Active Sensors: Calibration & Applications

by

Mehdi Faraji

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computing Science

University of Alberta

Abstract

Active sensors, such as active cameras and ultrasound transducers, are becoming more popular. One particular type of active camera, the Pan-Tilt-Zoom (PTZ) camera, has become ubiquitous in surveillance platforms. Given their active nature, active cameras are omnipresent in robotic systems as well. Intravascular Ultrasound (IVUS) is an intra-operative imaging modality that facilitates observing and appraising the vessel wall structures of the human coronary arteries. In this thesis, I propose novel algorithms for two tasks on images acquired by the above-mentioned modalities. First, I propose new mathematical equations for calibrating an active PTZ camera along with a novel algorithm called Simplified Active calibration (SAC). The proposed equations are closed-form and linear and hence can be used in real-time. Also, SAC needs only 4 pairs of images to calibrate the camera. It can estimate the focal length in each direction with only one point correspondence between each pair without any calibration pattern. I have extensively evaluated and analyzed the proposed equations using synthetic and real images. The results illustrate that SAC can be employed in practical real-world applications. In the second part of the thesis, I propose a novel segmentation method for delineating two different parts of the blood vessel walls from IVUS images. Segmentation of arterial wall boundaries from the IVUS images is not only crucial for quantitative analysis of the vessel walls and plaque characteristics, but is also necessary for reconstructing 3D models of the artery. Using a feature detection algorithm, namely, Extremal Region of Extremum Level (EREL), the proposed method delineates the luminal and media-adventitia borders in IVUS frames acquired by 20 MHz probes. Next, I propose a region selection strategy to label two ERELs as lumen and media based on the stability of their texture information. I extensively evaluated our selection strategy on the test set of a standard publicly available dataset containing 326 IVUS B-mode images. The results of my experiments reveal that our selection strategy is able to segment the lumen with $\leq$ 0.3 mm Hausdorff distance (HD) from the ground truth even though the images contain major artifacts, such as bifurcations, shadows, and side branches. Moreover, even

when there are no artifacts, my method is able to delineate media-adventitia boundaries with 0.31 mm HD. Furthermore, my segmentation method runs in linear time. Based on this work, by using a 20-MHz IVUS probe with controlled pullback, not only can we now more accurately analyze the internal structures of human arteries, we can also segment each frame during the pullback procedure in real-time.

Preface

This thesis describes novel work and research that I have accomplished during my PhD studies. All articles and their contents have gone through peer-review processes and have been published in prestigious journals and as conference papers.

Specifically, the content of Chapter 4 has been presented at and published for the International Conference of Smart Multimedia (ICSM) as *Faraji, Mehdi, and Anup Basu. A Simplified Active Calibration Algorithm for Focal Length Estimation. International Conference on Smart Multimedia. Springer, Cham, 2018*, in which I proposed a novel mathematical equations to estimate the focal length of an active camera along with some initial experiments and evaluations that seemed promising.

Encouraged by the initial results of the ICSM 2018 study, I extended the method and extensively tested and validated the results in a new article that thoroughly investigated the proposed Simplified Active Calibration equations. The new paper was published in the Elseveir Journal of Image and Vision Computing in 2019. Chapter 5 contains this article as *Faraji, Mehdi, and Anup Basu. Simplified Active Calibration. Image and Vision Computing 91 (2019): 103799.*

Chapter 6 of this thesis has been published as *Faraji, Mehdi, et al. Segmentation of arterial walls in intravascular ultrasound cross-sectional images using extremal region selection. Ultrasonics 84 (2018): 356-365.* This article illustrates my novel proposal to segment the images acquired by another active camera (i.e, Intravascular Ultrasound).

Not all of my accomplishments and publications during my PhD studies have been included in this thesis. The following is a comprehensive list of the work that I published during my PhD program:

Journal Publications

- [27] Faraji, Mehdi, and Anup Basu. Simplified Active Calibration. Image and Vision Computing 91 (2019): 103799
- [28] Faraji, Mehdi, et al. Segmentation of arterial walls in intravascular ultrasound cross-sectional images using extremal region selection. Ultrasonics 84 (2018): 356-365.
- [111] Yang, Ji, Mehdi Faraji, and Anup Basu. Robust segmentation of arterial walls in

intravascular ultrasound images using Dual Path U-Net. Ultrasonics 96 (2019): 24-33

Conference Publications

- [26] Faraji, Mehdi, and Anup Basu. A Simplified Active Calibration Algorithm for Focal Length Estimation. International Conference on Smart Multimedia. Springer, Cham, 2018
- [112] Yang, Ji, Mehdi Faraji, et al. IVUS-Net: an intravascular ultrasound segmentation network. International Conference on Smart Multimedia. Springer, Cham, 2018.
- [68] Li, Yuying, and Mehdi Faraji. Erel selection using morphological relation. International Conference on Smart Multimedia. Springer, Cham, 2018.

Awards

- Department of Computing Science **Early PhD Achievement Award**, 2018.

To Jasmine, Audrey and Bryan
with love.

Everything is perfect in the universe, even your desire to improve it.

– Wayne Dyer.

Acknowledgements

Throughout my PhD studies I received a great deal of support from my supervisor Prof. Anup Basu. I would like to wholeheartedly thank him for all his patience, kindness and his valuable supervision. Also, I would like to give special regards to my co-adviser, Prof. Irene Cheng, who supported my work. You have been a tremendous mentor for me.

In addition, I would like to thank the rest of my thesis committee, Prof. Pierre Boulanger, Prof. Herb Yang, Prof. Bruce Cockburn and Prof. Osmar Zaiane for all the great feedback they gave on my research and thesis.

I would also like to thank my friends in Cincinnati and Edmonton, Ed, Lisa, Greg, Ji, Gaurav, Niharika, Amir, Preet, Ava and Maryam. I am glad that I met you all during my PhD studies.

At last, I would like to give special thanks to my family. Words cannot express how grateful I am to my wife, Jasmine. You mean a lot to me.

Contents

List of Tables

List of Figures

List of Symbols

$\mathbf{x}$	A vector consists of a point $[u, v, 1]^T$ in image coordinate space.
u	Horizontal coordinates of a point in the image.
v	Vertical coordinates of a point in the image.
u_0	Horizontal coordinates of a point in the image that the principal axis of camera passes through.
v_0	Vertical coordinates of a point in the image that the principal axis of camera passes through.
u_d	Horizontal coordinates of a distorted point by lens in the image.
v_d	Vertical coordinates of a distorted point by lens in the image.
$\mathbf{X}$	A vector consists of a point $[x, y, z, 1]^T$ in world coordinate system.
x	Horizontal coordinates of a 3D point
y	Vertical coordinates of a 3D point
z	Depth coordinates of a 3D point
x_c	Horizontal coordinates of a 3D point in the camera coordinate frame
y_c	Vertical coordinates of a 3D point in the camera coordinate frame
z_c	Depth coordinates of a 3D point in the camera coordinate frame
x_w	Horizontal coordinates of a 3D point in the world coordinate frame
y_w	Vertical coordinates of a 3D point in the world coordinate frame
z_w	Depth coordinates of a 3D point in the world coordinate frame
f_u	Focal length of the camera in pixel in u direction
f_v	Focal length of the camera in pixel in v direction
f	Focal length of the camera in mm
α	Rotation angle of the camera around the X-axis
β	Rotation angle of the camera around the Y-axis
γ	Rotation angle of the camera around the Z-axis
t_x	Translation of the camera along the X-axis
t_y	Translation of the camera along the Y-axis
t_z	Translation of the camera along the Z-axis
$\mathbf{t}$	Translation vector of the camera
κ_1	First coefficient of the radial distortion polynomial of the lens
κ_2	Second coefficient of the radial distortion polynomial of the lens
$\mathbf{R}$	Rotation matrix of the camera
r_{ij}	An element of rotation matrix of the camera at row i and column j Rotation matrix of the camera
$\mathbf{P}$	Projection matrix of the camera
ϕ	The skew of the two image axes

List of Acronyms

AC	Active Calibration
ANGUS	Angiography and IVUS
ECG	Electrocardiogram
EREL	Extermal Regions of Extremum Levels
HD	Hausdorff Distance
IVUS	Intravascular Ultrasound
PTZ	Pan-Tilt-Zoom
RAC	Radial Alignment Constraint
SAC	Simplified Active Calibration

Chapter 1

Introduction

Humans are able to effortlessly perform many visual tasks. An extensive amount of research has been carried out so far to discover why humans can so easily perceive the visual world. Many factors are responsible for the smooth visual perception in the human brain, such as the neural architecture of the brain, hierarchical information processing that creates distinctive representations, binocular vision that enables us to perceive depth based on disparity, texture, shading and blur [105], memory that helps humans extract past experience to use it in fresh inferences, and so on. Among all of these capabilities, there is an important characteristic of human vision that often is ignored. Human vision is not passive. In contrast, humans are active observers with an *Active Vision* system.

According to [2], [3] "An observer is called active when engaged in some kind of activity whose purpose is to control the geometric parameters of the sensory apparatus. The purpose of the activity is to manipulate the constraints underlying the observed phenomena in order to improve the quality of the perceptual results." Specifically, for humans, geometric parameters of the sensory apparatus include adjusting the eyes to the level of illumination, focusing and defocusing, rotating the eyes and the head to have a different view of the scene, and etc. An example that shows humans' brain utilizes active vision is foveal vision. Only a small part of the retina contains high density of cones[1] cells that provide a very sharp central vision. Humans use this sharp and high quality vision to accomplish everyday tasks such as reading, driving and so on. To process further information, humans use head, eye and body movements to point the foveal vision towards the region of interest. Moving head, body and eyes are examples of geometric parameters of an active vision system that can significantly influence the level of the visual problem in ways that even an ill-posed problem[2] for a passive observer (who lacks the ability to change the geometric parameters of the sensory input) can be relaxed to a well-defined problem, including but not limited to shape

[1] Cones are a specific type of photoreceptors in the eye that are responsible for color and high acuity vision.

[2] A problem that is not well-posed in the sense of Hadamard is called ill-posed. Hadamard defines a well-posed problem as follows: If it has a solution, the solution is unique and its behaviors changes by changing the initial conditions. Usually inverse problems are ill-posed.

from shading, texture, contour and depth estimation, and structure from motion [2], [3].

1.1 Active Camera

The intricate system of human vision which are classified as active vision, can also be replicated through various cameras. These cameras which are called Active Cameras (or sensors), have the ability to be controlled automatically by an intelligent software algorithm. Their geometric parameters, such as location and orientation, the internal parameters such as zoom, focus, etc, can be modified automatically in real-time. This means that assuming that a sufficiently intelligent system (e..g the human brain) receives input images (visual stimuli), an active camera can adjust itself based on the control signals that come from the intelligent system for resolving the issues that the system faces in perceiving the last input. For example, if there is any occlusion in the last input, the system signals the active camera to rotate and move to a different location in space, thus a 3D object can be seen from a different viewpoint to avoid occlusion. Utilizing the advantages of active systems, in this thesis I present my research carried out on two type of active cameras (sensors), namely PTZ cameras and Intravascular Ultrasound.

1.1.1 PTZ Cameras

Pan-Tilt-Zoom (PTZ) cameras are becoming increasingly ubiquitous in surveillance and robotic platforms. This type of cameras provide the user with various functionalities, such as the ability to automatically change zoom, focus and rotate around the X-axis (tilt) and Y-axis (pan) by a specific angle. Most PTZ cameras are equipped with microcontrollers that allow the developer to control these functionalities by giving the instructions to a camera to get the status of the camera, rotate it by a specific angle, move the focus, adjust the exposure, and so forth.

Since the specific type of research that I conducted on active sensors required images from PTZ cameras that were taken for specific angles, I developed customized software that uses low-level instruction sets of a specific PTZ camera to issue commands to the camera to move to the required state. An screenshot of the software is illustrated in Fig.1.1(a). As can been seen, the software can rotate the camera by specific angles, zoom in and zoom out, change the focus to automatic or manual, and reduces the level of noise. More importantly, the software can accept a series of predefined rotations and zooms and automatically apply these predefined parameters to the camera and record the images and videos. The camera that was used for acquiring the real images used in this thesis is Canon vc-c50i which is shown in Fig.1.1(b). Four other PTZ cameras available in the market are depicted in Fig.1.1(c).

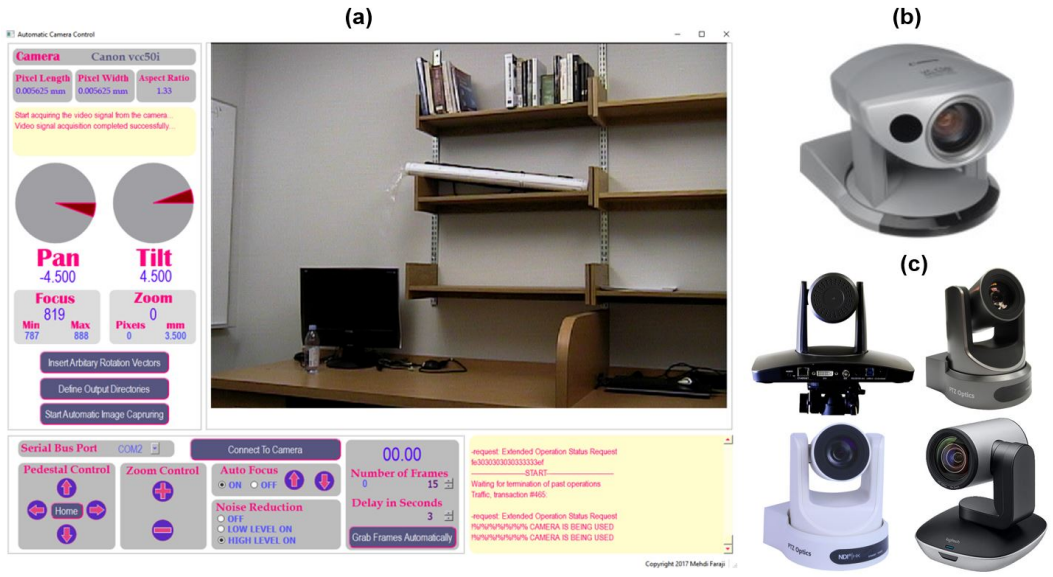


Figure 1.1: (a) A screenshot of the Automatic Camera Control application designed during my PhD studies for acquiring images to verify the proposed methods in this thesis. An automatic Camera Control application is able to rotate the camera by specific angles about Y-axis (pan) and X-axis (tilt). (b) Canon vc-c50i that is used in this thesis. (c) Images of other available PTZ cameras in the market.

1.1.2 Intravascular Ultrasound

Intravascular Ultrasound (IVUS) is another active sensor that I have studied during my PhD. IVUS is an invasive medical procedure in which the surgeon inserts a guidewire with a catheter that carries an ultrasound probe on its tip into the patient's arteries and navigates it to the desired destination by commanding the guidewire catheter. This imaging modality can also be considered as an active sensor because the geometric and photometric parameters of the sensory apparatus can be adjusted by guiding the catheter. A more detailed background on how this procedure is performed along with a review of the state-of-the-art can be found in Ch. 3.

1.2 Connecting the Dots

Parts of the research that I have done in my PhD studies have been included in this dissertation. The main theme of this PhD thesis encompasses *Active Sensors*. As stated in the title, the main theme is composed of two components:

1. **Calibration.** Calibration of a particular active sensor, i.e., an active camera in which I have proposed new mathematical equations, called Simplified Active Calibration, to calibrate an active camera (Section1.1.1).

2. **Application.** I proposed a novel segmentation method useful for the images of an-

other active sensor, i.e., Intravascular Ultrasound (Section 1.1.2).

The common denominator of these two research paths is in the characteristics of their sensory apparatus, i.e., *being active*. The ultimate goal of the research that I started at the beginning of my PhD was to design a pipeline of processes for 3D reconstruction of the vessel wall from Intravascular Ultrasound videos. The current methods in the literature have to employ another imaging modality (X-ray), most of the time, during the procedure to track the tip of a catheter during a surgery. This is harmful for patients because of radiation. Having a method that can reconstruct the 3D surface of a blood vessel in real-time during a surgery, without requiring another imaging modality, would be extremely useful and safe for patients. Although, the project was bigger than what can be done in one PhD dissertation, we decided to complete as many parts as we could and leave the rest as future work.

Theoretically speaking, the fact that IVUS imaging device is an active sensor should provide the researcher with ability to play around with the geometrical parameters of the sensor to estimate its location in space. This is the main component that I thought could help achieve the bigger goal. The first plausible step to achieve this is to have a method that can calibrate the sensor during a surgical procedure. Once a sensor is calibrated and the mapping from 2D to 3D is known, the guidewire catheter can be moved to the desired known location and tracking can be achieved.

In order to be able to calibrate the IVUS images the first task is to find matching points between IVUS images. The IVUS images look like an image containing many random textures. This textural characteristic of IVUS images prevents existing feature detection methods to detect feature points from the images with high repeatability. If we cannot match the points in a reference frame with subsequent frames, there is no way to calibrate an active sensor in real-time. One approach that would help finding correct matching points is to match regions instead of points. If we can segment parts of the vessel wall then we can correspond the segmented regions and use the correspondences in the calibration task. We need to assume that the segmentation method is fast and accurate and it does not require a large number of points because there are only two parts of the vessel anatomy that can be segmented in IVUS images, namely lumen and media. Fig. 1.2 illustrates IVUS images and the segmentation of lumen and media achieved by the method proposed in this thesis.

Once we have accurate segmentation methods, we should be able to use it in calibration. The fundamental idea that I pursued for calibrating an active sensor was to adjust the geometric parameters of the camera, take pictures and use it in calibration. For example, rotate the sensor around one axis, save the image and then do another motion. Corresponding these saved images can help calibrate a sensor in real-time. However, I faced a serious challenge. Applying the translation and rotation during IVUS pull back was not possible unless it is included as built-in functions in the device. To the best of my knowledge most

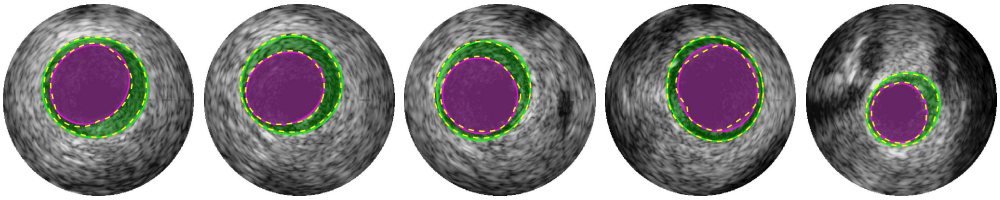


Figure 1.2: Images of 20 MHz IVUS frames with lumen and media segmentation results. Segmented lumen and media have been highlighted by magenta and green colours, respectively. The yellow dashed lines illustrate the gold standard that have been delineated by four clinical experts [4].

ultrasound devices, including the one in our lab, do not have this functionality. In other words, there was no way to move and rotate the guidewire by a specific angle manually nor by using software. Thus, I decided to evaluate the calibration idea with another active sensor. This time I chose a type of active camera called a PTZ camera. Fortunately, in contrast to an ultrasound device, the microcontroller of the PTZ camera responds to the instructions and hence I could design software to control the camera and create the required datasets for further research and experiments. An screenshot of the software is illustrated in Fig.1.1(a).

1.3 Thesis Contributions

My contribution in this dissertation is twofold:

1. I propose a Simplified Active Calibration (SAC) formulation for online calibration of an active camera. Specifically, I recommend and formulate:

   - Novel closed-formed equations for estimating the focal length;
   - Novel linear equations for estimating the principal points;
   - A novel algorithm for calibration;
   - A Simplified Active Calibration (SAC) algorithm that can estimate the focal length of a camera in each direction with only one point correspondence;
   - A SAC algorithm that only requires 4 images to calibrate a camera; and,
   - A SAC algorithm that can be used online because of its low complexity;

   In addition, the proposed SAC algorithm that has been evaluated extensively on synthetic and real data. Also, the behavior of proposed SAC under various uncertainties has been evaluated experimentally and mathematically through analytic differential analysis.

2. For segmentation of Intavascular Ultrasound images, I contributed the following:

- Adapted *Extremal Regions of Extremum Levels* for detecting the lumen and media regions;
- Proposed a region selection approach to automatically segment the arterial walls from Intravascular Ultrasound images;
- Achieved superior accuracy results over the standard publicly available IVUS dataset;
- Proposed a method that does not require any training steps; and,
- Proposed a method that has linear time complexity with respect to the number of pixels in an IVUS frame;

Furthermore, the proposed method is a general approach that works on any dataset, and the performance is not specific to the dataset used for testing.

1.4 Thesis Structure

The following Ph.D. dissertation is composed of a collection of 3 papers that I have written during my PhD, sandwiched between an introduction and a conclusion. The structure of the thesis is as follows:

Since related work might receive a shallow treatment due to the (multiple) paper format, I have included extended discussions on related work in two separate related works chapters. Specifically, Chapter 2 elaborates on calibration background and reviews the relevant literature that are not mentioned in the published papers. Chapter 3 reviews the literature on Intravascular Ultrasound (IVUS) segmentation methods.

Chapter 4 consists of a conference paper [26] in which I propose the basic idea of Simplified Active Calibration for focal length estimation.

Chapter 5 contains a journal paper [27] in which I propose a complete version of Simplified Active Calibration along with extensive experiments and mathematical analysis.

Chapter 6 consists of a journal paper [28] in which I propose a new fast and robust segmentation of Intravascular Ultrasound frames.

Finally, I conclude the dissertation and outline future work in Chapter 7.

Chapter 2

Related Work: Camera Calibration

General camera calibration methods tend to find an approximation of physical and optical behaviour of the camera by using a set of parameters [82] that are comprised of intrinsic and extrinsic components. The intrinsic parameters characterize the internal geometry and optical properties of the camera. The extrinsic parameters on the other hand, store the position and rotation of the camera with respect to the world coordinate system [82]. As we know, vision is an inverse problem, and this imposes an inherent difficulty in solving vision related problems, which raises the question of how these parameters can be estimated knowing that the vision is an ill-posed problem[1]? The main idea behind the estimation is to relate an object presented in the image to its known 3D representation. This mapping usually helps to estimate the camera parameters. Various methods have been proposed in the last few decades to approximate these parameters which use either an iterative process or a direct method. Mainly, these approaches can be classified into three categories.

1. ***Non-linear optimization techniques***
   Methods such as [81] that have to deal with lens distortion or any lens flaws, require a non-linear system of equations to be solved to approximate the camera parameters. To achieve this, an optimization method is employed, which iteratively minimizes a distance between the image points and the modelled projection or residual errors of some equations [82], [106]. Several classic papers in photogrammetry that have been designed based on this idea are [13], [25], [107].

   - ***Advantages***
     - This kind of optimization can optimize any parameter sets of the camera model because it is very common to cover many kinds of distortions [106],

[1] According to Wikipedia :"Problems that are not well-posed in the sense of Hadamard are termed ill-posed. Inverse problems are often ill-posed. For example, the inverse heat equation, deducing a previous distribution of temperature from final data, is not well-posed in that the solution is highly sensitive to changes in the final data."

and the accuracy can also be improved by increasing the iteration number [82]. In [59], this technique is utilized.

- If the model can establish a proper estimation, after the convergence, the algorithm may achieve highly accurate results [106].

- ***Disadvantages***
  - Because the optimization is iterative, convergence to a good solution is achieved only by providing the optimization process with a good initial guess. If no appropriate initial guess is provided for the method, the optimization may remain at the local minimum and this causes the calibration to fail [51].
  - Adding distortion coefficients to the model causes the optimization to be unstable (i.e. false solutions or divergence may be obtained) due to the interaction between the external and distortion parameters unless the iteration procedure has been carefully designed[106].

2. ***Linear techniques***

Methods such as [33], [41], [44], [58] have dealt with linear calibration. In order to estimate the transformation matrix, the linear techniques employ the *least square method*. So, the parameters are computed by a non-iterative algorithm based on a closed-form solution [106]. Usually, these methods define linear equations using several intermediate parameters. Then, the solution of the intermediate parameters helps solve for the final parameters.

- ***Advantages***

  The main benefit of these methods is that they are straightforward, which also means they run very fast.

- ***Disadvantages***
  - These methods estimate the transformation matrix implicitly. Therefore, extracting intrinsic and extrinsic parameters from the transformation matrix should also be performed separately.
  - No information about the lens imperfection is presented by these methods. Consequently, lens distortions cannot be fixed, unless another transformation is considered, such as in [86] whereby the coefficients were transformed into an eigenvalue problem.
  - Because the algorithms are supposed to be intrinsically non-iterative, the real constraints in the intermediate parameters are neglected, causing the intermediate solution to not satisfy the constraints in the presence of noise, as even a small inaccuracy in the transformation matrix can cause a noticeable amount of error [49], [106].

- Linear techniques are not able to optimally estimate the calibration parameters because they do not produce minimum variance estimates [49].

3. ***Two-stage techniques***
   Studies such as [96], [104], [106], [114], [115] proposed a two-stage method. First, an initial guess for several parameters can be obtained by a linear optimization process. In fact, a direct solution for most calibration parameters is obtained [106]. In the next step, the rest of the parameters are estimated iteratively. The two-stage technique is a combination of both aforementioned methods that aims to overcome their shortcomings. In fact, based on a linear initial guess, a converged solution is guaranteed through fewer iterations.

Also, I can classify the camera calibration methods from different perspectives, each of which emboldens one aspect of available approaches. For example, one can classify current methods as follows:

- Linear and non-linear methods
  These approaches try to formulate the calibration problem as a linear or non-linear problem. For example, one can ignore higher degree polynomials such as distortion parameters and formulate the problem as a linear problem.

- Extrinsic and intrinsic methods
  - Methods such as those used in [67], [77] calculate the intrinsic parameters of a camera.
  - In [100], only extrinsic parameters are computed for a camera mounted on a robot. Also, camera location is based on point correspondences or line correspondences calculated in [69].

- Explicit and implicit methods
  Some approaches only consider solving one set of camera parameters, such as the following:
  - Implicit methods do not use the physical parameters of the camera [104].
  - Explicit methods [10], [96] explicitly use the camera's physical parameters.

- Making use of prior knowledge of the camera or the 3D space
  These methods enjoy having partial information from either the camera or 3D world and try to formulate the problem in ways that are more beneficial once the prior knowledge is available.

1. Traditional calibration approaches (methods that use a calibration pattern)

- Methods that use known 3D points such as [33], [96]
- Methods that use a reduced set of 3D points [55]

2. Methods that do not use a calibration pattern but either use partially known camera information or use no prior information:
   - Self-calibration / auto-calibration [1], [32], [47], [48]
   - Active calibration [5]–[7], [9]
   - Methods that do not use a known set of points or a so-called calibration pattern [17], [24], [102]
   - Methods that are based on deep learning [11], [54], [65]

2.1 Camera Model

In computer vision, a camera is described by a mathematical equation to approximate how a a physical camera projects the points in the 3D world to the 2D sensor images. The camera, in fact, is a mapping from 3D space to a 2D space through a central projection device, which is a specialization of a general projective camera, as Hartley [46] stated.

2.1.1 Projective Camera

A general projective camera, as depicted in Fig. 2.1, maps the 3D points to 2D image points based on the following equation:

$$\mathbf{x} = P\mathbf{X} \tag{2.1}$$

where the $\mathbf{x}$ is a vector that consists of a point $[u, v, 1]^T$ in image coordinate space, and u and v are horizontal and vertical coordinates of a point in the image. In a homogeneous coordinate system, one can treat P as a 4×4 matrix. In this case, a fourth element with a value of 1 is added to X, which denotes the coordinates of a 3D point as follows:

$$s \begin{bmatrix} u \\ v \\ 11 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2.2}$$

where s is scale of the points and x, y and z are the coordinates of the 3D point.

2.2 General Camera Calibration in Computer Vision and Photogrammetry

In this section, we review the most important general camera calibration methods, namely those of Tsai [96], Hall [44], Weng [106], Heikkilä [51] and Zhang [114], [115] which are by

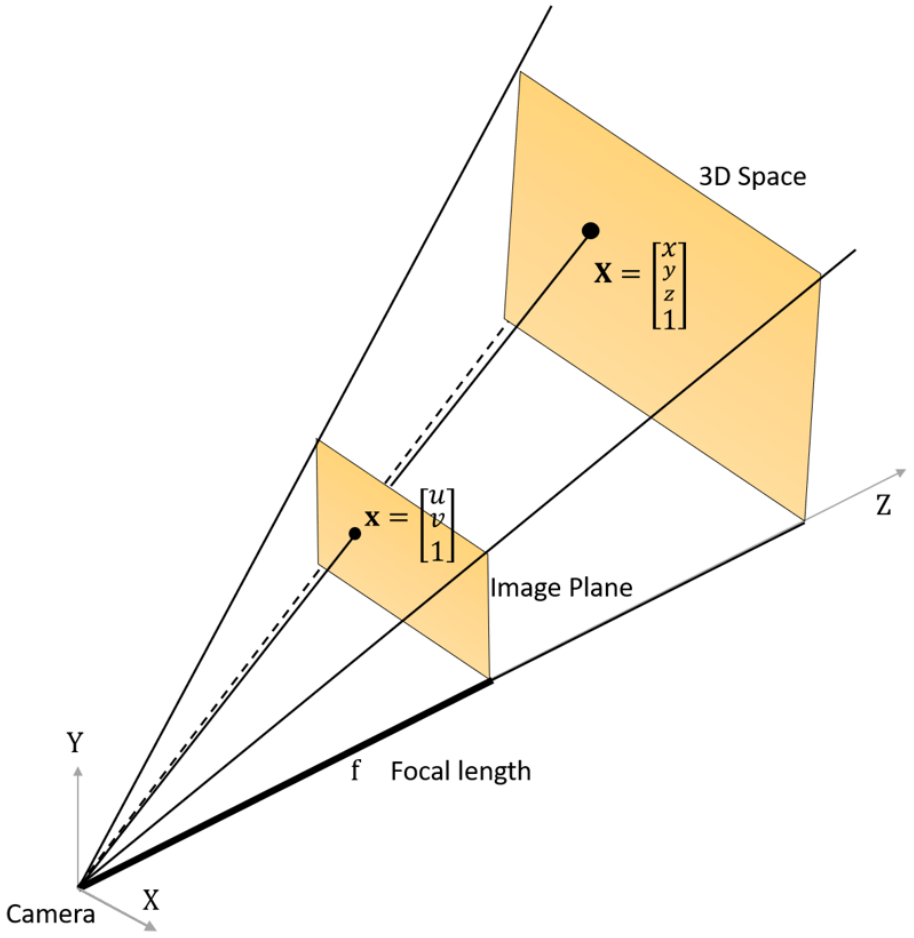


Figure 2.1: Illustration of a projective camera located at the center of the world coordinate system.

far the most popular methods that have been employed in numerous applications after their proposals. At the end of the chapter, we deeply investigate methods [6] and then explain how we can model the lens distortion.

2.3 Tsai's Method

The well-known two-stage calibration method of Tsai [96] was originally motivated by reducing the dimensionality of the parameter's space. To do this, some parameters needed to be constrained in a way that helps solve the equations for other parameters. Based on this idea, Tsai defined a real constraint function and named it the *radial alignment constraint* (RAC). Using the RAC, Tsai [96] calculated the *extrinsic parameters* (except for T_z) and one of the *intrinsic parameters*, namely s_x. As a result, the number of unknowns in the final linear equation was reduced from 7 to 5, which caused to the computation speed to increase and also eliminated the need for having an initial guess.

To begin the calibration or, as he called it *the physical setup*, he used a checkerboard pattern plane to facilitate having mono-view coplanar points [96]. Next, for the *coplanar case*, he chose the location of the 3D world origin in such a way that the points on the checkerboard plane were placed on the XY plane as well, or in other words, as $Z = 0$ for all of them. Also, he indicated that points should not be close to the Y axis to avoid having $T_Y = 0$. Then, using a two-fold algorithm, he first estimated the rotation matrix in stage one. Next, by ignoring the lens distortion parameters, he estimated focal length and translation of the camera on the z-axis. Having initial values for the focal length and depth of the camera, he could refine the estimation along with the lens distortion parameter through a closed-form equation.

2.3.1 Pros and Cons of Tsai's Method

- ***Advantages***
  - For most of the parameters, a closed-form solution is derived which is immune to the lens radial distortion [106].
  - The high number of computed parameters in the closed-form solution, decreases the number of parameters involved in the iterative process.
- ***Disadvantages***
  - Other types of lens distortions have not been modelled in his method. So, his method can only handle radial distortion [106].
  - His method completely discarded the *radial component* of the point and only modelled the *tangential component* of it since the radial components are not

reliable enough. However, neglecting even an unreliable radial component that is responsible for at least half of the observations results in a less reliable estimator [106].

2.4 Hall's Method

Hall [44] proposed a calibration model to measure surface points from an image. He used a recorded image of a projected pattern to relate the points of the image to the points in the 3D world. He tested his method by embedding it in a robot to locate a ball. So, he took the transformation between the 3D real world points and 2D points from the image coordinate system denoted in Eq.2.2 and derived a separate equation for each coordinate component in the image coordinate system as follows:

$$u = \frac{P_{11}x + P_{12}y + P_{13}z + P_{14}}{P_{31}x + P_{32}y + P_{33}z + P_{34}} \tag{2.3}$$

$$v = \frac{P_{21}x + P_{22}y + P_{23}z + P_{24}}{P_{31}x + P_{32}y + P_{33}z + P_{34}} \tag{2.4}$$

By rearranging the coefficients and variables, the following equations are obtained.

$$\begin{aligned} &P_{11}x - P_{31}xu + P_{12}y - P_{32}yu + P_{13}z - P_{33}zu + P_{14} - P_{34}u = 0 \\ &P_{21}x - P_{31}xv + P_{22}y - P_{32}yv + P_{23}z - P_{33}zv + P_{24} - P_{34}v = 0 \end{aligned} \tag{2.5}$$

The next step is to arrange unknowns represented by P into a separate matrix as following.

$$QA = 0 \tag{2.6}$$

where

$$A = [P_{11}\, P_{12}\, P_{13}\, P_{14}\, P_{21}\, P_{22}\, P_{23}\, P_{24}\, P_{31}\, P_{32}\, P_{33}\, P_{34}]^T \tag{2.7}$$

Every term of Eq. 2.5 must be included in the Q matrix. Hall [44] arranged matrix Q in a way that for every pair of point correspondences, two rows are added to the matrix: one for their X coordinates and one for their Y coordinates. So the dimension of matrix; Q is $2n \times 12$. Note that n should be greater or equal to 6 and points must be located in a non-coplanar manner.

$$Q_{2i-1} = \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -x_{wi}u_i \\ -y_{wi}u_i \\ -z_{wi}u_i \\ -u_i \end{bmatrix}^T , Q_{2i} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \\ -x_{wi}v_i \\ -y_{wi}v_i \\ -z_{wi}v_i \\ -v_i \end{bmatrix}^T \tag{2.8}$$

Because the equation $QA = 0$ is in a homogeneous form, the result contains a scale factor. To normalize the result, A_{34} should be unity ($A_{34} = 1$) [44]. To solve this linear system by least squares as a pseudo-inverse method, it is necessary to break the matrix A and Q and bring the last column of the Q to the other side of equation and then apply the least squares method. Therefore, we break the matrix as we obtain $Q'A' = B'$. where

$$A' = [A_{11}\, A_{12}\, A_{13}\, A_{14}\, A_{21}\, A_{22}\, A_{23}\, A_{24}\, A_{31}\, A_{32}\, A_{33}]^T \tag{2.9}$$

$$Q'_{2i-1} = \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -x_{wi}u_i \\ -y_{wi}u_i \\ -z_{wi}u_i \end{bmatrix}^T , Q'_{2i} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \\ -x_{wi}v_i \\ -y_{wi}v_i \\ -z_{wi}v_i \end{bmatrix}^T \tag{2.10}$$

$$B'_{2i-1} = [u_i] \qquad B'_{2i} = [v_i] \tag{2.11}$$

The final result of the unknown matrix A' can be computed by the following pseudo-inverse equation.

$$A' = (Q'^T Q')^{-1} Q'^T B' \tag{2.12}$$

2.5 Weng's Method

The approach put forth by Weng [106] is a two-stage camera calibration model. The first step consists of estimating the camera calibration parameters (except for distortion coefficients) using closed-form solutions. In the second step, the obtained results of the parameters in the first step and the distortion coefficients are then improved together through a non-linear optimization process.

Weng aimed to address two critical questions. First, is it necessary to include the lens distortion parameters in the calibration model? Second, how much improvement is achieved in general by adding the distortion parameters? In order to find the answers to these questions, he drew a comparison between a simplified version of his calibration method and a complex version of it, that was designed for dealing with the lens distortion. For the complex version, he considered two models to investigate. One model only had the radial distortion parameter. The other was equipped with the full model of lens distortions containing both radial and tangential components.

2.5.1 Camera Model

To have a more suitable linear model, Weng [106] simplified the pinhole camera model equations using several intermediate parameters. First, let's consider the relationship between the world coordinate system and the camera coordinate system.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \tag{2.13}$$

where R is a 3×3 rotation matrix and T is a translation vector. Consider that the principal point of the image is indicated by $O'(u_0, v_0)$, which may not be located on the exact geometrical center of the image plane. Every point on the image plane is computed by:

$$\begin{aligned} u &= f\frac{x_c}{z_c} \\ v &= f\frac{y_c}{z_c} \end{aligned} \tag{2.14}$$

The position of the corresponding image points (u, v) with respect to the image coordinate center (top left of the image) can be expressed by

$$\begin{aligned} u - u_0 &= s_u u \\ v - v_0 &= s_v v \end{aligned} \tag{2.15}$$

Combining Eq. 2.14 and Eq. 2.15 yields equations that relates pixel position, the world coordinates and other parameters together.

$$\begin{aligned} \frac{u}{f} = \frac{u - u_0}{f_u} &= \frac{r_{11}x + r_{12}y + r_{13}z + t_1}{r_{31}x + r_{32}y + r_{33}z + t_3} = \hat{u} \\ \frac{v}{f} = \frac{v - v_0}{f_u} &= \frac{r_{21}x + r_{22}y + r_{23}z + t_2}{r_{31}x + r_{32}y + r_{33}z + t_3} = \hat{v} \end{aligned} \tag{2.16}$$

where $(\hat{u}, \hat{v})$ represent the coordinates in the normalized image plane, located at $z_c = 1$ and $f_u = s_u f$ and $f_v = s_v f$, and are called row focal length and column focal length, respectively [106].

Due to several types of imperfections in the lenses, the aforementioned equations will not hold true unless we include the geometrical displacement of the distortion in them.

$$\begin{aligned} u_d &= u + \delta_u(u, v) \\ v_d &= u + \delta_v(u, v) \end{aligned} \tag{2.17}$$

where (u, v) is the coordinate of a non-observable and undistorted image point and (u_d, v_d) is the corresponding distorted coordinate. By substituting Eq. 2.15 in Eq. 2.17, the following equations are obtained.

$$\begin{aligned} u + \delta_u(u, v) &= \frac{u_d - u_0}{-s_u} \\ v + \delta_v(u, v) &= \frac{v_d - v_0}{-s_v} \end{aligned} \tag{2.18}$$

Two intermediate parameters are defined by:

$$\begin{aligned}\hat{u}_d &= \frac{(u_d - u_0)}{f_u} \\ \hat{v}_d &= \frac{(v_d - v_0)}{f_v}\end{aligned} \tag{2.19}$$

Based on the two new variables, Eq. 2.18 is redefined.

$$\begin{aligned}\frac{u}{f} &= \hat{u}_d - \frac{\delta_u(u,v)}{f} \\ \frac{v}{f} &= \hat{v}_d - \frac{\delta_v(u,v)}{f}\end{aligned} \tag{2.20}$$

It is obvious that since the image is contaminated with noise and distortion, the exact values of u and v cannot be obtained by observation. Based on this fact, instead of computing the distortion with respect to the u and v, the distortion δ'_u and δ'_v with respect to the intermediate variables $(\hat{u}_d, \hat{v}_d)$ is computed. Therefore, the following equation relates the distorted to the undistorted points.

$$\begin{aligned}\frac{u}{f} &= \hat{u}_d + \delta_u(\hat{u}_d, \hat{v}_d) \\ \frac{v}{f} &= \hat{v}_d + \delta_v(\hat{v}_d, \hat{u}_d)\end{aligned} \tag{2.21}$$

In the last step, all coefficients belonging to lens distortion should be conflated in one equation to obtain a total camera model. However, first by replacing four terms ($g_1 = s_1 + p_1$, $g_2 = s_2 + p_2$, $g_3 = 2p_1$, $g_4 = 2p_2$) in Eq. 2.141 and Eq. 2.142, the final total lens distortion model is obtained.

$$\begin{aligned}\delta_u &= (g_1 + g_3)u^2 + g_1 v^2 + g_4 uv + \kappa_1 u(u^2 + v^2) \\ \delta_v &= (g_2 + g_4)v^2 + g_2 u^2 + g_3 uv + \kappa_1 v(u^2 + v^2)\end{aligned} \tag{2.22}$$

The camera model is obtained by replacing Eq. 2.22 and combining the last four equations. Note that this equation is finally used to determine the back-projection line of a sensed point in the world coordinate system [106].

$$\begin{aligned}\frac{r_{11}x + r_{12}y + r_{13}z + t_1}{r_{31}x + r_{32}y + r_{33}z + t_3} &= \hat{u}_d + (g_1 + g_3)\hat{u}_d^2 + g_4\hat{u}_d\hat{v}_d + g_1\hat{v}_d^2 + \kappa_1\hat{u}_d(\hat{u}_d^2 + \hat{v}_d^2) \\ \frac{r_{21}x + r_{22}y + r_{23}z + t_2}{r_{31}x + r_{32}y + r_{33}z + t_3} &= \hat{v}_d + (g_1 + g_4)\hat{v}_d^2 + g_3\hat{u}_d\hat{v}_d + g_2\hat{u}_d^2 + \kappa_1\hat{v}_d(\hat{u}_d^2 + \hat{v}_d^2)\end{aligned} \tag{2.23}$$

The polynomial equations (Eq. 2.23) that have derived by considering the three types of lens distortions, can now be used in the optimization process. So, the calibration problem is changed to a polynomial fitting problem that fits a polynomial to the measured image data.

2.5.2 Optimization

The parameters that are unknown and should be estimated are divided into two groups. The first group consists of parameters of the linear equation that can be solved by a closed-form solution. Weng [106] called it vector m.

$$m = (u_0, v_0, f_u, f_v, t_x, t_y, t_z, \alpha, f_v, \gamma)^T \tag{2.24}$$

where α, f_v, γ are the independent rotation elements and t_x, t_y, t_z are components of the translation vector. The second group that should be estimated in a non-linear optimization consists of five lens distortion coefficients, combined in a vector named d.

$$d = (\kappa_1, g_1, g_2, g_3, g_4)^T \tag{2.25}$$

Prior to doing the optimization, the function that needs to be minimized has to be specified. Consider two types of points we have observed that in the calibration initialization. The real 3D points' set is denoted by Ω and its correspondents' set is specified by ω. The task of the optimization process is to minimize the function $F(\Omega, \omega, m, d)$. To minimize for coefficients grouped by vector m and d, Weng proposed the following procedure.

1. First, assume all terms in d are zero. This changes Eq.2.23 to a linear equation.

2. Compute m that minimizes function $F(\Omega, \omega, m, d)$. Note that d is fixed.

$$\min_m \; F(\Omega, \omega, m, d) \tag{2.26}$$

3. Fix m with the values calculated from the last step and minimize function $F(\Omega, \omega, m, d)$ based on different values for d. So, every set of values for d that minimizes $F(\Omega, \omega, m, d)$ is selected.

$$\min_d \; F(\Omega, \omega, m, d) \tag{2.27}$$

4. If the minimization error is less than a threshold, the process is stopped, otherwise the same process is repeated from step 2.

One of the reasons why m and d are computed in different steps is that minimizing both of them simultaneously causes the optimization to reach false minima since the distortion parameters (d) are able to significantly relocate the position of the points according to the image coordinate system[106]. The second reason is that parameters in m cannot be estimated well in the presence of the distortion coefficients because the equations are formulated as a distortion-free model. The third reason states that for estimating d, it is necessary to have an initial estimation of the distortion-free parameters. Therefore, the coefficient partitioning to m and d notably reduces the harmful interaction between them [106]. For the

first iteration of the optimization (which is when the initial solution has to be found), the position of the correspondent points does make a difference. Weng [106] chose the points around the center of the image where he believed that there was lower distortion. So, it correlated with his initial assumption to consider $d = 0$. However, points could not also be so close to the center because, in that case, the external parameters' values were not accurate. He selected points that were located in the radius from the center, equal to one fourth of the image's dimensions.

2.6 Faugerass' Method and Radial Distortion

Extension of Faugeras-Toscani [33] method proposed in[81], [83]. Since the Faugeras-Toscani model is a distortion-free model, the calibration results in real models cannot be trusted. Therefore, the radial distortion has been included in the camera model in the extended version. Consider the following equation derived from the projective camera mapping denoted in Eq. 2.2

$$\begin{aligned} x_d + x_d\kappa_1 r^2 &= f\frac{r_{11}x + r_{12}y + r_{13}z + t_x}{r_{31}x + r_{32}y + r_{33}z + t_z} \\ y_d + y_d\kappa_1 r^2 &= f\frac{r_{21}x + r_{22}y + r_{23}z + t_y}{r_{31}x + r_{32}y + r_{33}z + t_z} \end{aligned} \tag{2.28}$$

where $r = \sqrt{x_d^2 + y_d^2}$ is the radius of distortion.

Also, because of the use of the lens distortion coefficient in the camera model, it is no longer possible to employ a linear minimization method (least square). Instead, the camera parameters need to be iteratively calibrated using the Newton-Raphson or Levenberg-Marquardt methods [89].

$$\begin{aligned} U(x) &= f\frac{r_{11}x + r_{12}y + r_{13}z + t_x}{r_{31}x + r_{32}y + r_{33}z + t_z} - \frac{(u_d - u_0)}{-f_u} - \\ &\kappa_1((\frac{(u_d - u_0)}{-f_u})^2 + (\frac{(v_d - v_0)}{-f_v})^2)(\frac{(u_d - u_0)}{-f_u}) \\ V(x) &= f\frac{r_{21}x + r_{22}y + r_{23}z + t_y}{r_{31}x + r_{32}y + r_{33}z + t_z} - \frac{(v_d - v_0)}{-f_v} - \\ &\kappa_1((\frac{(u_d - u_0)}{-f_u})^2 + (\frac{(v_d - v_0)}{-f_v})^2)(\frac{(v_d - v_0)}{-f_v}) \end{aligned} \tag{2.29}$$

To solve this non-linear system according to the all n calibrating points, the Newton-Raphson minimization is employed.

2.7 Heikkilä's Method

Heikkilä's calibration method [51] paid special attention to the model-fitting problem of the available camera calibration pipelines which was among several neglected aspects of

calibration methods such as control point extraction from images, image correction, and errors arising from these stages. He proposed an extension to the two-stage calibration method and calibrated the camera using a four-step approach.

Later on, he proposed an improved version [49] of a camera model that used circular control points for calibration. In Heikkilä's model the points in the image are related to the points in the 3D world by the following equation:

$$\begin{bmatrix} \lambda u_d \\ \lambda v_d \\ \lambda \end{bmatrix} = \mathbf{F} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2.30}$$

where $\mathbf{F} = PM$ is the perspective transformation matrix.

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} sf & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ \mathbf{M} &= \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \\ \mathbf{t} &= \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T \end{aligned} \tag{2.31}$$

where $\mathbf{R}$ is a 3×3 orthonormal rotation matrix which is defined by the three Euler angles α, β, γ.

$$\begin{aligned} \alpha &= \sin^{-1}(r_{31}) \\ \beta &= \arctan 2(-\frac{r_{32}}{\cos\beta}, \frac{r_{33}}{\cos\beta}) \\ \gamma &= \arctan 2(-\frac{r_{21}}{\cos\beta}, \frac{r_{11}}{\cos\beta}) \end{aligned} \tag{2.32}$$

where r_{ij} is an element of the following rotation matrix.

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{2.33}$$

2.7.1 Calibrating the Camera

Heikkilä's proposed procedure [50] for calibrating the camera is mainly suitable for circular landmarks; however, it can also be used when the extracted points are shapeless, i.e. without any specific geometry.

His method calibrates eight intrinsic parameters $\gamma_{int}[s, f, u_0, v_0, \kappa_1, \kappa_2, p_1, p_2]^T$ and six extrinsic parameters $\gamma_{ext}[t_x, t_y, t_z, \alpha, \beta, \gamma]^T$. The radius r of every circle in pattern is assumed to be known in advance.

To find the optimal solution, an iterative search algorithm is employed to estimate the following vector.

$$\gamma = \begin{bmatrix} \gamma_{int}^T & \gamma_{ext}^T(1) & \gamma_{ext}^T(2) & \dots & \gamma_{ext}^T(K) \end{bmatrix}^T \tag{2.34}$$

The First Stage

Like all of the two-stage methods, in this stage, the method tries to find an initial guess for the linear equations while discarding the lens distortions coefficients. Assuming that the circular control points are not coplanar, all elements of the perspective transformation matrix $\mathbf{F}$ in Eq. 2.30 have non-zero values. In this stage, $\mathbf{F}$ is redefined as $\hat{\mathbf{F}}$ and is related to the intrinsic and extrinsic parameters by the following equation.

$$\hat{\mathbf{F}} = \mathbf{P}\hat{\mathbf{M}} = \begin{bmatrix}\mathbf{P}_{13}\hat{\mathbf{R}} & \mathbf{P}_{13}\hat{\mathbf{t}}\end{bmatrix} \tag{2.35}$$
$$\mathbf{P}_{13} = \begin{bmatrix} sf & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ are the estimates of $\mathbf{R}$ and $\mathbf{t}$, respectively. By solving Eq. 2.35 we have:

$$\hat{\mathbf{R}} = \mathbf{P}_{13}^{-1}\hat{\mathbf{F}}_{13} \tag{2.36}$$

$$\hat{\mathbf{t}} = \mathbf{P}_{13}^{-1}\hat{\mathbf{F}}_{4} \tag{2.37}$$

where $\hat{\mathbf{F}}_{13}$ is the first three columns of the matrix $\hat{\mathbf{F}}$ and $\hat{\mathbf{F}}_4$ is the fourth column of the matrix $\hat{\mathbf{F}}$. Since the estimation of the matrix $\hat{\mathbf{R}}$ does not hold the characteristics of an orthonormal matrix anymore, the obtained $\hat{\mathbf{R}}$ matrix is normalized and orthogonalized using the Singular Value Decomposition (SVD). So, the new rotation matrix is obtained.

$$\hat{\mathbf{R}}' = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{2.38}$$

The new rotation matrix now carries the orthonormal properties of a rotation matrix. Therefore, the three Euler angles α, β, γ can be calculated using Eq. 2.32.

At the end of the first stage, the value of the vector γ is

$$\hat{\gamma}_0 = [1, f_0, \frac{N_u}{2}, \frac{N_v}{2}, 0, 0, 0, 0, \gamma_{ext}^T(1), \gamma_{ext}^T(2), ..., \gamma_{ext}^T(K)]^T \tag{2.39}$$

The Second Stage

In the second stage, the calibration parameters are estimated by minimizing the sum of squared differences between the observation and the model [49]. The observed coordinates of the N circular control points of K images are specified by $e_o(n,k)$, and their corresponding vector obtained by the camera model is denoted by $e_d(n,k)$. These two vectors are combined to constitute the following vector.

$$\mathbf{y}(\gamma) = [(e_0(1,1) - e_d(1,1))^T, (e_0(2,1) - e_d(2,1))^T, ..., (e_0(N,K) - e_d(N,K))^T]^T \tag{2.40}$$

The combined vector is then iteratively minimized by the following objective function.

$$J(\gamma) = \mathbf{y}^T(\gamma)\mathbf{C}_\mathbf{e}^{-1}\mathbf{y}(\gamma) \tag{2.41}$$

where $\mathbf{C_e}$ is the covariance matrix of the observation error [49].

2.8 Zhang's Method

Zhang [114], [115] proposed a calibration method suitable for desktop computers which soon became very popular. His technique only required a planar pattern to be seen by the camera in a few different orientations (or at least two). The pattern or the camera can be moved by hand, and there is also no need to know the motion [115]. His method is very flexible because almost every one is able to print the calibration pattern, attach it to a rigid planar surface and take several photos of it with various orientations. The images can then be used to calibrate the camera. Zhang's method combines ideas from classic photogrametric calibration with concepts of self-calibration, which use implicit 3D information or motion rigidity[115].

In Zhang's model [115] the relationship between a 3D point and its image projection is expressed by

$$\begin{bmatrix} su_d \\ sv_d \\ s \end{bmatrix} = \begin{bmatrix} f_u & \phi & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2.42}$$

where s is an arbitrary scale factor. The first matrix represents the intrinsic parameters and is called A. The parameters (u_0, v_0) are the coordinates of the principal point in the image, and f_u and f_v are focal lengths (scale factors) in u and v axes of the image, respectively. The parameter ϕ expresses the skew of the two image axes [115].

$$A = \begin{bmatrix} f_u & \phi & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.43}$$

The extrinsic matrix, which describes the rotation and translation of the camera can also be written in a shorter manner.

$$[R|t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{2.44}$$

Therefore, Eq. 2.42 can be shortened as

$$\begin{bmatrix} su_d \\ sv_d \\ s \end{bmatrix} = A[R|t] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2.45}$$

If we assume that the calibration plane in the 3D world lies in plane $Z = 0$ of the world coordinate system, the following relation is obtained. Note that r_i represents the ith column of the rotation matrix.

$$\begin{bmatrix} su_d \\ sv_d \\ s \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \tag{2.46}$$

It can be seen that the image plane is related to a 2D plane in the 3D world, so it is reasonable to consider the transformation matrix as a homography matrix.

$$s\mathbf{x}_d = H\mathbf{X}$$
$$H = A\begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \tag{2.47}$$

For every point, there would be one equation; therefore, in total there are $2n \times 9$ equations. The value of each elements of homography matrix is then used in Eq. 2.47.

$$\begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} = \lambda A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \tag{2.48}$$

where λ is an arbitrary scaler. So, based on this equality, we can write

$$h_1 = Ar_1 \ \rightarrow r_1 = A^{-1}h_1 = h_1^T A^{-T}$$
$$h_2 = Ar_2 \ \rightarrow r_2 = A^{-1}h_2 = h_2^T A^{-T} \tag{2.49}$$

Since the rotation components r_1 and r_2 are orthonormal, the dot product of each pair of them should be zero. In addition, each vector has a unit length. So, the following relationships are obtained.

$$r_1.r_2 = 0 \ \rightarrow h_1^T A^{-T} A^{-1} h_2 = 0 \tag{2.50}$$

$$r_1.r_1 = r_2.r_2 = 1 \ \rightarrow h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 \tag{2.51}$$

Note that $A^{-T}A^{-1}$ describes the image of the absolute conic [115].

2.8.1 Calibrating the Camera

To calibrate the camera, Zhang employs a two-stage strategy. The first stage uses a closed-form solution as an initial guess for the second stage, which is a nonlinear maximum likelihood estimation. In fact, in the second stage the value of coefficients will be refined.

The First Stage

Consider $B = A^{-T}A^{-1}$ as a new intermediate parameter. By calculating the matrix product, the following equation is obtained.

$$B = A^{-T}A^{-1} = \begin{bmatrix} f_u & 0 & 0 \\ \phi & f_v & 0 \\ u_0 & v_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} f_u & \phi & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \left(\begin{bmatrix} f_u & \phi & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_u & 0 & 0 \\ \phi & f_v & 0 \\ u_0 & v_0 & 1 \end{bmatrix} \right)^{-1}$$
$$= \begin{bmatrix} f_u^2 + \phi^2 + u_0^2 & f_v\phi + u_0 v_0 & u_0 \\ f_v\phi + u_0 v_0 & f_v^2 & v_0 \\ u_0 & v_0 & 1 \end{bmatrix}^{-1}$$

$$= \frac{\begin{bmatrix} f_v^2 - v_0^2 & u_0v_0 - f_v\phi v_0 - u_0v_0^2 & f_v\phi v_0^2 + u_0v_0^2 - f_v^2u_0 \\ u_0v_0 - f_v\phi - u_0v_0 & f_u^2 + \phi^2 + u_0^2 - u_0^2 & f_v\phi u_0 - u_0^2v_0 \\ f_v\phi v_0 - f_v^2u_0 & f_v\phi u_0 - f_u^2v_0 + \phi^2v_0 + u_0^2v_0 & f_u^2f_v^2 + \phi^2f_v^2 + f_v^2u_0^2 - (f_v\phi + u_0v_0)^2 \end{bmatrix}}{(f_u^2 + \phi^2 + u_0^2)(f_v^2 - u_0v_0) - (f_v\phi + u_0v_0)(f_v\phi + u_0v_0 - u_0v_0) + (u_0)(f_v\phi v_0 + u_0v_0^2)(f_v^2u_0)}$$

$$\rightarrow= \begin{bmatrix} \frac{1}{f_u^2} & \frac{-\phi}{f_u^2f_v} & \frac{\phi v_0 - u_0f_v}{f_u^2f_v} \\ \frac{-\phi}{f_u^2f_v} & \frac{\phi^2}{f_u^2f_v^2} + \frac{1}{f_v^2} & -\frac{\phi(\phi v_0 - u_0f_v)}{f_u^2f_v^2} - \frac{v_0}{f_v^2} \\ \frac{\phi v_0 - u_0f_v}{f_u^2f_v} & \frac{\phi(\phi v_0 - u_0f_v)}{f_u^2f_v^2} - \frac{v_0}{f_v^2} & \frac{(\phi v_0 - u_0f_v)^2}{f_u^2f_v^2} + \frac{v_0^2}{f_v^2} + 1 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \quad (2.52)$$

Matrix B is symmetric and defines the image of the absolute conic, so only six elements should be calculated. The vector consisting of these six elements is called $\mathbf{b}$.

$$\mathbf{b} = \begin{bmatrix} B_{11} & B_{12} & B_{13} & B_{22} & B_{23} & B_{33} \end{bmatrix}^T \quad (2.53)$$

If the ith column vector of $\mathbf{H}$ expressed by $\mathbf{h}_i = [h_{i1}\ h_{i2}\ h_{i3}]$, then we can write

$$\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j = \mathbf{v}_{ij}^T\mathbf{b} \quad (2.54)$$

where

$$\mathbf{v}_{ij} = \begin{bmatrix} h_{i1}h_{j1} & h_{i1}h_{j2} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} + h_{i1}h_{j3} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{bmatrix} \quad (2.55)$$

Thus, Eq. 2.50 and Eq. 2.51 can be rewritten as two homogeneous equations:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0 \quad (2.56)$$

For n images of calibration patterns, $2n \times 6$ equations are stacked up into the matrix $\mathbf{V}$

$$\mathbf{V}\mathbf{b} = 0 \quad (2.57)$$

Finally, the solution of Eq. 2.57 is the eigenvector of $\mathbf{V}^\mathbf{V}$ associated to the smallest eigenvalue [114], [115]. By solving Eq. 2.57 and estimating vector $\mathbf{b}$, all camera intrinsic parameters can be calculated by the following equations.

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2} \quad (2.58)$$

$$\lambda = B_{33} - \frac{B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{13})}{B_{11}} \quad (2.59)$$

$$f_u = \sqrt{\frac{\lambda}{B_{11}}} \quad (2.60)$$

$$f_v = \sqrt{\frac{\lambda B_{11}}{B_{11}B_{22} - B_{12}^2}} \quad (2.61)$$

$$\phi = -\frac{B_{12}f_u^2f_v}{\lambda} \quad (2.62)$$

$$u_0 = \frac{\phi v_0}{f_u} - \frac{B_{13}f_u^2}{\lambda} \quad (2.63)$$

Using these equations, matrix $\mathbf{A}$ is formed and contributes to calculations of extrinsic parameters.

$$r_1 = \lambda \mathbf{A}^{-}1\mathbf{h}_1 \tag{2.64}$$

$$r_2 = \lambda \mathbf{A}^{-}1\mathbf{h}_2 \tag{2.65}$$

$$r_3 = r_1 \times r_2 \tag{2.66}$$

$$t = \lambda \mathbf{A}^{-}1\mathbf{h}_3 \tag{2.67}$$

The Second Stage

In this stage, the less meaningful solution obtained from the first stage is refined by maximum likelihood inference. The main idea behind this refinement is to minimize the distance of back-projected points from the 3D real world to the image. Note that n is the total number of images containing a calibration pattern (the model plane) and m is the number of points in each model plane.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{x}_{ij} - \hat{\mathbf{X}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{x}_j)\|^2 \tag{2.68}$$

where $\hat{\mathbf{X}}(\mathbf{A}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{x}_j)$ is the projection of $\mathbf{x}_j$ based on the transformation function specified in Eq. 2.42. Consequently, a Levenberg-Marquardt algorithm is employed to minimize Eq. 2.68 using the solution acquired from the first stage as the initial guess.

2.9 Single Image Calibration Methods Using Deep Learning

Deep-learning-based methods usually utilize images acquired from a wide field of view cameras to calibrate the camera as a process in which one tries to estimate some of the camera parameters without using any calibration patterns. For example, the method proposed in [108] can only estimate the focal length from a single view image by first regressing to a horizontal field of view. In contrast, the methods proposed in [80] could only estimate radial distortion without even calibrating the camera. There are several other studies that can estimate more parameters, such as [11] in which the method could estimate the focal length along with distortion parameters or [54] that can approximate the focal length as well as camera orientation by relying on horizon line estimation from the image. A common denominator of most of the methods of these kinds is generating the training data. Since training a CNN and regressors requires seeing millions of images with labelled ground truth values (camera parameters) and there is no dataset that can satisfy this need, most researchers have tried to synthetically create images with ground truth. The more accurate the synthetic data generation is, the better the results are.

2.10 Active Calibration

In this section, we analyze and evaluate active calibration from a mathematical point of view and explain the ambiguities of it. The goal of doing this comparison is to establish a deep understanding of the method that eventually helps detect the weak points of it.

The mathematical derivations in the papers [5]–[7], [9] can be interpreted as an assumption that has been proposed by Kantani [63]. In his proposal, Kantani has suggested that the rotation of the camera is equivalent to the rotation of the scene in the opposite sense.

Unlike in many other calibration methods such as [31], [32], [44], [81], [96], [106], [114], [115], all measurements in active calibration are happening on an image coordinate system, and hence, the computation is in pixel dimensions. That can work to the advantage of active calibration unless the acquisition of one or several input arguments becomes noisy and causes large variations in the output results.

2.10.1 Theoretical Derivation of Active Calibration

Before starting to use active calibration equations, we need to have a correct understanding of the rotation matrices and their directions. It is very important to correctly match the pictures taken from a scene with their correct rotation angles. In fact, the angle is based on the camera pose, but in the mathematical derivation, the rotation of the scene is taken into account. So, having a correct understanding from the rotated image is crucial at outset of the process.

As it has been depicted in the image of an active camera in an article on active calibration [6], and based on the arrows illustrated on the image and the fact that the defined matrix multiplications in [63] is column-wised, we can infer the following.

1. The camera is looking towards the scene along the z-direction. Because in none of the active camera's pictures in the papers is the direction of the rotation around the z-axis defined, we consider a *clockwise* rotation to represent the term *roll rotation*. So, every clockwise rotation has a positive sign and every counter-clockwise rotation is considered as negative. Based on this fact, the matrix that rotates an object around z-axis is calculated by the following factor.

$$R_z(\theta_r) = \begin{bmatrix} \cos(\theta_r) & \sin(\theta_r) & 0 \\ -\sin(\theta_r) & \cos(\theta_r) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.69}$$

2. For pan rotation, a counter-clockwise direction has been explicitly indicated in the paper. So, rotation of the camera to the left is considered as a positive direction. The

rotation matrix is then calculated by the following factor.

$$R_y(\theta_p) = \begin{bmatrix} \cos(\theta_p) & 0 & -\sin(\theta_p) \\ 0 & 1 & 0 \\ \sin(\theta_p) & 0 & \cos(\theta_p) \end{bmatrix} \tag{2.70}$$

3. Rotating around the x-axis has also been specified in the paper and needs to be done in a clockwise direction. It means that a positive angle of rotation is taken into account when the camera is moved downward. In this case, the rotation matrix for tilt is denoted by the following factor.

$$R_x(\theta_t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_t) & -\sin(\theta_t) \\ 0 & \sin(\theta_t) & \cos(\theta_t) \end{bmatrix} \tag{2.71}$$

In short, while the camera rotation around the z-axis (roll) and x-axis (tilt) is clockwise, its direction when it rotates around the y-axis (pan) is counter-clockwise. Note that in all of the three above-mentioned rotations, the camera is supposed to rotate; thus, the scene should be moved in the opposite sense.

In this section, we reconsider the mathematical derivation of the active calibrations equations by proving them from the beginning.

Rotation Transformation of Points in the Image Plane

In active calibration [5]–[7], [9], the camera is placed at the world coordinate center. So, the parameters are described in the pixel system rather than the metric system. Active calibration also assumes that the image plane is parallel to the XY plane. Since the rotation matrix is orthonormal, the projected points $R^{-1} = R^T$ can be easily computed. So, consider that by rotating the camera, a point $\mathbf{x}(u, v)$ is transformed to a new point $\mathbf{x}'(u', v')$ in the image space. In this case, and based on Eq. 2.2 we have

$$\mathbf{x} = \mathbf{PX} \tag{2.72}$$

where $\mathbf{P}$ is the projection matrix of the camera:

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \tag{2.73}$$

If the camera is only rotated about its center, the translation vector would be zero and we therefore have

$$\mathbf{P} = \mathbf{KRX} \tag{2.74}$$

One can calculate the coordinates of the 3D point by:

$$\mathbf{X} = (\mathbf{KR})^{-1}\mathbf{x} \tag{2.75}$$

Assuming the reference image has a zero rotation around each angle, we can replace the rotation matrix with identity, and hence

$$\mathbf{X} = \mathbf{K}^{-1}\mathbf{x} \tag{2.76}$$

Now, if the same point $\mathbf{X}(x, y, z)$ is projected on the sensor of the same camera that is purely rotated, we have

$$\mathbf{X} = (\mathbf{KR})^{-1}\mathbf{x}' \tag{2.77}$$

Combining Eq. 2.76 and Eq. 2.77 yields

$$\mathbf{K}^{-1}\mathbf{x} = (\mathbf{KR})^{-1}\mathbf{x}' \tag{2.78}$$

Finally, the relationship between two points on an image before and after rotating a camera around it center can be denoted by:

$$\mathbf{x} = \mathbf{KR}^{-1}\mathbf{K}^{-1}\mathbf{x}' \tag{2.79}$$

It should be noted that Eq. 2.79 expresses the relationship between a point in the image plane and its new location after the camera transformation in the new image plane.

Now consider a situation in which the camera is rotated by a small tilt (around the u axis) angle. By approximating the $\sin(\theta_t) \approx \theta_t$ and $\cos(\theta_t) \approx 1$, the values of the rotation matrix become

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_t) & -\sin(\theta_t) \\ 0 & \sin(\theta_t) & \cos(\theta_t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\theta_t \\ 0 & \theta_t & 1 \end{bmatrix}$$

By replacing the elements of the rotation matrix in Eq. 2.79, we have

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{u}{1 - \theta_t \dfrac{v}{f_v}} \\ \dfrac{v + \theta_t f_v}{1 - \theta_t \dfrac{v}{f_v}} \\ 1 \end{bmatrix} \tag{2.80}$$

This equation explains the relation between the projected points in the image plane before and after rotating the camera by a small tilt (upside) movement.

A similar procedure can be done for the small pan angle (rotation around the Y) axis. In this case, the rotation matrix is

$$\mathbf{R} = \begin{bmatrix} \cos(\theta_p) & 0 & -\sin(\theta_p) \\ 0 & 1 & 0 \\ \sin(\theta_p) & 0 & \cos(\theta_p) \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\theta_p \\ 0 & 1 & 0 \\ \theta_p & 0 & 1 \end{bmatrix} \tag{2.81}$$

In addition, the new point's coordinate is obtained by the following equation.

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{u + \theta_p f_u}{1 - \theta_p \dfrac{u}{f_u}} \\ \\ \dfrac{v}{1 - \theta_p \dfrac{u}{f_u}} \\ \\ f \end{bmatrix} \tag{2.82}$$

Tilt Movement of the Camera

By moving the camera downward, the following equations are derived from Eq. 2.80. Every point after the transformation is indicated by a subscript denoting the type of transformation. For example, (u_t, v_t) denotes the coordinate of a point in the images after tilt rotation of the reference image, and θ_t represents the tilt angle.

$$\begin{aligned} u_t &= \frac{u}{1 - \theta_t \dfrac{v}{f_v}} \\ v_t &= \frac{v + \theta_t f_v}{1 - \theta_t \dfrac{v}{f_v}} \end{aligned} \tag{2.83}$$

For each equation of u_t and v_t, we can use the exact same equation or estimate the denominator using the Taylor series. We consider both situations here.

$$u_t = \frac{u}{1 - \theta_t \dfrac{v}{f_v}} \rightarrow \begin{cases} u_t = u(1 - \theta_t \dfrac{v}{f_v})^{-1} \rightarrow f_v = \left(\dfrac{\theta_t}{u_t - u}\right) u_t v \\ \\ u_t \approx u(1 + \theta_t \dfrac{v}{f_v}) \rightarrow f_v \approx \left(\dfrac{\theta_t}{u_t - u}\right) uv \end{cases} \tag{2.84}$$

It can be seen that the only difference between these two equations is that the estimation is based on the multiplication of the current coordinates. Also, the term $u_t - u$ is still present in the estimation. For the v coordinate, similar relations can be derived but in a quadratic form.

$$v_t = \frac{v + \theta_t f_v}{1 - \theta_t \dfrac{v}{f_v}} \rightarrow \begin{cases} v_t = (v + \theta_t f_v)(1 - \theta_t \dfrac{v}{f_v})^{-1} \rightarrow f_v^2 - \left(\dfrac{v_t - v}{\theta_t}\right) f_v + v v_t = 0 \\ \\ v_t \approx (v + \theta_t f_v)(1 + \theta_t \dfrac{v}{f_v}) \rightarrow f_v^2 + \left(\dfrac{v(1 + \theta_t^2) - v_t}{\theta_t}\right) f_v + v^2 \approx 0 \end{cases} \tag{2.85}$$

In the second equation, which is the result of the approximation, the value of v has been multiplied by one plus a small value. This might help in cases when the variation in the v direction is small, which it is not here, since the movement is alongside the v axis (tilt). Thus, the difference between v coordinates is not small and can be sensitive to noise.

2.10.2 Pan Movement of the Camera

To move the camera around the v axis (pan), we use the rotation matrix explained in the beginning of the section. The direction for the positive movement is counter-clockwise and literally causes the camera to point to the left. From Eq. 2.82, we have

$$\begin{aligned} u_p &= \frac{u + \theta_p f_u}{1 - \theta_p \frac{u}{f_u}} \\ v_p &= \frac{v}{1 - \theta_p \frac{u}{f_u}} \end{aligned} \tag{2.86}$$

where θ_p is the pan angle and (u_p, v_p) denotes the coordinate of a point in the images after pan rotation of the reference image. A similar formulation to tilt transformation can be obtained for pan transformation. We can use the exact equations or use the estimation.

$$u_p = \frac{u + \theta_p f_u}{1 - \theta_p \frac{u}{f_u}} \rightarrow \begin{cases} u_p = (u + \theta_p f_u)(1 - \theta_p \frac{u}{f_u})^{-1} \rightarrow f_x^2 - \left(\frac{u_p - u}{\theta_p}\right) f_u + u u_p = 0 \\ \\ u_p \approx (u + \theta_p f_u)(1 + \theta_p \frac{u}{f_u}) \rightarrow f_u^2 + \left(\frac{u(1+\theta_p^2) - u_p}{\theta_p}\right) f_u + u^2 \approx 0 \end{cases} \tag{2.87}$$

And for the v coordinate after pan, we have:

$$v_p = \frac{v}{1 - \theta_p \frac{u}{f_u}} \rightarrow \begin{cases} v_p = v(1 - \theta_p \frac{u}{f_u})^{-1} \rightarrow f_u = \left(\frac{\theta_p}{v_p - v}\right) u v_p \\ \\ v_p \approx v(1 + \theta_p \frac{u}{f_u}) \rightarrow f_u \approx \left(\frac{\theta_p}{v_p - v}\right) u v \end{cases} \tag{2.88}$$

2.10.3 Roll Movement of the Camera

When a camera rolls (rotates around the z-axis) by angle θ_r, then

$$\begin{aligned} u_r &= \cos(\theta_r) u + \sin(\theta_r) \frac{f_u}{f_v} v \\ v_r &= -\sin(\theta_r) \frac{f_v}{f_u} u + \cos(\theta_r) v \end{aligned} \tag{2.89}$$

2.10.4 Estimating the Principal Point

The center of the image is estimated based on a difference in each direction. To calculate the centers, every u and v is replaced by $u - \delta_u$ and $v - \delta_v$. Because the angle of rotation is small, it is reasonable that we assume that the centers of both images (the reference and the transformed) are very close, or in fact, equal. After replacement the equations are changed as

For tilt movement the following equations are obtained

$$\begin{cases} f_v = \left(\dfrac{\theta_t}{u_t - u}\right) u_t v \rightarrow f_v = \left(\dfrac{\theta_t}{u_t - u}\right)(u_t - \delta_u)(v - \delta_v) \\ \\ f_v \approx \left(\dfrac{\theta_t}{u_t - u}\right) uv \rightarrow f_v \approx \left(\dfrac{\theta_t}{u_t - u}\right)(u - \delta_u)(v - \delta_v) \end{cases} \tag{2.90}$$

$$\begin{cases} f_v^2 - \left(\dfrac{v_t - v}{\theta_t}\right) f_v + v v_t = 0 \rightarrow f_v^2 - \left(\dfrac{v_t - v}{\theta_t}\right) f_v + (v - \delta_v)(v_t - \delta_v) = 0 \\ \\ f_v^2 + \left(\dfrac{v(1+\theta_t^2) - v_t}{\theta_t}\right) f_v + v^2 \approx 0 \rightarrow f_v^2 + \left(\dfrac{v(1+\theta_t^2) - v_t}{\theta_t}\right) f_v + (v - \delta_v)^2 \approx 0 \end{cases} \tag{2.91}$$

Therefore, in order to find the image centers using f_v, the following four equations can be used.

$$f_v = \left(\frac{\theta_t}{u_t - u}\right)(u_t v - u_t \delta_v - v\delta_u + \delta_u \delta_v) \tag{2.92}$$

$$f_v \approx \left(\frac{\theta_t}{u_t - u}\right)(uv - u\delta_v - v\delta_u + \delta_u \delta_v) \tag{2.93}$$

$$f_v^2 - \left(\frac{v_t - v}{\theta_t}\right) f_v + (v v_t - v\delta_v - v_t \delta_v + \delta_v^2) = 0 \tag{2.94}$$

$$f_v^2 + \left(\frac{v(1+\theta_t^2) - v_t}{\theta_t}\right) f_v + (v^2 - 2v\delta_v + \delta_v^2) \approx 0 \tag{2.95}$$

For pan movement, the equations are as follows

$$\begin{cases} f_u^2 - \left(\dfrac{u_p - u}{\theta_p}\right) f_u + u u_p = 0 \rightarrow f_u^2 - \left(\dfrac{u_p - u}{\theta_p}\right) f_u + (u - \delta_u)(u_p - \delta_u) = 0 \\ \\ f_u^2 + \left(\dfrac{u(1+\theta_p^2) - u_p}{\theta_p}\right) f_u + u^2 \approx 0 \rightarrow f_u^2 + \left(\dfrac{u(1+\theta_p^2) - u_p}{\theta_p}\right) f_u + (u - \delta_u)^2 \approx 0 \end{cases} \tag{2.96}$$

$$\begin{cases} f_u = \left(\dfrac{\theta_p}{v_p - v}\right) u v_p \rightarrow f_u = \left(\dfrac{\theta_p}{v_p - v}\right)(u - \delta_u)(v_p - \delta_v) \\ \\ f_u \approx \left(\dfrac{\theta_p}{v_p - v}\right) uv \rightarrow f_u \approx \left(\dfrac{\theta_p}{v_p - v}\right)(u - \delta_u)(v - \delta_v) \end{cases} \tag{2.97}$$

In order to use f_u for estimating the image centers we have the following equations

$$f_u^2 - \left(\frac{u_p - u}{\theta_p}\right) f_u + (uu_p - u\delta_u - u_p\delta_u + \delta_u^2) = 0 \tag{2.98}$$

$$f_u^2 + \left(\frac{u(1+\theta_p^2) - u_p}{\theta_p}\right) f_u + (u^2 - 2u\delta_u + \delta_u^2) \approx 0 \tag{2.99}$$

$$f_u = \left(\frac{\theta_p}{v_p - v}\right)(uv_p - u\delta_v - v_p\delta_u + \delta_u\delta_v) \tag{2.100}$$

$$f_u \approx \left(\frac{\theta_p}{v_p - v}\right)(uv - u\delta_v - v\delta_u + \delta_u\delta_v) \tag{2.101}$$

It is worth noting that in order to have a linear system of equations that is more straightforward to solve, the higher order terms of δ_u and δ_v are neglected during the calculations.

Estimation Using Three Points

By replacing three different points $\{(\bar{u}^{(1)}, \bar{v}^{(1)}), (\bar{u}^{(2)}, \bar{v}^{(2)}), (\bar{u}^{(3)}, \bar{v}^{(3)})\}$ on three different contours into Eq. 2.93, the following equations are obtained. Note that the symbol "$\bar{}$" denotes the average taken over the relevant image contour [6].

$$f_v \approx \frac{\theta_t}{\bar{u}_t^{(1)} - \bar{u}^{(1)}}(\bar{u}^{(1)}\bar{v}^{(1)} - \delta_u\bar{v}^{(1)} - \delta_v\bar{u}^{(1)}) \tag{2.102}$$

$$f_v \approx \frac{\theta_t}{\bar{u}_t^{(2)} - \bar{u}^{(2)}}(\bar{u}^{(2)}\bar{v}^{(2)} - \delta_u\bar{v}^{(2)} - \delta_v\bar{u}^{(2)}) \tag{2.103}$$

Combing Eq. 2.102 and Eq. 2.103 results in:

$$\frac{\theta_t}{\bar{u}_t^{(1)} - \bar{u}^{(1)}}(\bar{u}^{(1)}\bar{v}^{(1)} - \delta_u\bar{v}^{(1)} - \delta_v\bar{u}^{(1)}) \approx \frac{\theta_t}{\bar{u}_t^{(2)} - \bar{u}^{(2)}}(\bar{u}^{(2)}\bar{v}^{(2)} - \delta_u\bar{v}^{(2)} - \delta_v\bar{u}^{(2)}) \tag{2.104}$$

$$\bar{u}^{(2)}\bar{v}^{(2)} - K_1\bar{u}^{(1)}\bar{v}^{(1)} \approx \delta_u(\bar{v}^{(2)} - K_1\bar{v}^{(1)}) + \delta_v(\bar{u}^{(2)} - K_1\bar{u}^{(1)}) \tag{2.105}$$

Where $K_1 = \dfrac{\bar{u}_t^{(2)} - \bar{u}^{(2)}}{\bar{u}_t^{(1)} - \bar{u}^{(1)}}$

The same result can be derived for the third point:

$$\bar{u}^{(3)}\bar{v}^{(3)} - K_2\bar{u}^{(1)}\bar{v}^{(1)} \approx \delta_u(\bar{v}^{(3)} - K_2\bar{v}^{(1)}) + \delta_v(\bar{u}^{(3)} - K_2\bar{u}^{(1)}) \tag{2.106}$$

where $K_2 = \dfrac{\bar{u}_t^{(3)} - \bar{u}^{(3)}}{\bar{u}_t^{(1)} - \bar{u}^{(1)}}$

Regarding the same term for K_1 and K_2, the following equation can be derived from Eq. 2.92.

$$\bar{u}_t^{(2)}\bar{v}^{(2)} - K_1\bar{u}_t^{(1)}\bar{v}^{(1)} = \delta_u(\bar{v}^{(2)} - K_1\bar{v}^{(1)}) + \delta_v(\bar{u}_t^{(2)} - K_1\bar{u}_t^{(1)}) \tag{2.107}$$

$$\bar{u}_t^{(3)}\bar{v}^{(3)} - K_2\bar{u}_t^{(1)}\bar{v}^{(1)} = \delta_u(\bar{v}^{(3)} - K_2\bar{v}^{(1)}) + \delta_v(\bar{u}_t^{(3)} - K_2\bar{u}_t^{(1)}) \tag{2.108}$$

For pan movement, the effective term is certainly f_u. Therefore, two other formulations can be obtained from Eq. 2.101.

$$\bar{u}^{(2)}\bar{v}^{(2)} - K_3\bar{u}^{(1)}\bar{v}^{(1)} \approx \delta_u(\bar{v}^{(2)} - K_3\bar{v}^{(1)}) + \delta_v(\bar{u}^{(2)} - K_3\bar{u}^{(1)}) \tag{2.109}$$

$$\bar{u}^{(3)}\bar{v}^{(3)} - K_4\bar{u}^{(1)}\bar{v}^{(1)} \approx \delta_u(\bar{v}^{(3)} - K_4\bar{v}^{(1)}) + \delta_v(\bar{u}^{(3)} - K_4\bar{u}^{(1)}) \tag{2.110}$$

And from the accurate formulation of Eq. 2.100, the following equations are obtained.

$$\bar{u}^{(2)}\bar{v}_p^{(2)} - K_3\bar{u}^{(1)}\bar{v}_p^{(1)} = \delta_u(\bar{v}_p^{(2)} - K_3\bar{v}_p^{(1)}) + \delta_v(\bar{u}^{(2)} - K_3\bar{u}^{(1)}) \tag{2.111}$$

$$\bar{u}^{(3)}\bar{v}_p^{(3)} - K_4\bar{u}^{(1)}\bar{v}_p^{(1)} = \delta_u(\bar{v}_p^{(3)} - K_4\bar{v}_p^{(1)}) + \delta_v(\bar{u}^{(3)} - K_4\bar{u}^{(1)}) \tag{2.112}$$

where $K_3 = \dfrac{\bar{v}_p^{(2)} - \bar{v}^{(2)}}{\bar{v}_p^{(1)} - \bar{v}^{(1)}}$ and $K_4 = \dfrac{\bar{v}_p^{(3)} - \bar{v}^{(3)}}{\bar{v}_p^{(1)} - \bar{v}^{(1)}}$

2.10.5 Focal Length Estimation

Considering the situation in which the camera is placed at the origin of the world and the center of the image acquired by a slightly rotated camera, is very close to the center of the reference image, f_u and f_v can be estimated using Eq. 2.98,Eq. 2.99,Eq. 2.94, and Eq. 2.95. We have:

$$f_u = \frac{(u_p - u)}{2\theta_p} + \frac{1}{2}\sqrt{\left(\frac{u_p - u}{\theta_p}\right)^2 - 4uu_p} \tag{2.113}$$

$$f_u = \frac{(u(1+\theta_p^2) - u_p)}{-2\theta_p} + \frac{1}{2}\sqrt{\left(\frac{(u(1+\theta_p^2) - u_p)}{\theta_p}\right)^2 - 4u^2} \tag{2.114}$$

A similar equation can be obtained for f_v by considering the tilt rotation.

$$f_v = \frac{(v_t - v)}{2\theta_t} + \frac{1}{2}\sqrt{\left(\frac{v_t - v}{\theta_t}\right)^2 - 4vv_t} \tag{2.115}$$

$$f_v = \frac{(v(1+\theta_t^2) - v_t)}{-2\theta_t} + \frac{1}{2}\sqrt{\left(\frac{(v(1+\theta_t^2) - v_t)}{\theta_t}\right)^2 - 4v^2} \tag{2.116}$$

Estimation Using Roll Movement

If there is a prior knowledge about f_u and f_v, two linear equations can be obtained to calculate the values of δ_u and δ_v. By including the centers into Eq. 2.89, we have

$$\begin{aligned} u_r - \delta_u &= \cos(\theta_r)(u - \delta_u) + \sin(\theta_r)\frac{f_u}{f_v}(v - \delta_v) \\ v_r - \delta_v &= -\sin(\theta_r)\frac{f_v}{f_u}(u - \delta_u) + \cos(\theta_r)(v - \delta_v) \end{aligned} \tag{2.117}$$

$$\delta_u\big(1-\cos(\theta_r)\big)-\sin(\theta_r)\frac{f_u}{f_v}\delta_v = u_r-\cos(\theta_r)u-\sin(\theta_r)\frac{f_u}{f_v}v$$
$$\delta_u\sin(\theta_r)\frac{f_v}{f_u}+\delta_v\big(1-\cos(\theta_r)\big) = v_r-\cos(\theta_r)v+\sin(\theta_r)\frac{f_v}{f_u}u \tag{2.118}$$

For a roll angle equal to 180° the values of δ_u and δ_v can be calculated by the following equations.

$$\delta_u = \frac{u_r+u}{2}$$
$$\delta_v = \frac{v_r+v}{2} \tag{2.119}$$

2.10.6 Active Calibration Strategies

To use active calibration, four strategies have been proposed in several papers [5]–[7], [9]. The following sections explain these four strategies.

Strategy A

1. Calculate the error of the estimated lens center (δ_u, δ_v) using the three points from the image contours by Eq. 2.105 and Eq. 2.106.
2. Use the estimated results for (δ_u, δ_v) in order to calculate the (f_u, f_v) by Eq. 2.101 and Eq. 2.93.
3. Repeat items one and two until the desired accuracy is obtained.

Strategy B

Since for each rotation, the terms $\bar{u}_t - \bar{u}$ and $\bar{v}_p - \bar{v}$ are very small, they are prone to being unstable in the presence of noise. To overcome this problem, strategy B was proposed.

1. Estimate f_u and f_v from Eq. 2.114 and Eq. 2.116, respectively, using a single contour.
2. Use the estimation from step one in Eq. 2.93 and compute δ_u and δ_v using another independent contour.
3. Repeat items one and two until the desired accuracy is obtained.

Strategy C

1. Use the procedure from strategy B to calculate the values of f_u and f_v.
2. Calculate the values of δ_u and δ_v by Eq. 2.118 regarding the roll movement information.

Strategy D

1. Having considered the roll movement of 180° and a single image contour, calculate the estimate for δ_u and δ_v using the Eq. 2.119.

2. Replace these values into Eq. 2.95 to solve for f_u and f_v, which includes the ignored items.

2.11 Lens Distortion Models

Estimating the distortion parameters of lenses, is one of the main barriers of camera calibration methods [101]. One of the first efforts to correct the lens distortion was performed by Brown in 1966 [14]. He proposed the Brown-Conrady method, which was built based on the old decentering distortion of A. Conrady in 1919.

To have a camera model capable of dealing with various types of distortions, Faig's accurate model [25] has taken four different distortions into account, namely radial symmetric distortion, decentering distortion, affinity distortion and the distortion caused by non-perpendicularity of axes. Others considered either using different parameters [96] (only radial distortion) or no parameters for lens distortion [33].

In order to geometrically measure objects in the image, a distorted camera is the main obstacle rather than the image quality. For instance, when the image is blurred, the position of a point can still be found by locating the center of the point [106]. However, the inaccurate position of a point will cause wrong results for any kind of computations that depends on image coordinates.

Fisheye lenses are able to present real-world visual information in various resolutions at the same time. This ability resembles the human visual system (HVS) which enjoys the foveal visual signals at very high spatial resolution and peripheral information at low spatial resolution [8]. The fisheye lenses form strongly distorted images due to the inherent characteristics of their structure. To deal with this distortion, Basu proposed a polynomial fitting model for the fisheye transform (PFET) [8]. Mathematically speaking, every image can be transformed into its fisheye version or cortical polar (FET) coordinates (ρ, θ^*) by using the following equations:

$$r = \sqrt{x^2 + y^2} \tag{2.120}$$

$$\theta = \tan^{-1}(\frac{y}{x}) \tag{2.121}$$

$$\rho = s \log(1 + \lambda r) \tag{2.122}$$

$$\theta^* = \theta \tag{2.123}$$

where s represents the scale factor of the transformation and λ controls the severity of the

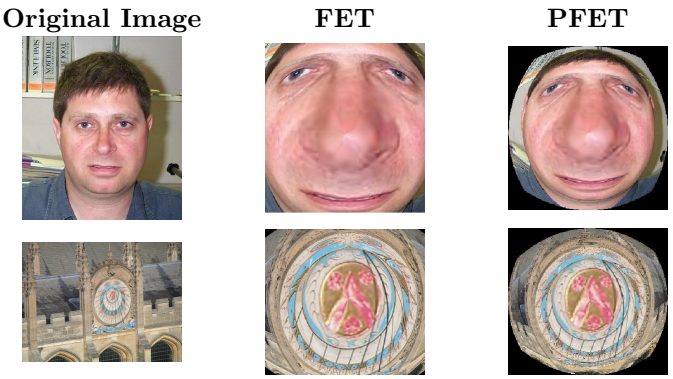


Figure 2.2: Examples of FET and PFET transformations on three image of Caltech101 [34].

distortion. As a result, the new coordinates (x^*, y^*) are obtained by the following equations.

$$x^* = \rho \cos(\theta^*) \tag{2.124}$$

$$y^* = \rho \sin(\theta^*) \tag{2.125}$$

Also, another formulation of the distortion can be achieved by replacing ρ with a polynomial function.

$$\rho = G(r) = \kappa_0 + \kappa_1 r + \kappa_2 r^2 + \kappa_3 r^3 + ... = \sum_{i=0}^{n} \kappa_i r^i \tag{2.126}$$

Fig.2.11 illustrates the difference between the Fisheye Transformation (FET) and Polynomial Fisheye Transformation (PFET) on three sample images from [34]. Considering that PFET as a common model of fisheye lenses, the Basu method [8] calibrates the distortion caused by fisheye lenses. To achieve this, a polynomial model should be fitted on the image. Initially, the optical center of distortion or focus of distortion (FD) is found by taking several pictures with a fisheye lens from a grid, and then lines having the minimum curvature are selected from the acquired images. The intersection between extracted orthogonal lines indicates the focus of distortion. To obtain a better estimate, an average of FDs of all images is considered.

Having the focus of distortion facilitates the calculation of distortion in every direction. Therefore, in the next step, an average of the distortion between every direction (up, down, left, right) is taken. So, based on these observations from the images, the least square method is employed to fit a polynomial Eq. 2.127 or logarithmic Eq. 2.128 model to the average of distortion of n data pairs by minimizing the following summations.

$$\chi = \sum_{i=0}^{n} (\rho_i - G(r_i))^2 \tag{2.127}$$

$$\chi = \sum_{i=0}^{n} (\rho_i - s \log(1 + \lambda r_i))^2 \tag{2.128}$$

As a result, they approximate every necessary coefficients for modeling the logarithmic or polynomial distortion of the lens.

Usually in camera calibration literature, lens distortion can be explained mathematically based on either a camera center view or the view from an image coordinate system in non-metric calibration. The former can be defined as follows

$$x_d = x + \delta_x \tag{2.129}$$

$$y_d = y + \delta_y \tag{2.130}$$

In the non-metric lens modeling, the general equation denoted as

$$u = u_d + \delta_u \tag{2.131}$$

$$v = v_d + \delta_v \tag{2.132}$$

Specifically, lens distortions can be classified into three different types

1. ***Radial distortion***

   The flawed radial curvature of a lens can create radial distortions. Radial distortion in the literature is calculated by the following equation.

   $$\delta_{xr} = x(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + ...) \tag{2.133}$$

   $$\delta_{yr} = y(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + ...) \tag{2.134}$$

   where r is a distance from point (x_d, y_d) to the center of the radial distortion and equals the value of $\sqrt{x_d^2 + y_d^2}$ and κ_1,κ_2, and κ_3 are the radial distortion coefficients. In addition, in many methods, the high-order parameters higher than r^4 have been neglected.

2. ***Decentering distortion***

   If the optical center of the lens has not been correctly aligned with the center of the camera, decentering distortion happens [106]. It is mathematically denoted as follows

   $$\delta_{xd} = p_1(3\ {}^C X_u^2 + {}^C Y_u^2) + 2p_2\ u\ v \tag{2.135}$$

   $$\delta_{yd} = p_2({}^C X_u^2 + 3\ {}^C Y_u^2) + 2p_1\ u\ v \tag{2.136}$$

3. ***Thin prism distortion***

   This type of distortion is due to poor lens design and manufacturing, and inaccurate camera assembly [82]. Particularly, it arises when there is a tilt angle between the lens and the image sensor array [101].

   $$\delta_{xp} = s_1(x^2 + y^2) \tag{2.137}$$

   $$\delta_{yp} = s_2(x^2 + y^2) \tag{2.138}$$

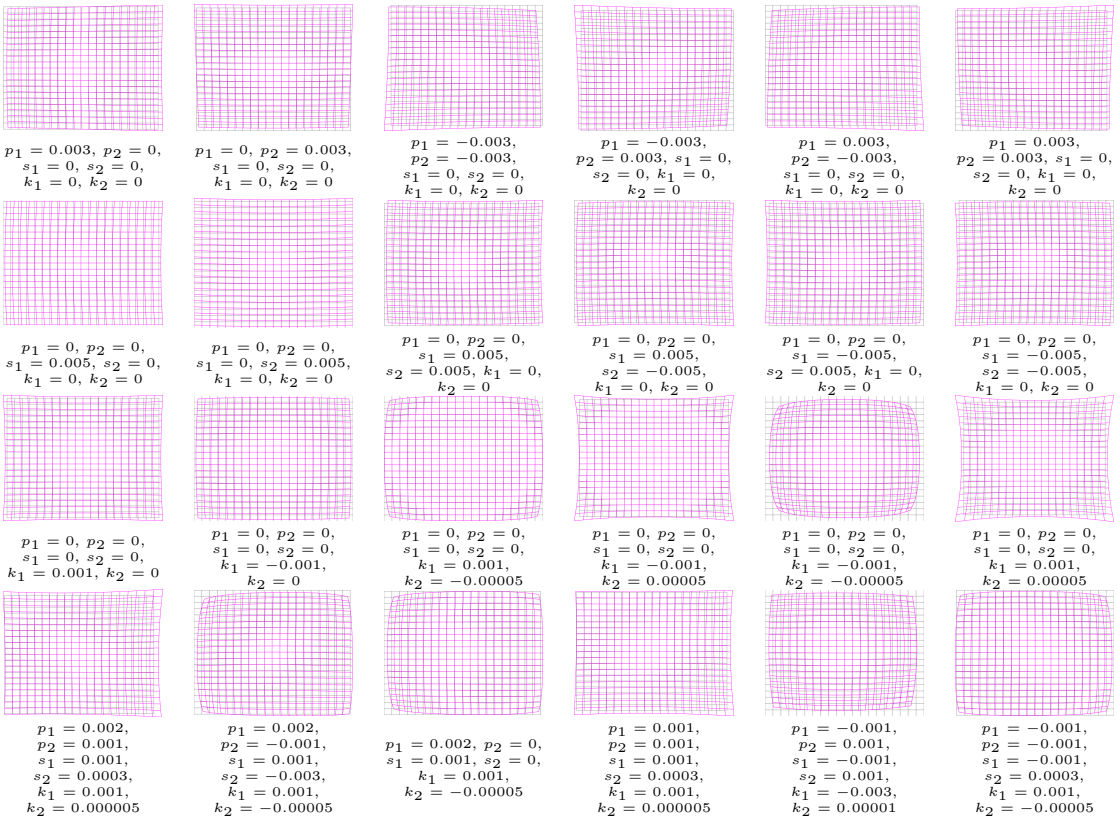


Figure 2.3: Different types of distortions derived by the combination of three distortion models (Eq. 2.141 and Eq. 2.142).

In many models, such as that of Weng [106], the total lens distortion is computed by simply adding all three components of lens distortion.

$$\delta_x = \delta_{xr} + \delta_{xd} + \delta_{xp} \tag{2.139}$$

$$\delta_y = \delta_{yr} + \delta_{yd} + \delta_{yp} \tag{2.140}$$

By substituting Eq. 2.133, Eq. 2.135, and Eq. 2.137 in Eq. 2.139, we have:

$$\begin{aligned} \delta_x &= x(\kappa_1 r^2 + \kappa_2 r^4) + p_1(3\ x^2 + y^2) + 2p_2\ u\ v + s_1(x^2 + y^2) \\ \delta_x &= (x^2 + y^2)(u(\kappa_1 + \kappa_2(x^2 + y^2)) + p_1 + s_1) + 2p_1\ u + 2p_2\ v \end{aligned} \tag{2.141}$$

By substituting the Eq. 2.134, Eq. 2.136, and Eq. 2.138 in the Eq. 2.140, we have:

$$\begin{aligned} \delta_y &= y(\kappa_1 r^2 + \kappa_2 r^4) + p_2(x^2 + 3\ y^2) + 2p_1\ u\ v + s_2(x^2 + y^2) \\ \delta_y &= (x^2 + y^2)(v(\kappa_1 + \kappa_2(x^2 + y^2)) + p_2 + s_2) + 2p_1\ u + 2p_2\ v \end{aligned} \tag{2.142}$$

The abovementioned coefficients of the lens distortion model, for both x and y directions (κ_1, κ_2, s_1, s_2, p_1, p_2) create various structures of distortions, such as barrel and pincushion distortions. Mathematically speaking, these coefficients are independent from each other.

However, this raises the question of whether these coefficients are physically independent as well. The answer to this question is the main topic of Wang's paper [101] in which he argues that, in reality, all of the lens coefficients are dependent. By considering the relation among the radial, decentering, and thin prism distortions, he has cited several manufacturing situations that might contribute to this phenomenon. For instance, while manufacturing radial symmetry of lenses can be insured due to the high-tech equipment used in the process, the eccentricity and tilt of the optical components with respect to each other during assembly is difficult to achieve [101]. Under these circumstances, the radial distortion will be affected and the radial symmetry will be broken. As a result, a part of the radial distortion will manifest itself in the form of decentering distortion or prism distortion [101]. So, by this analysis, he reported an improvement in the total lens distortion model by expressing a tangential distortion using a rotation matrix and a translation vector.

Chapter 3

Related Work: Intravascular Ultrasound Imaging

Intravascular Ultrasound (IVUS) is the preeminent choice for studying the morphology and biomechanical properties of vessels [53]. IVUS visualizes the inner parts of the coronary arteries using a catheter-based diagnostic method. Generally, a sequence of cross-sectional real-time views of the arteries is generated by IVUS that reveals the *lumen* and the *atheroma* that are concealed within the artery wall. More specifically, inside the IVUS we can observe 1) the adventitia, which is the outer covering of the artery, 2) the media, which is the actual wall of the artery, 3) the intima, which is a layer of endothelial and other cells that make direct contact with blood inside the artery, and 4) the lumen, which is the actual open channel of the artery through which the blood flows.[1]

The procedure of acquiring the IVUS sequence can take more than an hour. It should be noted that it is an invasive procedure and has its own risks. The entire procedure consists of seven steps: 1) To insert the catheter, the surrounding area of the groin should be shaved and cleaned. 2) To help the patient to be relaxed, a mild sedative should be injected. 3) A local anesthetic should also be injected into the catheter. 4) The imaging physician inserts the catheter through the arteries until it reaches the area to be studied. 5) A guide wire with an ultrasound probe on its tip is inserted into the catheter and guided to the furthest position to be imaged. 6) Sound waves are emitted from the probe. The probe also receives and returns the echo information, sending images to a computer. 7) The guide wire is held in place and the probe is slid backwards, usually under steady, smooth, and motorized control while sending and receiving ultrasound images along the way.[2]

Fig. 3.1 shows a picture of the IVUS device used to acquire our data. According to all indications from the available research, the device's pullback is manual. If the pullback were to be automatic, we would need to have a number of pullback steps, which would determine

[1] http://www.ptca.org/ivus/ivus.html

[2] https://www.cedars-sinai.edu/Patients/Programs-and-Services/Womens-Heart-Center/Services/Intravascular-Ultrasound-IVUS.aspx

Figure 3.1: Picture of the IVUS device used to acquire our data.

the pullback length. This is very important for tracking and 3D reconstruction because there is no supporting spatial data to help us with registering and reconstructing the 3D coordinates.

3.1 3D Reconstruction Using ANGUS

In [87], 3D reconstructions of coronary arteries are obtained by fusing IVUS and angiography information. At the first step, IVUS volumes are acquired during an R-wave-triggered, motorized, stepped pullback. Then the lumen contour is extracted from IVUS images and fused with angiogram data. Finally, the obtained reconstruction is optimized by quantitative matching of the silhouettes of the 3D reconstruction with actual biplane images.

3.2 Rigid In-Plane Motion Estimation

An integrative framework for exploring vessel dynamics and structures has been proposed in [53]. This study aimed to build stable models of arteries using IVUS in order to show the effectiveness of using the in-plane rigid dynamic to analyze and correct itself and also to retrieve the cardiac phase and segment the adventitia layer. Using the rigid in-plane dynamic, the authors were able to correct the IVUS image misalignment. It has also been shown that the estimation can be used to retrieve the cardiac phase. In addition, they tried to segment the adventitia using the rigid in-plane dynamics. The obtained results showed that they could decrease the computational time; however, the accuracy error was still kept within the range of inter-observer variability [53]. They concluded that their method had a high potential to be used in clinical practice and would reduce the computation time since it could be parallelized.

3.3 3D Intravascular Visualization Using Shape-based Interpolation

A novel volumetric 3D IVUS reconstruction algorithm using a nonlinear shape-based interpolation is proposed in [79]. Because 3D IVUS is able to mitigate the limitations of 2D IVUS in terms of the complex spatial distribution of arterial morphology and acoustic backscatter information, the authors proposed a volumetric 3D IVUS reconstruction algorithm by utilizing both IVUS signal data and a shape-based non-linear interpolation. The reconstruction algorithm created intermediary slices between original 2D IVUS slices. The newly added slices were obtained using natural cubic spline interpolation because of the inherent nonlinear vascular structure geometry and acoustic backscatter in the arterial wall. The results of the reconstruction were validated by in vitro and in vivo studies that demonstrated the robustness of the reconstructed model.

3.4 Image-Based Cardiac Gating for 3D IVUS

Because the artifacts caused by cardiac motion and vessel wall pulsation restrict accurately visualizing coronary lumen and plaque volume , an image-based gating method has been preposed in [76] in order to reconstruct a 3D IVUS image with negligible cardiac motion and vessel pulsation artifacts. Although using an ECG-gated image acquisition helps with decreasing the effects of the artifacts, it not only increases the image acquisition time, but also affects the intensity of the recorded signals. The method first tracks the lumen contour variation over the cardiac cycle and then extracts IVUS images belonging to the same cardiac phase from an asynchronously-acquired series. The method achieved a reduction in the artifacts by 86% and 80% for the in vitro and in vivo studies respectively.

3.5 Real-Time Gating Based on Motion Blur Analysis

The general problem that IVUS pullback is confronted with is the swinging fluctuation of the probe position along the vessel axis. In [42], a gating algorithm based on the analysis of motion blur variations during the IVUS sequence has been proposed. The accuracy of the algorithm has been evaluated on an in vitro ground truth database and superior results were reported by the author of that study.

3.6 Image-Based Device Tracking for the Co-Registration of Angiography and Intravascular Ultrasound Images

In [103], a method is proposed to accurately track the catheters and transducers during a coronary intervention using biplane angiograms. The authors of that study proposed using an online tracking system. The system receives a selection of angiographic coronary branches as an input and tracks the location of the medical devices (IVUS transducers and catheter tips) during the subsequent IVUS pullback. A combination of learning-based detection methods with model-based tracking and geodesic registration is performed to achieve the desired tracking. They claimed that their method was the first reported system able to automatically establish a robust correspondence between the angiography and IVUS images, thus providing clinicians with a comprehensive view of the coronaries [103].

3.7 3D Fusion of IVUS and Coronary CT

This study [43] presented a framework to obtain the 3D reconstruction of human coronary arteries by the fusion of Intravascular Ultrasound (IVUS) and coronary computed tomography angiography (CT). The authors of that study, acquired IVUS and CT from 23 patients. Then the obtained lumen and wall from IVUS were positioned on the 3D centerline that had been extracted from CT. After transformation to a surface template, they calculated shear stress and plaque thickness. The authors claimed that this new framework could therefore successfully be applied to shear stress analysis in human coronary arteries [43].

3.8 3D Reconstruction of Coronary Artery to Assess ESS

In order to assess endothelial shear stress (ESS), Bourantas et al. [12] developed a methodology to reconstruct the coronary artery using IVUS and angiographic data during routine catheterization. They combined the conventional method and a new approach to estimate the orientation of IVUS planes. To achieve this, they utilized the luminal 3D center lines and

several anatomical landmarks. They claimed that their method provided a geometrically correct model and permitted reliable ESS computation.

3.9 A Review of Intravascular Ultrasound-Based Multimodal Intravascular Imaging

A recently published review paper [71] for studying the various catheter-based intravascular imaging modalities appears as though it may be useful for further investigation. The author states the importance of IVUS in enhancing the characterization of vulnerable plaques. He also discusses current scientific innovations, technical challenges, and prospective strategies in the development of IVUS-based multi-modality intravascular imaging systems aimed at assessing atherosclerotic plaque vulnerability.

3.10 Real Time Co-Registration of IVUS and Coronary Angiography

A method for real-time co-registration of intravascular ultrasound (IVUS) and coronary angiography has been proposed in [38]. The study tried to mitigate the inaccuracies involving the uncorrelated location of IVUS and angiogram. They used phantoms simulating the coronary tree to test the accuracy and potential of co-registration. As a result, they found that the novel IVUS and coronary angiography co-registration method was accurate, easy to use, fast and user-friendly.

3.11 IVUS Angio Tool

A fast and accurate 3D reconstruction of coronary arteries using IVUS and biplane angiography data is reported in [20]. The author published free software, containing a graphical user interface that is capable of doing the reconstruction. A similar work to [87] has been performed here in terms of fusing the detected arterial boundaries in IVUS images and the catheter path derived from biplane angiography. The software is able to perform several other tasks including the automatic selection of the end-diastolic IVUS images, semi-automatic and automatic IVUS segmentation, vascular morphometric measurements, graphical visualization of the 3D model and can export the model in a format compatible with other computer-aided design applications. The author claims that the software can significantly reduce the total processing time for 3D coronary reconstruction.

Chapter 4

A Simplified Active Calibration Algorithm for Focal Length Estimation

Abstract

We introduce new linear mathematical formulations to calculate the focal length of a camera in an active platform. Through mathematical derivations, we show that the focal lengths in each direction can be estimated using only one point correspondence that relates images taken before and after a degenerate rotation of the camera. The new formulations will be beneficial in robotic and dynamic surveillance environments when the camera needs to be calibrated while it freely moves and zooms. By establishing a correspondence between only two images taken after slightly panning and tilting the camera and a reference image, our proposed Simplified Calibration Method is able to calculate the focal length of the camera. We extensively evaluate the derived formulations on a simulated camera, 3D scenes and real-world images. Our error analysis over simulated and real images indicates that the proposed Simplified Active Calibration formulation estimates the parameters of a camera with low error rates.

4.1 Introduction

Many 3D computer vision applications require knowledge of the camera parameters to relate the 3D world to the acquired 2D image(s). The process of estimating the camera parameters is called *camera calibration* in which two groups of parameters (intrinsic and extrinsic) are estimated.

In order to calibrate a camera, conventional calibration methods need to acquire some information from the real 3D world using calibration objects such as grids, wands, or LEDs. This imposes a major limitation on the calibration task since the camera can be calibrated only in off-line and controlled environments. To address this issue, Maybank and Faugeras [32], [72] proposed the so-called *self-calibration* approach in which they used the information of matched points in several images taken by the same camera from different views instead of using known 3D points (calibration objects). In their two-step method, they first estimated the epipolar transformation from three pairs of views, and then linked it to the image of an absolute conic using the Kruppa equations [72]. Not long after the seminal work of Maybank and Faugeras, Basu proposed the idea of Active Calibration [5], [6] in which he included the concept of active camera motions and eliminated point-to-point correspondences.

The main downside of the Active Calibration strategies (A and B) in [5]–[7] is that it calculates the camera intrinsics using a component of the projection equation in which a constraint is imposed by the degenerate rotations. For example, after panning the camera, the equation derived from vertical variations observed in the new image plane is unstable. Furthermore, the small angle approximation using $\sin(\theta) = \theta$ and $\cos(\theta) = 1$ decreases the accuracy of strategies when the angle of rotation is not very small. Also, rolling the camera [9] is impractical (without having a precise mechanical device) because it creates translational offsets in the camera center. In this chapter, we propose a Simplified Active Calibration (SAC) formulation in which the equations are closed-form and linear. To overcome the instability caused by using degenerate rotations in Active Calibration, we calculate focal length in each direction separately. In addition, we do not use small angle approximation by replacing $\sin(\theta) = \theta$ and $\cos(\theta) = 1$. Hence, in our formulation we only refer to the elements of the rotation matrix. Moreover, the proposed method is more practical because it does not require a roll rotation of the camera; only pan and tilt rotations, which can be easily acquired using PTZ cameras, are sufficient.

The rest of the chapter is organized as follows. In Section 4.2 we present our proposed Simplified Active Calibration formulation. Section 4.3 reports and analyzes the results of the proposed method on simulated and real scenes. Finally, conclusions are drawn in Section 4.4.

4.2 Simplified Active Calibration

Simplified Active Calibration (SAC) was inspired by the novel idea of approximating the camera intrinsics using small angle rotations of the camera, which was initially proposed in [5], [6] and extended in [7], [9]. Imposing three constraints on the translation of the camera generates a pure rotation motion. In addition, using small angle rotations allows us to ignore some non-linear terms in order to estimate the remaining linear parameters. The estimated intrinsics can then be used as an initial guess in the non-linear refinement processes.

In general, SAC can be used in any platform in which information about the camera motion is provided by the hardware, such as in robotic applications where the rotation of the camera can be extracted from the inertial sensors or in the surveillance control softwares that are able to rotate the PTZ cameras by specific angles. Having access to the rotation of the camera, we propose a 2-step process to estimate the focal length of the camera. In the first step, we present a closed-form solution to calculate an approximation of the focal length in the v direction (f_v) using an image taken after a pan rotation of the camera, assuming that v and u represent the two major axes of the image plane. In the second step, we estimate the focal length of the camera in the u direction (f_u) using an image taken after tilt rotation of the camera. Therefore, to estimate the two main components (focal length) of the intrinsic matrix, namely f_v, f_u, two pairs of images are required, one taken before and after a small pan rotation, and another taken before and after a small tilt rotation.

4.2.1 Focal Length in the v Direction

We assume that the camera is located at the origin of the Cartesian coordinate system and is looking at distance $z = f$ where the principal point is specified. Every 3D point $\mathbf{X} = [X\ Y\ Z]^T$ in the world that is visible to the camera can be projected onto a specific point $\mathbf{u} = [v\ u\ 1]^T$ of the image plane where the coordinates of the principal points are denoted by $[v_0\ u_0]^T$. With modern cameras it is reasonable to assume that image pixels are square so that the value of the camera skew is zero.

Every point $\mathbf{u} = [v\ u]^T$ in an image seen by a stationary camera (that freely rotates but stays in a fixed location) is transformed to a point $\mathbf{u'} = [v'\ u']^T$ in another image taken after camera rotation. The mathematical relationship between $\mathbf{u}$ and $\mathbf{u'}$ when the camera is panned is denoted by $w\mathbf{u'} = \mathbf{K}\mathbf{R}_y^T\mathbf{K}^{-1}\mathbf{u}$ and after expanding the equation, the relationship is thus represented by:

$$v' = \frac{r_{11}(v - v_0) + r_{31}f_v}{r_{13}\dfrac{v - v_0}{f_v} + r_{33}} + v_0 \tag{4.1}$$

$$u' = u_0 - \frac{u_0 - u}{r_{13}\dfrac{v - v_0}{f_v} + r_{33}} \tag{4.2}$$

where r_{ij} is an element of the rotation matrix around Y-axis at row i and column j. After simplification of Eq. 4.2:

$$\frac{v - v_0}{f_v} = \frac{\frac{u_0 - u}{u_0 - u'} - r_{33}}{r_{13}} \tag{4.3}$$

Note that after a pure pan rotation, the u coordinates of the new image will not be affected by the transformation. (The reader is referred to [62] for a detailed explanation and analysis of this fact.) In other words, image pixels only move horizontally. Thus, the rate of change in the u direction before and after the pan rotation is close to one, viz:

$$\frac{u_0 - u}{u_0 - u'} \approx 1 \tag{4.4}$$

Substituting Eq. 4.4 into Eq. 4.3 and then replacing the equation obtained for the term $\frac{v - v_0}{f_v}$ in the Eq. 4.1, we have:

$$v' \approx \frac{r_{11}(v - v_0) + r_{31} f_v}{r_{13}\frac{1 - r_{33}}{r_{13}} + r_{33}} + v_0 \tag{4.5}$$

The above substitution changes the value of the denominator to 1 and hence simplifies the whole projection equation.

$$v' - r_{11} v \approx r_{31} f_v + (1 - r_{11}) v_0 \tag{4.6}$$

Knowing that the principal point is close to the center of the image ($c_u = h/2, c_v = w/2$), where h and w represent the image height and width respectively, we replace v_0 with c_v in Eq.4.6. Thus, we can derive a suitable linear equation to estimate the focal length in the x direction from an image taken after a pan rotation.

$$f_v \approx \frac{v' - r_{11} v - (1 - r_{11}) c_v}{r_{31}} \tag{4.7}$$

Eq. 4.7 needs only one point v in the reference image that corresponds to v' in the transformed image. If there are more point correspondences, we can easily use the average of these points to obtain more robust results.

4.2.2 Focal Length in the u Direction

So far, we could estimate f_v by the information provided from an image taken after a pan rotation. We repeat the same procedure to approximate f_u. This time, we need an image taken after a pure tilt rotation of the camera. Thus, the projection equation is characterized by replacing $\mathbf{R}$ with the proper rotation matrix that describes rotation of the camera around X-axis. Following the same reasoning as in Section 4.2.1, a closed-form solution to estimate the focal length of the camera in the u direction is obtained by:

$$f_u \approx \frac{r_{22} u - u' + (1 - r_{22}) c_u}{r_{32}} \tag{4.8}$$

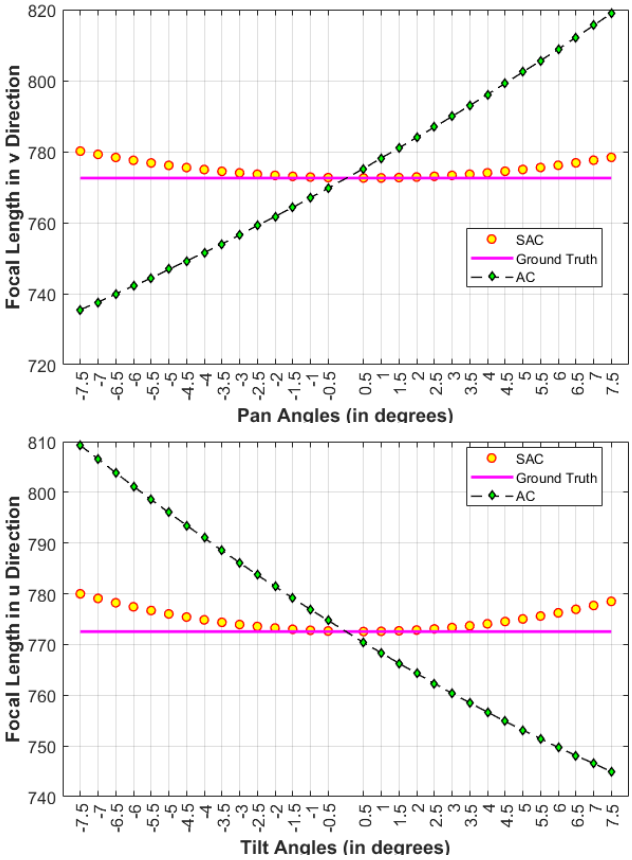


Figure 4.1: Focal lengths calculated in the v and u directions using Active Calibration Strategy B (AC)[9] versus SAC for various angles of rotations. In SAC we only use one point correspondence.

4.3 Results and Analysis

Based on our proposed method, the focal length in the v and u directions can be estimated using Eq. 4.7 and Eq. 4.8, respectively. Only one point correspondence is required to calculate the focal length. Fig. 4.1 shows the estimated focal lengths using various pan and tilt angles on a 3D synthetic scene of a teapot taken by a simulated camera. It can be seen that when the pan and tilt angles are small, the estimated focal lengths are very close to the ground truth.

In another experiment, we calculate the proposed simplified active calibration formulation on 1000 different runs of 500 randomly generated 3D points for small pan and tilt angles. The mean and standard deviation of the results obtained are shown in Table 4.1. As we can see, our proposed active calibration formulation attains results very close to the ground truth. Specifically, the error in focal length estimates is less than 1 pixels.

Table 4.1: Results of the proposed simplified active calibration on 1000 separate 3D random points for various small pan and tilt angles. In the table, GT denotes the Ground Truth, SD represents the Standard Deviation. The error values are in pixels.

Pan	Tilt		f_v	f_u
		GT	772.55	772.55
		Mean	772.61	772.76
$1°$	$-1°$	SD	0.02	0.09
		Error	0.06	0.21
		Mean	773.02	772.73
$-1.5°$	$1.5°$	SD	0.13	0.07
		Error	0.47	0.19

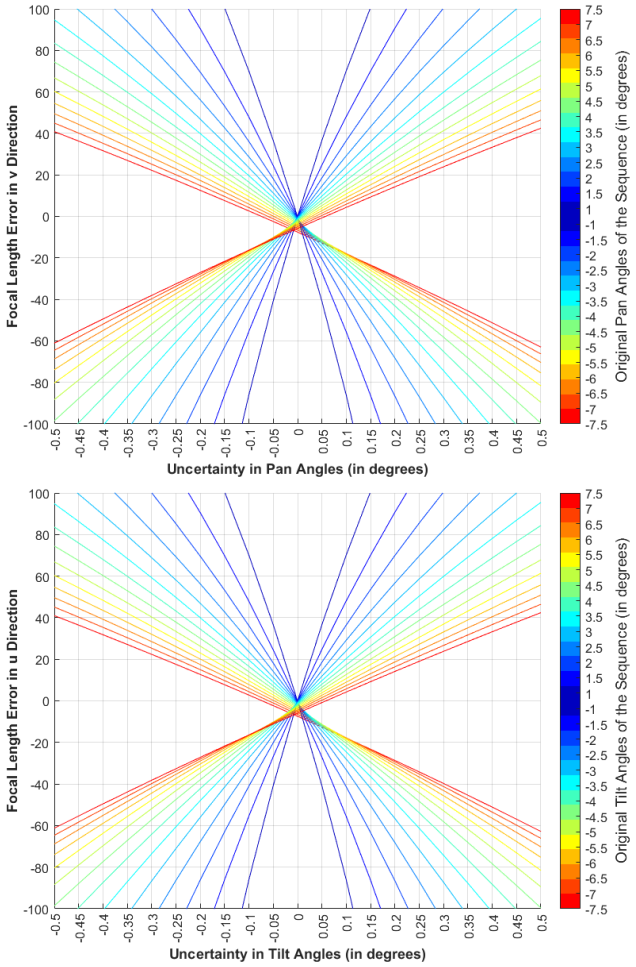


Figure 4.2: The error caused by uncertainty in determining the angle of the camera. Top: The effects of the uncertainty of the camera pan rotation on calculating the focal length in the v direction by SAC. Bottom: The effects of the uncertainty of the camera tilt rotation on calculating the focal length in the u direction by SAC.

4.3.1 Angular Uncertainty

Acquiring the rotation angles requires either specific devices, such as gyroscopes, or a specially designed camera called a PTZ camera. Even using these devices does not guarantee

that the extracted rotation angles are noise-free. To simulate the noisy conditions of a real-world application, we contaminated the angles of the above-mentioned teapot sequences with increasing angular errors.

While the point correspondences are kept fixed for all of the pan and tilt rotations, we calculate the focal length (Eq. 4.7 and Eq. 4.8) using contaminated pan and tilt angles. The results are shown in Fig. 4.2. Specifically, Fig. 4.2(a) and Fig. 4.2(b) show the error of our proposed formula for estimating the focal length when the pan and tilt angles are not accurate. Every sequence has been colored based on its rotation angle, ranging from blue indicating smaller angles to red for larger angles. For focal length estimation, Fig. 4.2(a) and Fig. 4.2(b) illustrate that the sequences taken with smaller angles have steeper slopes than the sequences acquired with larger rotation angles. This shows that focal lengths are more sensitive to angular noise when the camera is rotated by smaller angles rather than larger angles.

Overall, when the camera is rotated by small angles, the influence of the angular noise on SAC equations is higher. On the other hand, SAC tends to use the benefit of rotating the camera by small angles. Therefore, to avoid magnifying the effect of noise it is important not to rotate the camera by very small angles.

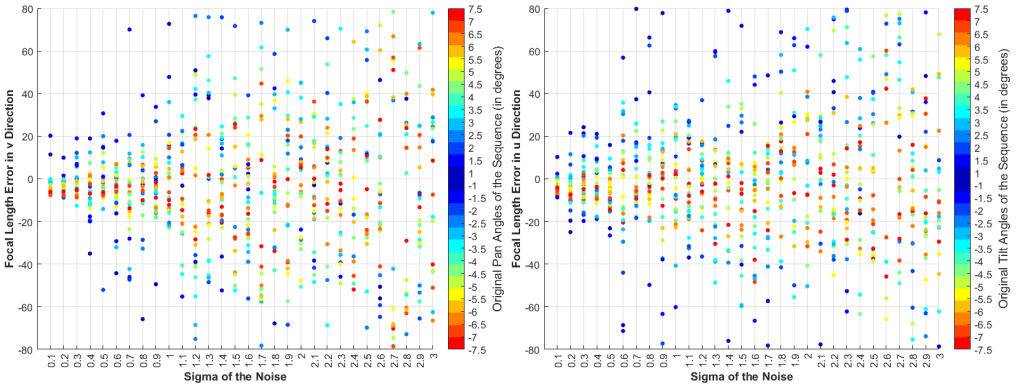
4.3.2 Point Correspondence Noise

Another type of noise that affects the SAC equations is the noise in the location of features used for matching. To simulate such conditions, we assume that the location of every teapot point, is disturbed by a Gaussian noise with zero mean and variance σ_{pixel}. Then, we calibrate the camera using SAC for all σ_{pixel} in the range of 0 to 3. The intrinsic parameters obtained are illustrated in Fig. 4.3.

Fig. 4.3(a) and Fig. 4.3(b) illustrate the influence of pixel noise on the estimation of focal length (Eq. 4.7 and Eq. 4.8). Colors are distributed based on the rotation angles of the camera and, hence, the distribution of the colors reveals how noise affects the SAC equations. In fact, the high concentration of red, yellow, and orange points around the zero error line in Fig. 4.3(a) to (b) reveals that when the angle of the camera rotation is not very small, SAC achieves low-error estimates for focal lengths. This corroborates the claim that very small camera rotations can cause results from the SAC formulations to have high error.

4.3.3 Real Images

We studied the proposed SAC formulations on real images as well. We used the Canon VC-C50i PTZ camera that is able to freely rotate around Y-axis (pan) and X-axis (tilt). The camera can be controlled by a host computer using a standard RS-232 serial communication.



(a) **(b)**

Figure 4.3: The error caused by uncertainty in location of points. **a)** Error of the estimated focal length in the v direction using SAC when the location of the teapot points are disturbed by different values of σ_{pixel}. **b)** Error of the estimated focal length in the u direction using SAC under the same conditions as in (a).

(a) **(b)**

Figure 4.4: A sequence of real images taken for SAC. **a)** Image taken after panning the camera by $0.5625°$. **b)** Image taken after tilting the camera by $-0.675°$.

Therefore, the required pan and tilt rotation angles can be set in a specific packet and then be written into the camera serial buffer to rotate the camera based on the assigned rotation angles.

Using the above-mentioned procedure, we captured four sequences of images for evaluating the proposed SAC formulations. Fig. 4.4 shows a sequence of our bookshelf scene. All sequences were taken using a fixed zoom. While keeping the zoom of the camera unchanged, another 30 images were acquired from various viewpoints of a checkerboard pattern. The ground truth of intrinsic parameters were found by applying the method of Zhang [114] on the checkerboard images.

The performance of the SAC formulations on the four sequences of real images is reported in Table 4.2. For every sequence, we only used the images in the sequence. For example, to calculate the focal length in the v direction of Sequence 1, we found the point correspondence between a reference image and the image taken after the pan rotation of

the camera (Fig. 4.4(a)). Then, we used only one of the matched points that is closer to the center of the image. Although we did not include the lens distortion parameter into the SAC formulation (because it creates non-linear equations), we decrease the inaccuracy of the focal length estimates by using a matched point that is closer to the center of the image and, thus, is less affected by the lens distortion. A similar procedure was adopted with the image taken after a tilt rotation of the camera (Fig. 4.4(b)) for calculating the focal length in the u direction of Sequence 1.

The errors reported by applying SAC on four different sequences of real images in Table 4.2 show that despite the presence of various types of noise, such as angular uncertainties, point correspondence noise and lens distortion, focal lengths estimated by SAC are close to the results of the method of Zhang [114], except when the angles of rotations are very small ($< 1^\circ$).

4.4 Conclusion

Inspired by the idea of calibrating a camera through active movements of the camera, in this chapter we presented a Simplified Active Calibration formulation. Our study provides closed-form and linear equations to estimate the parameters of the camera using two image pairs taken before and after panning and tilting the camera.

A basic assumption about the rotation of a fixed camera was made; i.e., to solve the proposed equations, knowing the rotation angles of the camera is necessary. The proposed formulation can be used in practical applications such as surveillance because in PTZ and mobile phone cameras, accessing the camera motion information is straightforward.

The proposed closed-form formulations for estimating the focal lengths can be solved with only one point correspondence. Finding the correspondence point is straightforward. Due to the recent developments in feature extractors, one may use [29], [30] to extract repeatable regions from a pair of images. This is especially useful for applications that prefer no point correspondence; where instead of the reference and transferred points in Eq. 4.8 and Eq. 4.7, the average of the edge points or the centroid of the regions can be used.

The results of solving our proposed formulations on randomly simulated 3D scenes indicated a very low error rate in estimating the focal lengths. We evaluated our proposed SAC formulation for two different noise conditions, namely angular and pixel noise. The simulated results showed that if the angle of rotation is not very small, the SAC formulation can robustly estimate the focal lengths. This conclusion was later verified in our experiment with real images. Our future work will focus on deriving linear equations for calculating the location of the principal point and also including lens distortion parameters into the Simplified Active Calibration equations.

Table 4.2: Results of the proposed simplified active calibration on four sequences of real images. All angles are in degrees. δ_{f_v},δ_{f_u} are the percentage errors from the corresponding ground truth acquired by the method of Zhang [114].

#	Pan	Tilt	f_v	f_u	δ_{f_v}	δ_{f_u}
1	0.5625°	-0.675°	880.42	-999.07	15.3	5.33
2	-1.8°	2.025°	1052.1	-966.35	1.18	1.88
3	-4.6125°	-4.1625°	1067.9	-970	2.70	2.26
4	-7.9875°	-7.425°	1069.6	-986.58	2.87	4.01

Chapter 5

Simplified Active Calibration

Abstract

We present a new mathematical formulation to estimate the intrinsic parameters of a camera in active or robotic platforms. We show that the focal lengths can be estimated using only one point correspondence that relates images taken before and after a degenerate rotation of the camera. The estimated focal lengths are then treated as known parameters to obtain a linear set of equations to calculate the principal point. Assuming that the principal point is close to the image center, the accuracy of the linear equations is increased by integrating the image center into the formulation. We extensively evaluate the formulations on a simulated camera, 3D scenes and real-world images. Our error analysis over simulated and real images indicates that the proposed Simplified Active Calibration method estimates the parameters of a camera with low error rates that can be used as an initial guess for further non-linear refinement procedures. Simplified Active Calibration can be employed in real-time environments for automatic calibrations given the proposed closed-form solutions.

5.1 Introduction

Intrinsic camera calibration is an essential step in many 3D computer vision applications where we need to calculate how the 3D world is projected onto a 2D image. In general, intrinsic camera calibration aims not only to estimate the focal lengths of the camera in each direction, but also the center of projection, pixel skew and aspect ratio.

In order to calibrate a camera, conventional calibration methods need to acquire some information from the real 3D world using calibration objects such as grids, wands, LEDs, or even by adding augmented reality markers to a camera [116]. This imposes a major limitation on the calibration task since the camera can be calibrated only in off-line and controlled environments. To address this issue, Maybank and Faugeras [32], [72] proposed the so-called *self-calibration* approach in which they used the information of matched points in several images taken by the same camera from different views instead of using known 3D points (calibration objects). In their two-step method, they first estimated the epipolar transformation from three pairs of views, and then linked it to the image of an absolute conic using the Kruppa equations [72]. Not long after the seminal work of Maybank and Faugeras, Basu proposed the idea of Active Calibration [5], [6] in which he included rotations of a camera and eliminated point-to-point correspondences.

An active environment can change the characteristics of a problem. For instance, an ill-posed and nonlinear problem for a passive observer can become well-defined and linear for an active observer [2]. Thus, to successfully calibrate a camera, Active Calibration needs to control the camera motion. This makes it a perfect choice in on-line platforms like robotics or surveillance, where the internal parameters might change due to focusing, zooming, or mechanical and thermal variations of the environment surrounding a camera. Therefore, knowing the motion of the camera is essential in Active Calibration, and as Hartley stated [48], "it simplifies the calibration task enormously." Another advantage of Active Calibration is its closed-form strategies that calculate the intrinsic parameters through only two pairs of images taken after panning and tilting the camera.

Other works [22], [23], [88] that used known camera motions have also been published almost at the same time. Since the method of Maybank and Faugeras needed high accuracy in the computations [47], [48], complicated rectification processes, and also because of the unavailability of the epipolar structure in the scenes taken from a fixed point, Hartley [47] proposed a method for self-calibrating a camera with constant intrinsics using projective distortions of several pure camera rotations. Inspired by Hartley's work, Agapito et al. proposed a self-calibration method for cameras that freely rotate while changing their internal parameters by zooming [1]. The notion of varying intrinsics has also been considered in [78] but no assumption about the camera motion has been made. Research in this area has expanded since the emergence of cell phone cameras capable of measuring the cam-

Table 5.1: Main Equations of Active Calibration. f_u and f_v denote the focal lengths in the u and v directions. The principal point is represented as a point with (δ_u, δ_y) distance to the center of the image. θ_p, θ_t and θ_r are angles of pan, tilt and roll rotations, respectively. (u_p, v_p), (u_t, v_t) and (u_r, v_r) are corresponding points after pan, tilt and roll rotations, respectively.

$$f_u \approx \left(\frac{\theta_p}{v_p - v}\right)(vu - v\delta_u - u\delta_v + \delta_u\delta_v) \tag{5.1}$$

$$f_v \approx \left(\frac{\theta_t}{u_t - u}\right)(uv - u\delta_v - v\delta_u + \delta_u\delta_v) \tag{5.2}$$

$$f_v^2 + \left(\frac{v(1+\theta_t^2) - v_t}{\theta_t}\right)f_v + (v^2 - 2v\delta_v + \delta_v^2) \approx 0 \tag{5.3}$$

$$f_u^2 + \left(\frac{u(1+\theta_p^2) - u_p}{\theta_p}\right)f_u + (u^2 - 2u\delta_u + \delta_u^2) \approx 0 \tag{5.4}$$

$$\delta_u\big(1 - \cos(\theta_r)\big) - \sin(\theta_r)\frac{f_u}{f_v}\delta_v = u_r - \cos(\theta_r)u - \sin(\theta_r)\frac{f_u}{f_v}v \tag{5.5}$$

$$\delta_u \sin(\theta_r)\frac{f_v}{f_u} + \delta_v\big(1 - \cos(\theta_r)\big) = v_r - \cos(\theta_r)v + \sin(\theta_r)\frac{f_v}{f_u}u \tag{5.6}$$

$$\delta_u = \frac{u_r + u}{2} \tag{5.7}$$

$$\delta_v = \frac{v_r + v}{2} \tag{5.8}$$

era motions with Gyroscope and Inertial Measurement Unit (IMU), and cameras such as Pan-Tilt-Zoom (PTZ). Given a close approximation of the camera motion, several papers proposed new formulations for calibrating the camera [35]–[37]. Some studies calibrate the camera by having specific type of control over camera rotations [56], [61], [66], [99]. More recent methods proposed self-calibration formulations that include camera lens distortions [40], [91], [109]. Also, some researchers expanded the self-calibration formulation to robotic camera networks [52] or used a camera rotation observed by another camera as the pattern to calibrate the observing camera [15]. Human motion has also been considered as a way to deduce the camera parameters [95].

The original Active Calibration strategies A, B, C, and D employ the equations reported in Table 5.1. Each strategy uses different combinations of these equations in order to calibrate the camera. Specifically, Strategy A first calculates the principal point using Eq. 5.2 and then estimates the focal length using Eq. 5.1 and Eq. 5.2. On the other hand, Strategy B first estimates focal lengths in both directions using Eq. 5.3 and Eq. 5.4. It then calculates the principal point location by Eq. 5.1. Strategies C and D require roll rotation of the camera. In particular, Strategy C first utilizes Strategy B to estimate focal lengths and later Eq. 5.6 to compute the center of projection. Strategy D follows a similar procedure for its first step. However, it calculates the principal point location using Eq. 5.7 and Eq. 5.8

which have been derived from rolling the camera by 180°.

The main downside of the Active Calibration Strategies A and B in [5]–[7] is that it calculates the camera intrinsics using a component of the projection equation in which a constraint is imposed by the degenerate rotations. For example, after panning the camera the equation derived from vertical variations observed in the new image plane is unstable. Furthermore, the small angle approximation using $\sin(\theta) = \theta$ and $\cos(\theta) = 1$ decreases the accuracy of the strategies when the angle of rotation is not very small. Also, rolling the camera [9] is impractical (without having a precise mechanical device) because it creates translational offsets in the camera center. In this chapter, we propose a Simplified Active Calibration (SAC) formulation in which the equations are closed-form and linear. To overcome the instability caused by using degenerate rotations in Active Calibration, we calculate focal lengths in each direction separately [26]. Then, through a mathematical derivation we remove the corresponding degenerate component from the equation. In addition, we do not use small angle approximation by replacing $\sin(\theta) = \theta$ and $\cos(\theta) = 1$. Hence, in our formulation we only refer to the elements of the rotation matrix. Moreover, the proposed method is more practical because it does not require a roll rotation of the camera; only pan and tilt rotations, which can be easily acquired using PTZ cameras, are sufficient.

The rest of the chapter is organized as follows. In Section 5.2 we present our proposed Simplified Active Calibration formulation. Section 5.3 reports and analyzes the results of the proposed method on simulated and real scenes. Finally, our conclusions are drawn in Section 5.4.

5.2 Simplified Active Calibration

Simplified Active Calibration (SAC) has been inspired by the novel idea of approximating the camera intrinsics using small rotations of the camera which was initially proposed in [5], [6] and extended in [7], [9]. Imposing three constraints on the translation of the camera generates a pure rotation motion. In addition, using small rotation angles provides a condition suitable for ignoring some non-linear terms in order to estimate the remaining linear parameters. The estimated intrinsics can then be used as an initial guess in a non-linear refinement process.

In general, SAC can be used in any platform in which information about the camera motion is provided by the hardware, such as in robotic applications where the rotation of the camera can be extracted from the inertial sensors or in surveillance control softwares that are able to rotate the PTZ cameras by specific angles. Having access to the rotation of the camera, we propose a 3-step process to calibrate a camera. In the first step, we present a closed-form solution to calculate an approximation of the focal length in u direction (f_u) using an image taken after a pan rotation of the camera, assuming that u and v represent

the two major axes of the image plane. In the second step, we estimate the focal length of the camera in the v direction (f_v) using an image taken after a tilt rotation of the camera. The third step consists of forming a system of linear equations to estimate the location of the principal point (u_0, v_0) in the image. Now, we have estimates for the four main components of the intrinsic matrix, namely f_u, f_v, u_0 and v_0. Thus, we require three pairs of images, one taken before and after a small pan rotation, one taken before and after a small tilt rotation, and one taken before and after a small pan-tilt rotation.

5.2.1 Rotation Formulation

Throughout the rest of the chapter, we formulate the rotation of the camera using the Euler angles in which every angle in the 3D coordinate system represents the amount of rotation about one of the coordinate axes and is denoted by a separate 3×3 matrix. The final rotation matrix is thus computed using $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$ where $\mathbf{R}_x$, $\mathbf{R}_y$, and $\mathbf{R}_z$ denote the rotations about the x, y, and z axes respectively. This formulation implies that the resulting matrix $\mathbf{R}$ has three degrees of freedom. Also, the elements of the final rotation matrix are represented as:

$$\mathbf{R} = [r_{ij}]_{3\times3} \tag{5.9}$$

where i denotes the row-wise element indices and j represents the column-wise element indices.

In SAC, it is crucial to know the correct direction of the rotation matrix and its handedness since it has to correspond to the acquired images. For example, "which rotation matrix corresponds to an image acquired after panning the camera to the left." Due to the importance of this issue in having an elegant formulation and obtaining realistic results, we briefly explain every rotation matrix and its direction which is used throughout the chapter.

Roll

Roll is a rotation about the z-axis, used only in Strategies C and D of the original Active Calibration [9]. However, SAC does not need a rolled image because rolling the camera and keeping the principal point fixed at the same time is impractical with current cameras. (Imposing a constraint on u or v while rotating about z is very difficult and creates a translational offset [60]). Therefore, a 3×3 identity matrix is used to calculate the final rotation matrix.

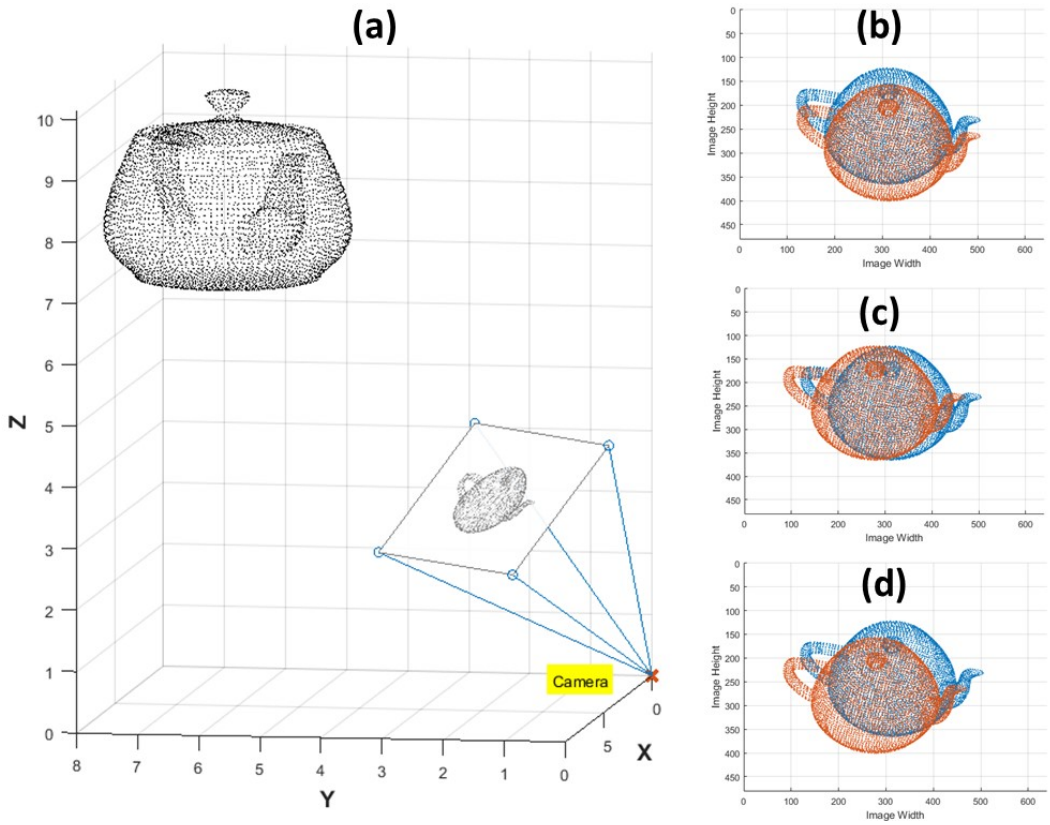


Figure 5.1: 3D scene and the simulated camera. **a)** A teapot in the 3D scene and its projected image on the simulated camera. **b)** The projected image of the teapot on the camera before (blue teapot) and after (red teapot) tilting the camera by 2.5°. **c)** The projected image of the teapot on the camera before (blue teapot) and after (red teapot) panning the camera by 2.5°. **d)** The projected image of the teapot on the camera before (blue teapot) and after (red teapot) panning the camera by 2.5° and then tilting the camera by 2.5°.

Pan

Pan rotation of the camera represents a rotation about the y-axis and is computed using the following equation.

$$\mathbf{R}_y = \begin{bmatrix} \cos(\theta_p) & 0 & -\sin(\theta_p) \\ 0 & 1 & 0 \\ \sin(\theta_p) & 0 & \cos(\theta_p) \end{bmatrix} \tag{5.10}$$

The direction implied by this rotation matrix in our predefined camera model is a clockwise orientation if one uses the right-hand rule. Therefore, rotating the camera to the right indicates a positive angle value. On the other hand, for a rotation to the left side the angle should have a negative sign.

Tilt

By tilt rotation we mean a rotation of the camera about the x-axis which can be achieved by the following matrix.

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_t) & \sin(\theta_t) \\ 0 & -\sin(\theta_t) & \cos(\theta_t) \end{bmatrix} \tag{5.11}$$

Unlike the pan rotation, tilt orientation is counter-clockwise considering the right-hand rule. So, if the camera rotates upward the angle is positive and if it rotates downward the angle of rotation is negative.

5.2.2 Camera Model

We assume that the camera is located at the origin of the Cartesian coordinate system and is looking at a distance of $z = f$ where the principal point is specified. It should be noted that f represents the focal length of the camera. Furthermore, the principal axis coincides with the z-axis, and the image plane is perpendicular to the principal axis. A point on the normalized camera coordinates is denoted by $\mathbf{x} = [x\ y\ 1]^T$. Also, the column (u) and row (v) coordinate axes of the reference image plane are parallel to the x-axis and the y-axis of the camera, respectively. The relation between points in the normalized camera coordinates and the image points is as follows:

$$u = m_u x + u_0 \tag{5.12}$$

$$v = -m_v y + v_0 \tag{5.13}$$

where m_u and m_v represent the width and height of the pixels, respectively, and (u_0, v_0) are the location coordinates of the principal point in the image.

Every 3D point $\mathbf{X} = [X\ Y\ Z]^T$ in the world that is visible to the camera can be projected onto a specific point $\mathbf{v} = [u\ v\ 1]^T$ of the image plane and can be calculated using the camera

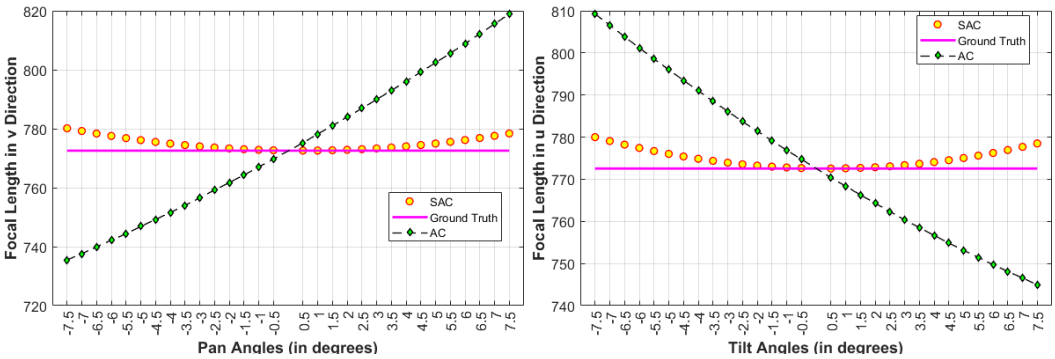


Figure 5.2: Focal length calculated in the u and v directions using Active Calibration Strategy B (AC) versus SAC for various angles of rotations. In SAC we only use one point correspondence.

intrinsic matrix.

$$\mathbf{K} = \begin{bmatrix} f_u & s & u_0 \\ 0 & -f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.14}$$

where $f_u = fm_u$ is the focal length of the camera in the u direction (in pixels), $f_v = fm_v$ represents the focal length of the camera in the v direction. With modern cameras it is reasonable to assume that sensor elements are perpendicular and so the value of the camera skew (s) is zero [46].

Note that any camera transformation is equivalent to a similar transformation of the scene but in the opposite direction [63]. For stationary cameras that freely rotate but stay in a fixed location, the camera transformation is only modeled by its rotation. As stated in [46], for cameras, such as PTZ cameras, which are fixed in location but free to rotate and zoom, the effect of camera translation is insignificant and hence the translation of the camera can be considered to be zero. Therefore, every point $\mathbf{v}$ in an image seen by a stationary camera is transformed to a point $\mathbf{v}'$ in another image taken after camera rotation. The mathematical relationship between $\mathbf{v}$ and $\mathbf{v}'$ is thus represented by:

$$w\mathbf{v}' = \mathbf{K}\mathbf{R}^T\mathbf{K}^{-1}\mathbf{v} \tag{5.15}$$

where w is the scale of the projection and represents the depth of the point. It should be noted that $\mathbf{R}^T = \mathbf{R}^{-1}$ because the rotation matrix is orthonormal.

5.2.3 Focal Length in the u Direction

An image taken after a pan rotation of a camera provides a very straightforward formulation to estimate the focal length of the camera. In fact, it imposes two constraints on the camera rotations around the x and z axes. The resulting projection equation after substituting $\mathbf{R_y}$ for $\mathbf{R}$ is:

$$w\mathbf{v}' = \mathbf{K}\mathbf{R}_y^T\mathbf{K}^{-1}\mathbf{v} \tag{5.16}$$

$\mathbf{R}_y^T$ has only one DoF, which is the angle of rotation around the y-axis. After expanding and simplifying Eq.5.16 and eliminating the scale of projection, the following direct projection equations are obtained.

$$u' = \frac{r_{11}(u - u_0) + r_{31} f_u}{r_{13}\dfrac{u - u_0}{f_u} + r_{33}} + u_0 \tag{5.17}$$

$$v' = v_0 - \frac{v_0 - v}{r_{13}\dfrac{u - u_0}{f_u} + r_{33}} \tag{5.18}$$

where r_{ij} is an element of $\mathbf{R}_y^T$ at row i and column j. After simplification of Eq.5.18:

$$\frac{u - u_0}{f_u} = \frac{\dfrac{v_0 - v}{v_0 - v'} - r_{33}}{r_{13}} \tag{5.19}$$

Note that after a small pure pan rotation, the v coordinates of the new image will not be affected by the transformation. (The reader is referred to [62] for a detailed explanation and analysis about this fact.) In other words, image pixels only move horizontally when the angle of rotation is small and so the rate of change in the v direction before and after the pan rotation is close to one, viz:

$$\frac{v_0 - v}{v_0 - v'} \approx 1 \tag{5.20}$$

Substituting Eq.5.20 into Eq.5.19 and then replacing the equation obtained for the term $\dfrac{u - u_0}{f_u}$ in the Eq.5.17, we have:

$$u' = \frac{r_{11}(u - u_0) + r_{31} f_u}{r_{13}\dfrac{1 - r_{33}}{r_{13}} + r_{33}} + u_0 \tag{5.21}$$

The above substitution changes the value of the denominator to 1 and hence simplifies the whole projection equation.

$$u' - r_{11}u = r_{31} f_u + (1 - r_{11})u_0 \tag{5.22}$$

Since Eq.5.22 is linear, one might think that it can be solved by constructing a linear system of equations using the matched points from two images taken after the pan rotations of the camera. Unfortunately, the equation is numerically unstable because the value of $1 - r_{11} \approx 0$ which causes ambiguity in calculating the shift in the principal point [1]. In short, we cannot calculate the location of the principal point in the u direction from a camera rotated purely around the y axis. Knowing that the principal point is close to the center of the image ($c_v = h/2, c_u = w/2$), where h and w represent the image height and width respectively, we replace u_0 with c_u in Eq.5.22. Thus, we can derive a suitable linear equation to estimate the focal length in x direction from an image taken after a pan rotation.

$$f_u = \frac{u' - r_{11}u - (1 - r_{11})c_u}{r_{31}} \tag{5.23}$$

Eq.5.23 needs only one point u in the reference image that corresponds to u' in the transformed image. If there are more point correspondences, we can easily use the average of these points to obtain more robust results.

5.2.4 Focal Length in v Direction

So far, we could estimate f_u by the information provided from an image taken after a pan rotation. We repeat the same procedure to approximate f_v. This time we need an image taken after a pure tilt rotation of the camera. Thus, the projection equation is characterized by replacing $\mathbf{R}$ with $\mathbf{R}_x$ and relating the coordinates of a point in the reference image $\mathbf{v}$ and a point in the tilted image $\mathbf{v}'$ by:

$$u' = \frac{u - u_0}{r_{23}\dfrac{v_0 - v}{f_v} + r_{33}} + u_0 \tag{5.24}$$

$$v' = v_0 - \frac{r_{22}(v_0 - v) + r_{32}f_v}{r_{23}\dfrac{v_0 - v}{f_v} + r_{33}} \tag{5.25}$$

Following the same reasoning as in Section 5.2.3, a closed-form solution to estimate the focal length of the camera in the v direction is obtained by:

$$f_v = \frac{r_{22}v - v' + (1 - r_{22})c_v}{r_{32}} \tag{5.26}$$

5.2.5 Principal Point

To estimate the location of the principal point we need to impose one constraint on the rotation matrix which can be achieved by preventing the camera from rotating around the z-axis. In real applications the easiest way is to mount the camera on a tripod. In case of working in a robotic environment or with a PTZ camera, controlling roll rotation is straightforward since the camera has already been mounted or fixed. Therefore, we match an image taken after a pan and tilt rotation of the camera with the reference image and use the acquired point correspondences to estimate the location of the principal point.

Following the general projection equation in Eq.5.15, the direct equations for relating the location of a point in the reference image to its matched point in the transformed image are described by:

$$u' = \frac{r_{11}(u - u_0) + r_{21}(v_0 - v)\dfrac{f_u}{f_v} + r_{31}f_u}{r_{13}\dfrac{u - u_0}{f_u} + r_{23}\dfrac{v_0 - v}{f_v} + r_{33}} + u_0 \tag{5.27}$$

$$v' = v_0 - \frac{r_{12}(u - u_0)\dfrac{f_v}{f_u} + r_{22}(v_0 - v) + r_{32}f_v}{r_{13}\dfrac{u - u_0}{f_u} + r_{23}\dfrac{v_0 - v}{f_v} + r_{33}} \tag{5.28}$$

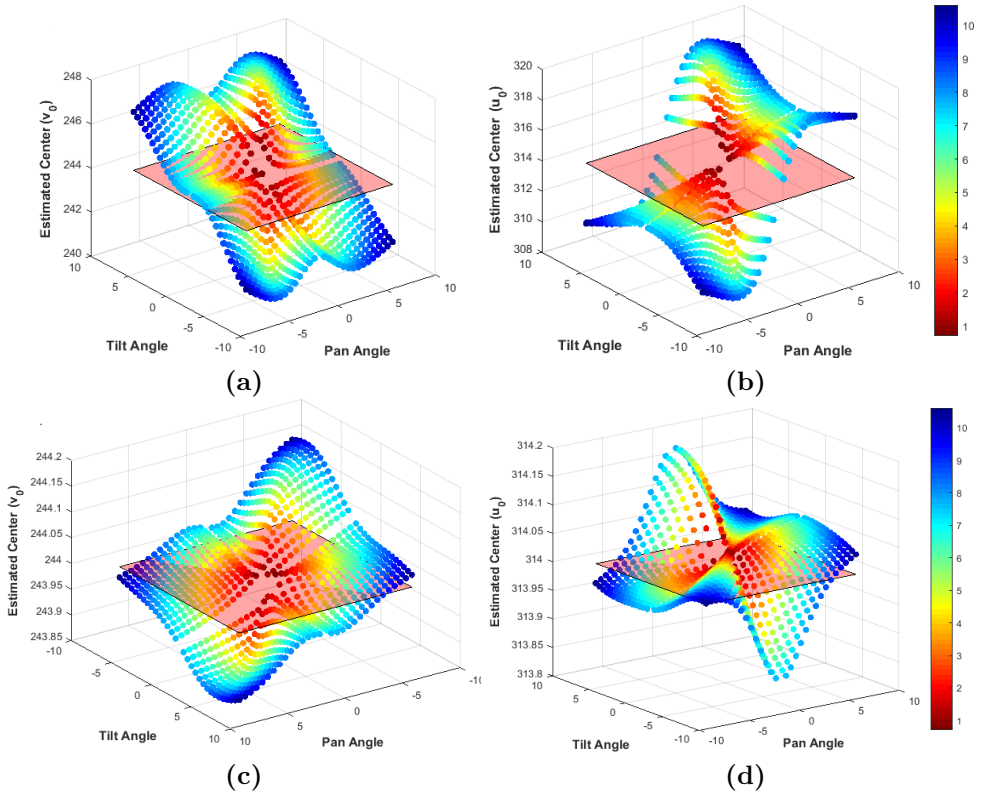


Figure 5.3: Coordinates of the principal points calculated after various pan/tilt rotations of random 3D points. Colors are distributed based on the L^2 norm of the pan and tilt angles.The red plane represents the ground truth. **a)** Shows the values obtained for v_0 when inaccurate focal lengths ($f_v = 774.71$ and $f_u = 771.18$) are used. $MSE(v_0) = 1.49$ pixels for all combinations of pan and tilt angles. **b)** Shows the values obtained for u_0 when inaccurate focal lengths ($f_v = 774.71$ and $f_u = 771.18$) are used. $MSE(u_0) = 2.30$ pixels for all combinations of pan and tilt angles. **c)** Shows the values obtained for v_0 when accurate (ground truths denoted by F) focal lengths ($F_u = F_v = 772.55$) are used. $MSE(v_0) = 0.05$ pixels for all combinations of pan and tilt angles. **d)** Shows the values obtained for u_0 when accurate (ground truths denoted by F) focal lengths ($F_u = F_v = 772.55$) are used. $MSE(u_0) = 0.04$ pixels for all combinations of pan and tilt angles. The red plane specifies the ground truth.

where except for u_0 and v_0, all other terms are known. After simplifying the equations and collecting the coefficients of various powers of u_0 and v_0, we see that the equations are nonlinear due to the presence of the two terms $-r_{13}f_u^{-1}u_0^2 - r_{23}f_v^{-1}u_0v_0$ and $-r_{23}f_v^{-1}v_0^2 - r_{13}f_u^{-1}u_0v_0$ in Eq.5.27 and Eq.5.28, respectively. Nevertheless, the equations can be solved using any nonlinear solver such as Levenberg-Marquardt. But, in order to let the nonlinear solver converge towards the true global minimum, we first need a reasonable initial guess. Here, we propose a method to linearize Eq.5.27 and Eq.5.28 to achieve a close estimation of the location of the principal point when the focal lengths and the camera rotations are known.

A feasible approach to linearize the projection equations is to decrease the contributions of the two above-mentioned nonlinear terms in the equations and then neglect them entirely in the equations. Decreasing the value of the nonlinear terms depends on two factors, namely r_{13} and r_{23} which are the elements of the rotation matrix and the value of the unknowns which are u_0 and v_0. The former coefficients have already been reduced due to our initial assumption of rotating the camera by small angles. On the other hand, we will show that we can reduce u_0 and v_0 to smaller values by manipulating the scale of these points' coordinates.

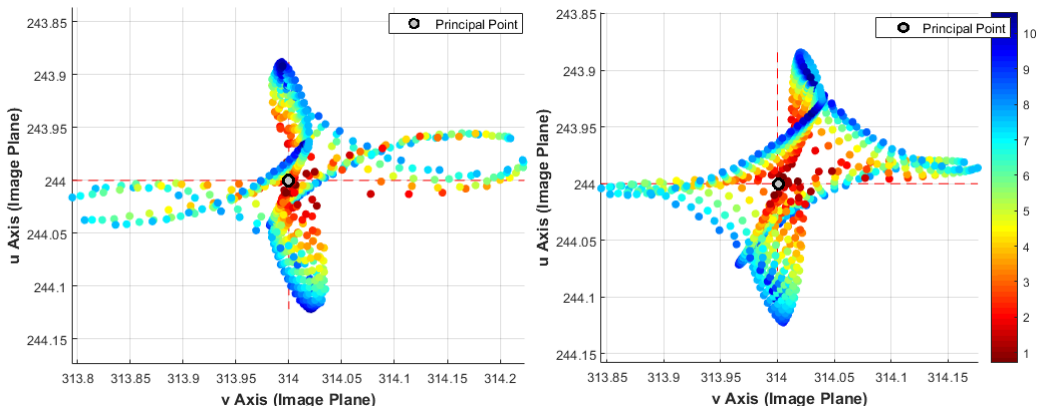
Estimating the principal point by a nonlinear algorithm is known to be arduous since it tends to fit to noise [1]. Due to this sensitivity to noise, researchers have taken advantage of including some prior knowledge about the distribution of the principal point. It is reasonable to expect that the principal point is close to the center of the image. This prior knowledge is the basis of the *Maximum a Posteriori Estimation* employed in [1] to alter the cost function of the minimization problem. In order to arrive at a linear system, we employ the same idea. Specifically, we assume that the principal point is only slightly shifted from the center of the image.

$$\begin{aligned} u_0 &= c_u + \delta_u \\ v_0 &= c_v + \delta_v \end{aligned} \tag{5.29}$$

where δ_u and δ_v represent the amount of shift in the u and v directions, respectively and (c_u, c_v) are the coordinates of the center of the image. Following this change, we replace each $u - u_0$ term with $\hat{u} - \delta_u$ and every $v_0 - v$ with $\hat{v} + \delta_v$ in Eq.5.27 and Eq.5.28 where $\hat{u} = u - c_u$ and $\hat{v} = c_v - v$. Therefore, after some simplifications, the general projection equations can be rewritten based on the new variable substitutions.

$$G\delta_u^2 + H\delta_u\delta_v + (A + I - G\hat{u}')\delta_u + (B - H\hat{u}')\delta_v = I\hat{u}' - C \tag{5.30}$$

$$-H\delta_v^2 - G\delta_u\delta_v + (D - G\hat{v}')\delta_u + (E - I - Hv')\delta_v = I\hat{v}' - F \tag{5.31}$$



(a) **(b)**

Figure 5.4: The estimated locations of the principal point on the image plane for combinations of various rotation angles (from $-7.5°$ to $7.5°$) using Eq.5.32. Colors are distributed based on the L^2 norm of the pan and tilt angles. **a)** Results for solving with only four point correspondences of the teapot point cloud. **b)** Results for solving with 500 point correspondences of the random 3D points. The actual principal point location is (314,244).

where,

$$
\begin{aligned}
A &= -r_{11} \quad , \quad B = r_{21}\frac{f_u}{f_v} \\
C &= r_{11}\hat{u} + r_{21}\hat{v}\frac{f_u}{f_v} + r_{31}f_u \\
D &= -r_{12}\frac{f_v}{f_u} \quad , \quad E = r_{22} \\
F &= r_{12}\hat{u}\frac{f_v}{f_u} + r_{22}\hat{v} + r_{32}f_v \\
G &= -\frac{r_{13}}{f_u} \quad , \quad H = \frac{r_{23}}{f_v} \\
I &= r_{13}\frac{\hat{u}}{f_u} + r_{23}\frac{\hat{v}}{f_v} + r_{33}
\end{aligned}
$$

By including our prior knowledge about the image center into the equations, we significantly reduce the values of nonlinear terms and allow them to be ignored. Once the nonlinear terms are removed, a linear system of equations can be constructed using the detected point correspondences.

$$
\begin{bmatrix}
A + I_1 - Gu'_1 & B - Hu'_1 \\
\vdots & \vdots \\
A + I_n - Gu'_n & B - Hu'_n \\
D - \hat{v}'_1 G & E - I_1 - Hv'_1 \\
\vdots & \vdots \\
D - \hat{v}'_n G & E - I_n - Hv'_n
\end{bmatrix}
\begin{bmatrix} \delta_u \\ \delta_v \end{bmatrix}
=
\begin{bmatrix}
\hat{u}'_1 I_1 - C_1 \\
\vdots \\
\hat{u}'_n I_n - C_n \\
\hat{v}'_1 I_1 - F_1 \\
\vdots \\
\hat{v}'_n I_n - F_n
\end{bmatrix}
\tag{5.32}
$$

where $(.)_i$ represents using coordinates of the ith point in the corresponding term and n is the number of correspondences. As shown above, the system of equations is constructed in

Table 5.2: Results of the proposed simplified active calibration on 1000 separate 3D random points for various small pan and tilt angles. In the table, GT denotes the Ground Truth and SD represents the Standard Deviation. The error values are in pixels.

Pan	Tilt		f_u	f_v	u_0	v_0
		GT	772.55	772.55	314	244
		Mean	772.68	772.57	314.005	244.02
$-0.5°$	$0.5°$	SD	0.07	0.01	0.01	0.01
		Error	0.13	0.02	0.005	0.02
		Mean	772.68	772.62	314.03	244.06
$-0.5°$	$1°$	SD	0.07	0.04	0.07	0.06
		Error	0.13	0.07	0.03	0.06
		Mean	772.61	772.76	314.23	243.61
$1°$	$-1°$	SD	0.02	0.09	0.12	0.15
		Error	0.06	0.21	0.23	0.38
		Mean	773.02	772.73	314.21	244.44
$-1.5°$	$1.5°$	SD	0.13	0.07	0.17	0.17
		Error	0.47	0.19	0.21	0.44

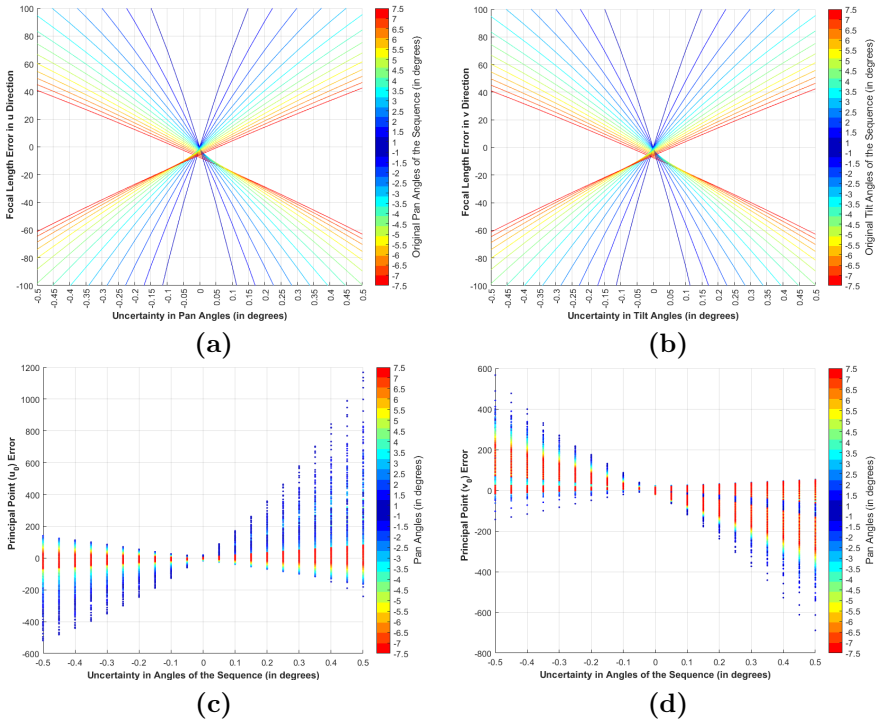


Figure 5.5: The error caused by uncertainty in determining the angle of the camera. **a)** The effects of the uncertainty of the camera pan rotation on calculating the focal length in u direction by SAC. **b)** The effects of the uncertainty of the camera tilt rotation on calculating the focal length in v direction by SAC. **c)** The effects of the uncertainty of the camera pan and tilt rotation on calculating the u coordinate of the principal points by SAC. **d)** The effects of the uncertainty of the camera pan and tilt rotation on calculating the v coordinate of the principal points by SAC.

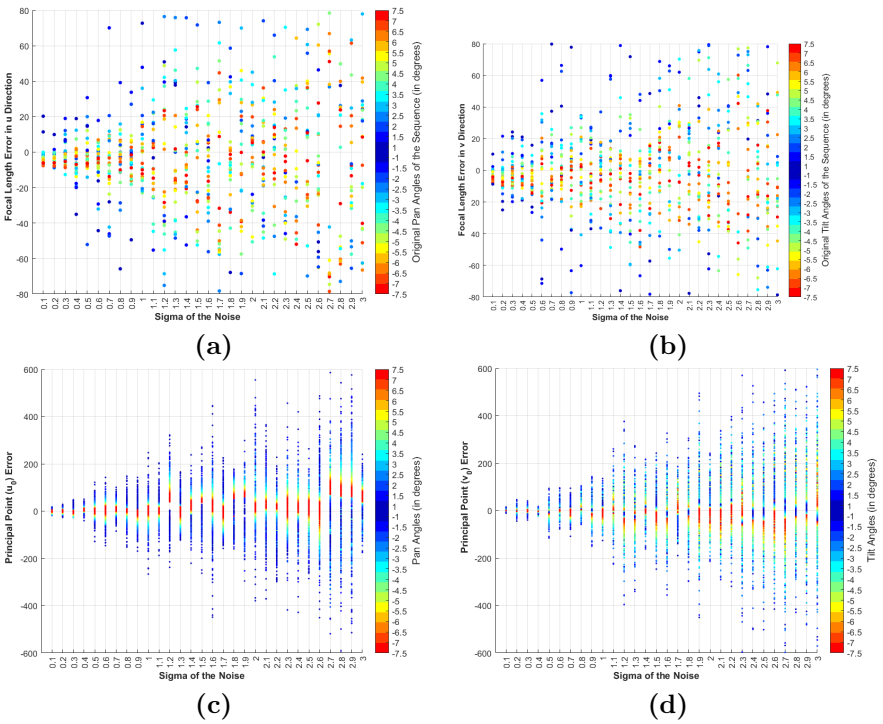


Figure 5.6: The error caused by uncertainty in location of points. **a)** Error of the estimated focal length in u direction using SAC when the location of the teapot points are disturbed by different values of σ_{pixel}. **b)** Error of the estimated focal length in v direction using SAC under the same conditions as in (a). **c)** Error of the estimated u_0 using SAC under the same conditions as in (a). **d)** Error of the estimated v_0 using SAC under the same conditions as in (a).

the form $A_{[n\times 2]}\mathbf{x} = \mathbf{b}_{[n\times 1]}$ and can be easily solved using a least square method or any linear solver. However, since there are two unknowns, by detecting only one point correspondence we are able to solve the system for δ_x and δ_y. For better estimates, one can use more point correspondences.

5.2.6 Algorithm

In this section, we provide a recipe that can guide the reader how to calibrate an active camera with SAC. In order to use SAC equations, there should exist an active camera and a software that can extract the value of motions from the camera or can instruct the camera to rotate by specific angles. In our experiments we used a Canon vc-c50i that is controlled by our own software which can be seen in Fig.5.7. Having these requirements, one can calibrate the camera for a specific zoom position the following steps:

1. Rotate the camera to its home position or any position that needs to be the reference frame and take a picture.

2. Pan the camera and take a picture. The absolute value of pan angle preferably can

be bounded from above by approximately 7° and from below by approximately 2°.

3. Tilt the camera and take a picture. The absolute value of tilt angle preferably can be bounded from above by approximately 7° and from below by approximately 2°.

4. Pan and tilt the camera and take a picture. The absolute value of pan and tilt angles preferably can be bounded from above by approximately 7° and from below by approximately 2°.

5. Find correspondences between pan image and reference frame, tilt image and reference frame, and pan-tilt image and reference frame.

6. Among all correspondences between pan and reference image, find the closest matched point to center of the image and use it in Eq.5.23 and estimate f_u. Instead of using the closest point to the center of the image, one can also calculate f_u for each matched point separately and take an average between all focal length estimates at the end.

7. Among all correspondences between tilt and reference image, find the closest matched point to center of the image and use it in Eq.5.26 and estimate f_v. Instead of using the closest point to the center of the image, one can also calculate f_v for each matched point separately and take an average between all focal length estimates at the end.

8. Use all correspondences of pan-tilt and reference image and form a linear system of equations based on Eq.5.32, solve the system [1] for (u_0, v_0).

5.3 Results and Analysis

In order to better understand the performance of the proposed simplified active calibration, we perform several experiments using synthetic and real data to clarify how the method works for various rotation angles. Our synthetic data was acquired from a synthetic 3D scene with a camera looking at the object (a teapot) in the scene. Figure 5.1 illustrates the 3D scene. We show that rotating by a small angle is crucial to obtain good results.

Based on our proposed method, the focal length in u and v directions can be estimated using Eq.5.23 and Eq.5.26, respectively. Only one point correspondence is required to calculate the focal length. Fig.5.2 shows the focal lengths estimated using various pan and tilt angles. It can be seen that when the pan and tilt angles are small, the estimated focal lengths are very close to the ground truth. Note that we have not yet considered the impact of noise on the equations.

The magnitude of the pan and tilt angles affect estimating the principal point location as well. In fact, two successive rotations (pan and tilt) are required to calculate the center of

[1] MATLAB can solve the equation system by the command $A\backslash\mathbf{b}$.

projection. Thus, in another experiment, we rotate the camera by 900 combinations of pan and tilt rotations (from -7.5° to 7.5°), and then use the projected points on the image plane to estimate the principal point location by calculating the proposed formulation (Eq.5.32). Note that we use either the estimated value of f_u and f_v that were calculated in the previous step or the actual focal length. The results are shown in Fig.5.3. We can see that for small rotations, the results obtained by our formulation are close to the real location which has been identified in the figure by a red plane. Even when the focal length is not accurate, a good estimate of the principal point can be found if the pan and tilt rotations are small assuming that the camera works in ideal conditions and so there is no noise.

In addition, Fig.5.3 shows that the error caused by the angle variation is more negligible than the error caused by inaccurate values of the focal length. Specifically, the Mean Square Error of v_0 ($MSE(v_0)$) is 1.49 pixels when inaccurate focal lengths are used (Fig.5.3(a)). By contrast, when the actual focal length is used in Eq.5.32, $MSE(v_0)$ decreases to 0.05 pixels (Fig.5.3(c)). This reveals the significance of having accurate focal length in calculating the principal point location. A similar analysis is valid for the other axis (u) which is shown in Fig.5.3(b) and Fig.5.3(d).

Knowing how many point correspondences are required to calculate the principal point location (by Eq.5.32) is crucial. Based on our experiments, with only four points that are uniformly distributed in the image, a good estimate of the principal point location can be obtained. Fig.5.4(a) illustrates principal point locations (on the image plane) obtained by solving Eq.5.32 with inaccurate focal lengths and only four point correspondences of the teapot point cloud. Fig.5.4(b) shows the estimated principal points on the image obtained by solving Eq.5.32 with inaccurate focal lengths and 500 point correspondences of random 3D points. Both experiments are carried out on 900 combinations of pan and tilt angles which range from -7.5° to 7.5°. Pan and tilt angles are included in Fig.5.4 by calculating the L^2 norm of the angles ($\sqrt{\theta_p^2 + \theta_t^2}$) and assigning meaningful colors to them that range from red (closer to zero) to blue (bigger angles). As can be seen in Fig.5.4(a), when the pan and tilt rotations are small, even with four point correspondences a principal point that is very close to the actual principal point (specified by a red cross) can be calculated. By using more point correspondences we can obtain almost similar error distribution, which is shown in Fig.5.4(b).

In another experiment, we calculate the proposed simplified active calibration formulation on 1000 different runs of 500 randomly generated 3D points for small pan ($\theta_p = -0.5^\circ$) and tilt ($\theta_t = 0.5^\circ$) angles. The order of calculating the intrinsics was specified earlier. The mean and standard deviation of the results obtained are shown in Table 5.2. As we can see, our proposed active calibration formulation attains results very close to the ground truth. Specifically, the error in the principal point location is less than one pixel and the error in

focal length estimates is less than 2 pixels.

All things considered, we assessed the proposed Simplified Active Calibration formulation on simulated scenes in ideal situations, i.e., when the 3D rays are not altered due to camera lens distortions and when there is no noise in the scene. We showed that the proposed formulation can estimate the camera intrinsics when the camera rotation is small and pure. In fact, for calculating the focal length we used the so-called "degenerate camera configuration." Moreover, we demonstrated that using small rotations one can compensate for the error (in ideal conditions) caused by inaccuracy in the estimated focal length to find the principal point. In other words, rotating the camera by small angles lessens the influence of inaccurate focal length in calculating the principal point location assuming that there is no noise.

5.3.1 Noise Analysis

All of the above-mentioned experiments were done in ideal situations where the angles acquired from the camera and the location of matched points were assumed to be exact. In real-world conditions, however, angles and point correspondences are noisy. In the following sections we try to understand how the proposed method works in real-world conditions where parameters are contaminated by various types of noise.

5.3.2 Angular Uncertainty

Acquiring the rotation angles requires either specific devices such as gyroscopes or a specially designed camera called a PTZ camera. Even using these devices does not guarantee that the extracted rotation angles are noise-free.

Using the analytic differential analysis, we can derive a formula that demonstrates how the angular noise affects the proposed equations. We assume that the noise affects pan and tilt rotation of the camera separately and so the noise in each axis is independent from the other axis. In this case, the elements of the rotation matrix will be affected by the noise and the proposed equation for estimating f_u will be changed to:

$$f_u \approx \frac{(c_u - u)cos(\theta_p + \delta_{\theta_p}) + u' - c_u}{sin(\theta_p + \delta_{\theta_p})} \tag{5.33}$$

where δ_{θ_p} is the error caused by pan rotation of the camera. The sensitivity of the equation can then be denoted as:

$$\frac{\partial f_u}{\partial \delta_{\theta_p}} \approx -\frac{(u' - c_u)cos(\theta_p + \delta_{\theta_p}) - u + c_u}{sin^2(\theta_p + \delta_{\theta_p})} \tag{5.34}$$

If we use the angle sum identities, we have:

$$\frac{\partial f_u}{\partial \delta_{\theta_p}} \approx -\frac{(u' - c_u)(cos(\theta_p)cos(\delta_{\theta_p}) - sin(\theta_p)sin(\delta_{\theta_p})) - u + c_u}{(sin(\theta_p)cos(\delta_{\theta_p}) + cos(\theta_p)sin(\delta_{\theta_p}))^2} \tag{5.35}$$

Assume that δ_{θ_p} is very small and hence, $cos(\delta_{\theta_p}) \approx 1$ and $sin(\delta_{\theta_p}) \approx \delta_{\theta_p}$. In this case the equation becomes:

$$\frac{\partial f_u}{\partial \delta_{\theta_p}} \approx -\frac{(u' - c_u)(cos(\theta_p) - \delta_{\theta_p} sin(\theta_p)) - u + c_u}{sin^2(\theta_p) + \delta_{\theta_p}^2 cos^2(\theta_p) + \delta_{\theta_p} sin(2\theta_p)} \tag{5.36}$$

As we can see in Eq.5.36, the impact of the angular noise depends on several parameters including the error in angle of pan rotation δ_p, the coordinates of matched points u and u' and the center of the image c_u. However, it is clear from Eq.5.36 that if $\theta_p \gg \delta_{\theta_p}$ the equation will be less affected by the noise element since $sin^2(\theta_p)$ and $cos(\theta_p)$ become the most dominant terms. Also, a similar equation and analysis can be made for f_v.

From a practical perspective, we can do some experiments to see how the angular noise affects the equations. To do so, we simulate the noisy conditions of a real-world application by contaminating the angles of the above-mentioned teapot sequences with increasing angular errors. While the point correspondences are kept fixed for all of the pan and tilt rotations, we calculate the focal length (Eq.5.23 and Eq.5.26) and principal point coordinates (Eq.5.32) using contaminated pan and tilt angles. The results are shown in Fig.5.5. Specifically, Fig.5.5(a) and Fig.5.5(b) show the error of our proposed formula for estimating the focal length when the pan and tilt angles are not accurate. Every sequence has been colored based on its rotation angle, ranging from blue indicating smaller angles to red for larger angles. For focal length estimation, Fig.5.5(a) and Fig.5.5(b) illustrate that the sequences taken with smaller angles have steeper slopes than the sequences acquired with larger rotation angles. This shows that focal lengths are more sensitive to angular noise when the camera is rotated by smaller angles rather than larger angles.

Fig.5.5(c) and Fig.5.5(d) demonstrate how uncertainty in angular values affects estimating the principal point coordinates. Similar to the effect of noise on focal lengths, the distribution of the red colors (greater angles) around the zero line in Fig.5.5(c) and Fig.5.5(d) indicates that the estimates of the principal point coordinates are less affected by the angular noise when the angle of rotation is not very small.

Overall, when the camera is rotated by very small angles, the influence of the angular noise on SAC equations is significant. On the other hand, SAC tends to use the benefit of rotating the camera by small angles. Therefore, to avoid magnifying the effect of noise it is important not to rotate the camera by very small angles. If the pan and tilt angle of the camera is not very small (usually $> 2°$), the difference in estimated focal lengths will be less than 50 pixels which are still considered as close initial guesses for further non-linear refinements. Nonetheless, our experiments with real images in Section 5.3.4 reveal that SAC can be used in real situations and the angular noise makes the estimation slightly inaccurate.

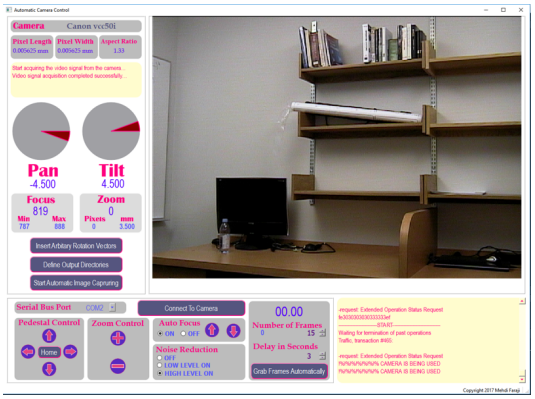


Figure 5.7: A screenshot of the designed Automatic Camera Control application that is able to rotate the camera by specific angles about Y-axis (pan) and X-axis (tilt). The camera is a Canon vc-c50i.

5.3.3 Point Correspondence Noise

Another type of noise that affects the SAC equations is the noise in the location of features used for matching. To simulate such conditions, we assume that the location of every teapot point is disturbed by a Gaussian noise with zero mean and standard deviation σ_{pixel}. In terms of an analytical analysis we can calculate the partial derivatives with respect to the noise to figure out the rate at which the function will be changed based on the noise contamination. Let $\mathbf{u} = [u \;\; v]^T$ correspond to $\mathbf{u} = [u' + \delta_u \;\; v' + \delta_v]^T$ where δ_u and δ_v are displacements in u and v due to noise, respectively. We have:

$$f_u \approx \frac{(u' + \delta_u) - r_{11}u - (1 - r_{11})c_u}{r_{31}} \tag{5.37}$$

$$f_v \approx \frac{r_{22}v - (v' + \delta_v) + (1 - r_{22})c_v}{r_{32}} \tag{5.38}$$

Therefore, the rate of change in f_u and f_v can be formulated as:

$$\frac{\partial f_u}{\partial \delta_u} = \frac{1}{r_{31}} = \frac{1}{sin(\theta_p)} \tag{5.39}$$

$$\frac{\partial f_v}{\partial \delta_v} = \frac{-1}{r_{32}} = \frac{1}{sin(\theta_t)} \tag{5.40}$$

As we can see, the point correspondence noise of each axis affects the focal length equations based on the angle of the rotation. So, if we rotate the camera by higher angles, the equations become more robust. However, one should note that due to the approximation that we made in Eq.5.20 to linearize the equation, rotating by a large angle would be problematic and decreases the accuracy of focal length estimation.

To further analyze the sensitivity of the method experimentally, we calibrate the camera using SAC for all σ_{pixel} in the range of 0 to 3. The intrinsic parameters obtained are illustrated in Fig.5.6.

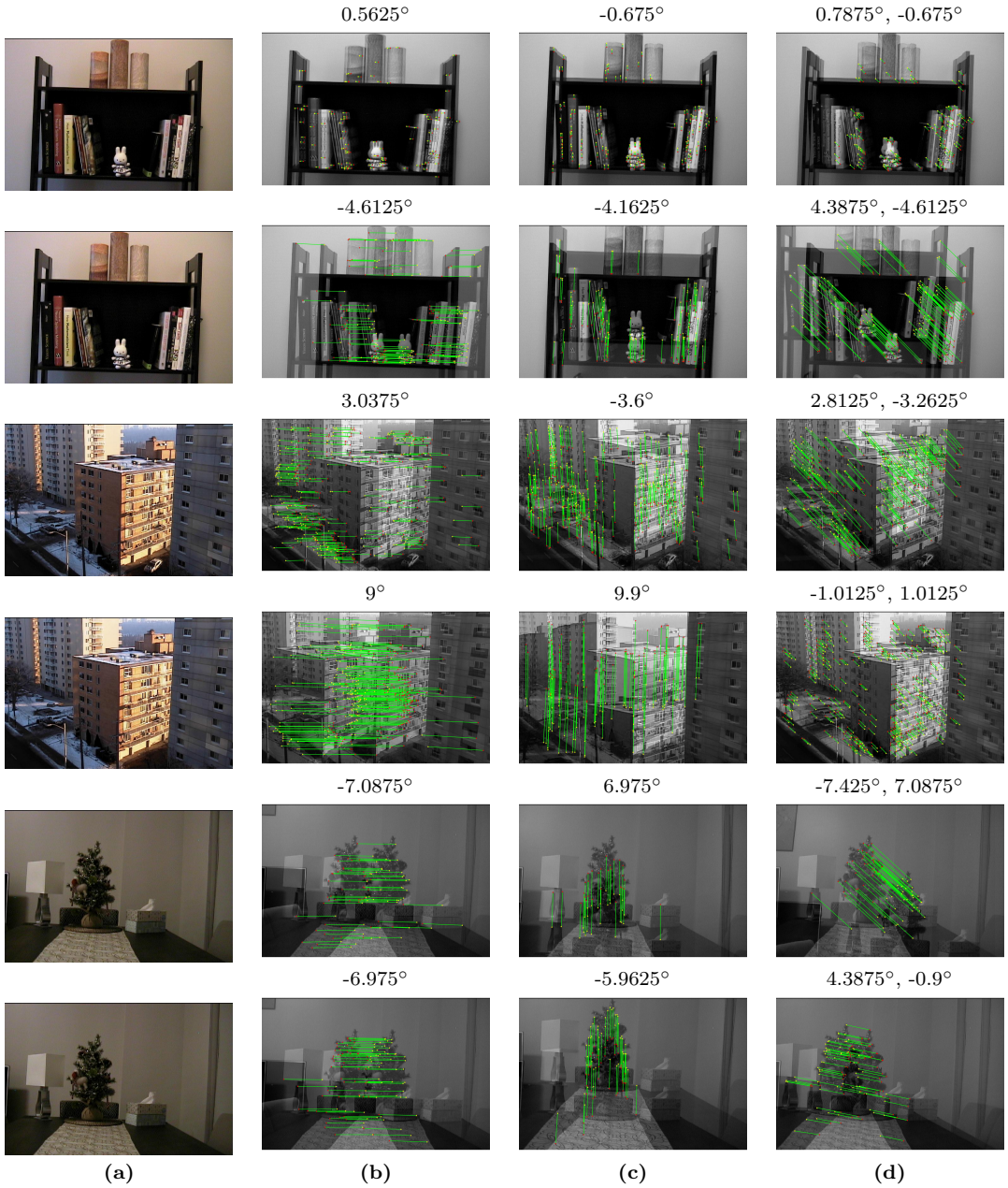


Figure 5.8: Six sequences of real images used for SAC taken with 2 different cameras. Matched points are shown by overlaying the new image on the reference image. SAC only uses one of these correspondences to estimate the focal length. Every row represents one sequence. Camera angles are shown on top of each image. **a)** Reference images. **b)** Image taken after panning the camera. **c)** Image taken after tilting the camera. **d)** Image taken after first panning the camera and then tilting the camera.

Table 5.3: Results of the proposed Simplified Active Calibration for 9 sequences of real images taken from 2 different cameras pointed at 3 scenes. All angles are in degrees. The "Pan" column indicates the pan angle of the camera for the first image of the sequence. The "Tilt" column represents the tilt angle of the camera for the second image of the sequence. The "PT Pan" and "PT Tilt" columns denote the pan and tilt angles of the camera for the third image of the sequence, respectively. δ_{f_u},δ_{f_v},δ_{u_0}, and δ_{v_0} are the pixel errors from the corresponding ground truth acquired from calibrating the camera using the method of Zhang [114]. Note that we only used one matched point for estimating the focal length in both direction; however, for estimating principal points we used 50 to 200 correspondences depending on the image content.

Camera 1 -Indoor scene Zhang's calibration results (checkerboard used) [114]: $F_u = 1039$, $F_v = -948$, $U_0 = 357.7$, $V_0 = 252.8$												
#	Pan	Tilt	PT Pan	PT Tilt	f_u	f_v	u_0	v_0	δ_{f_u}	δ_{f_v}	δ_{u_0}	δ_{v_0}
1	0.5625°	-0.675°	0.7875°	-0.675°	906.55	-1001.65	333.68	245.27	133.17	53.18	24.06	7.58
2	-1.8°	2.025°	-1.575°	1.8°	1034.68	-960.84	341.21	241.56	5.03	12.37	16.52	11.29
3	3.0375°	-3.6°	2.8125°	-3.2625°	1094.46	-982.49	408.27	187.27	54.74	34.01	50.53	65.59
4	-4.6125°	-4.1625°	4.3875°	-4.6125°	1065.70	-967.17	447.95	141.51	25.98	18.70	90.22	111.35
5	-7.0875°	6.975°	-7.425°	7.0875°	1094.02	-983.66	350.41	239.23	54.30	35.19	7.32	13.62
6	-7.9875°	-7.425°	1.35°	-0.7875°	1066.96	-979.01	351.81	226.68	27.25	30.53	5.92	26.17
7	9°	9.9°	-1.0125°	1.0125°	1098.68	-998.20	459.99	239.88	58.96	49.73	102.25	12.97
8	-6.975°	-5.9625°	4.3875°	-0.9°	1079.24	-986.78	358.91	224.26	39.52	38.30	1.18	28.59
9	7.7625°	6.525°	-0.9°	5.175°	1083.45	-991.52	416.62	238.42	43.73	43.05	58.88	14.43

Camera 2 - Indoor scene Zhang's calibration results (checkerboard used) [114]: $F_u = 1097$, $F_v = -1002$, $U_0 = 365.96$, $V_0 = 234.5$												
#	Pan	Tilt	PT Pan	PT Tilt	f_u	f_v	u_0	v_0	δ_{f_u}	δ_{f_v}	δ_{u_0}	δ_{v_0}
1	0.5625°	-0.675°	0.7875°	-0.675°	1151.02	-857.33	266.48	322.84	53.81	145.18	99.49	88.26
2	-1.8°	2.025°	-1.575°	1.8°	1082.42	-962.19	294.99	269.11	14.79	40.32	70.97	34.53
3	3.0375°	-3.6°	2.8125°	-3.2625°	1060.35	-987.40	367.30	255.09	36.86	15.11	1.34	20.51
4	-4.6125°	-4.1625°	4.3875°	-4.6125°	1103.63	-968.62	86.05	449.70	6.42	33.89	279.91	215.12
5	-7.0875°	6.975°	-7.425°	7.0875°	1101.20	-988.23	338.94	217.83	3.98	14.28	27.02	16.75
6	-7.9875°	-7.425°	1.35°	-0.7875°	1052.45	-1004.32	320.45	265.73	44.76	1.81	45.51	31.15
7	9°	9.9°	-1.0125°	1.0125°	1074.06	-992.89	332.90	230.75	23.16	9.63	33.06	3.83
8	-6.975°	-5.9625°	4.3875°	-0.9°	1059.75	-1002.19	363.85	261.32	37.46	0.33	2.11	26.74
9	7.7625°	6.525°	-0.9°	5.175°	1063.63	-990.24	374.22	220.99	33.59	12.27	8.26	13.59

Camera 1 - Outdoor scene Zhang's calibration results (checkerboard used) [114]: $F_u = 1123$, $F_v = -1024$, $U_0 = 341.1$, $V_0 = 247.1$												
#	Pan	Tilt	PT Pan	PT Tilt	f_u	f_v	u_0	v_0	δ_{f_u}	δ_{f_v}	δ_{u_0}	δ_{v_0}
1	0.5625°	-0.675°	0.7875°	-0.675°	977.86	-1002.11	363.64	239.05	145.82	22.74	22.54	8.08
2	-1.8°	2.025°	-1.575°	1.8°	1066.52	-965.03	413.11	181.81	57.16	59.82	72.01	65.32
3	3.0375°	-3.6°	2.8125°	-3.2625°	1064.21	-954.54	336.49	222.01	59.47	70.31	4.61	25.12
4	-4.6125°	-4.1625°	4.3875°	-4.6125°	1057.19	-967.33	149.22	296.56	66.49	57.52	191.88	49.43
5	-7.0875°	6.975°	-7.425°	7.0875°	1053.42	-971.67	377.53	227.67	70.26	53.18	36.42	19.46
6	-7.9875°	-7.425°	1.35°	-0.7875°	1057.40	-963.15	327.79	248.48	66.28	61.70	13.31	1.35
7	9°	9.9°	-1.0125°	1.0125°	1061.47	-986.99	439.00	217.58	62.21	37.86	97.90	29.55
8	-6.975°	-5.9625°	4.3875°	-0.9°	1054.45	-971.12	293.22	278.69	69.23	53.72	47.88	31.56
9	7.7625°	6.525°	-0.9°	5.175°	1055.13	-974.99	415.87	230.18	68.55	49.86	74.77	16.95

Fig.5.6(a) and Fig.5.6(b) illustrate the influence of pixel noise on the estimation of focal length (Eq.5.23 and Eq.5.26). Also, Fig.5.6(c) and Fig.5.6(d) show how SAC estimates the coordinate of the principal point (Eq.5.32) in noisy conditions. Colors are distributed based on the rotation angle of the camera. Hence, the distribution of the colors reveals how noise affects the SAC equations. In fact, the high concentration of red, yellow, and orange points around the zero error line in Fig.5.6(a) to (d) reveals that when the angle of the camera rotation is not very small, SAC achieves low-error estimates for focal lengths. This corroborates the claim that very small camera rotations can cause results from the SAC formulations to have high errors.

5.3.4 Real Images

We studied the proposed SAC formulations on real images as well. We used a Canon VC-C50i PTZ camera that is able to freely rotate around the Y-axis (pan) and the X-axis (tilt). The camera can be controlled by an application called Automatic Camera Control (created by the first author) that uses a standard RS-232 serial communication to control the camera. Therefore, the required pan and tilt rotation angles can be set in a specific packet and then be written into the camera serial buffer to cause the camera to rotate based on the assigned rotation angles. A screenshot of the ACC application can be seen in Fig.5.7.

Using the above-mentioned procedure, we took 9 sequences of images from 3 different scenes for evaluating the proposed SAC formulations. Fig.5.8 shows 6 sequences of these scenes. All sequences were taken using a fixed zoom. For each scene, while keeping the zoom of the camera unchanged, another 30 images were acquired from various viewpoints of a checkerboard pattern. The ground truth for the intrinsic parameters were calculated by applying the method of Zhang [114] on the checkerboard images.

The performance of SAC formulations on the 9 sequences of real images taken from 2 cameras from 3 different scenes, is reported in Table 5.3. For every sequence, we only used the images in the sequence. For example, to calculate the focal length in the u direction of Sequence 1, we found the point correspondence using [29], [30] between the reference image (Fig.5.8(a)) and the image taken after the pan rotation of the camera (Fig.5.8(b)). Then, we used only one of the matched points that is closer to the center of the image. Although, we did not include the lens distortion parameter into the SAC formulation (because it creates non-linear equations), we decrease the inaccuracy of the focal length estimates by using a matched point that is closer to the center of the image. Thus, the results are less affected by the lens distortion. A similar procedure was adopted with the image taken after a tilt rotation of the camera (Fig.5.8(c)) for calculating the focal length in the v direction of Sequence 1. The location of the principal point was estimated by corresponding the reference image (Fig.5.8(a)) to the image taken after panning and tilting the camera (Fig.5.8(d)). To solve the linear system of equations of SAC (Eq.5.32), we used all of the matched points.

The errors reported by applying SAC on 9 different sequences of real images taken by 2 cameras from 3 different scenes in Table 5.3 show that in spite of the presence of various types of noise, such as angular uncertainties, point correspondence noise and lens distortion, focal lengths estimated by SAC are close to the results of the method of Zhang [114], except when the angles of rotations are very small ($< 2^\circ$). Also, we calibrated more than one camera to show that our method can be used in real scenarios with different cameras. As we discussed earlier, the SAC formulation for estimating the coordinates of the principal point is sensitive to angular noise. Therefore, we can see that sometimes the error in the principal point estimate is increased; e.g., in Sequences 3 and 4 of Table 5.3. One can

decrease this error by including more matched points (in Eq.5.32) taken after panning and tilting the camera by various angles.

In addition, as it can be seen in Fig.5.8 that we have not only tested SAC on images taken from indoor scenes but also on an outdoor scene. Since in outdoor scenes usually a camera has a greater distance to objects than a camera looking at an indoor scenes, we evaluated the SAC formula on an outdoor environment as well. The calibration results of SAC on this type of images are reported in the 3^{rd} sequence of Table 5.3. According to the reported results, the errors are almost similar to indoor scene errors. This shows that the SAC equations can be employed in applications where cameras look at different types of scenes.

In real scenarios, there are often various types noise. One way to reduce the effect of noise is to refine the estimations. In order to do this, one can estimate camera parameters by SAC across several sequences with fixed zoom as we did in this study for 9 sequences. Then, we can form 4 separate vectors consisting of the estimated focal lengths and principal point locations. For example, all 9 estimated f_u of a scene can be put into a vector. Next, through a process of first removing the outliers from each vector and then taking an average from the output of outlier removal, a better estimate for the camera parameters can be found. This process helps remove noisy estimates by treating them as outliers.

5.4 Conclusion

In this chapter we presented a new Simplified Active Calibration formulation. Our derivations provided closed-form and linear equations to estimate the parameters of a camera using three image pairs taken before and after panning, tilting, and panning-tilting the camera.

A basic assumption about the rotation of a fixed camera was made; i.e., to solve the proposed equations, knowing the rotation angles of the camera is necessary. The proposed formulation can be used in practical applications such as surveillance, because in PTZ cameras accessing the camera motion information is straightforward.

The proposed closed-form formulations for estimating the focal lengths can be solved with only one point correspondence. Finding the correspondence point is straightforward. Following recent developments in feature extractors, one can extract repeatable regions from a pair of images. This is especially useful for applications that favor no point correspondences but instead in contour-to-contour correspondences that was done in the original Active Calibration [9]; where instead of the reference and transferred points in Eq.5.23 and Eq.5.26, the average of contour points or the centroid of the region can be used.

The results of solving our proposed formulations on randomly simulated 3D scenes indicated a very low error rate in estimating the focal lengths and the principal point location in ideal conditions. We evaluated our proposed SAC formulation for two different noise

conditions, namely angular and pixel noise. The simulated results showed that if the absolute value of the rotation angle is bounded from above by approximately 7° and from below by approximately 2°, the error caused by using the SAC formulation is low. This conclusion was later verified in our experiment with real images. Our future work will focus on including non-linear parameters into the Simplified Active Calibration equations and use the result of the current study as a close initial guess for an optimization procedure.

Chapter 6

Segmentation of Arterial Walls in Intravascular Ultrasound Cross-Sectional Images Using Extremal Region Selection

Abstract

Intravascular Ultrasound (IVUS) is an intra-operative imaging modality that facilitates observing and appraising the vessel wall structure of the human coronary arteries. Segmentation of arterial wall boundaries from the IVUS images is not only crucial for quantitative analysis of the vessel walls and plaque characteristics, but is also necessary for generating 3D reconstructed models of the artery. The aim of this study is twofold. First, we investigate the feasibility of using a recently proposed region detector, namely Extremal Region of Extremum Level (EREL), to delineate the luminal and media-adventitia borders in IVUS frames acquired by 20-MHz probes. Second, we propose a region selection strategy to label two ERELs as lumen and media based on the stability of their textural information. We extensively evaluated our selection strategy on the test set of a standard publicly available dataset containing 326 IVUS B-mode images. We showed that in the best case, the average Hausdorff Distances (HD) between the extracted ERELs and the actual lumen and media were 0.22 mm and 0.45 mm, respectively. The results of our experiments revealed that our selection strategy was able to segment the lumen with ≤ 0.3 mm HD to the gold standard even though the images contained major artifacts such as bifurcations, shadows, and side branches. Moreover, when there was no artifact, our proposed method was able to delineate media-adventitia boundaries with 0.31 mm HD to the gold standard. Furthermore, our proposed segmentation method runs in time that is linear in the number of pixels in each frame. Based on the results of this work, by using a 20-MHz IVUS probe with controlled pullback, not only can we now analyze the internal structure of human arteries more accurately, but also segment each frame during the pullback procedure because of the low run time of our proposed segmentation method.

6.1 Introduction

Catheter-based Intravascular Ultrasound (IVUS) has captured considerable attention in the last two decades. This worldwide attention is mostly due to the ability of the imaging method to picture the inside of the human coronary arteries and, hence, provide an opportunity to diagnose and treat cardiovascular diseases such as atherosclerosis (e.g., thin-cap fibroatheroma) that causes a heart attack and a brain stroke [39]. Aside from this, the IVUS technique can be helpful in visualizing some internal structures of the human coronary such as the lumen, and thickness and distribution of the plaques [64]. Therefore, IVUS is regularly used to locate the atherosclerosis lesions in the coronary arteries to study the lumen and plaque dimensions, and to guide intervention and stent deployment [71].

A typical IVUS imaging system consists of four parts: catheter, transducer, pullback device, and scanning console. The catheter is composed of a 150 cm long guidewire and a tip of 1.2-1.5 mm in size. It is usually inserted in the femoral artery and proceeds toward the coronary arteries. The catheter is responsible for carrying the ultrasound transducer, or other necessary devices, such as inflatable balloons and stents [64]. The transducer is a miniaturized ultrasound probe that emits ultrasound pulses and listens for the backscattered signal. After the catheter has reached the distal end of the coronary, it needs to be manually or automatically pulled back. The speed of the pullback varies between 0.5-1 mm/s [64]. The scanning console is essentially a computer used to post-process the acquired signals (using amplification, filtering, etc.) to provide a user-friendly environment for the surgeon to control the device.

Segmentation of the acquired IVUS images is among the most challenging tasks in medical image analysis. In particular, delineating the interior (lumen) and exterior (media) vessel walls is problematic due to the presence of various artifacts such as motion of the catheter after a heart contraction, guide wire effects, bifurcation and side-branches or similar echogenicity between the vessel wall and some plaques. In some cases even the difference in transducer frequencies affect the segmentation results [64].

The intrinsic difficulty of IVUS segmentation has attracted many researchers to study and develop solutions using different methodologies, such as intensity-based, statistics and probability-based, active contour and graph search-based approaches. In addition, several methods have been proposed to segment either the lumen or the media or both. A great number of approaches in the literature have utilized the 2D information provided as cross-sectional frames to segment the lumen and media. These 2D cross-sectional gray-scale images are formed after digitization of the backscattered RF signals and are called IVUS B-mode frames. To the best of our knowledge, recent approaches have mostly worked on the B-mode frames which will be reviewed in the following paragraph. For more in-depth reviews of the methods published before 2013, please refer to [4], [64].

The lumen and media-adventitia border variations have been modeled within a shape space in [97]. The lumen segmentation is then performed by maximizing a nonparametric probability density energy. Also, the edge information has been used to segment the media-adventitia. A physics-based model of the IVUS signal scattered by the structure of the vessel has been used in [74] to estimate the differential backscattering cross-sections from the IVUS RF signal. The segmentation curve is obtained after training a Support Vector Machine (SVM) model using the annotated data. Deformable models have been used in [73], [94], [117] to detect the border of lumen/intima and media/adventitia. In [94] anisotropic diffusion followed by an edge detector are used to create an initial segmentation which is then corrected using both geometric and parametric deformable models. In [117] an improved version of gradient vector flow (iGVF) has been proposed which includes a balloon force in the snake model that lets the contour pass over leaks and bifurcations. A probabilistic method that formulates the deformation of a lumen contour curve and can be minimized has been proposed in [73]. However, every first frame of the sequence needs user interactions to manually segment the lumen and media. Following this, a SVM is trained over the annotated data to compute the probability that each pixel is blood or non-blood. A fourfold algorithm based on a deterministic statistical strategy for segmenting the media has been proposed in [113]. Their method consists of preprocessing, initial contour detection, active contour segmentation, and contour refinement. First, a sparse binary image is constructed using the local appearance model and the initial contour is elicited. To achieve this, a feature vector is built for each pixel. It includes gray-level values of the pixel's neighbours, the average intensity of neighbours and the gray-level values of the pixel's neighbours in a contour-enhanced version of the image[113]. The K-SVD method is utilized to classify the extracted feature vector. The initial contour is then refined. Next, an active contour model is used to delineate the media border in polar coordinates. The detected contours are then refined using the information provided by identifying the calcification and shadow regions. Artificial Neural Networks have been employed in [90] to represent the spatial and neighbouring features of the IVUS image data. As a result, two different vascular structures for lumen and media are extracted and optimized using two ANNs. The borders obtained are then refined and smoothed by an active contour model. In [110] the lumen is segmented by a combination of image gradient and fuzzy connectedness model and the media-adventitia border is extracted by a fast marching model. A sequential forward selection process using SVMs and PR curve has been employed to conduct an in-depth analysis of several image features in [98]. It has been shown that the median filtered image and Haralick's texture features [45] provide stronger discrimination capabilities for arterial structures. A limitation of their analysis is that it only works for artifact-free IVUS sequences.

As we can see, most approaches have employed either a type of energy minimization

method or require annotated data in order to train a classifier or an ANN. However, in this paper we propose a straightforward approach that not only does not require training but also does not use any variational method or deformable model. We show that by extracting EREL features [29], [30] the problem of the IVUS segmentation can be relaxed to a region selection. In particular, we illustrate that it is very likely to find regions similar to lumen and media among the extracted ERELs. Therefore, we propose a selection procedure that efficiently chooses ERELs that are most similar to lumen and media.

The rest of this paper is organized as follows. In Section 6.2 we present our proposed method and describe the sequence of Intravascular Ultrasound images that we use throughout this paper. Section 6.3 illustrates the segmentation results of our proposed method. In Section 6.4 we discuss the advantages and weaknesses of the proposed method. Finally, the concluding remarks are given in Section 6.5.

6.2 Materials and Method

6.2.1 Materials

Our proposed method has been evaluated on the test set of a publicly available dataset consisting of 326 in-vivo pullbacks of the human coronary artery frames that were acquired by the Si5 (Volcano Corporation), equipped with a 20-MHz Eagle Eye monorail catheter [4]. The dataset includes a multi-frame 3D context that has between 20 to 50 gated frames acquired using a full pullback at the end-diastolic cardiac phase from 10 patients. Manual annotations for IVUS images are available in the dataset. The annotations have been provided by four clinical experts who work regularly with IVUS echograph. The experts were not aware of other expert's manual annotations and two of them repeated the annotations after about one week from their first delineation [4].

The test set contains several types of common artifacts. Specifically, it includes 44 images containing bifurcation, 93 images with a side vessel artifact, and 96 images that have been contaminated by a shadow artifact (some images contain more than one artifact). There are also 143 images that do not contain any serious artifacts except for plaque.

6.2.2 Proposed Method

In this paper we present a segmentation approach for 20 MHz Intravascular Ultrasound images based on a region detection strategy. Particularly, we investigate whether a recently proposed novel feature extraction method called Extremal Regions of Extremum Levels (EREL) [29], [30] can segment the most essential regions of interest (lumen and media) from the IVUS images required to establish the atherosclerotic plaque area [19]. The proposed method consists of four steps. We first remove the typical artifacts of IVUS frames, such as ring-down effects and calibration squares. Then ERELs are extracted and the obtained

regions are filtered based on their types. Next, we perform a region selection procedure to specify two regions as lumen and media. Finally, the contour of the two selected regions is traced and smoothed by an ellipse fitting algorithm.

6.2.3 Preprocessing

Most of the IVUS images have been contaminated by speckle noise[4]. Speckle is a multiplicative noise that imposes difficulties in processing the Ultrasound images [70]. Therefore, to decrease the sensibility of our method to speckle noise, we first use a non-linear median operator to filter the IVUS images.

One of the main identifiable artifacts in IVUS images is the ring-down effects of the catheter that need to be eliminated from the B-mode frame or from its polar image. Otherwise, there is a high risk of obtaining an erroneous segmentation. To remove the ring-down effects of the catheter we employ the method proposed in [97] which is a very fast and straightforward procedure. Detecting the ring-down artifact can be done by processing the whole volume since the artifact is almost available in all of the IVUS frames. Therefore, taking the minimum over all the frames generates an image where there is a significant contrast between the artifact and the non-artifact pixels.

$$I_{min}(x, y) = \min_{i \in \lambda} I_i(x, y) \tag{6.1}$$

where λ is a set of available frames in a particular IVUS sequence. We can then locate the artifacts' coordinates by subtracting the artifact zones from every frame, as in [97]. Figure 6.1(b) and Figure 6.1(f) show the resulted minimum image in B-mode and polar frame, respectively.

Another type of artifact that we can detect using Eq. 6.1 is the calibration square artifact. These small squares have a very bright constant intensity in all frames that remains bright in the minimum image. In several longitudinal cuts of the IVUS volume the effects of these artifacts are revealed as horizontal lines. Figure 6.1 illustrates both the artifacts and the resulting image after removing them.

6.2.4 Extremal Regions of Extremum Levels

EREL [29], [30] is a region detector that employs a union-find structure [84] in conjunction with the edge information to detect a series of connected pixels from the image. The edge information of the image is included in the method by using the Maxima of Gradient Magnitude (MGM) points. The idea underlying EREL is to binarize the image with all possible integer thresholds and analyze the results obtained based on their global criterion and their local edge information. The regions belonging to the globally distinguished levels (Extremum Levels) are then extracted from the union-find tree.

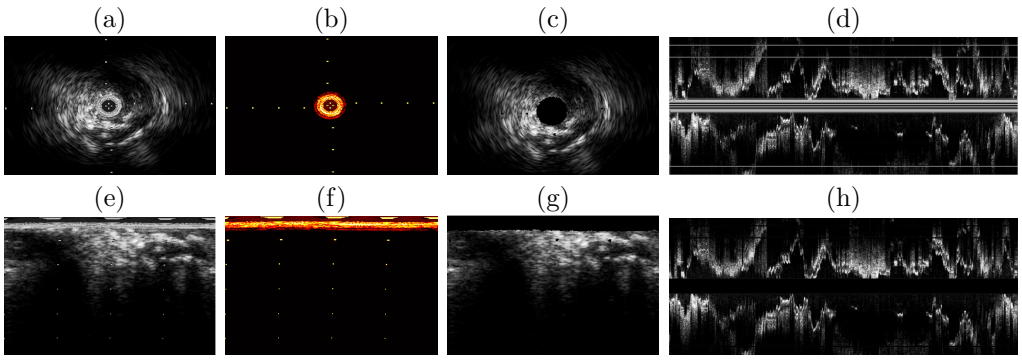


Figure 6.1: Artifact removal in B-mode and polar frames. a) A 40 MHz IVUS B-mode frame. b) Computed minimum image of (a). The yellow colour demonstrates higher values and the red colour represents lower values. c) Result of the IVUS frame shown in (a) after the artifact removal. d) A longitudinal cut of the whole volume. The horizontal lines are the effects of the artifact revealed after cutting. e) Corresponding polar frame of (a). f) Calculated minimum image of (e). The yellow colour demonstrates higher values and the red colour represents lower values. g) Corresponding polar frame of (e) after artifact removal. h) Result of artifact removal in all frames of the volume illustrated in a longitudinal view that is cut by the same plane as the one used to cut (d).

Generally, two types of regions can be extracted from a gray-level image. The first type includes regions that evolve from brighter surfaces to darker boundaries which are identified by Q^-. The superscript '$-$' emphasizes the fact that the intensity values are decreased from the surface of the regions towards the boundaries. The second type consists of regions that evolve from darker surfaces to brighter boundaries and are denoted by Q^+. This type of the region is consistent with the inherent characteristics of the lumen and media visualized by the backscattered 20-MHz IVUS signals. Therefore, we only need to extract Q^+ regions to obtain the lumen and media because both regions evolve from darker surfaces to brighter boundaries.

To use EREL, we need to set several initialization parameters, namely A_{min}, A_{max}, α, β. These parameters define the functionality of the detector and can be tuned based on the application [29], [30]. In particular, we use A_{min}, A_{max} to set the minimum and maximum area of the extracted regions. To better separate small regions from bigger ones, we choose a value for A_{min} that correlates with image dimensions. Specifically, we set $A_{min} = (R \times C)/100$ and $A_{max} = (R \times C)/3$ where R and C represent the number of rows and columns of the IVUS image, respectively. The parameter α is usually in [0 2.5] and represents the strength of the resulting interest points [29]. In this study we set $\alpha = 0.5$. Finally, $\beta = 1$ denotes the width of the moving window over the global criterion vector [29]. The extracted EREL regions using the above-mentioned values are illustrated in Figure 6.2.

However, not all four types of the regions depicted in Figure 6.2 encompass lumen and media regions. In fact, we only need to extract large area Q^+ regions as illustrated in Figure 6.2(d). Since large area Q^+ regions contain the actual lumen and media segments,

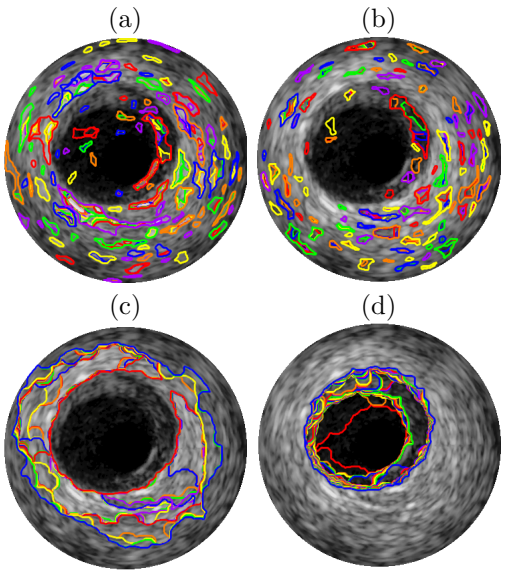


Figure 6.2: Extracted ERELs from a 20-MHz IVUS frame belonging to dataset [4]. The initial parameters of EREL are: $\alpha = 0.5$, $\beta = 1$, $A_{min} = (R \times C)/100 = 1474$ and $A_{max} = (R \times C)/3 = 49152$. a) Q^- regions with small area. b) Q^+ regions with small area. c) Q^- regions with large area. d) Q^+ regions with large area. Contour colours have been randomly assigned and are only for visualization purposes.

detectors need not track Q^- regions and, therefore, omit unnecessary computations which eventually helps to have a faster detector.

6.2.5 EREL Selection

The goal of this section is to address the problem of finding the most appropriate ERELs to be designated as lumen and media. By the most appropriate we mean the closest regions to the gold standard. As can be seen in Figure 6.2(d), although ERELs are nested regions, it is clear that there is at least one EREL that is very close to the true lumen and similarly there is at least one EREL that corresponds to the true media. Therefore, we can relax the problem of lumen and media segmentation to only a selection procedure, i.e., assigning two nested ERELs to lumen and media.

Local maxima searching is the approach that we employ to select lumen and media from the nested set of ERELs. Assume that we have a vector (denoted by V) representing the evolution of the Q^+ regions. The index of each element of the vector V corresponds to a Q^+ region. Since the extracted ERELs start from the smallest enclosing region and end with the largest enclosing region, the actual region corresponding to the lumen should be found among the regions located at the early indices of the vector. Likewise, a region representing the media is expected to be found among the regions belonging to the end section of the vector V.

Our aim is to construct a vector that ultimately gives us the stability of the regions in

terms of the length of the boundary, average intensity and entropy variation. As we can see in Figure 6.2(d), the boundaries of the Q^+ regions are not smooth and are subject to large variations. Therefore, lumen and media regions should be selected among those Q^+ regions that have more stable boundary length variations. Also, the average intensity of the Q^+ regions should be stable enough (i.e., they should not change much over several subsequent regions). Entropy measure can be used in order to create a feature vector that is sensitive to textural information of the regions.

The calculation of these three features are straightforward. The boundary lengths of the regions are available as an output of the EREL algorithm and are calculated based on a bottom-up tracking of boundary pixels along a parametric curve $C(p)$.

$$\mathcal{L} = \int_0^1 |\frac{\partial C(p)}{\partial p}| dp \tag{6.2}$$

Additionally, the calculation of the average intensity of ERELs can be readily done.

$$E = \frac{\sum_{i=1}^{N} Q_i^+}{N} \tag{6.3}$$

where N represents the area of the region and Q_i^+ is the intensity value of pixel i. The entropy measure of a grayscale region [45] is denoted as follows.

$$\mathcal{H} = -\sum_{i}^{K} p_i \log_2 p_i \tag{6.4}$$

where p_i is the value of the bin i of the normalized histogram of the region and captures the probability of having a pixel with a certain gray-value. K is the number of available bins in the normalized histogram.

Afterwards, for every IVUS image, we create a vector (V) where each element is obtained from the product of the above-mentioned measures for each region.

$$V = \begin{bmatrix} \mathcal{L}_1 E_1 \mathcal{H}_1 & \mathcal{L}_2 E_2 \mathcal{H}_2 & \dots & \mathcal{L}_n E_n \mathcal{H}_n \end{bmatrix} \tag{6.5}$$

where n is the number of Q^+ regions extracted by the EREL algorithm.

Vector V illustrates the variation of textural information of regions through the extraction and evolution of Q^+ regions. It is strictly increasing because the Q^+ regions are nested and non-repetitive. More stable sequences of vector V are more likely to represent lumen and media since these stable sequences shows that the extremal regions are subject to saturation and the subsequent regions might contain a noticeable change. The best way to find the stable regions is to create a vector describing the stability score of regions. Every element of the stability vector is calculated as follows.

$$\Omega_i = \frac{V_i}{V_{i+1} - V_{i-1}} \tag{6.6}$$

where i specifies a specific element of the vector V and varies from one to the number of the detected Q^+ regions for an IVUS image.

The local maxima of the stability score points to regions with high stability because the ratio of their current value to the change among their two neighbours is larger than the other surrounding elements. So, we select lumen and media from the detected local maxima. Specifically, a Q^+ region with higher prominence value among the first two peaks is considered as lumen. If the IVUS image contains no artifacts, the media will be represented by the last detected peak. Based on our observation, the stability score of the images that contain serious artifacts have none or a small number of peaks since the presence of the artifacts interferes with the natural extraction of Q^+ regions (see Figure 6.3(ii)). Therefore, when a small number of local maxima is detected, we consider the last extracted region as media. This process for an IVUS with no particular artifact is illustrated in Figure 6.3. The local maxima of the stability score indicates the regions for which the variation of the textural characteristics is more stable than their surrounding regions. As can be seen in Figure 6.3(i)(a), the second peak is selected as a suitable region for the lumen since it has a higher prominence than the first peak. Also, the region corresponding to the last peak of Ω which is chosen as the media has been shown in Figure 6.3(i)(b).

It is important to note that before searching for the local maxima we need to remove outliers. To find outliers we employ the modified Z-score normalization suggested in [57].

$$M_i = \frac{0.6745(A_i - \tilde{A})}{MAD} \tag{6.7}$$

where A_i represents the area of the region i, $\tilde{A}$ is the median of a vector of all region areas, and MAD denotes the Median Absolute Deviation and is calculated by the following equation.

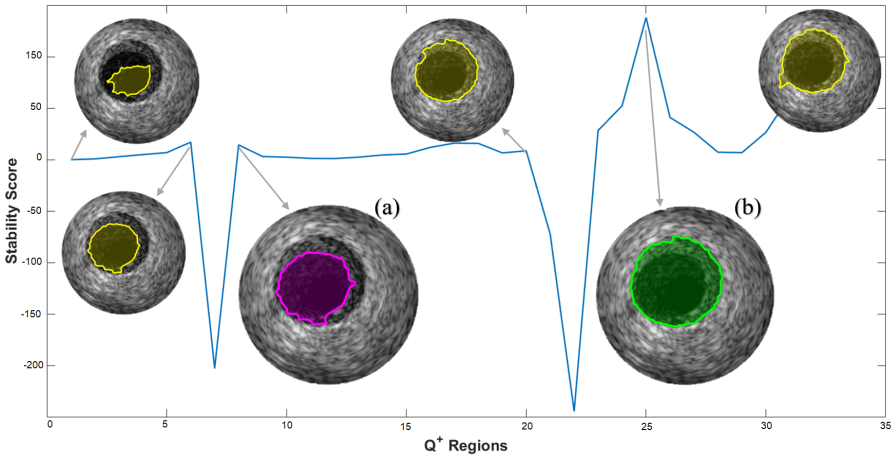
$$MAD = median(|A_i - \tilde{A}|) \tag{6.8}$$

Therefore, the regions that have a modified Z-score less than $Z_{min} = -3$ and greater than $Z_{max} = 3$ are unlikely to represent lumen and media and hence can be removed from the selection process.

6.2.6 Contour Extraction

Since the general shape of the lumen and media regions of the vessel are very similar to conic sections, we propose to represent lumen and media by ellipses. To find the pixels inside and on the ellipse border, it is sufficient to find the orientation of the ellipse, the major and minor axis length. Generally, EREL outputs all pixels belonging to each extracted region in addition to its shape description parameters which are three coefficients of ellipse equation [29], [30]. Specifically, if a region Q is described by an ellipse with three coefficients, namely

(i) No Artifact



(ii) Plaque and Shadow

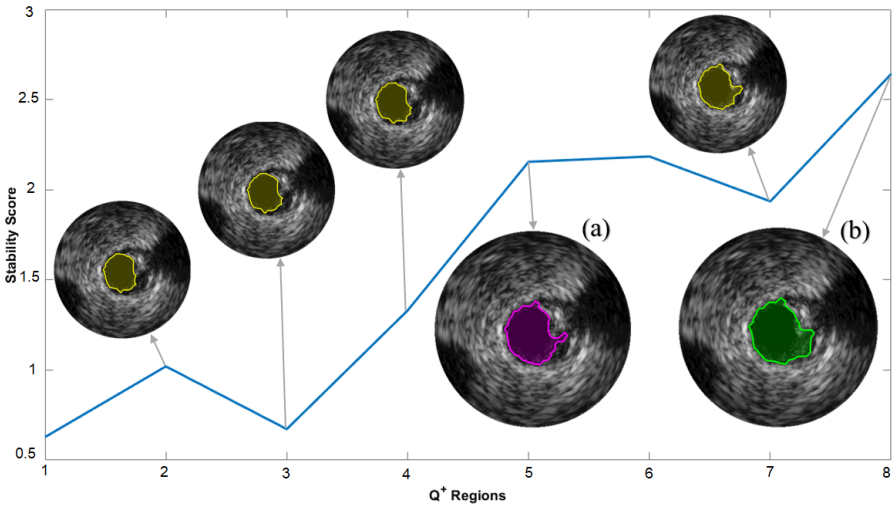


Figure 6.3: The evolution of the Q^+ regions and their stability criteria in the absence of artifacts vs. plaque and shadow artifact. a) The best candidate region representing the lumen. b) The best region representing the media. The neglected regions are highlighted by the yellow colour and the selected regions for lumen and media are indicated by magenta and green colours.

c,d and e then:

$$\forall x, y \in Q : cx^2 + 2dxy + ey^2 \leq 1 \tag{6.9}$$

There is a direct relationship between the parameters of the fitted ellipse and the second central moments [85]:

$$M = \begin{bmatrix} c & d \\ d & e \end{bmatrix} = \frac{1}{4(\mu_{yy}\mu_{xy} - \mu_{xx}^2)} \begin{bmatrix} \mu_{yy} & \mu_{xy} \\ \mu_{xy} & \mu_{xx} \end{bmatrix} \tag{6.10}$$

where μ_{xx}, μ_{xy} and μ_{yy} are the second order central moments and are calculated as follows [85]:

$$\mu_{xx} = \frac{1}{A} \sum_{(x,y)\in Q} (x - \bar{x})^2 \tag{6.11}$$

$$\mu_{yy} = \frac{1}{A} \sum_{(x,y)\in Q} (y - \bar{y})^2 \tag{6.12}$$

$$\mu_{xy} = \frac{1}{A} \sum_{(x,y)\in Q} (x - \bar{x})(y - \bar{y}) \tag{6.13}$$

where A represents the area of the region and $(\bar{x}, \bar{y})$ specifies the coordinates of the region's centroid.

After having obtained matrix M (reported by EREL), finding pixels belonging to ellipse border of the regions is straightforward. We assume that $M = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ where $\mathbf{V}$ denotes a matrix containing the eigenvectors and $\mathbf{\Lambda}$ indicates a matrix of eigenvalues. The minimum (λ_{min}) and maximum (λ_{max}) eigenvalues in $\mathbf{\Lambda}$ are used to calculate the length of the minor(b) and major (a) axes of the ellipse, respectively [93].

$$a = \frac{1}{\sqrt{\lambda_{max}}} \tag{6.14}$$

$$b = \frac{1}{\sqrt{\lambda_{min}}} \tag{6.15}$$

The orientation of the ellipse (θ) can be recovered by calculating the angle between the major axis and the x axis. Let $\mathbf{V} = [\mathbf{u}\ \mathbf{v}]$ where $\mathbf{u} = [u_1\ u_2]^T$ and $\mathbf{v} = [v_1\ v_2]^T$ are two column vectors corresponding to the minimum and maximum eigenvalues, respectively. The orientation of the ellipse is then obtained as follows.

$$\theta = \tan^{-1} \frac{v_1}{v_2} \tag{6.16}$$

6.2.7 Computational Cost

The run time of the proposed method depends only on the complexity of EREL, which is $O(N)$ [29], where N is the total number of pixels in the image. The subsequent operation proposed in this paper in Section 6.2.5 works on a constant number of regions. The maximum number of nested regions that can be extracted from the same root pixel is at most 256 for an 8-bit image [29]. In our experiments with IVUS images, the number of candidate ERELs

(Q^+) rooted from the center of image are even less (lower than 75). Therefore, the region properties denoted in Eq. 6.2, Eq. 6.3, Eq. 6.4 and Eq. 6.6 and the second moment matrix are calculated on a constant number of regions containing far less than N pixels. Therefore, the overall run time of the proposed method is $O(N)$ in the worst case. A comparison between the actual run time of the proposed method and the methods reported in [4] is shown in Table 6.1.

Table 6.1: Comparison of each method's run time (required for segmenting a frame) reported in [4] and the proposed method.

	Category	**Semi/Auto**	**2D/3D**	**Time per frame**	**Hardware used**
Proposed Method	Lumen and media	AUTO	2D	0.19 s	Core i7-4700HQ, 2.4 GHz
Unal et al. [97]	Lumen and media	SEMI	2D	3.25 s	Pentium 6200 2.13 GHz
Wang et al.[4]	Lumen	SEMI	2D	1 m 40 s	Xeon 2.67 GHz
Destrempes et al.[4]	Lumen and media	SEMI	2D	8.64 a	Core i7 Q740 @ 1.73 GHz
Downe et al. [21]	Lumen and media	AUTO	3D	0.16 s	Core 2, 2.4 GHz
Alberti et al.[4]	Lumen	AUTO	3D	13 s	Core 2, Duo 2.13 GHz
Ciompi et al. [18]	Media	AUTO	2D	20 s	Core i7, 2.8 GHz
Mendizabal et al. [73]	Lumen	SEMI	2D	4.96 s	Core i7, 2 GHz
Exarchos et al. [4]	Lumen and media	AUTO	2D	0.5 s	Core 2, Duo 3.33 GHz

6.3 Results

In this section, we present the segmentation results of our method. Also, by showing the best case results, we demonstrate that regions very close to the lumen and media exist among the extracted ERELs and a proper selection strategy (the proposed method) can distinctively select the lumen and media regions from the extracted ERELs. Furthermore, we present extensive evaluation results of our method based on three standard evaluation metrics on IVUS frames containing various artifacts.

6.3.1 Evaluation Measures

To assess the segmentation obtained by our method, we employ three evaluation metrics, namely Jaccard Measure (JM), Hausdorff Distance (HD), and Percentage of Area Difference (PAD). Using these metrics that have been also used in [4] to evaluate the results of 8 state-of-the-art methods on the same dataset makes it possible to draw a fair comparison between our method and the results reported in [4].

The Jaccard Measure is calculated based on the comparison of the automatic segmentation result and the manual segmentation delineated by experts. It quantifies the overlap area between the automatic and manual segmentation as computed by the following equation.

$$JM = \frac{R_{auto} \cap R_{man}}{R_{auto} \cup R_{man}} \tag{6.17}$$

Table 6.2: The best case performance results of the proposed method. Measures represent the mean and standard deviation (std) evaluated on 435 frames of dataset [4]. The measures are categorized based on the presence of a specific artifact in each frame. The evaluation measures are Jaccard Measure (JM), Hausdorff Distance (HD), and Percentage of Area Difference (PAD).

		Lumen			**Media**		
		HD	**JM**	**PAD**	**HD**	**JM**	**Pad**
General Performance	*EREL*	0.22 (0.12)	0.91 (0.04)	0.03 (0.03)	0.50 (0.45)	0.83 (0.15)	0.13 (0.15)
	Intra-obs	0.28 (0.13)	0.88 (0.05)	0.11 (0.08)	0.24 (0.12)	0.92 (0.03)	0.06 (0.04)
	Inter-obs	0.17 (0.13)	0.93 (0.05)	0.04 (0.06)	0.14 (0.12)	0.95 (0.03)	0.03 (0.03)
No Artifact	*EREL*	0.20 (0.09)	0.92 (0.03)	0.02 (0.02)	0.23 (0.17)	0.92 (0.05)	0.03 (0.05)
	Intra-obs	0.28 (0.13)	0.88 (0.05)	0.11 (0.08)	0.24 (0.12)	0.92 (0.03)	0.06 (0.04)
	Inter-obs	0.17 (0.13)	0.93 (0.05)	0.04 (0.06)	0.14 (0.12)	0.95 (0.03)	0.03 (0.03)
Bifurcation	*EREL*	0.36 (0.22)	0.85 (0.07)	0.07 (0.07)	0.95 (0.49)	0.67 (0.16)	0.27 (0.17)
	Intra-obs	0.30 (0.12)	0.88 (0.04)	0.09 (0.06)	0.24 (0.09)	0.92 (0.02)	0.06 (0.03)
	Inter-obs	0.18 (0.21)	0.92 (0.07)	0.05 (0.09)	0.15 (0.13)	0.95 (0.04)	0.03 (0.03)
Side Vessels	*EREL*	0.20 (0.09)	0.90 (0.04)	0.03 (0.03)	0.62 (0.53)	0.78 (0.16)	0.18 (0.16)
	Intra-obs	0.30 (0.13)	0.88 (0.05)	0.10 (0.08)	0.24 (0.11)	0.92 (0.04)	0.06 (0.04)
	Inter-obs	0.20 (0.11)	0.91 (0.05)	0.06 (0.05)	0.15 (0.10)	0.95 (0.03)	0.03 (0.04)
Shadow	*EREL*	0.22 (0.11)	0.89 (0.04)	0.03 (0.03)	1.10 (0.38)	0.64 (0.13)	0.31 (0.14)
	Intra-obs	0.31 (0.13)	0.88 (0.05)	0.11 (0.08)	0.27 (0.15)	0.92 (0.04)	0.06 (0.05)
	Inter-obs	0.18 (0.14)	0.93 (0.05)	0.04 (0.06)	0.14 (0.10)	0.96 (0.03)	0.02 (0.02)

where R_{auto} is the vessel region segmented by the method and R_{man} represents the region that has been segmented manually by experts.

The Hausdorff Distance between the automatic (C_{auto}) and manual (C_{man}) curves is the greatest distance of all points belonging to C_{auto} to the closest point in C_{man} and is defined as follows [75].

$$HD = \max\{d(C_{man}, C_{auto}), d(C_{auto}, C_{man})\} \tag{6.18}$$

To calculate $d(C_{auto}, C_{man})$ first the minimum of all Euclidean distances from each point belonging to C_{auto} to all points in C_{man} is obtained. Then, $d(C_{auto}, C_{man})$ is computed by taking the maximum of all the minimum distances. Similarly, $d(C_{man}, C_{auto})$ is computed by taking the maximum of all minimum distances from C_{man} to C_{auto} [75].

The Percentage of Area Difference calculates the segmentation area difference between the automatic (A_{auto}) and manual (A_{man}) segmentation and is computed as follows.

$$PAD = \frac{|A_{auto} - A_{man}|}{A_{man}} \tag{6.19}$$

6.3.2 Best Case Results

In order to show that the extracted EREL regions have the potential to represent lumen and media regions, we evaluate all of the extracted ERELs by calculating the evaluation metrics (Section 6.3.1) for the contours of each EREL and the manually annotated contours. Then, the ERELs that correspond to the maximum JM of lumen and media are selected as the best extracted ERELs for that frame. The quantitative results in comparison with the intra-observer and inter-observer variability are reported in Table 6.2.

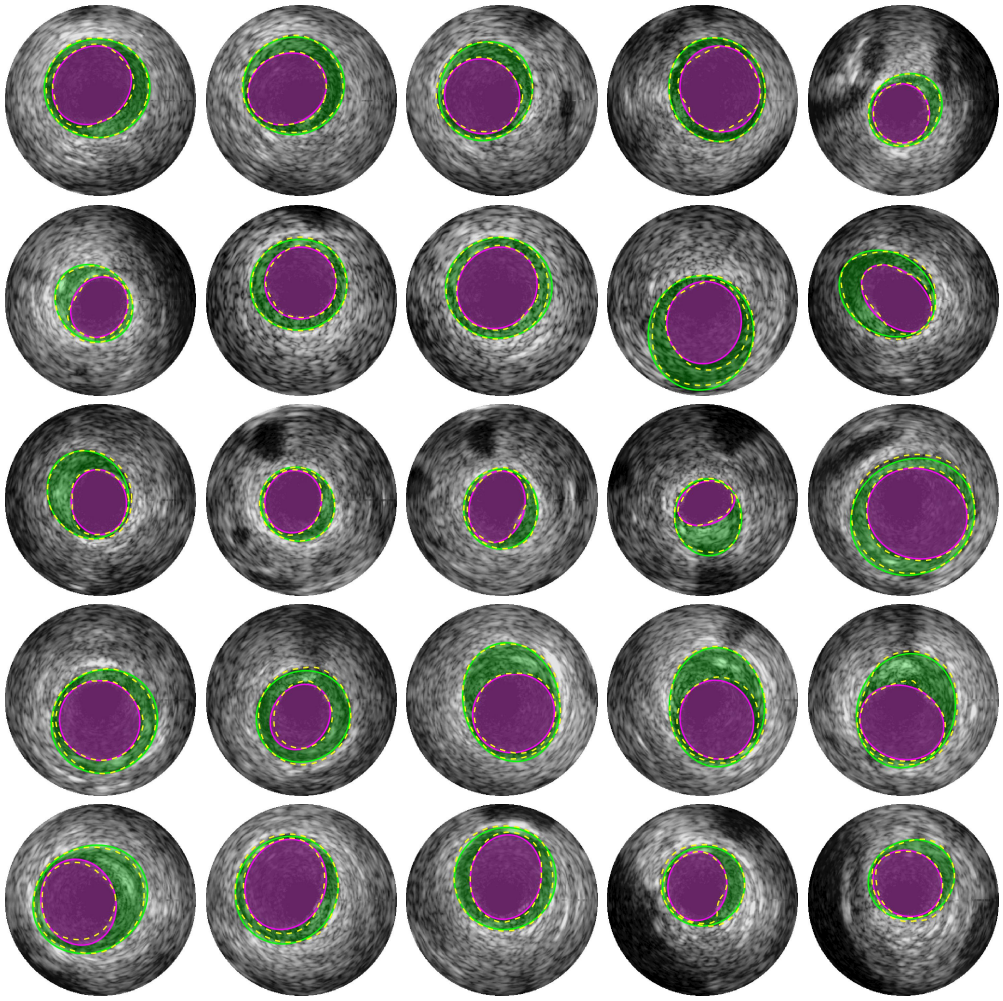


Figure 6.4: Lumen and media segmentation results. Segmented lumen and media have been highlighted by magenta and green colours, respectively. The yellow dashed lines illustrate the gold standard that have been delineated by four clinical experts [4].

Table 6.3: Performance of the proposed EREL selection strategy. Measures represent the mean and standard deviation evaluated on 435 frames of dataset B [4] and categorized based on the presence of a specific artifact in each frame. The evaluation measures are Jaccard Measure (JM), Hausdorff Distance (HD), and Percentage of Area Difference (PAD).

		Lumen			**Media**		
		HD	**JM**	**PAD**	**HD**	**JM**	**Pad**
General Performance	Proposed Method	0.30 (0.20)	0.87 (0.06)	0.08 (0.09)	0.67 (0.54)	0.77 (0.17)	0.19 (0.18)
	Unal et al. [97]	0.47 (0.39)	0.81 (0.12)	0.14 (0.13)	0.64 (0.48)	0.76 (0.13)	0.21 (0.16)
	Wang et al.[4]	0.51 (0.25)	0.83 (0.08)	0.14 (0.12)	–	–	–
	Destrempes et al.[4]	0.34 (0.14)	0.88 (0.05)	0.06 (0.05)	0.31 (0.12)	0.91 (0.04)	0.05 (0.04)
	Downe et al. [21]	0.47 (0.22)	0.77 (0.09)	0.15 (0.12)	0.76 (0.48)	0.74 (0.17)	0.23 (0.19)
	Alberti et al.[4]	0.46 (0.30)	0.79 (0.08)	0.16 (0.09)	–	–	–
	Ciompi et al. [18]	–	–	–	0.57 (0.39)	0.84 (0.10)	0.12 (0.12)
	Mendizabal et al. [73]	0.38 (0.26)	0.84 (0.08)	0.11 (0.12)	–	–	–
	Exarchos et al. [4]	0.42 (0.22)	0.81 (0.09)	0.11 (0.11)	0.60 (0.28)	0.79 (0.11)	0.19 (0.19)
No Artifact	Proposed Method	0.29 (0.17)	0.88 (0.05)	0.08 (0.07)	0.31 (0.23)	0.89 (0.07)	0.07 (0.08)
	Lo Vercio et al. [98]	–	0.83 (0.05)	0.18 (0.06)	–	–	–
Bifurcation	Proposed Method	0.53 (0.34)	0.79 (0.10)	0.15 (0.17)	1.22 (0.45)	0.57 (0.13)	0.32 (0.19)
	Unal et al. [97]	0.65 (0.47)	0.76 (0.14)	0.18 (0.15)	0.57 (0.49)	0.78 (0.13)	0.19 (0.15)
	Wang et al.[4]	0.54 (0.27)	0.81 (0.11)	0.14 (0.13)	–	–	–
	Destrempes et al.[4]	0.42 (0.18)	0.85 (0.06)	0.08 (0.06)	0.32 (0.13)	0.91 (0.03)	0.06 (0.04)
	Downe et al. [21]	0.64 (0.27)	0.70 (0.11)	0.21 (0.15)	0.79 (0.53)	0.71 (0.19)	0.24 (0.21)
	Alberti et al.[4]	0.61 (0.43)	0.75 (0.10)	0.20 (0.10)	–	–	–
	Ciompi et al. [18]	–	–	–	0.52 (0.29)	0.85 (0.07)	0.09 (0.07)
	Mendizabal et al. [73]	0.53 (0.36)	0.79 (0.12)	0.17 (0.18)	–	–	–
	Exarchos et al. [4]	0.47 (0.23)	0.80 (0.09)	0.10 (0.09)	0.63 (0.25)	0.78 (0.11)	0.23 (0.23)
Side Vessels	Proposed Method	0.24 (0.11)	0.87 (0.05)	0.06 (0.05)	0.74 (0.18)	0.73 (0.60)	0.21 (0.18)
	Unal et al. [97]	0.51 (0.39)	0.79 (0.12)	0.17 (0.14)	0.57 (0.39)	0.78 (0.11)	0.18 (0.12)
	Wang et al.[4]	0.59 (0.23)	0.80 (0.10)	0.16 (0.13)	–	–	–
	Destrempes et al.[4]	0.36 (0.15)	0.87 (0.04)	0.07 (0.04)	0.31 (0.12)	0.91 (0.04)	0.04 (0.04)
	Downe et al. [21]	0.46 (0.19)	0.77 (0.08)	0.15 (0.11)	0.76 (0.47)	0.74 (0.16)	0.22 (0.20)
	Alberti et al.[4]	0.47 (0.24)	0.79 (0.07)	0.17 (0.09)	–	–	–
	Ciompi et al. [18]	–	–	–	0.53 (0.37)	0.85 (0.09)	0.10 (0.13)
	Mendizabal et al. [73]	0.38 (0.19)	0.84 (0.07)	0.11 (0.11)	–	–	–
	Exarchos et al. [4]	0.53 (0.24)	0.77(0.09)	0.16 (0.12)	0.63 (0.31)	0.78 (0.12)	0.18 (0.16)
Shadow	Proposed Method	0.29 (0.20)	0.86 (0.07)	0.08 (0.09)	1.24 (0.39)	0.58 (0.13)	0.37 (0.15)
	Unal et al. [97]	0.57 (0.39)	0.78 (0.12)	0.17 (0.12)	0.66 (0.50)	0.77 (0.13)	0.19 (0.15)
	Wang et al.[4]	0.59 (0.27)	0.81 (0.10)	0.18 (0.16)	–	–	–
	Destrempes et al.[4]	0.39 (0.18)	0.87 (0.05)	0.06 (0.05)	0.33 (0.14)	0.92 (0.03)	0.05 (0.04)
	Downe et al. [21]	0.55 (0.26)	0.76 (0.11)	0.14 (0.13)	0.77 (0.48)	0.74 (0.16)	0.22 (0.19)
	Alberti et al.[4]	0.53 (0.29)	0.78 (0.08)	0.18 (0.09)	–	–	–
	Ciompi et al. [18]	–	–	–	0.58 (0.36)	0.84 (0.09)	0.11 (0.11)
	Mendizabal et al. [73]	0.43 (0.27)	0.83 (0.09)	0.12 (0.11)	–	–	–
	Exarchos et al. [4]	0.46 (0.19)	0.80 (0.10)	0.12 (0.12)	0.57 (0.28)	0.82 (0.11)	0.14 (0.17)

6.3.3 EREL Selection Results

Qualitative evaluations are illustrated in Figure 6.4 and show the successful segmentation results of the proposed EREL selection strategy for 20 IVUS frames. The lumen areas are highlighted by the magenta colour while the media regions are green. Also, the manually annotated contours for both lumen and media are drawn as yellow dashed lines. As we can see, the chosen frames contain a variety of lumen and media morphologies.

A detailed evaluation result and comparison with 9 recently published IVUS segmentation methods are reported in Table 6.3 where the performance of the proposed EREL selection strategy in the presence of various artifacts is shown as well.

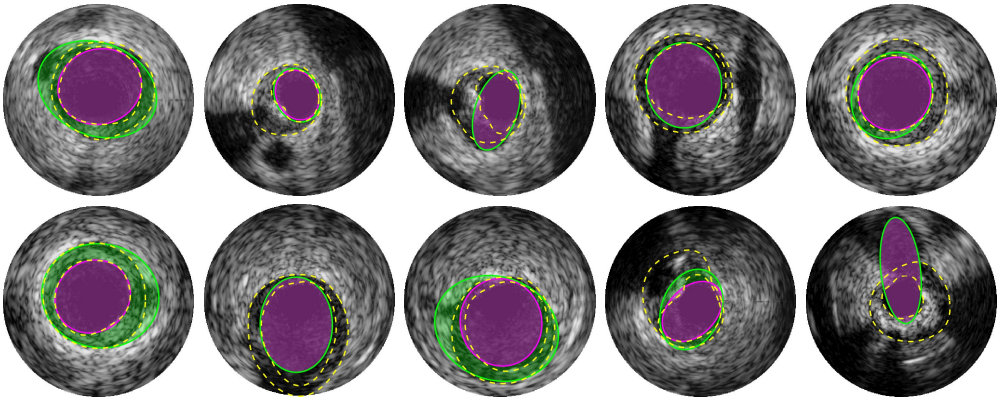


Figure 6.5: Several inaccurate lumen and media segmentation results in the presence of various artifacts. Segmented lumen and media have been highlighted by magenta and green colours, respectively. The yellow dashed lines illustrate the gold standard that have been delineated by four clinical experts [4].

6.4 Discussion

The present study is twofold. We first investigate whether the extracted ERELs [29], [30] are proper regions to be considered as lumen and media delineations of 20-MHz Intravascular Ultrasound frames taken from the inside of the human coronary arteries. We quantitatively (Table 6.2) validate that very close candidates can be found from the small number (almost less than 50) of extracted ERELs for each IVUS frame. Specifically, the best case results demonstrate the average of the closest extracted EREL to the lumen and media among all 435 images of the dataset [4]. As the results of the best case study suggest, if we design a proper selection strategy we are able to find two regions out of the extracted ERELs that are very close to the actual lumen and media structure.

Second, we propose an ERELs selection strategy based on creating a vector that represents several textural characteristics of the regions and search for the regions which have higher textural stability scores. This method works very well when no artifact is present in the IVUS image or more accurately, when no major artifact is attached to the media. In this case, the average Hausdorff distance for both lumen and media is less than 0.3 mm, which shows that the proposed method works very well for images without any strong artifact. Furthermore, the general performance of our proposed method for lumen segmentation is superior to other available methods in terms of the average Hausdorff distance, which is 0.3 mm. Likewise, our method is able to segment the lumen when the image contains shadows ($HD = 0.29$ mm) or side vessel ($HD = 0.24$ mm) artifacts, which is by far the lowest distance to the gold standard.

If artifacts that are usually dark (low intensity value) regions, join the lumen and media regions, then a type of leak will be created. This leak significantly increases the size and the mean intensity of the regions and generates irregular patterns that are very difficult

to distinguish by the EREL extraction algorithm. Some of the inaccurate segmentations that are caused by the presence of the artifacts are shown in Figure 6.5. As can be seen in Figure 6.5, although the presence of the artifacts disrupts the detection of the media regions, most of the lumen regions can still be segmented accurately. This is also supported by the quantitative results presented in Table 6.3. Evidently, even in the presence of artifacts, such as bifurcations, side vessels, and shadows, the lumen segmentation performance remains high, though the accuracy of the media segmentation drops. In fact, even in the worst case condition of lumen segmentation that happens when the image contains bifurcation artifacts, the average Hausdorff distance to the gold standard is only 0.53 mm which is still lower than most of the other methods. On the other hand, the media segmentation is more sensitive to the presence of artifacts. For instance, the Hausdorff metric for media in some cases increases to 1.22 mm for bifurcation and to 1.24 mm for shadow artifact. Accordingly, it can be concluded that the lumen segmentation is robust to the common artifacts of the IVUS images.

The success of the proposed method depends strictly on the correct extraction of ERELs. In cases when continuous and clear boundaries do not exist between extremal regions, EREL cannot extract correct candidate regions for lumen and media. For example, in IVUS images taken with a 40-MHz catheter probe, some lumen pixels located on the boundaries adjacent to intima do not have clear connections to each other. A similar situation occurs for the boundaries between media and adventitia. This creates leakages that attach media region to the adventitia. So, the sequential process of EREL's enumeration connects pixels from two semantically different regions and constructs a big connected component that includes both regions. Eventually, the leakage obstructs EREL's region extraction capability for detecting Q^+ regions. This leakage can also be created in IVUS 20-MHz images by various artifacts that somehow attach the media regions to the adventitia as illustrated in Figure 6.5.

Apart from the above, our proposed method offers several benefits. Not only is our method automatic, but works also solely on the B-mode images (in contrast to some methods available in the literature that work on the polar space [73], [94], [97], [113]) and, thus, it eliminates the need for transformation to the polar space. In addition, some methods [16], [92] require the whole volume to be presented in order to segment the arterial walls. This makes it impossible to segment the lumen and media in online segmentation applications (during the in-vivo pullback procedures). On the other hand, our method delineates lumen and media contours using only the information available in the current frame and, hence, performs no redundant processing. Furthermore, due to the low computational complexity of the EREL [29], [30], every frame can be segmented in linear time based on the number of pixels in the frame. The average run time per frame in the proposed method over the test set of [4] is 0.19 seconds.

6.5 Conclusion

In this study, we investigated the suitability of a recently proposed region detector called *Extremal Regions of Extremum Levels* for detecting the lumen and media regions. The results of our experiments reveal that among the ERELs extracted from a single IVUS frame, there exist regions with very low difference from the manually annotated data. Therefore, we presented a region selection approach to automatically segment the arterial walls from Intravascular Ultrasound images taken with a 20-MHz catheter probe. We embedded the mean intensity value, the entropy, and the boundary length of the regions into a single feature vector representing the texture information of the region. Then, our selection method looks for regions with higher stability score. The region that is most stable at the beginning section of the texture vector is considered as the lumen. Similarly, the last region with a stable textural variation is labeled as media.

We evaluated our segmentation method on the test set of a standard publicly available IVUS dataset using three common evaluation metrics; namely, JM, HD, and PAD. Extensive quantitative comparisons show the high accuracy of our proposed method. In general, the average HD for our lumen segmentation results is 0.3 mm which is the closest distance to the gold standard among all existing methods. In addition, when no artifact was present, our method segments both lumen and media achieving average HD lower than 0.29 mm and 0.31 mm, respectively.

As reported in the best case results, among the extracted EREL regions, there exist regions very close to the actual structure of the arterial walls. In future work, we would like to devise a better selection strategy that chooses ERELs more accurately.

Since IVUS is completely radiation-free, unlike X-ray and CT, in the future we would also like to investigate the feasibility of combining the segmentation results from individual slices to obtain 3D views of collateral arteries.

6.6 Acknowledgment

We gratefully acknowledge the assistance of Prof. Simone Balocco at the University of Barcelona for providing the labeling information on the existing artifacts in the dataset.

Chapter 7

Conclusion and Future Work

My research reported in my thesis was twofold:

1. During the first year of my PhD, I proposed a novel method to segment Intravascular Ultrasound (IVUS) frames. The proposed method utilized a region detector called *Extremal Regions of Extremum Levels* for detecting the lumen and media regions. Using the ERELs extracted from a single IVUS frame followed by a region selection approach, the proposed method could automatically segment the arterial walls from Intravascular Ultrasound images captured with a 20-MHz catheter probe. We embedded the mean intensity value, the entropy, and the boundary length of the regions into a single feature vector representing the texture information of the region. Then our selection method looked for regions with a higher stability score. The region that was most stable at the initial section of the texture vector was considered as the lumen. Similarly, the last region with a stable textural variation was labeled as media.

   We evaluated our segmentation method on the test set of a standard publicly available IVUS dataset using three common evaluation metrics: namely, JM, HD, and PAD. Extensive quantitative comparisons showed the high accuracy of our proposed method. In general, the average HD for our lumen segmentation results was 0.3 mm, which was the closest distance to the gold standard among all existing methods. In addition, when no artifact was present, our method segmented both lumen and media achieving average HD lower than 0.29 mm and 0.31 mm, respectively.

   The proposed segmentation method is a general approach that can be run over any dataset and does not require any training step. More importantly it segments an IVUS frame in linear time with respect to the number of pixels in the image. These advantages of the proposed segmentation method make it a good choice for real-time segmentation and 3D reconstruction applications.

2. In the other part of my studies, I proposed new equations to calibrate an active camera, that is called Simplified Active Calibration. I provided closed-form and linear

equations to estimate the parameters of a camera using three image pairs (utilizing 4 images overall) taken before and after panning, tilting, and panning-tilting the camera.

A basic assumption about the rotation of a fixed camera was made; i.e., to solve the proposed equations, knowing the rotation angles of the camera is necessary. The proposed formulation can be used in practical applications such as surveillance because in PTZ cameras accessing the camera motion information is straightforward.

The proposed closed-form formulations for estimating the focal lengths can be solved with only one point correspondence. This is especially useful for applications that favor no point correspondences, but instead utilize contour-to-contour correspondences.

The results of solving our proposed formulations on randomly simulated 3D scenes indicated a very low error rate in estimating the focal lengths and the principal point location in ideal conditions. We evaluated our proposed SAC formulation for two different noise conditions, namely angular and pixel noise. The simulated results showed that if the absolute value of the rotation angle is bounded from above by approximately 7° and from below by approximately 2°, the error caused by using the SAC formulation is very small. This conclusion was later verified in our experiment with real images.

7.1 Future Work

- In Active Calibration, a basic assumption about the rotation of a fixed camera was made; i.e., to solve the proposed equations, knowing the rotation angles of the camera is necessary. This formulation can be employed in practical applications, such as surveillance using PTZ and mobile phone cameras, where the camera motion information can be adjusted. However, the extracted motion parameters from mechanical devices, gyroscope and IMUs (Inertial Measurement Units) are always noisy and inaccurate. Although, one can alleviate the effects of these angular noises by avoiding small rotations, having another method that can find and offset the effect of the noise would be extremely helpful as a pre-processing step. Therefore, I believe finding a way to de-noise the angular motions is important to pursue.

- Recently, some studies such as [11], [54], [65], have employed deep learning in order to remove the role of point correspondences. One disadvantage of these methods is that they can only estimate the focal length from outdoor scenes where the horizon line can be seen. Basically, to formulate the focal length they used a different formulation, rather than the camera projection matrix. Since in SAC we are mostly dealing with indoor scenes, it might be possible to extract the scene structures and use them

to define a new deep learning model and remove the requirements of finding point correspondences.

- Including non-linear parameters (such as lens distortion) into SAC equations and using the result of the study as a close initial guess for an optimization procedure could be studied. A mathematical optimization process such as the *Levenberg Marquardt* approach, can be helpful in refining the results and obtaining more accurate estimates.

- In a more general sense, trying SAC on mobile cameras is hard. The reason for this is the difficulty for users to take pictures with a mobile camera while they maintain fixed values of rotations around 2 axes and 3 translations. If we try to remove the constraints imposed by SAC on the camera projection matrix in order to make it easier use in mobile applications, it would no longer be an active method. However, one can further investigate this idea to see if it is feasible to modify the equations and make them easier to use in mobile phones.

- Both active methods that are proposed in this dissertation, with little adjustments, can be combined and integrated into an Image Guided Surgery (IGS) pipeline for 3D reconstruction of the vessel wall in Intravascular Ultrasound frames. Assuming that the guidewire catheter of a device can be directly controlled (with a low level microcontroller interface), then one can follow an approach similar to SAC to calibrate the Ultrasound[1] and then use our proposed online segmentation method for both matching the points and reconstruct the surface.

- Registration of IVUS frames is a required step in 3D reconstruction of the vessel. The registration can be done over the results of the segmentation method proposed in this thesis by comparing the shape of the segmented regions (contour-to-contour) instead of matching the point correspondences.

[1]The equations should be adjusted to be consistent with physical characteristics Ultrasound image formation.

References

[1] L. Agapito, E. Hayman, and I. Reid, "Self-calibration of rotating and zooming cameras," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 107–127, 2001 (cit. on pp. 10, 57, 64, 67).

[2] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *International journal of computer vision*, vol. 1, no. 4, pp. 333–356, 1988 (cit. on pp. 1, 2, 57).

[3] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988 (cit. on pp. 1, 2).

[4] S. Balocco, C. Gatta, F. Ciompi, A. Wahle, P. Radeva, S. Carlier, G. Unal, E. Sanidas, J. Mauri, X. Carillo, *et al.*, "Standardized evaluation methodology and reference database for evaluating ivus image segmentation," *Computerized Medical Imaging and Graphics*, vol. 38, no. 2, pp. 70–90, 2014 (cit. on pp. 5, 83, 85, 86, 88, 93–98).

[5] A. Basu, "Active calibration," in *Robotics and Automation, 1993. Proceedings., 1993 IEEE International Conference on*, IEEE, 1993, pp. 764–769 (cit. on pp. 10, 25, 26, 33, 46, 47, 57, 59).

[6] ——, "Active calibration: Alternative strategy and analysis," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, IEEE, 1993, pp. 495–500 (cit. on pp. 10, 12, 25, 26, 31, 33, 46, 47, 57, 59).

[7] ——, "Active calibration of cameras: Theory and implementation," *IEEE Transactions on Systems, man, and cybernetics*, vol. 25, no. 2, pp. 256–265, 1995 (cit. on pp. 10, 25, 26, 33, 46, 47, 59).

[8] A. Basu and S. Licardie, "Alternative models for fish-eye lenses," *Pattern Recognition Letters*, vol. 16, no. 4, pp. 433–441, 1995 (cit. on pp. 34, 35).

[9] A. Basu and K. Ravi, "Active camera calibration using pan, tilt and roll," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 27, no. 3, pp. 559–566, 1997 (cit. on pp. 10, 25, 26, 33, 46, 47, 49, 59, 60, 79).

[10] J. Batista, H. Araújo, and A. T. de Almeida, "Iterative multistep explicit camera calibration," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 5, pp. 897–917, 1999 (cit. on p. 9).

[11] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin, "Deepcalib: A deep learning approach for automatic intrinsic calibration of wide field-of-view cameras," in *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, ACM, 2018, p. 6 (cit. on pp. 10, 24, 101).

[12] C. V. Bourantas, M. I. Papafaklis, L. Athanasiou, F. G. Kalatzis, K. K. Naka, P. K. Siogkas, S. Takahashi, S. Saito, D. I. Fotiadis, C. L. Feldman, *et al.*, "A new methodology for accurate 3-dimensional coronary artery reconstruction using routine intravascular ultrasound and angiographic data: Implications for widespread assessment of endothelial shear stress in humans," *EuroIntervention*, vol. 9, no. 5, pp. 582–93, 2013 (cit. on p. 42).

[13] D. C. Brown, "Decentering distortion of lenses," *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966 (cit. on p. 7).

[14] ——, "Close-range camera calibration," in *Photogrammetric Engineering*, Citeseer, 1971 (cit. on p. 34).

[15] M. Brückner, F. Bajramovic, and J. Denzler, "Intrinsic and extrinsic active self-calibration of multi-camera systems," *Machine vision and applications*, vol. 25, no. 2, pp. 389–403, 2014 (cit. on p. 58).

[16] M.-H. Cardinal, J. Meunier, G. Soulez, R. L. Maurice, É. Therasse, and G. Cloutier, "Intravascular ultrasound image segmentation: A three-dimensional fast-marching method based on gray level distributions," *IEEE transactions on medical imaging*, vol. 25, no. 5, pp. 590–601, 2006 (cit. on p. 98).

[17] S.-Y. Chen and W.-H. Tsai, "A systematic approach to analytic determination of camera parameters by line features," *Pattern Recognition*, vol. 23, no. 8, pp. 859–877, 1990 (cit. on p. 10).

[18] F. Ciompi, O. Pujol, C. Gatta, M. Alberti, S. Balocco, X. Carrillo, J. Mauri-Ferre, and P. Radeva, "Holimab: A holistic approach for media–adventitia border detection in intravascular ultrasound," *Medical image analysis*, vol. 16, no. 6, pp. 1085–1100, 2012 (cit. on pp. 93, 96).

[19] F. Destrempes, M.-H. R. Cardinal, L. Allard, J.-C. Tardif, and G. Cloutier, "Segmentation method of intravascular ultrasound images of human coronary arteries," *Computerized Medical Imaging and Graphics*, vol. 38, no. 2, pp. 91–103, 2014 (cit. on p. 85).

[20] C. Doulaverakis, I. Tsampoulatidis, A. P. Antoniadis, Y. S. Chatzizisis, A. Giannopoulos, I. Kompatsiaris, and G. D. Giannoglou, "Ivusangio tool: A publicly available software for fast and accurate 3d reconstruction of coronary arteries," *Computers in biology and medicine*, vol. 43, no. 11, pp. 1793–1803, 2013 (cit. on p. 43).

[21] R. Downe, A. Wahle, T. Kovarnik, H. Skalicka, J. Lopez, J. Horak, and M. Sonka, "Segmentation of intravascular ultrasound images using graph search and a novel cost function," in *Proc. 2nd MICCAI workshop on computer vision for intravascular and intracardiac imaging*, Citeseer, 2008, pp. 71–9 (cit. on pp. 93, 96).

[22] L. Dron, "Dynamic camera self-calibration from controlled motion sequences," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, IEEE, 1993, pp. 501–506 (cit. on p. 57).

[23] F. Du and M. Brady, "Self-calibration of the intrinsic parameters of cameras for active vision systems," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, IEEE, 1993, pp. 477–482 (cit. on p. 57).

[24] T. Echigo, "A camera calibration technique using three sets of parallel lines," *Machine Vision and Applications*, vol. 3, no. 3, pp. 159–167, 1990 (cit. on p. 10).

[25] W. Faig, "Calibration of close-range photogrammetric systems: Mathematical formulation," *Photogrammetric engineering and remote sensing*, vol. 41, no. 12, 1975 (cit. on pp. 7, 34).

[26] M. Faraji and A. Basu, "A simplified active calibration algorithm for focal length estimation," in *International Conference on Smart Multimedia*, Springer, 2018, pp. 381–390 (cit. on pp. v, 6, 59).

[27] ——, "Simplified active calibration," *Image and Vision Computing*, vol. 91, p. 103 799, 2019 (cit. on pp. iv, 6).

[28] M. Faraji, I. Cheng, I. Naudin, and A. Basu, "Segmentation of arterial walls in intravascular ultrasound cross-sectional images using extremal region selection," *Ultrasonics*, vol. 84, pp. 356–365, 2018 (cit. on pp. iv, 6).

[29] M. Faraji, J. Shanbehzadeh, K. Nasrollahi, and T. Moeslund, "Extremal regions detection guided by maxima of gradient magnitude," *Image Processing, IEEE Transactions on*, 2015 (cit. on pp. 53, 78, 85–87, 90, 92, 97, 98).

[30] M. Faraji, J. Shanbehzadeh, K. Nasrollahi, and T. B. Moeslund, "Erel: Extremal regions of extremum levels," in *Image Processing (ICIP), 2015 IEEE International Conference on*, IEEE, 2015, pp. 681–685 (cit. on pp. 53, 78, 85–87, 90, 97, 98).

[31] O. Faugeras and G. Toscani, "Camera calibration for 3d computer vision," in *International Workshop on Machine Vision and Machine Intelligence*, 1987, pp. 240–247 (cit. on p. 25).

[32] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank, "Camera self-calibration: Theory and experiments," in *European conference on computer vision*, Springer, 1992, pp. 321–334 (cit. on pp. 10, 25, 46, 57).

[33] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 86, 1986, pp. 15–20 (cit. on pp. 8, 10, 18, 34).

[34] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007 (cit. on p. 35).

[35] J.-M. Frahm and R. Koch, "Camera calibration and 3d scene reconstruction from image sequence and rotation sensor data.," in *VMV*, 2003, pp. 79–86 (cit. on p. 58).

[36] ——, "Camera calibration with known rotation.," in *ICCV*, 2003, pp. 1418–1425 (cit. on p. 58).

[37] ——, "Robust camera calibration from images and rotation data," in *Joint Pattern Recognition Symposium*, Springer Berlin Heidelberg, 2003, pp. 249–256 (cit. on p. 58).

[38] A. Frimerman, E. Abergel, D. S. Blondheim, A. Shotan, S. Meisel, M. Shochat, P. Punjabi, and A. Roguin, "Novel method for real time co-registration of ivus and coronary angiography," *Journal of interventional cardiology*, 2016 (cit. on p. 43).

[39] J. Frostegård, "Sle, atherosclerosis and cardiovascular disease," *Journal of internal medicine*, vol. 257, no. 6, pp. 485–495, 2005 (cit. on p. 83).

[40] R. Galego, A. Bernardino, and J. Gaspar, "Auto-calibration of pan-tilt cameras including radial distortion and zoom," in *International Symposium on Visual Computing*, Springer, 2012, pp. 169–178 (cit. on p. 58).

[41] S. Ganapathy, "Decomposition of transformation matrices for robot vision," *Pattern Recognition Letters*, vol. 2, no. 6, pp. 401–412, 1984 (cit. on p. 8).

[42] C. Gatta, S. Balocco, F. Ciompi, R. Hemetsberger, O. R. Leor, and P. Radeva, "Real-time gating of ivus sequences based on motion blur analysis: Method and quantitative validation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2010, pp. 59–67 (cit. on p. 42).

[43] A. G. van der Giessen, M. Schaap, F. J. Gijsen, H. C. Groen, T. van Walsum, N. R. Mollet, J. Dijkstra, F. N. van de Vosse, W. J. Niessen, P. J. de Feyter, *et al.*, "3d fusion of intravascular ultrasound and coronary computed tomography for in-vivo wall shear stress analysis: A feasibility study," *The International Journal of Cardiovascular Imaging*, vol. 26, no. 7, pp. 781–796, 2010 (cit. on p. 42).

[44] E. L. Hall, J. B. Tio, C. A. McPherson, and F. A. Sadjadi, "Measuring curved surfaces for robot vision," *Computer*, no. 12, pp. 42–54, 1982 (cit. on pp. 8, 10, 13, 14, 25).

[45] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979 (cit. on pp. 84, 89).

[46] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003 (cit. on pp. 10, 63).

[47] R. I. Hartley, "Self-calibration from multiple views with a rotating camera," in *European Conference on Computer Vision*, Springer, 1994, pp. 471–478 (cit. on pp. 10, 57).

[48] ——, "Self-calibration of stationary cameras," *International journal of computer vision*, vol. 22, no. 1, pp. 5–23, 1997 (cit. on pp. 10, 57).

[49] J. Heikkila, "Geometric camera calibration using circular control points," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 10, pp. 1066–1077, 2000 (cit. on pp. 8, 9, 19, 20).

[50] J. Heikkila and O. Silven, "Calibration procedure for short focal length off-the-shelf ccd cameras," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, IEEE, vol. 1, 1996, pp. 166–170 (cit. on p. 19).

[51] J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, IEEE, 1997, pp. 1106–1112 (cit. on pp. 8, 10, 18).

[52] L. Heng, G. H. Lee, and M. Pollefeys, "Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle," *Autonomous Robots*, vol. 39, no. 3, pp. 259–277, 2015 (cit. on p. 58).

[53] A. Hernàndez-Sabaté and D. Gil, *The Benefits of IVUS Dynamics for Retrieving Stable Models of Arteries*. INTECH Open Access Publisher, 2012 (cit. on pp. 39, 41).

[54] Y. Hold-Geoffroy, K. Sunkavalli, J. Eisenmann, M. Fisher, E. Gambaretto, S. Hadap, and J.-F. Lalonde, "A perceptual measure for deep single image camera calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2354–2363 (cit. on pp. 10, 24, 101).

[55] Z.-Q. Hong and J.-Y. Yang, "An algorithm for camera calibration using a three-dimensional reference point," *Pattern Recognition*, vol. 26, no. 11, pp. 1655–1660, 1993 (cit. on p. 10).

[56] L. Hua, W. Fu-Chao, and H. Zhan-Yi, "A new linear camera self-calibration technique," *Chinese J. Computers*, vol. 23, no. 11, pp. 1121–1129, 2000 (cit. on p. 58).

[57] B. Iglewicz and D. C. Hoaglin, *How to detect and handle outliers*. Asq Press, 1993, vol. 16 (cit. on p. 90).

[58] M. Ito, "Robot vision modelling-camera modelling and camera calibration," *Advanced robotics*, vol. 5, no. 3, pp. 321–335, 1990 (cit. on p. 8).

[59] A. Izaguirre, P. Pu, and J. Summers, "A new development in camera calibration: Calibrating a pair of mobile cameras," *The International Journal of Robotics Research*, vol. 6, no. 3, pp. 104–116, 1987 (cit. on p. 8).

[60] Q. Ji and S. Dai, "Self-calibration of a rotating camera with a translational offset," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 1, pp. 1–14, 2004 (cit. on p. 60).

[61] I. N. Junejo and H. Foroosh, "Practical ptz camera calibration using givens rotations," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, IEEE, 2008, pp. 1936–1939 (cit. on p. 58).

[62] ——, "Optimizing ptz camera calibration from two images," *Machine Vision and Applications*, vol. 23, no. 2, pp. 375–389, 2012 (cit. on pp. 48, 64).

[63] K.-I. Kanatani, "Camera rotation invariance of image characteristics," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 328–354, 1987 (cit. on pp. 25, 63).

[64] A. Katouzian, E. D. Angelini, S. G. Carlier, J. S. Suri, N. Navab, and A. F. Laine, "A state-of-the-art review on segmentation algorithms in intravascular ultrasound (ivus) images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 5, pp. 823–834, 2012 (cit. on p. 83).

[65] A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5974–5983 (cit. on pp. 10, 101).

[66] J. Knight, A. Zisserman, and I. Reid, "Linear auto-calibration for ground plane motion," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, IEEE, vol. 1, 2003, pp. I–I (cit. on p. 58).

[67] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-d machine vision metrology," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 10, no. 5, pp. 713–720, 1988 (cit. on p. 9).

[68] Y. Li and M. Faraji, "Erel selection using morphological relation," in *International Conference on Smart Multimedia*, Springer, 2018, pp. 437–447 (cit. on p. v).

[69] Y. Liu, T. S. Huang, and O. D. Faugeras, "Determination of camera location from 2-d to 3-d line and point correspondences," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 1, pp. 28–37, 1990 (cit. on p. 9).

[70] C. P. Loizou and C. S. Pattichis, "Despeckle filtering algorithms and software for ultrasound imaging," *Synthesis lectures on algorithms and software in engineering*, vol. 1, no. 1, pp. 1–166, 2008 (cit. on p. 86).

[71] T. Ma, B. Zhou, T. K. Hsiai, and K. K. Shung, "A review of intravascular ultrasound-based multimodal intravascular imaging: The synergistic approach to characterizing vulnerable plaques," *Ultrasonic imaging*, vol. 38, no. 5, pp. 314–331, 2016 (cit. on pp. 43, 83).

[72] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 123–151, 1992 (cit. on pp. 46, 57).

[73] E. G. Mendizabal-Ruiz, M. Rivera, and I. A. Kakadiaris, "Segmentation of the luminal border in intravascular ultrasound b-mode images using a probabilistic approach," *Medical image analysis*, vol. 17, no. 6, pp. 649–670, 2013 (cit. on pp. 84, 93, 96, 98).

[74] G. Mendizabal-Ruiz and I. A. Kakadiaris, "A physics-based intravascular ultrasound image reconstruction method for lumen segmentation," *Computers in biology and medicine*, vol. 75, pp. 19–29, 2016 (cit. on p. 84).

[75] F. Molinari, K. M. Meiburger, G. Zeng, A. Nicolaides, and J. S. Suri, "Caudles-ef: Carotid automated ultrasound double line extraction system using edge flow," *Journal of digital imaging*, vol. 24, no. 6, pp. 1059–1077, 2011 (cit. on p. 94).

[76] S. K. Nadkarni, D. Boughner, and A. Fenster, "Image-based cardiac gating for three-dimensional intravascular ultrasound imaging," *Ultrasound in medicine & biology*, vol. 31, no. 1, pp. 53–63, 2005 (cit. on p. 41).

[77] M. A. Penna, "Camera calibration: A quick and easy way to determine the scale factor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 12, pp. 1240–1245, 1991 (cit. on p. 9).

[78] M. Pollefeys, R. Koch, and L. Van Gool, "Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 7–25, 1999 (cit. on p. 57).

[79] Y. Rim, D. D. McPherson, and H. Kim, "Volumetric three-dimensional intravascular ultrasound visualization using shape-based nonlinear interpolation," *Biomedical engineering online*, vol. 12, no. 1, p. 39, 2013 (cit. on p. 41).

[80] J. Rong, S. Huang, Z. Shang, and X. Ying, "Radial lens distortion correction using convolutional neural networks trained with synthesized images," in *Asian Conference on Computer Vision*, Springer, 2016, pp. 35–49 (cit. on p. 24).

[81] J. Salvi *et al.*, *An approach to coded structured light to obtain three dimensional information*. Universitat de Girona, 1998 (cit. on pp. 7, 18, 25).

[82] J. Salvi, X. Armangué, and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern recognition*, vol. 35, no. 7, pp. 1617–1635, 2002 (cit. on pp. 7, 8, 36).

[83] J. Salvi, J. Batlle, and E. Mouaddib, "A robust-coded pattern projection for dynamic 3d scene measurement," *Pattern Recognition Letters*, vol. 19, no. 11, pp. 1055–1065, 1998 (cit. on p. 18).

[84] R. Sedgewick and K. Wayne, *Algorithms, 4th Edition*. Addison-Wesley, 2011, pp. I–XII, 1–955, ISBN: 978-0-321-57351-3 (cit. on p. 86).

[85] L. Shapiro and G. C. Stockman, "Computer vision. 2001," *ed: Prentice Hall*, 2001 (cit. on p. 92).

[86] S.-W. Shih, Y.-P. Hung, and W.-S. Lin, "Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem," *Optical Engineering*, vol. 32, no. 1, pp. 138–149, 1993 (cit. on p. 8).

[87] C. J. Slager, J. J. Wentzel, J. C. Schuurbiers, J. A. Oomen, J. Kloet, R. Krams, C. Von Birgelen, W. J. Van Der Giessen, P. W. Serruys, and P. J. De Feyter, "True 3-dimensional reconstruction of coronary arteries in patients by fusion of angiography and ivus (angus) and its quantitative validation," *Circulation*, vol. 102, no. 5, pp. 511–516, 2000 (cit. on pp. 40, 43).

[88] G. P. Stein, "Accurate internal camera calibration using rotation, with analysis of sources of error," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, IEEE, 1995, pp. 230–236 (cit. on p. 57).

[89] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*. Springer Science & Business Media, 2013, vol. 12 (cit. on p. 18).

[90] S. Su, Z. Hu, Q. Lin, W. K. Hau, Z. Gao, and H. Zhang, "An artificial neural network method for lumen and media-adventitia border detection in ivus," *Computerized Medical Imaging and Graphics*, 2016 (cit. on p. 84).

[91] Q. Sun, X. Wang, J. Xu, L. Wang, H. Zhang, J. Yu, T. Su, and X. Zhang, "Camera self-calibration with lens distortion," *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 10, pp. 4506–4513, 2016 (cit. on p. 58).

[92] Z. Sun and C. Liu, "A parallel method for segmenting intravascular ultrasound image sequence," in *Applied Mechanics and Materials*, Trans Tech Publ, vol. 130, 2012, pp. 2051–2055 (cit. on p. 98).

[93] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010 (cit. on p. 92).

[94] A. Taki, Z. Najafi, A. Roodaki, S. K. Setarehdan, R. A. Zoroofi, A. Konig, and N. Navab, "Automatic segmentation of calcified plaques and vessel borders in ivus images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, no. 3-4, pp. 347–354, 2008 (cit. on pp. 84, 98).

[95] P. A. Tresadern and I. D. Reid, "Camera calibration from human motion," *Image and Vision Computing*, vol. 26, no. 6, pp. 851–862, 2008 (cit. on p. 58).

[96] 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*, vol. 3, no. 4, pp. 323–344, 1987 (cit. on pp. 9, 10, 12, 25, 34).

[97] G. Unal, S. Bucher, S. Carlier, G. Slabaugh, T. Fang, and K. Tanaka, "Shape-driven segmentation of the arterial wall in intravascular ultrasound images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 3, pp. 335–347, 2008 (cit. on pp. 84, 86, 93, 96, 98).

[98] L. L. Vercio, J. I. Orlando, M. del Fresno, and I. Larrabide, "Assessment of image features for vessel wall segmentation in intravascular ultrasound images," *International journal of computer assisted radiology and surgery*, vol. 11, no. 8, pp. 1397–1407, 2016 (cit. on pp. 84, 96).

[99] D. Wan and J. Zhou, "Self-calibration of spherical rectification for a ptz-stereo system," *Image and Vision Computing*, vol. 28, no. 3, pp. 367–375, 2010 (cit. on p. 58).

[100] C.-C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 2, pp. 161–175, 1992 (cit. on p. 9).

[101] J. Wang, F. Shi, J. Zhang, and Y. Liu, "A new calibration model of camera lens distortion," *Pattern Recognition*, vol. 41, no. 2, pp. 607–615, 2008 (cit. on pp. 34, 36, 38).

[102] L.-L. Wang, W.-H. Tsai, *et al.*, "Camera calibration by vanishing lines for 3-d computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 370–376, 1991 (cit. on p. 10).

[103] P. Wang, T. Chen, O. Ecabert, S. Prummer, M. Ostermeier, and D. Comaniciu, "Image-based device tracking for the co-registration of angiography and intravascular ultrasound images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2011, pp. 161–168 (cit. on p. 42).

[104] G.-Q. Wei and S. De Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 469–480, 1994 (cit. on p. 9).

[105] A. E. Welchman, "The human brain in depth: How we see in 3d," *Annual review of vision science*, vol. 2, pp. 345–376, 2016 (cit. on p. 1).

[106] J. Weng, P. Cohen, M. Herniou, *et al.*, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 10, pp. 965–980, 1992 (cit. on pp. 7–10, 12–18, 25, 34, 36, 37).

[107] K. W. Wong, "Mathematical formulation and digital analysis in close-range photogrammetry," *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 11, 1975 (cit. on p. 7).

[108] S. Workman, C. Greenwell, M. Zhai, R. Baltenberger, and N. Jacobs, "Deepfocal: A method for direct focal length estimation," in *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015, pp. 1369–1373 (cit. on p. 24).

[109] Z. Wu and R. J. Radke, "Keeping a pan-tilt-zoom camera calibrated," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1994–2007, 2013 (cit. on p. 58).

[110] J. Yan, D. Lv, and Y. Cui, "A novel segmentation approach for intravascular ultrasound images," *Journal of Medical and Biological Engineering*, vol. 37, no. 3, pp. 386–394, 2017 (cit. on p. 84).

[111] J. Yang, M. Faraji, and A. Basu, "Robust segmentation of arterial walls in intravascular ultrasound images using dual path u-net," *Ultrasonics*, vol. 96, pp. 24–33, 2019 (cit. on p. iv).

[112] J. Yang, L. Tong, M. Faraji, and A. Basu, "Ivus-net: An intravascular ultrasound segmentation network," in *International Conference on Smart Multimedia*, Springer, 2018, pp. 367–377 (cit. on p. v).

[113] F. S. Zakeri, S. K. Setarehdan, and S. Norouzi, "Automatic media-adventitia ivus image segmentation based on sparse representation framework and dynamic directional active contour model," *Computers in Biology and Medicine*, 2017 (cit. on pp. 84, 98).

[114] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Ieee, vol. 1, 1999, pp. 666–673 (cit. on pp. 9, 10, 21, 23, 25, 52–54, 77, 78).

[115] ——, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000 (cit. on pp. 9, 10, 21–23, 25).

[116] F. Zhao, T. Tamaki, T. Kurita, B. Raytchev, and K. Kaneda, "Marker-based non-overlapping camera calibration methods with additional support camera views," *Image and Vision Computing*, vol. 70, pp. 46–54, 2018 (cit. on p. 57).

[117] X. Zhu, P. Zhang, J. Shao, Y. Cheng, Y. Zhang, and J. Bai, "A snake-based method for segmentation of intravascular ultrasound images and its in vivo validation," *Ultrasonics*, vol. 51, no. 2, pp. 181–189, 2011 (cit. on p. 84).