rotor component of the gyroscopic instrument such as the object KK12-09. Currently, the object is owned by the Chair of Flight Measuring Technology, University of Stuttgart. Within the collection, KR05-17 is of the same type as KR04-17.

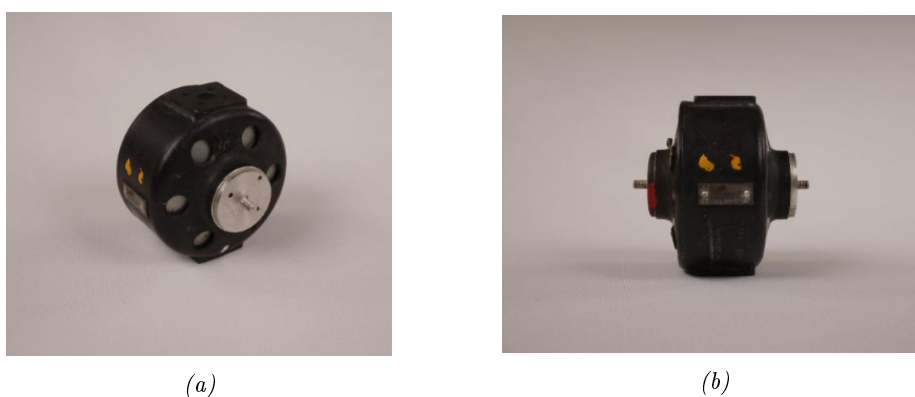


Figure 7.10: Gyroscopic rotor KR04-17: (a) general view; (b) side view.

### MFK01-09

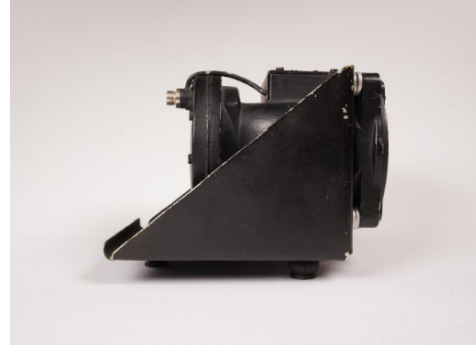
The inventory number MFK is the abbreviation for magnetic and long-range compasses. This object is produced by Bendix. This compass is a device that is used in American aircrafts during World War II. However, it could only be verified for the P40 Kittyhawk.

Table 7.9: Dimensional information for the gyroscopic rotor KR04-17

Parameter	Value
Height (mm)	58
Diameter (mm)	58
Weight (kg)	0.30



(a)



(b)

Figure 7.11: Magnetic and long-range compasses MFK01-09: (a) general view; (b) side view.

Table 7.10: Dimensional information for the magnetic and long-range compasses MFK01-09

Parameter	Value
Height (mm)	100
Width (mm)	98
Length (mm)	131
Weight (kg)	0.80

## KH09-09

These devices were a gift from Prof. H. de Mares for the assistance of the Stuttgart Institute in the development of systems for measuring equilibrium tests on a rowing boat. The SYP820 platform system was produced before the September of 1988 and was used in the Fiat G91 jet.



(a)



(b)

Figure 7.12: Operator unit KH09-09: (a) general view; (b) side view.

*Table 7.11:* Dimensional information for the Operator unit KH09-09

Parameter	Value
Height (mm)	44
Width (mm)	145
Length (mm)	148
Weight (kg)	0.55

# Acknowledgements

Before I came to Germany for my Master studies, having graduated with a Bachelor at Wuhan University, I never thought that I would be so interested in the profession and decided to pursue a Ph.D. During my master program at the University of Stuttgart, I felt great enthusiasm in the research of photogrammetry. This interest led me to the start of a Ph.D., and is also the inexhaustible motivation that supported me to overcome many difficulties during that period. First of all, I have to express my heartfelt gratitude to my doctoral supervisor Prof. Dieter Fritsch for leading me to explore the academic road, and for giving me all kinds of support at all times, both academically and in life. His professional and grand vision has benefited me a lot and will help me continue to make contributions in the academic field in the future.

During two and a half years working in the Gyrolog project within my Ph.D., I am also very grateful to my co-supervisor Jörg F. Wagner for his support, help and encouragement. Thanks also to Prof. Sven Simon for being a co-supervisor. His knowledge of CT has enabled the innovative research of a combination between photogrammetry and CT. In addition, I would like to dedicate my gratitudes to my colleagues Maria Niklaus, Sarah Peter, Benjamin Greiner and Benedikt Györfi. Although we are in different academic disciplines, we support and encourage each other and share experiences, which benefits us a lot in our Ph.D. stage. I would also like to thank the secretary of our institute Gerlinde Ott, who has been always helping me and encouraging me regardless of difficulties. Thanks also to Mr. Michael Klein of 7reasons in Vienna, Austria, who has given me comprehensive training on the topic of 3D modeling. The intense study as well as the beautiful city have left a beautiful and lifelong memory in my mind. I owe a particular debt of thanks to my senior student from Geoenigne master study as well as my good friend Yifan Song, who has also supported me and given me valuable academic advice all the way. In addition, thanks to his recommendation, I also had the chance to have a research stay in the Oceanic Machine Vision Group of GEOMAR Helmholtz Center in Kiel. The guidance from Dr. Kevin Köser as well as the discussion with Mengkun She have inspired me a lot for my research.

My gratitude also goes to my long-term sports partner Alexander Staib, Cheng Zuo, with whom I have spent a lot of happy times doing sport and kept me energetic status facing all the challenges. Finally, I also want to thank my parents and my brother, who gave me unconditional support, selfless love and help when I was studying abroad, which enabled me to overcome many difficulties and focus more on my studies. Without them, my studies would never be completed.

## Relevant Publications

1. Zhan, K., Fritsch, D., and Wagner, J. F.: Integration of Photogrammetry, Computed Tomography and Endoscopy for Gyroscope 3d Digitization, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-M-1-2021, 925–931, <https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-925-2021>, 2021.
2. Zhan, K., Fritsch, D. and Wagner, J.F., 2021. Photogrammetry and Computed Tomography Point Cloud Registration Using Virtual Control Points. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp.265-270.
3. Fritsch, D., Wagner, J.F., Ceranski, B., Simon, S., Niklaus, M., Zhan, K. and Mammadov, G., 2021. Making Historical Gyroscopes Alive—2D and 3D Preservations by Sensor Fusion and Open Data Access. *Sensors*, 21(3), p.957.
4. Zhan, K., Fritsch, D. and Wagner, J.F., 2021. Stability analysis of intrinsic camera calibration using probability distributions. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1048, No. 1, p. 012010). IOP Publishing.
5. Zhan, K., Song, Y., Fritsch, D., Mammadov, G. and Wagner, J., 2020. Computed tomography data colouring based on photogrammetric images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, pp.361-368.
6. Niklaus, Maria ; Zhan, Kun ; Wagner, J. F.: Gyrolog – Creating a 3-Dimensional Digital Collection of Classical Gyro Instruments. In: Hecker, P. (Hrsg.) ; Hecker, P. (Hrsg.): 2019 DGON Inertial Sensors and Systems (ISS), 2019 DGON Inertial Sensors and Systems (ISS) : IEEE, 2019 — ISBN 978-1-7281-1935-9, S. 1.1–1.23
7. Fritsch, D.; Wagner, J.F.; Simon, S.; Ceranski, B.; Niklaus, M.; Zhan, K.; Wang, Z. Gyrolog—Towards VR Preservations of Gyro Instruments for Historical and Didactical Research. In *Proceedings of the 2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, San Francisco, CA, USA, 27–30 October 2018; pp. 1–7

# Curriculum Vitae

## Personal

Name: Kun Zhan

Birthday: March 27, 1992

Birth Place: Huanggang, China

Nationality: Chinese

## Educational Background

09/2011 - 06/2015 Bachelor in School of Geodesy and Geomatics, Wuhan University

09/2014 - 01/2015 Exchange Student in Sun Yat-sen University Taiwan

10/2015 - 01/2018 Geoengine Master Program, University of Stuttgart

07/2018 - Present PhD student at the Institute for Photogrammetry and the Chair of Flight Measuring Technology, University of Stuttgart