STRUCTURE-FROM-MOTION FOR SYSTEMS WITH PERSPECTIVE AND OMNIDIRECTIONAL CAMERAS

YALIN BAŞTANLAR

JULY 2009

STRUCTURE-FROM-MOTION FOR SYSTEMS WITH PERSPECTIVE AND OMNIDIRECTIONAL CAMERAS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YALIN BAŞTANLAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JULY 2009

Approval of the Graduate School of Informatics:

_____

Prof. Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

_____

Prof. Dr. Yasemin YARDIMCI
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy.

_____

Asst. Prof. Dr. Alptekin TEMİZEL
Co-Supervisor

_____

Prof. Dr. Yasemin YARDIMCI
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Aydın ALATAN (METU, EEE) _____

Prof. Dr. Yasemin YARDIMCI (METU, II) _____

Asst. Prof. Dr. Alptekin TEMİZEL (METU, II) _____

Assoc. Prof. Dr. Uğur GÜDÜKBAY (Bilkent Univ., CS) _____

Asst. Prof. Dr. Altan KOÇYİĞİT (METU, II) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name, Last name : **Yalın Başţanlar**

Signature : _____

ABSTRACT

STRUCTURE-FROM-MOTION FOR SYSTEMS WITH PERSPECTIVE AND OMNIDIRECTIONAL CAMERAS

BAŞTANLAR, Yalın
Ph.D., Department of Information Systems
Supervisor: Prof. Dr. Yasemin YARDIMCI

July 2009, 100 pages

In this thesis, a pipeline for structure-from-motion with mixed camera types is described and methods for the steps of this pipeline to make it effective and automatic are proposed. These steps can be summarized as calibration, feature point matching, epipolar geometry and pose estimation, triangulation and bundle adjustment. We worked with catadioptric omnidirectional and perspective cameras and employed the sphere camera model, which encompasses single-viewpoint catadioptric systems as well as perspective cameras.

For calibration of the sphere camera model, a new technique that has the advantage of linear and automatic parameter initialization is proposed. The projection of 3D points on a catadioptric image is represented linearly with a 6×10 projection matrix using lifted coordinates. This projection matrix is computed with an adequate number of 3D-2D correspondences and decomposed to obtain intrinsic and extrinsic parameters. Then, a non-linear optimization is performed to refine the parameters.

For feature point matching between hybrid camera images, scale invariant feature transform (SIFT) is employed and a method is proposed to improve the SIFT

matching output. With the proposed approach, omnidirectional-perspective matching performance significantly increases to enable automatic point matching. In addition, the use of virtual camera plane (VCP) images is evaluated, which are perspective images produced by unwarping the corresponding region in the omnidirectional image.

The hybrid epipolar geometry is estimated using random sample consensus (RANSAC) and alternatives of pose estimation methods are evaluated. A weighting strategy for iterative linear triangulation which improves the structure estimation accuracy is proposed. Finally, multi-view structure-from-motion (SfM) is performed by employing the approach of adding views to the structure one by one. To refine the structure estimated with multiple views, sparse bundle adjustment method is employed with a modification to use the sphere camera model.

Experiments on simulated and real images for the proposed approaches are conducted. Also, the results of hybrid multi-view SfM with real images are demonstrated, emphasizing the cases where it is advantageous to use omnidirectional cameras with perspective cameras.

Keywords: Catadioptric omnidirectional camera, mixed camera system, camera calibration, feature point matching, structure-from-motion.

ÖZ

PERSPEKTİF VE TÜMYÖNLÜ KAMERA KULLANAN SİSTEMLER İLE HAREKETTEN YAPI ÇIKARIMI

BAŞTANLAR, Yalın
Doktora, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Prof. Dr. Yasemin YARDIMCI

Temmuz 2009, 100 sayfa

Bu tezde, karma kameralı sistemler ile hareketten yapı çıkarımı yapmak için basamaklı bir şema tanımlanmış ve bu basamaklar için yöntemler geliştirilerek yapı çıkarımı işleminin daha gürbüz ve otomatik olması sağlanmıştır. Tanımlanan basamaklar, kamera kalibrasyonu, nokta eşleştirme, epipolar geometri ve kameraların duruş (konum ve yönelim) kestirimi, üçgenleme ve toplu düzenleme olarak özetlenebilir. Katadioptrik tümyönlü kameralar ve perspektif kameralar ile çalıştık ve hem tek görüş noktalı (single viewpoint) tümyönlü kameraları hem de perspektif kameraları kapsayan küresel kamera modelini kullandık.

Küresel kamera modelinin kalibrasyonu için doğrusal ve otomatik parametre ilklendirmesi avantajına sahip yeni bir yöntem geliştirilmiştir. 3B uzaydaki noktaların katadioptrik imgeye düşürülmesi, yükseltilmiş koordinatlar kullanılarak, 6×10 boyutunda bir izdüşüm matrisi ile doğrusal olarak ifade edilmiştir. Yeterli sayıda 3B-2B nokta eşleniği ile bu izdüşüm matrisi hesaplanabilmekte, içsel ve dışsal parametreler bu matristen elde edilebilmektedir. Ardından doğrusal olmayan bir optimizasyon ile parametreler iyileştirilmektedir.

Karma kamera imgeleri arasında nokta eşlemesi için, ölçekten bağımsız öznitelik

dönüşümü (SIFT) kullanılmış ve eşleme başarısını artırmak üzere bir yöntem önerilmiştir. Önerilen yöntemle tümyönlü-perspektif eşleme performansı otomatik eşlemeyi mümkün kılacak ölçüde artmıştır. Ayrıca, tümyönlü imgelerden bükme sonucu elde edilen ve perspektif imge özellikleri taşıyan sanal kamera düzlemi (VCP) görüntülerini kullanarak eşleme yapılması da araştırılmıştır.

Karma epipolar geometri kestirimi rasgele örnek onaylaşımı (RANSAC) ile gerçeklenmiş ve kamera duruş (konum ve yönelim) tespiti için alternatifler değerlendirilmiştir. Yapı kestirimi doğruluğunu artırmak üzere döngülü doğrusal üçgenleme için bir ağırlıklandırma yöntemi geliştirilmiştir. Son olarak, çok imgeli hareketten yapı çıkarımı için, yapıya her defasında yeni bir imge ekleme yaklaşımına dayalı yöntem uygulanmış ve kestirilen yapının doğruluğunun artırılması için nadir verili toplu düzenleme (sparse bundle adjustment) yöntemi küresel kamera modeli için modifiye edilerek kullanılmıştır.

Önerilen yöntemler için benzetimli ve gerçek imgeler üzerinde deneyler yapılmıştır. Ayrıca, perspektif kameralarla beraber tümyönlü kameraları kullanmanın avantajlı olduğu durumlar vurgulanarak, karma kameralarla çok imgeli hareketten yapı çıkarımı gösterimleri yapılmıştır.

Anahtar Kelimeler: Katadioptrik tümyönlü kamera, karma kameralı sistemler, kamera kalibrasyonu, nokta eşleme, hareketten yapı çıkarımı.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Dr. Yasemin Yardımcı for her valuable supervision and friendly attitute throughout the research. Her guidance positively influenced my academic activities and I have been able to complete this thesis.

I have conducted some of my studies in INRIA Rhone-Alpes with Prof. Peter Sturm. I would like to thank him for this precious support. It is a pleasure to work with him.

I wish to thank Asst. Prof. Dr. Alptekin Temizel, Assoc. Prof. Dr. Aydın Alatan and Asst. Prof. Dr. Erhan Eren for their advises concerning the research held in this thesis. I also thank Assoc. Prof. Dr. Uğur Güdükbay and Asst. Prof. Dr. Altan Koçyiğit for reviewing this work.

I would like to express my deepest gratitude to my family for their unconditional support during my life. Special thanks goes to my beloved girlfriend Evren for her continuous support and understanding. Finally, I would like to thank my colleagues Luis, Mustafa, Koray, Habil and Medeni for helping me in various ways.

This thesis work is supported by The Scientific and Technical Research Council of Turkey (TÜBİTAK) under the grant EEEAG-105E187 and with the researcher support programs 2211 and 2214.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

The 3D computer vision studies for omnidirectional cameras started about a decade ago. Omnidirectional cameras provide 360° horizontal field of view in a single image, which is an important advantage in many application areas such as navigation, surveillance and 3D reconstruction [1–8]. With this enlarged view, fewer omnidirectional cameras may substitute many perspective cameras. Moreover, point correspondences from a variety of angles provide more stable structure estimation [9] and degenerate cases like viewing only a planar surface are avoided. Major drawback of these images is that they have lower resolution than perspective images. Using perspective cameras together with omnidirectional ones could improve the resolution while preserving the enlarged view advantage. A possible scenario is 3D reconstruction in which omnidirectional cameras provide low resolution background modeling whereas images of perspective cameras are used for modeling foreground or specific objects. In such scenarios, since the omnidirectional camera views a common scene with different perspective cameras which do not have a common view in between, omnidirectional view is able to combine the partial 3D structures obtained by different perspective cameras.

Considering surveillance applications, hybrid systems were proposed where pan-tilt-zoom cameras are directed according to the information obtained by an omnidirectional camera which performs event detection [4–6]. Such systems can be enhanced by adding 3D structure and location estimation algorithms without increasing the number of cameras.

While working with such hybrid camera systems, for consecutive steps of structure-from-motion such as point matching, pose estimation, triangulation and

bundle adjustment, we need to modify the approaches that are used in systems using one type of camera.

1.2 Previous Work on Hybrid Systems

Structure-from-motion (SfM) with perspective cameras have been studied for a few decades and an extensive summary of algorithms can be found in [10]. For omnidirectional cameras, SfM is performed by several researchers [9, 11–13] and these studies include both calibrated and uncalibrated systems.

There are comparatively fewer studies on hybrid systems. Chen *et al.* [14] worked on the exterior calibration of a perspective-catadioptric camera system. They first calibrated the catadioptric camera, then using pre-measured 3D points in the scene, they performed the exterior calibration of perspective cameras viewing the same scene. Adorni *et al.* [15] used a hybrid system for obstacle detection problem in robot navigation. Chen and Yang [16] developed a region matching algorithm for hybrid views based on planar homography.

Epipolar geometry between hybrid camera views was explained by Sturm [17] for mixtures of paracatadioptric (catadioptric camera with a parabolic mirror) and perspective cameras. Barreto and Daniilidis showed that the framework can also be extended to cameras with lens distortion [18]. Recently, Sturm and Barreto [19] extended these relations to the general catadioptric camera model, which is valid for all central catadioptric cameras.

Puig *et al.* [20] worked on feature point matching and fundamental matrix estimation between perspective and catadioptric camera images. For point matching, they first applied a catadioptric-to-panoramic conversion and directly applied Scale Invariant Feature Transform (SIFT, [21]) between panoramic and perspective views. To eliminate the false matches they employed random sample consensus (RANSAC, [22]) based on satisfying the epipolar constraint. They also compared the representation capabilities of 3x4, 3x6 and 6x6 hybrid fundamental matrices (with different coordinate lifting) for mirrors with varying parameters.

Ramalingam *et al.* [23] conducted a study on hybrid SfM. They used manually selected feature point correspondences to estimate epipolar geometry and mentioned that directly applying SIFT did not provide good results for their fisheye-perspective image pairs. They employed midpoint method for triangu-

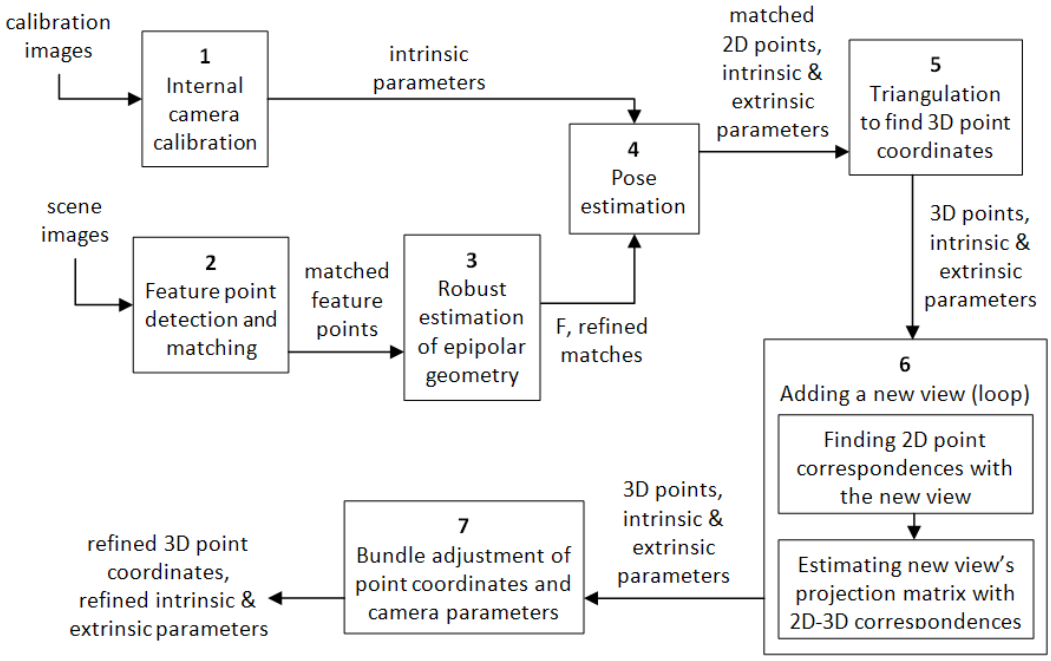


Figure 1.1: Steps of the implemented structure-from-motion pipeline.

lation to estimate 3D point coordinates and tested two different bundle adjustment approaches, one minimizing the distances between projection rays and 3D points and the other minimizing reprojection error. Their conclusion is that both approaches are comparable to each other. They employed a highly generic non-parametric imaging model, by which cameras are modeled with sets of projection rays. Internal calibration of cameras was performed by the method given in [24].

In the next section, the scope of this thesis is explained and related to the previous studies. Also, in the succeeding chapters, literature summaries regarding the context of the present chapter are given.

1.3 Thesis Study

In this thesis, a pipeline for structure-from-motion with mixed camera types is described and methods are proposed for the steps of this pipeline to make it robust and automatic. These steps can be summarized as camera calibration, point matching, epipolar geometry and pose estimation, triangulation and bundle adjustment (Fig. 1.1).

To represent mixed types of cameras, we employ the sphere camera model [25]

which is able to cover single viewpoint catadioptric systems as well as perspective cameras. Therefore, the SfM pipeline described here is generic for all the cameras that can be modeled with the sphere model. Moreover, the proposed methods for point matching and triangulation should work successfully also with the cameras beyond the scope of the sphere camera model, since their applicability is not related to the camera model.

The general imaging model used in [23] is more general than the sphere model because it encompasses non-central catadioptric cameras as well. However, there are some challenges such as defining a reprojection error for non-parametric model since there is no analytical projection equation. Another disadvantage is that, since the projection rays are represented with Plücker coordinates, the essential matrix extends to a 6x6 matrix with a special form requiring 17 point correspondences to be estimated linearly.

For calibration, we developed a calibration technique that is valid for all single-viewpoint catadioptric cameras. We are able to represent the projection of 3D points on a catadioptric image linearly with a 6x10 projection matrix, which uses lifted coordinates for the image and 3D points. This projection matrix can be computed with an adequate number of 3D-2D correspondences. We show how to decompose it to obtain intrinsic and extrinsic parameters. Moreover, we use this parameter estimation followed by a non-linear optimization to calibrate various types of cameras. When compared to the alternative sphere model calibration method [26], the proposed algorithm brings the advantage of linear and automatic parameter initialization.

For feature matching between the images of different camera types, widely accepted matching methods (eg. Scale Invariant Feature Transform: SIFT [21], Maximally Stable Extremal Regions: MSER [27]) do not perform well when they are directly employed for hybrid camera images [20,23]. In this thesis, we employ SIFT to match feature points between omnidirectional and perspective images and we propose a method to improve the SIFT matching result by preprocessing the input images. We performed tests on both catadioptric and fish-eye cameras and it is observed that, with the proposed approach, omnidirectional-to-perspective matching performance significantly increases. We also evaluate the use of virtual camera plane (VCP) images and observe that, for catadioptric cameras, VCP-to-perspective matching is more robust to increasing baseline.

Computation of hybrid fundamental matrix (F) was previously explained [17, 18] and random sample consensus (RANSAC) was implemented for catadioptric-perspective image pairs [20]. We define normalization matrices for lifted coordinates so that normalization and denormalization can be performed linearly. We present the results of our experiments on robust estimation of F to evaluate outlier elimination and effect of normalization.

We compare two options for pose estimation (extraction of motion parameters), one is directly estimating the essential matrix (E) with the calibrated 3D rays, the other option is estimating hybrid F and then extracting E from it. We performed experimental analysis to compare the effectiveness of these options.

For triangulation, we propose a weighting strategy for *iterative linear-Eigen* triangulation method to improve its 3D location estimation accuracy when employed for hybrid image pairs. The only previous study including hybrid camera triangulation was using the midpoint method [23]. However it has been shown that iterative linear methods are superior to midpoint method and non-iterative linear methods [28].

We perform multi-view SfM by employing the approach of adding views to the structure one by one [29]. Sparse bundle adjustment method [30] has become popular in the community due to its capability of solving enormous minimization problems (with many cameras and 3D points) in a reasonable time. We employed this method for multi-view hybrid SfM by modifying the projection function with sphere model projection and intrinsic parameters with sphere model parameters.

We also demonstrate the complete hybrid multi-view SfM with real images including the bundle adjustment step. We emphasize the case where the omnidirectional camera is able to combine structures estimated with different perspective cameras which do not have a view in common.

1.4 Contributions of the Thesis

• We developed a camera calibration technique for the sphere camera model. Its advantage: Initialization of intrinsic parameters is performed linearly without requiring a user input.

• We propose an improved feature point matching method which enables automatic omnidirectional-perspective matching.

• We propose a weighting strategy for iterative linear-Eigen triangulation method to improve its 3D structure estimation accuracy.

• We modified and tested existing approaches for hybrid SfM: Normalization, Pose estimation, Multi-view SfM, Sparse bundle adjustment.

1.5 Road Map

Chapter 2 provides background information on omnidirectional vision. After giving an introduction to omnidirectional vision and catadioptric cameras, sphere camera model is introduced to the reader, which is the model we employ to represent our hybrid cameras.

Chapter 3 begins with the literature survey on catadioptric camera calibration. Then, it presents the details of the developed calibration method for the sphere camera model together with the experiment results.

Chapter 4 presents the proposed feature point matching algorithm for mixed camera images. It first explains why SIFT is not effective when applied directly and how we solve the problem. Then the results of experiments with real images of catadioptric and fish-eye cameras are presented.

Chapter 5 focuses on the epipolar geometry and pose estimation steps of the SfM pipeline. The implementation of RANSAC on hybrid epipolar geometry and normalization of *lifted coordinates* are explained. Moreover, the experimental comparison of the options for pose estimation (extraction of motion parameters) is given.

Chapter 6 presents the proposed weighting strategy for *iterative linear-Eigen* triangulation method and shows its effectiveness to increase the 3D structure estimation performance with simulated images. Also, a two-view hybrid SfM experiment is presented in this chapter in order to evaluate the proposed triangulation approach.

Chapter 7 gives the details of the work on multi-view SfM with hybrid images. The improvement gained by sparse bundle adjustment is also given in this chapter. The proof of concept given in this thesis is presented with experiments where the complete pipeline of hybrid SfM is realized.

Finally, Chapter 8 presents conclusions and suggests future works.

chapter 2

Background on Omnidirectional Imaging

In this chapter, we introduce the omnidirectional vision to the reader and briefly explain image formation in catadioptric omnidirectional cameras, which will serve as a background information for the following chapters. We also explain the sphere camera model (Section 2.2) which is able to represent all single-viewpoint catadioptric cameras.

2.1 Introduction to Omnidirectional Vision

The term "omnidirectional" is used for the cameras that have very large fields of view. An omnidirectional viewing device ideally has the capability of viewing 360° in all directions. It is not practical to produce a "true" omnidirectional sensor, therefore manufactured cameras usually provide 360° horizontal view and a sufficient field of vertical view. Fish-eye lenses also have extended field of views up to a hemisphere and are used for omnidirectional viewing. However, most of the omnidirectional cameras are catadioptric systems which means they use combinations of mirrors and lenses. The term "catadioptrics" comprises "catoptrics"; the science of reflecting surfaces (mirrors) and "dioptrics"; the science of refracting elements (lenses). Rees [31] is the first to patent a catadioptric omnidirectional capturing system using a hyperboloidal mirror and a normal perspective camera in 1970. Since then, considerable amount of effort have been spent on the design of mirrors with enlarged vertical field of views, low cost and varying resolution properties. Among them, Nayar and Peri [32] worked on folded mirror systems that use multiple mirrors in order to obtain smaller omnidirectional devices with wider views. Conroy and Moore [33] derived mirror surfaces that are resolution invariant vertically, so that adjacent pixels in omnidirectional image correspond

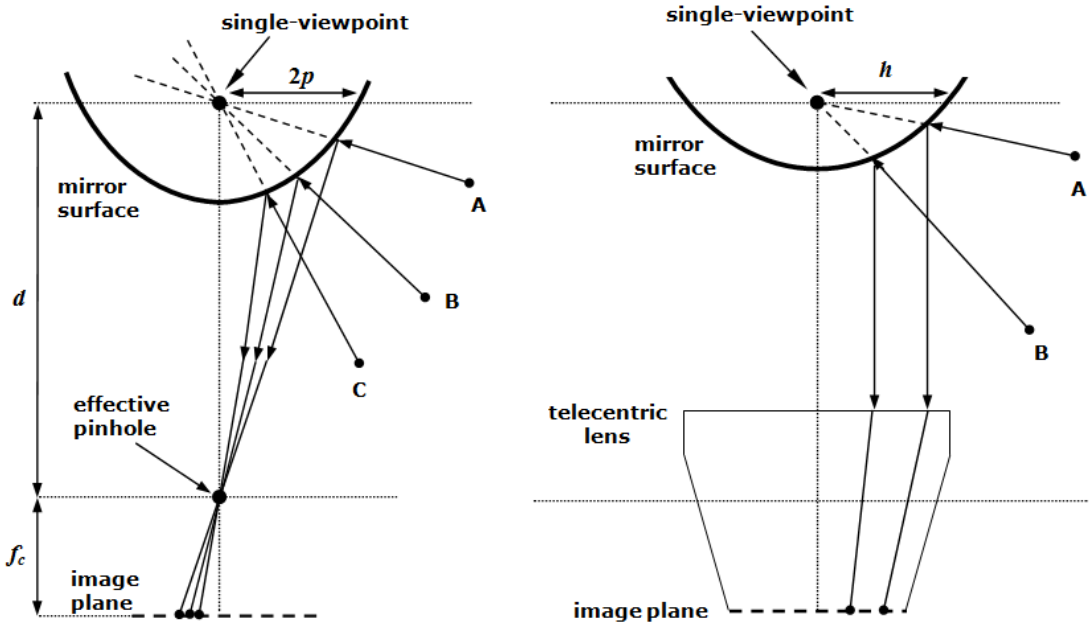
to the real world points that are vertically equi-distant from each other. In that paper, stereo omnidirectional systems are also introduced. They are constructed by two coaxial, axially symmetric mirror profiles. Hicks and Bajcsy [34] exhibited a mirror design that views wide horizontal area under the mirror and reflects an undistorted (perspective) omnidirectional image of this area. Gaspar *et al.* [35] summarizes the constant horizontal, vertical and angular resolution issues and presents a mirror that achieves uniform resolution when used with a specific log-polar camera. Swaminathan *et al.* [36] discusses the design issues of mirrors that minimize image errors.

2.1.1 Single-viewpoint Property

Catadioptric systems, combinations of camera lenses and mirrors, are able to provide single-viewpoint property if the mirror has a focal point which can behave like an effective pinhole. For instance, in the mirror shown in Fig. 2.1a, light rays coming from the world points A, B and C and targeting the focal point (single viewpoint) of the hyperboloidal mirror are reflected on the mirror surface so that they will pass through the pinhole (camera center). This single viewpoint acts a virtual pinhole through which the scene is viewed as in regular perspective cameras. Paraboloidal mirrors also have single-viewpoint property, but the rays targeting that viewpoint are reflected orthogonally, which requires the use of a telecentric lens to collect the parallel rays (Figure 2.1b). Single viewpoint constraint provides quick conversion of geometrically correct panoramic and perspective images because they are generated as seen from the mentioned viewpoint.

Since the cross sectional profiles of the mirrors of catadioptric sensors do not change when rotated around the optical axis, the cross sections of the mirrors are shown in the figures. Usually mirrors are referred with the names of these 2D cross-sections such as parabolic and hyperbolic mirrors. Catadioptric systems are often referred with the associated mirror type such as para-catadioptric or hyper-catadioptric systems.

For the system shown in Fig. 2.2, directions of the light rays that are used for image formation do not intersect at a certain point as in the single-viewpoint case, therefore a single point through which the scene can be viewed cannot be



(a) (b)

Figure 2.1: Directions of light rays in two single-viewpoint systems, (a) hyper-catadioptric and (b) para-catadioptric.

defined. Although single-viewpoint mirrors are desirable for efficient projection generation, non single-viewpoint mirrors may be preferred to achieve uniform resolution and wider field of view and/or ease of manufacturing. Spherical and conical mirrors are two examples which are in practical use in the catadioptric systems. In addition to the non single-viewpoint mirrors, errors in manufacturing and incorrectly aligned systems could cause single-viewpoint mirrors behave as non single-viewpoint.

Geometric properties of single-viewpoint cameras were examined by Baker and Nayar [37]. Swaminathan *et al.* [38] conducted a detailed study on the geometry of non-single-viewpoint systems. There also exist studies for approximating a viewpoint in non-single-viewpoint systems as Derrien and Konolige proposed for spherical mirrors [39].

2.1.2 Fish-eye Lenses

A fish-eye lens is a dioptric system comprises several lenses to reduce the incident angle of rays (US Patent 4,412,726 designed by M. Horimoto at Minolta). Fish-eye lenses sometimes provide convenient and practical omnidirectional vision

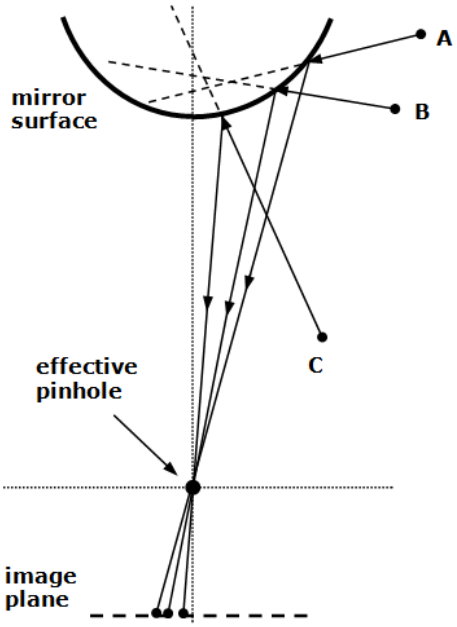


Figure 2.2: Directions of light rays in a non single-viewpoint system.

for computer vision applications. Fig. 2.3 shows the projection geometry of a fish-eye lens. Light rays passing through the adjacent pixels in the image belong to the diverging rays coming from outside. With this ability, a lens can capture more than a hemisphere (Eg. Fujinon Fisheye 185, FE185C046HA-1).

Fish-eye lenses belong to the non single-viewpoint family. A simple and common way to model the projection of these lenses is employing equi-distance projection model. It preserves equal distances in image plane for equal vertical angles between the diverging light rays:

$$r = f\theta$$

where θ is the angle between the optical axis and the incoming light ray, r is the distance between the image point and the principal point and f is the focal length. However, produced lenses do not exactly conform to the model and for accurate calibration polynomial models are used to map image pixels (r) to the incoming light rays (θ). Ho *et al.* [40] gives more information about history and physics of fisheye lenses. They also present their test results with different projection models.

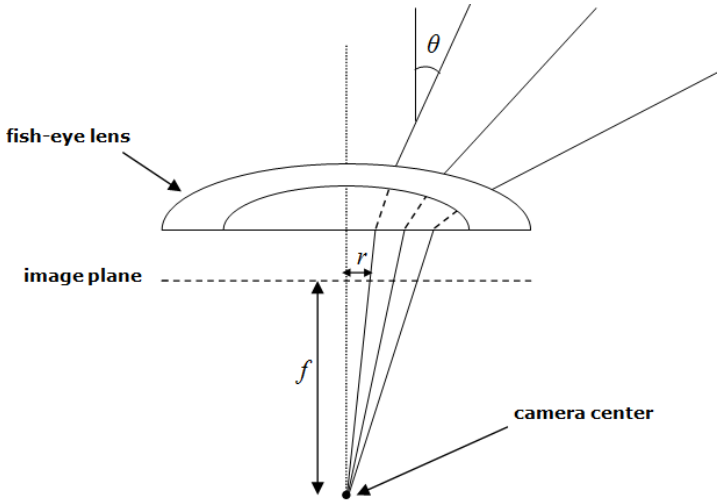


Figure 2.3: Projection geometry of a fish-eye lens.

2.2 Sphere Camera Model

Now, we briefly explain the sphere model for catadioptric projection introduced by Geyer and Daniilidis [25]. Later, Barreto and Daniilidis [18] showed that the framework can also be extended to cameras with lens distortions. In the following, matrices are represented by symbols in sans serif font, e.g. M and vectors by bold symbols, e.g. $\mathbf{Q}$, $\mathbf{q}$. Equality of matrices or vectors up to a scalar factor is written as $\sim$.

According to this model, all central catadioptric cameras can be modeled by a unit sphere and a perspective camera, such that the projection of 3D points can be performed in two steps (Fig. 2.2). First one is the projection of point $\mathbf{Q}$ in 3D space onto a unitary sphere and second one is the projection from the sphere to the image plane. First projection gives rise to two intersection points on the sphere, $\mathbf{r}_{\pm}$. The one that is visible to us is $\mathbf{r}_{+}$ and its projection on the image plane is $\mathbf{q}_{+}$. This model covers all central catadioptric cameras, denoted by ξ, which is the distance between the camera center and the center of the sphere. $\xi = 0$ for perspective, $\xi = 1$ for para-catadioptric, $0 < \xi < 1$ for hyper-catadioptric cameras.

Let the unit sphere be located at the origin and the optical center of the perspective camera be located at the point $\mathbf{C}_p = (0, 0, -\xi)^{\mathsf{T}}$ making z-axis positive downwards. The perspective projection from the sphere to the image plane is

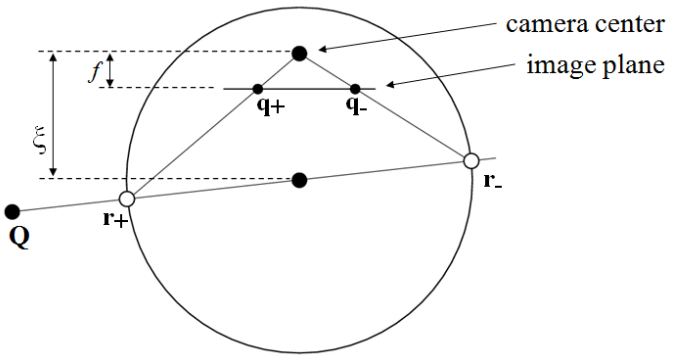


Figure 2.4: Projection of a 3D point to two image points in sphere camera model. Camera is looking down, accordingly z-axis of the camera coordinate system is positive downwards.

modeled by the projection matrix $\mathsf{P} \sim \mathsf{K}\begin{pmatrix}\mathtt{I} & -\mathbf{C}_p\end{pmatrix}$, where K is the calibration matrix of perspective camera embedded in the sphere model. To explain the projection in detail, let the intersection point of the sphere and the line joining its center and $\mathbf{Q}$ be

$$\mathbf{r}_+ = \left(Q_1, Q_2, Q_3, \sqrt{Q_1^2 + Q_2^2 + Q_3^2}\right)^\mathsf{T}$$

in 4-vector homogeneous coordinates. The same point represented with respect to the camera center, $\mathbf{C}_p$, in non-homogeneous coordinates is

$$\mathbf{b}_+ = \left(Q_1, Q_2, Q_3 + \xi\sqrt{Q_1^2 + Q_2^2 + Q_3^2}\right)^\mathsf{T}$$

Then, the image of the point in the perspective camera is

$$\mathbf{q}_+ \sim \mathsf{K}\mathbf{b}_+ \sim \mathsf{K}\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 + \xi\sqrt{Q_1^2 + Q_2^2 + Q_3^2}\end{pmatrix} \tag{2.1}$$

Therefore, intrinsic parameters of this model are ξ and K. Please note that in this formulation no rotation is modeled between sphere axis and perspective camera inside. This fact is referred as *tilting* and discussed in Section 2.2.1 where we also give the relation between the focal length of the sphere model and the actual camera focal length.

Back-projection.

Computing the outgoing 3D ray corresponding to an image point can be performed in two steps. First step is obtaining $\mathbf{b}$ from $\mathbf{q}$ and similar to any perspective camera it is computed by $\mathbf{b} = \mathsf{K}^{-1}\mathbf{q}$. Second step is carrying back the origin of coordinate system from the camera center to the center of the sphere, i.e. passing from $\mathbf{b}$ to $\mathbf{r}$. Let (x, y, z) represent the 3D ray $\mathbf{b}$, then the sphere centered 3D ray, $\mathbf{r}$, can be computed with the below equation [41]:

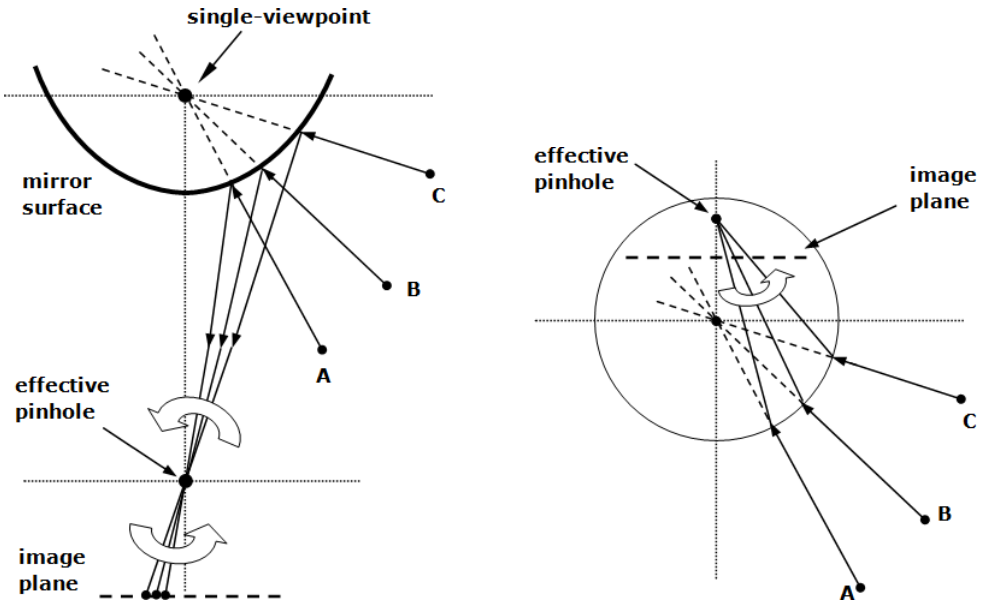
$$\mathbf{r} = \begin{pmatrix} \frac{z\xi+\sqrt{z^2+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+z^2}x \\ \frac{z\xi+\sqrt{z^2+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+z^2}y \\ \frac{z\xi+\sqrt{z^2+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+z^2}z - \xi \end{pmatrix} \tag{2.2}$$

2.2.1 Relation between the Real Catadioptric System and the Sphere Camera Model

In this section we analyze the relation between the parameters present in a real catadioptric system and their representation in the sphere camera model. The objective of this analysis is to observe if it is possible to recover the intrinsic parameters of the real catadioptric system from their counterpart in the sphere camera model.

Tilting.

Tilting in a camera can be defined as a rotation of the image plane w.r.t. the pinhole. This is also equivalent to tilting the incoming rays since both have the same pivoting point: the pinhole. In Fig. 2.5a, the tilt in a catadioptric camera is represented. Similarly, tilt in sphere model corresponds to tilting the rays coming to the perspective camera in the sphere model (Fig. 2.5b). Although the same image is generated by both models, the angles of the rays going through the effective pinholes are not the same, they are not even proportional to each other. So, it is also not possible to obtain the real system tilt amount by multiplying the sphere model tilt by a coefficient.



(a) (b)

Figure 2.5: Tilt in a real system (a) and in the sphere model (b).

Focal length.

The compositions of para-catadioptric and hyper-catadioptric systems are different. The first one uses a parabolic mirror and an orthographic camera with a telecentric lens. In this case the focal length of the real system, f_c, is infinite. To represent this system with the sphere model, we equalize f, the focal length of camera in the sphere model, to h, the mirror parameter in terms of pixels.

For hyper-catadioptric systems, we are able to relate f with the focal length of the perspective camera in the real system, f_c. We start with defining explicitly the projection matrix K of Eq. 2.1. Assuming image skew is zero and principal point is $(0, 0)$, K is given in [41] as

$$\mathsf{K} = \begin{pmatrix} (\psi - \xi) f_c & 0 & 0 \\ 0 & (\psi - \xi) f_c & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.3}$$

where ψ is defined as the distance between the camera center and image plane. The relation between focal lengths is $f = (\psi - \xi) f_c$. From the same study [41] we get

$$\xi = \frac{d}{\sqrt{d^2+4p^2}} \qquad \psi = \frac{d+2p}{\sqrt{d^2+4p^2}}. \tag{2.4}$$

where d is the distance between the foci of hyperbola and $4p$ equals to the latus rectum (Fig. 2.1). Developing the equations we obtain p in terms of d and ξ, $2p = \frac{d\sqrt{1-\xi^2}}{\xi}$ which is used to obtain $\psi = \xi + \sqrt{1-\xi^2}$. With this final relation we can write

$$f = (\sqrt{1-\xi^2})f_c \tag{2.5}$$

which shows that computing sphere model parameters, f and ξ, gives us the focal length of the perspective camera in the real system.

CHAPTER 3

DLT-Based Calibration of Sphere Camera Model

In this chapter, we present a calibration technique that is valid for all single-viewpoint catadioptric cameras. First we review the literature on catadioptric camera calibration in Section 3.1. Then, based on the introduction given in Section 2.2 for the sphere camera model, we explain the proposed calibration technique which estimates the parameters of the sphere camera model. Finally, in Sections 3.3 and 3.4, we present the results of experiments for the proposed calibration approach using both simulated and real images.

3.1 Literature on Catadioptric Camera Calibration

Several methods were proposed for calibration of catadioptric systems. Some of them consider estimating the parameters of the parabolic [42, 43], hyperbolic [44] and conical [45] mirrors together with the camera parameters. Calibration of outgoing rays based on a radial distortion model is another approach. Kannala and Brandt [46] used this approach to calibrate fisheye cameras. Scaramuzza *et al.* [47] extended the approach to include central catadioptric cameras as well. Mei and Rives [26], on the other hand, developed another Matlab calibration toolbox that estimates the parameters of the sphere camera model. Parameter initialization is performed with user input. The user defines the location of the principal point and depicts a real world straight line in the omnidirectional image which is used for focal length estimation.

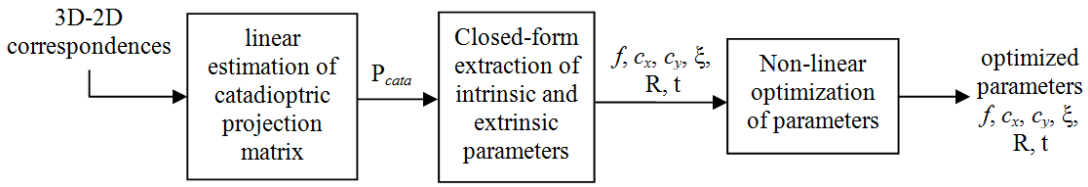


Figure 3.1: Block diagram of the proposed calibration technique.

3.2 Proposed Calibration Technique

Recently, Sturm and Barreto [19] showed that employing the sphere camera model, the catadioptric projection of a 3D point can be modeled using a projection matrix of size 6×10. The calibration method presented here puts this theory into practice. We compute the generic projection matrix, P_{cata}, with 3D-2D correspondences, using a straightforward Direct Linear Transform (DLT) [48] approach which is based on a set of linear equations. Then, we decompose P_{cata} to estimate intrinsic and extrinsic parameters. With these estimates as initial values of system parameters, we optimize the parameters by minimizing the reprojection error. The steps of the algorithm are shown in Fig. 3.1. When compared to the technique of Mei and Rives [26], the only previous work on calibration of sphere camera model, our approach has the advantages of not requiring input for parameter initialization and being able to calibrate perspective cameras as well. On the other hand, our algorithm needs a 3D calibration object.

3.2.1 Mathematical Background on Coordinate Lifting

Lifted coordinates from symmetric matrix equations.

The derivation of (multi-) linear relations for catadioptric imagery requires the use of lifted coordinates. The Veronese map $V_{n,d}$ of degree d maps points of $\mathcal{P}^n$ into points of an m dimensional projective space $\mathcal{P}^m$, with $m = \binom{n+d}{d} - 1$.

Consider the second order Veronese map $V_{2,2}$, that embeds the projective plane into the 5D projective space, by lifting the coordinates of point $\mathbf{q}$ to

$$\hat{\mathbf{q}} = \begin{pmatrix} q_1^2 & q_1 q_2 & q_2^2 & q_1 q_3 & q_2 q_3 & q_3^2 \end{pmatrix}^{\mathsf{T}}$$

Vector $\hat{\mathbf{q}}$ and matrix $\mathbf{q}\mathbf{q}^\mathsf{T}$ are composed by the same elements. The former can be derived from the latter through a suitable re-arrangement of parameters. Define $\mathbf{v}(\mathsf{U})$ as the vector obtained by stacking the columns of a generic matrix U [49]. For the case of $\mathbf{q}\mathbf{q}^\mathsf{T}$, $\mathbf{v}(\mathbf{q}\mathbf{q}^\mathsf{T})$ has several repeated elements because of the matrix symmetry. By left multiplication with a suitable permutation matrix P that adds the repeated elements, it follows that

$$\hat{\mathbf{q}} = \mathsf{D}^{-1} \underbrace{\begin{pmatrix} 1&0&0&0&0&0&0&0&0 \\ 0&1&0&1&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&0&1&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0&1&0 \\ 0&0&0&0&0&0&0&0&1 \end{pmatrix}}_{\mathsf{P}} \mathbf{v}(\mathbf{q}\mathbf{q}^\mathsf{T}), \tag{3.1}$$

with D a diagonal matrix, $D_{ii} = \sum_{j=1}^{9} P_{ij}$.

If U is symmetric, then it is uniquely represented by $\mathbf{v}_{sym}(\mathsf{U})$, the row-wise vectorization of its lower left triangular part:

$$\mathbf{v}_{sym}(\mathsf{U}) = \mathsf{D}^{-1}\mathsf{P}\mathsf{U} = (U_{11}, U_{21}, U_{22}, U_{31}, \cdots, U_{nn})^\mathsf{T}$$

Lifted matrices.

Let us now discuss the lifting of linear transformations. Consider a matrix A to transform a vector $\mathbf{q}$ such that $\mathbf{r} = \mathsf{A}\mathbf{q}$. The relation $\mathbf{r}\mathbf{r}^\mathsf{T} = \mathsf{A}(\mathbf{q}\mathbf{q}^\mathsf{T})\mathsf{A}^\mathsf{T}$ can be written as a vector mapping

$$(\mathbf{r}\mathbf{r}^\mathsf{T}) = (\mathsf{A} \otimes \mathsf{A})(\mathbf{q}\mathbf{q}^\mathsf{T}),$$

with $\otimes$ denoting the Kronecker product [49]. Using the symmetric vectorization, we have $\hat{\mathbf{q}} = \mathbf{v}_{sym}(\mathbf{q}\mathbf{q}^\mathsf{T})$ and $\hat{\mathbf{r}} = \mathbf{v}_{sym}(\mathbf{r}\mathbf{r}^\mathsf{T})$, thus:

$$\hat{\mathbf{r}} = \underbrace{\mathsf{D}^{-1}\mathsf{P}(\mathsf{A} \otimes \mathsf{A})\mathsf{P}^\mathsf{T}}_{\widehat{\mathsf{A}}} \hat{\mathbf{q}} \tag{3.2}$$

where $\widehat{\mathsf{A}}$ represents the lifted linear transformation.

Another representation for $\widehat{\mathsf{A}}$ is given in the following. Let $\mathbf{a}_i$ be the columns of A. Then, employing Eq. 3.1,

$$\widehat{\mathsf{A}} = \mathsf{D}^{-1}\mathsf{P} \begin{pmatrix} \mathbf{v}(\mathbf{a}_1\mathbf{a}_1^\mathsf{T}) & 2\mathbf{v}(\mathbf{a}_1\mathbf{a}_2^\mathsf{T}) & \mathbf{v}(\mathbf{a}_2\mathbf{a}_2^\mathsf{T}) & 2\mathbf{v}(\mathbf{a}_1\mathbf{a}_3^\mathsf{T}) & 2\mathbf{v}(\mathbf{a}_2\mathbf{a}_3^\mathsf{T}) & \mathbf{v}(\mathbf{a}_3\mathbf{a}_3^\mathsf{T}) \end{pmatrix}$$

A few useful properties of the lifting of transformations are [49, 50]:

$$\widehat{\mathsf{AB}} = \widehat{\mathsf{A}}\widehat{\mathsf{B}} \quad \widehat{\mathsf{A}^{-1}} = \widehat{\mathsf{A}}^{-1} \quad \widehat{\mathsf{A}^{\mathsf{T}}} = \mathsf{D}^{-1}\widehat{\mathsf{A}}^{\mathsf{T}}\mathsf{D} \tag{3.3}$$

In our work, we use the following liftings: 3-vectors $\mathbf{q}$ to 6-vectors $\hat{\mathbf{q}}$ and 4-vectors $\mathbf{Q}$ to 10-vectors $\hat{\mathbf{Q}}$. Analogously, 3×3 matrices are lifted to 6×6 and 3×4 matrices to 6×10.

3.2.2 Generic Projection Matrix

As explained in Section 2.2, a 3D point is mathematically projected to two image points. Sturm and Barreto [19] represented these two 2D points via the degenerate dual conic generated by them, i.e. the dual conic containing exactly the lines going through at least one of the two points. Let the two image points be $\mathbf{q}_+$, $\mathbf{q}_-$, and the dual conic is given by

$$\Omega \sim \mathbf{q}_+\mathbf{q}_-^{\mathsf{T}} + \mathbf{q}_-\mathbf{q}_+^{\mathsf{T}}$$

The vectorized matrix of the conic can be computed as shown below using the lifted 3D point coordinates, intrinsic and extrinsic parameters.

$$\mathbf{v}_{sym}(\Omega) \sim \widehat{\mathsf{K}}_{6\times6}\mathsf{X}_\xi\widehat{\mathsf{R}}_{6\times6}\begin{pmatrix}\mathtt{I}_6 & \mathsf{T}_{6\times4}\end{pmatrix}\hat{\mathbf{Q}}_{10} \tag{3.4}$$

Here, R represents the rotation of the catadioptric camera. X_ξ and $\mathsf{T}_{6\times4}$ depend only on the sphere model parameter ξ and position of the catadioptric camera $\mathbf{C} = (t_x, t_y, t_z)$ respectively, as shown here:

$$\mathsf{X}_\xi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -\xi^2 & 0 & -\xi^2 & 0 & 0 & 1-\xi^2 \end{pmatrix} \qquad \mathsf{T}_{6\times4} = \begin{pmatrix} -2t_x & 0 & 0 & t_x^2 \\ -t_y & -t_x & 0 & t_xt_y \\ 0 & -2t_y & 0 & t_y^2 \\ -t_z & 0 & -t_x & t_xt_z \\ 0 & -t_z & -t_y & t_yt_z \\ 0 & 0 & -2t_z & t_z^2 \end{pmatrix}$$

Thus, a 6×10 **catadioptric projection matrix**, P_{cata}, can be expressed by its intrinsic and extrinsic parameters, as in the case of a perspective camera:

$$\mathsf{P}_{cata} = \underbrace{\widehat{\mathsf{K}}\mathsf{X}_{\xi}}_{\mathsf{A}_{cata}} \underbrace{\widehat{\mathsf{R}}_{6\times 6}\begin{pmatrix}\mathtt{I}_6 & \mathsf{T}_{6\times 4}\end{pmatrix}}_{\mathsf{T}_{cata}} \tag{3.5}$$

3.2.3 Computation of the Generic Projection Matrix

Here we show the way used to compose the equations using 3D-2D correspondences to compute P_{cata}. Analogous to the perspective case ($[\mathbf{q}]_{\times}\mathsf{P}\mathbf{Q} = \mathbf{0}$), we write the constraint based on the lifted coordinates [19]:

$$\widehat{[\mathbf{q}]_{\times}}\ \mathsf{P}_{cata}\ \hat{\mathbf{Q}} = \mathbf{0}$$

where $[\mathbf{q}]_{\times}$ denotes the skew-symmetric matrix associated with the cross product of 3-vector $\mathbf{q}$. This is a set of 6 linear homogeneous equations in the coefficients of P_{cata}. Using the Kronecker product, this can be written in terms of the 60-vector $\mathbf{p}_{cata}$ containing the 60 coefficients of P_{cata}:

$$\left(\widehat{[\mathbf{q}]_{\times}} \otimes \hat{\mathbf{Q}}^{\mathsf{T}}\right)\mathbf{p}_{cata} = \mathbf{0}_6$$

Stacking these equations for n 3D-2D correspondences gives an equation system of size $6n \times 60$, which we solve to least squares. Note that the minimum number of required correspondences is 20: a 3×3 skew symmetric matrix has rank 2, its lifted counterpart rank 3. Therefore, each correspondence provides only 3 independent linear constraints.

Another observation is that the 3D points should be distributed on at least three different planes. Here follows a proof of why points on two planes are not sufficient to compute P_{cata} using linear equations [51]. Let $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ be the two planes. Hence, each calibration point $\mathbf{Q}$ satisfies $\left(\mathbf{\Pi}_1^{\mathsf{T}}\mathbf{Q}\right)\left(\mathbf{\Pi}_2^{\mathsf{T}}\mathbf{Q}\right) = 0$. This can be written as a linear constraint on the lifted calibration points: $\mathbf{p}^{\mathsf{T}}\hat{\mathbf{Q}} = 0$, where the 10-vector $\mathbf{p}$ depends exactly on the two planes. Thus, if P_{cata} is the true 6×10 projection matrix, then adding some multiple of $\mathbf{p}^{\mathsf{T}}$ to any row of P_{cata} gives another 6×10 projection matrix, $\bar{\mathsf{P}}_{cata}$, which maps the calibration points to the same image entities as the true projection matrix.

$$\bar{\mathsf{P}}_{cata} = \mathsf{P}_{cata} + \mathbf{v}\mathsf{p}^{\mathsf{T}}$$

where $\mathbf{v}$ is a 6-vector and represents the 6-dof on P_{cata} that can not be recovered using only linear projection equations and calibration points located in only two planes. For three planes, there is no linear equation as above that holds for all calibration points.

3.2.4 Decomposition of the Generic Projection Matrix

The calibration process consists of getting the intrinsic and extrinsic parameters of a camera. Our purpose is to decompose P_{cata} as in Eq. (3.5). Consider first the leftmost 6×6 submatrix of P_{cata}:

$$\mathsf{P}_s \sim \widehat{\mathsf{K}}\mathsf{X}_\xi\widehat{\mathsf{R}}$$

Let us define $\mathsf{M} = \mathsf{P}_s\mathsf{D}^{-1}\mathsf{P}_s^\mathsf{T}$. Using the properties given in Eq. (3.3) and knowing that for a rotation matrix $\mathsf{R}^{-1} = \mathsf{R}^\mathsf{T}$, we can write $\widehat{\mathsf{R}}^{-1} = \mathsf{D}^{-1}\hat{\mathsf{R}}^\mathsf{T}\mathsf{D}$. And from that we obtain $\mathsf{D}^{-1} = \hat{\mathsf{R}}\mathsf{D}^{-1}\hat{\mathsf{R}}^\mathsf{T}$ which we use to eliminate the rotation parameters:

$$\mathsf{M} \sim \widehat{\mathsf{K}}\mathsf{X}_\xi\widehat{\mathsf{R}}\,\mathsf{D}^{-1}\widehat{\mathsf{R}}^\mathsf{T}\mathsf{X}_\xi^\mathsf{T}\widehat{\mathsf{K}}^\mathsf{T} = \widehat{\mathsf{K}}\mathsf{X}_\xi\,\mathsf{D}^{-1}\mathsf{X}_\xi^\mathsf{T}\widehat{\mathsf{K}}^\mathsf{T} \tag{3.6}$$

The above equation holds up to scale, i.e. there is a λ with $\mathsf{M} = \lambda\widehat{\mathsf{K}}\mathsf{X}_\xi\,\mathsf{D}^{-1}\mathsf{X}_\xi^\mathsf{T}\widehat{\mathsf{K}}^\mathsf{T}$. We use some elements of M to extract the intrinsic parameters:

$$\begin{aligned}
\mathsf{M}_{16} &= \lambda\left(-(f^2\xi^2) + c_x^2(\xi^4 + c_x(1-\xi^2)^2\right)\\
\mathsf{M}_{44} &= \lambda\left(\frac{f^2}{2} + c_x^2(2\xi^4 + (1-\xi^2)^2)\right)\\
\mathsf{M}_{46} &= \lambda c_x(2\xi^4 + (1-\xi^2)^2)\\
\mathsf{M}_{56} &= \lambda c_y(2\xi^4 + (1-\xi^2)^2)\\
\mathsf{M}_{66} &= \lambda\left(2\xi^4 + (1-\xi^2)^2\right)
\end{aligned}$$

Note that for the initial computation of intrinsic parameters, we suppose that there is no tilt in the catadioptric camera, i.e., the perspective camera is not rotated away from the mirror. We compute the following 4 intrinsic parameters: ξ, f, c_x, c_y. The last three are the focal length and principal point coordinates of the perspective camera in the sphere model. After initialization, the parameters of tilt and distortion are also estimated by non-linear optimization (Section 3.2.5).

Since we obtained M up to a scale, to compute the parameters we should use ratios between the entries of matrix M. The intrinsic parameters are computed as follows:

$$c_x = \frac{\mathsf{M}_{46}}{\mathsf{M}_{66}} \qquad c_y = \frac{\mathsf{M}_{56}}{\mathsf{M}_{66}} \qquad \xi = \sqrt{\frac{\frac{M_{16}}{M_{66}} - c_x^2}{-2(\frac{M_{44}}{M_{66}} - c_x^2)}}$$

$$f = \sqrt{2(2\xi^4 + (1-\xi^2)^2)\left(\frac{M_{44}}{M_{66}} - c_x^2\right)}$$

After extracting the intrinsic part A_{cata} of the projection matrix, we are able to obtain the 6×10 extrinsic part T_{cata} by multiplying P_{cata} with the inverse of A_{cata}:

$$\mathsf{T}_{cata} = \widehat{\mathsf{R}}_{6\times 6}\,(\,\mathtt{I}_6\,\mathsf{T}_{6\times 4}\,) \sim \left(\widehat{\mathsf{K}}\mathsf{X}_\xi\right)^{-1}\mathsf{P}_{cata} \tag{3.7}$$

So, the leftmost 6×6 part of T_{cata} will be the estimate of the lifted rotation matrix. And if we multiply the inverse of this $\widehat{\mathsf{R}}_{est}$ with the rightmost 6×4 part of T_{cata}, we obtain an estimate for the translation ($\mathsf{T}_{6\times 4}$). This translation should have an ideal form as given in Eq. (3.4) and we are able to identify translation vector elements (t_x, t_y, t_z) from it.

We extract the rotation angles around x, y and z axes one by one using $\widehat{\mathsf{R}}_{est}$. First, we recover the rotation angle around the z axis, $\gamma = tan^{-1}\left(\frac{\widehat{\mathsf{R}}_{est,51}}{\widehat{\mathsf{R}}_{est,41}}\right)$.

Then, $\widehat{\mathsf{R}}_{est}$ is modified by being multiplied by the inverse of rotation around z axis, $\widehat{\mathsf{R}}_{est} = \widehat{\mathsf{R}}_{z,\gamma}^{-1}\,\widehat{\mathsf{R}}_{est}$. Then, rotation angle around y axis, β, is estimated and $\widehat{\mathsf{R}}_{est}$ is modified $\beta = tan^{-1}\left(\frac{-\widehat{\mathsf{R}}_{est,52}}{\widehat{\mathsf{R}}_{est,22}}\right)$, $\widehat{\mathsf{R}}_{est} = \widehat{\mathsf{R}}_{y,\beta}^{-1}\,\widehat{\mathsf{R}}_{est}$

Finally, rotation angle around x axis, α, is estimated by $\alpha = tan^{-1}\left(\frac{\widehat{\mathsf{R}}_{est,42}}{\widehat{\mathsf{R}}_{est,22}}\right)$.

3.2.5 Other Parameters of Non-linear Calibration

The intrinsic and extrinsic parameters extracted linearly in Section 3.2.4 are not always adequate to model a real camera. Extra parameters are needed to correctly model the catadioptric system, namely, tilting and lens distortions.

In Section 2.2.1, we explained why tilt in real catadioptric camera is not equal to the tilt in the sphere camera model. However, we can still estimate tilting parameters to remove the effect of such an imperfection. To do this, we define a rotation, R_p, between camera center and sphere center. Tilting has only

R_x and R_y components, because rotation around optical axis, R_z, is merged with the external rotation around the z axis.

As well known, imperfections due to lenses are modeled as distortions for camera calibration. Radial distortion models contraction or expansion with respect to the image center and tangential distortion models lateral effects. To add these distortion effects to our calibration algorithm, we employed the approach of Heikkila and Silven [52].

Radial distortion is given as

$$\Delta x = x(k_1 r^2 + k_2 r^4 + k_3 r^6 + ..)\,, \qquad \Delta y = y(k_1 r^2 + k_2 r^4 + k_3 r^6 + ..) \tag{3.8}$$

where $r = \sqrt{x^2 + y^2}$ and $k_1, k_2..$ are the radial distortion parameters. We observed that estimating two parameters was sufficient for an adequate modeling.

Tangential distortion is given as

$$\Delta x = 2p_1 xy + p_2(r^2 + 2x^2)\,, \qquad \Delta y = p_1(r^2 + 2y^2) + 2p_2 xy \tag{3.9}$$

where $r = \sqrt{x^2 + y^2}$ and p_1, p_2 are the tangential distortion parameters.

We applied distortion after projecting the 3D points on the sphere and modifying the coordinates with ξ (also after applying tilting, if modeled). Thus, in a sense we distort the rays outgoing from the camera center. Next, distorted rays are projected to the image by employing camera intrinsic parameters.

Once we have identified all the parameters to be estimated we perform a non-linear optimization to compute the whole model. We use the Levenberg-Marquardt (LM) method provided by the function **lsqnonlin** in Matlab. The minimization criterion is the root mean square (RMS) of distance error between a measured image point and its reprojected correspondence. Since the projection equations we use (cf. Eq. 3.4) map 3D points to dual image conics, we have to extract the two potential image points from it; the one closer to the measured point is selected and then the reprojection error is measured. We take as initial values the parameters obtained from P_{cata} and initialize the additional distortion parameters with zero.

3.3 Calibration Experiments in a Simulated Environment

A simulated calibration object of 3 planar faces which are perpendicular to each other was used. Each face has 11x11 points and the distance between points is 5cm. So size of a face is 50x50 cm. and a total of 363 points exist. The omnidirectional image fits in a 1 Megapixel square image. To represent the real world points we expressed the coordinates in meters, so they were normalized in a sense. This is important because we observed that using large numerical values causes bad estimations with noisy data in the DLT algorithm. Normalization of image coordinates was also performed since we observed a positive effect both on estimation accuracy and the convergence time. Therefore, in presented experiments, 3D point coordinates are in meters and image coordinates are normalized.

We performed experiments for different settings of intrinsic parameters, different amounts of noise and varying position of the calibration grid. Concerning the latter, we first place the grid in an "ideal" position, such that it well fills the image. Then, we successively move the grid downwards, parallel to the axis of the catadioptric camera. This causes the grid to appear smaller and smaller in the image. These different vertical positions of the grid are referred to by the vertical viewing angle of the topmost calibration points, e.g., $+15^\circ$ means that the highest of the points corresponds to an angle of 15 degrees above the horizontal line containing the sphere center in Fig. 2.2. Examples of simulated images are given in Fig. 3.2.

In Table 3.1, we listed the results for two (ξ, f) pairs, (0.96,360) and (0.80,270). We observe that errors in linear estimates, ξ_{DLT} and f_{DLT}, are biased (smaller than they should be) and the errors increase as the grid is lowered. For all the cases, the true intrinsic parameters were reached after non-linear optimization modulo errors due to noise.

Since the grid covers a smaller area in the image for its lowered positions, same amount of noise (in pixels) affects the non-linear optimization more and errors in non-linear results increase as expected. These errors were depicted in Table 3.1 as $err_\xi = 100 \cdot |\xi_{nonlin} - \xi_{real}| / \xi_{real}$ and $err_f = 100 \cdot |f_{nonlin} - f_{real}| / f_{real}$ and plotted as shown in Fig. 3.3 for the two (ξ, f) pairs. We observe the importance of a good placement of the calibration grid, i.e. such that it fills the image as

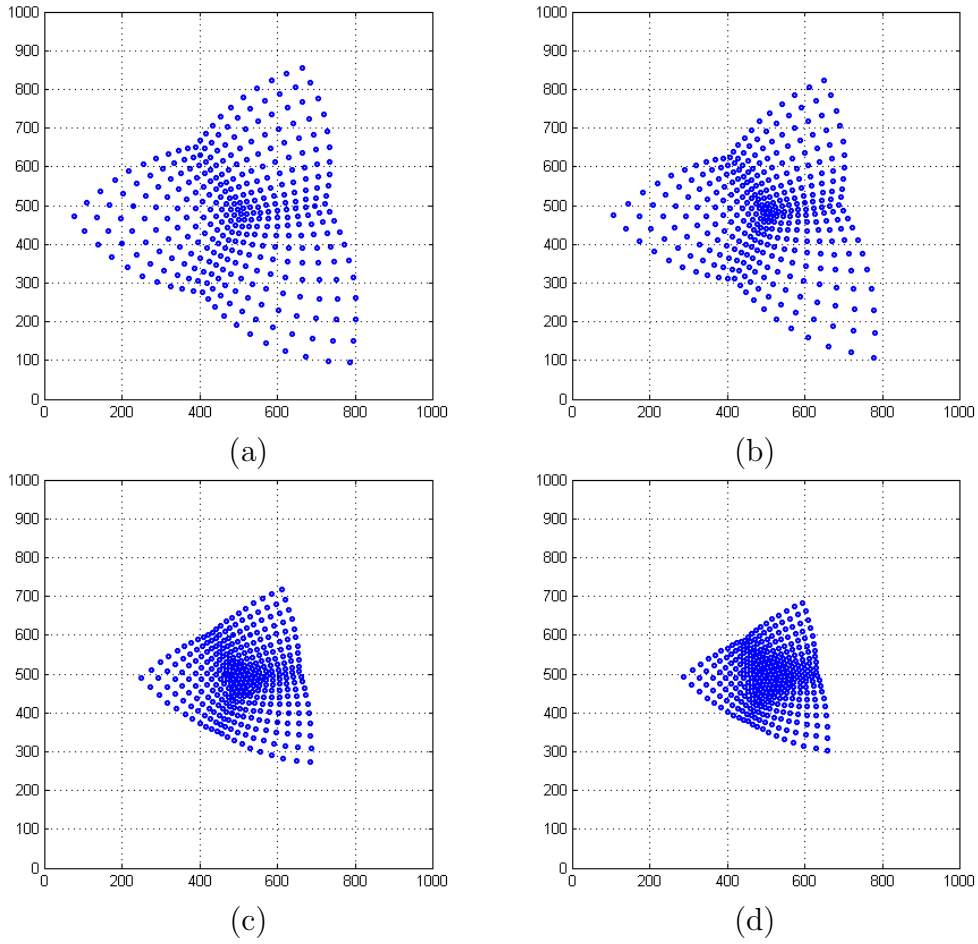


Figure 3.2: Examples of the simulated images with varying values of ξ, f and vertical viewing angle (in degrees) of the highest point in 3D calibration grid (θ). (a) (ξ, f, θ)=(0.96,360,15) (b) (ξ, f, θ)=(0.80,270,15) (c) (ξ, f, θ)=(0.96,360,-15) (d) (ξ, f, θ)=(0.80,270,-15)

much as possible. We also observe that larger ξ and f values produced slightly better results since errors in Fig. 3.3a are smaller.

3.3.1 Estimation Errors for Different Camera Types

Here we discuss the intrinsic and extrinsic parameter estimation for the two most common catadioptric systems: hyper-catadioptric and para-catadioptric, with hyperbolic and parabolic mirror respectively. We also present our observation for experiments on perspective cameras.

Table 3.1: Calibration experiment with simulated images. Initial and optimized estimates of parameters for varying grid heights and (ξ, f) values.

	Vertical viewing angle of the topmost grid points									
	$+15^\circ$		0°		-15°		-30°		-45°	
ξ_{real}	0.96	0.8	0.96	0.8	0.96	0.8	0.96	0.8	0.96	0.8
f_{real}	360	270	360	270	360	270	360	270	360	270
ξ_{DLT}	0.544	0.405	0.151	0.152	0.084	0.053	0.012	0.043	0.029	0.050
f_{DLT}	361	268	296	230	251	198	223	175	211	169
ξ_{nonlin}	0.960	0.800	0.955	0.793	0.951	0.810	0.991	0.780	0.912	0.750
f_{nonlin}	360	270	359	271	362	271	365	266	354	261
err_ξ	0.0	0.0	0.5	0.8	0.9	1.2	3.2	2.5	5.0	6.3
err_f	0.0	0.1	0.4	0.3	0.6	0.3	1.4	1.3	1.6	3.2

For all columns, c_x=c_y=500, and α= -0.628, β= 0.628 and γ= 0.175. Amount of noise: $\sigma = 1$ pixel. ξ_{DLT},f_{DLT} and ξ_{nonlin},f_{nonlin} are the results of DLT algorithm and non-linear optimization respectively, err_ξ and err_f are the relative errors, in percent.

Hyper-catadioptric system.

Table 3.2 shows non-linear optimization experiment results for two different noise levels ($\sigma = 0.5$, $\sigma = 1$), when the described 3D pattern is used and maximum vertical angle of pattern points is $+15^\circ$.

Para-catadioptric system.

Parabolic mirror has a $\xi = 1$, which has a potential to destabilize the estimations because X_ξ becomes a singular matrix. We observed that the results of DLT algorithm were not close to the actual values when compared to hyper-catadioptric system (initial values in Table 3.2). However, non-linear optimization was able to estimate the parameters as successful as the hyper-catadioptric examples given in Table 3.2.

Perspective camera.

In sphere camera model, $\xi = 0$ corresponds to the perspective camera. Our estimation in linear and non-linear steps are as successful as the hyper-catadioptric case.

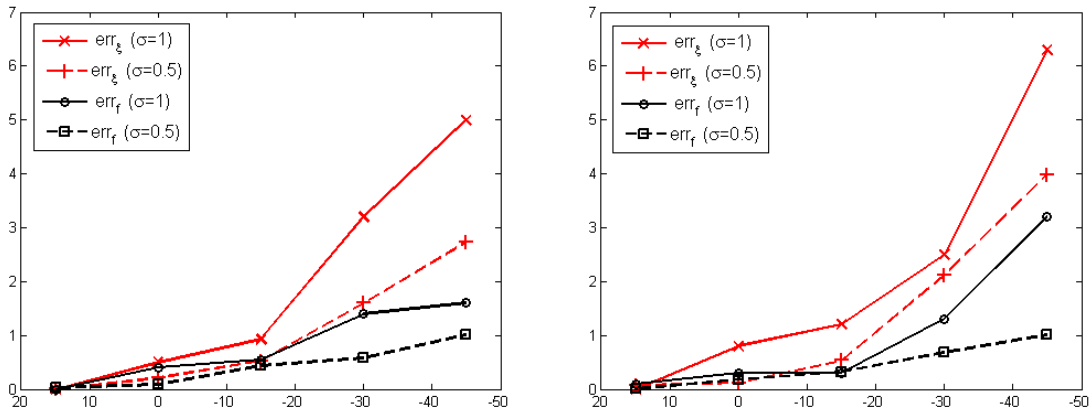


Figure 3.3: Errors for ξ and f for increasing vertical viewing angle of the highest 3D pattern point (x-axis) after non-linear optimization. (a) (ξ, f)=(0.96,360) (b) (ξ, f)=(0.80,270).

Table 3.2: Initial values (DLT) and non-linear optimization estimates of intrinsic and extrinsic parameters for two different amounts of noise: $\sigma = 0.5$ and $\sigma = 1.0$. The Matlab method *lsqnonlin* is employed, using Levenberg-Marquardt algorithm for 11 parameters (rotation, translation and intrinsic).

	Real values	$\sigma = 0.5$		$\sigma = 1$	
		Initial	Nonlinear Estimate	Initial	Nonlinear Estimate
f	360	361	360	354	360
c_x	500	503	500	505	500
c_y	500	498	500	509	500
ξ	0.96	0.848	0.960	0.530	0.961
$R_x(\alpha)$	-0.628	-0.604	-0.628	-0.405	-0.628
$R_y(\beta)$	0.628	0.625	0.628	0.654	-0.628
$R_z(\gamma)$	0.175	0.155	0.175	0.188	0.174
t_x	0.30	0.386	0.300	0.456	0.300
t_y	0.30	0.402	0.300	0.443	0.301
t_z	0.20	0.050	0.200	0.008	0.200
RMSE			0.70		1.42

3.3.2 Tilting and Distortion

It seems intuitive that small amounts of tangential distortion and tilting have similar effect on the image and in our simulations we observed that simultaneous estimation of both is not beneficial. Therefore, we investigated if we can estimate tangential distortion existing in the system by tilt parameters or tilt in the system

Table 3.3: Estimating tangential distortion with tilt parameters. $\sigma = 0.5$ pixels.

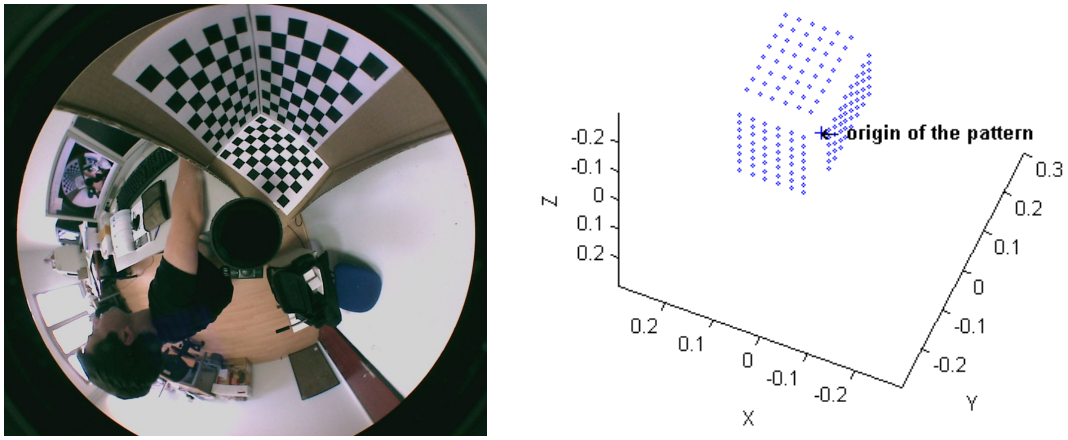
	Real values	$p_1 = p_2 = 0.006$		$p_1 = p_2 = -0.003$	
		Initial	Nonlin. Estimate	Initial	Nonlin. Estimate
f	180	196.5	179.4	195.6	180.3
c_x	500	508.2	495.6	507.0	501.9
c_y	500	486.9	495.9	485.7	502.5
ξ	0.96	1.037	0.9579	1.041	0.9617
$tilt_x$		0	-0.0349	0	0.0187
$tilt_y$		0	0.0367	0	-0.0171
k_1	-0.06	0	-0.061	0	-0.059
k_2	0.006	0	0.006	0	0.0058
RMSE			0.85		0.7

by tangential distortion parameters.

When there is no tilt but only tangential distortion and we estimate tilting parameters, we observed that the direction and amount of $tilt_x, tilt_y, c_x$ and c_y changes proportional to the tangential distortion applied and RMSE decreases. Table 3.3 shows the results for this experiment. We observe that having a tangential distortion of $p_1 = 0.006$ results in $\sim 0.036^\circ$ change in $tilt_x$ and ~ 4.5 pixels change in c_x.

However, RMSE does not reach the values when there is no distortion. In noiseless case, for example, final RMSe are 0.48 for $p_1 = p_2 = 0.006$ and 0.27 for $p_1 = p_2 = -0.003$.

Hence, we conclude that tilt parameters compensate the tangential distortion effect up to an extent, but not perfectly. We also investigated if tilting can be compensated by tangential distortion parameters and we had very similar results. Thus, tangential distortion parameters have the same capability to estimate tilting. Also knowing from Section 2.2.1 that the tilt in the sphere camera model is not equivalent to the tilt in real catadioptric camera, we decided to continue with estimating tangential distortion parameters.



(a) (b)

Figure 3.4: (a) Omnidirectional image of the 3D pattern (1280×960 pixels, flipped horizontal). (b) Constructed model of the 3D pattern.

3.4 Calibration with Real Images using a 3D Pattern

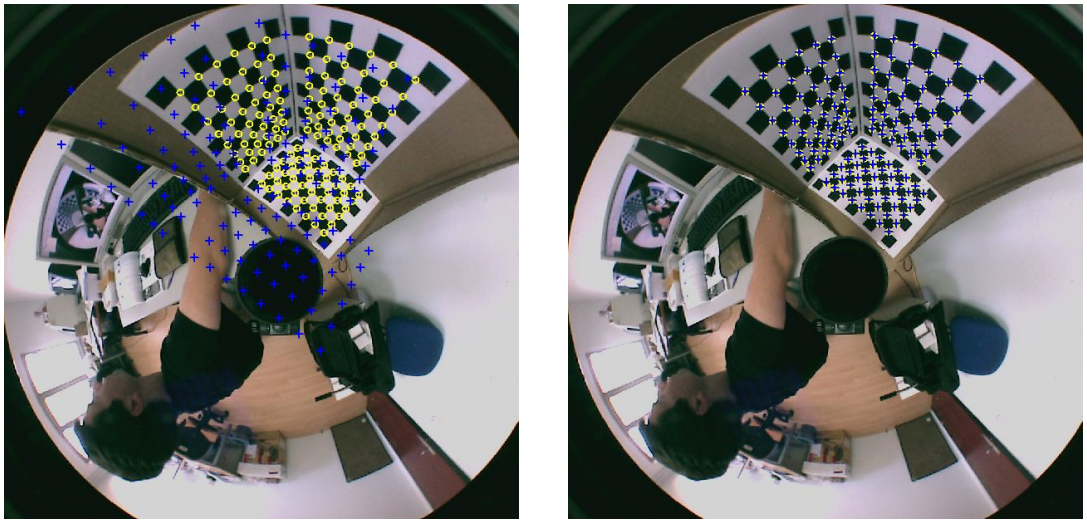
In this section, we perform calibration using a 3D pattern. We use an omnidirectional image viewing the 3D pattern (Fig. 3.4a) and we construct the 3D model of the pattern knowing the distances between the corners of the pattern (Fig. 3.4b). A one megapixel omnidirectional image was acquired using a catadioptric system comprising a 1/2 inch CCD camera (Imaging Source DFK 41F02) and a omnidirectional viewing apparatus with a parabolic mirror (Remote Reality S80).

We computed, from a total of 144 3D-2D correspondences, the projection matrix P_{cata} and extracted the intrinsic and extrinsic parameters as explained in Section 3.2. From the simulations, we observed that we have better estimations if the 3D-2D correspondences are in the same order of magnitude. Therefore, 3D points are given in meters and 2D points are normalized so that the centroid of the reference points is at the origin of coordinates and mean distance of the points from the origin is equal to $\sqrt{2}$.

The experiment is focused on obtaining the intrinsic parameters from P_{cata} with DLT approach to get initial estimates of these values and optimizing these parameters together with distortion parameters with a non-linear optimization step based on the reprojection error. Table 3.4 shows the estimation results and

Table 3.4: Intrinsic parameters estimated with the proposed calibration approach.

parameters	Initial estimates	Final estimates	Final estimates (ξ=1 restricted)
f	314.7	550.3	435.0
c_x	563.0	543.1	542.4
c_y	347.5	511.0	509.5
ξ	0.339	1.475	1.0
k_1	0	0.180	-0.060
k_2	0	0.235	0.008
t_1	0	0.024	0.006
t_2	0	-0.006	-0.003
RMSE	134.0	0.368	0.395



(a) (b)

Figure 3.5: Reprojections with the estimated parameters after initial (DLT) step (a) and after non-linear optimization step (b).

in Fig. 3.5 one can see the reprojections with the estimated parameters for initial values (a) and values after nonlinear estimation (b).

We observe from the third column of Table 3.4 that, when unrestricted, final ξ estimate is much larger than its theoretical value of 1.0. Focal length (f) and radial distortion parameters (k_1,k_2) compensate this high ξ value resulting in a low reprojection error seen in Fig. 3.5b. Alternatively, we can apply a restriction such as $\xi = 1.0$ in optimization step. In this case we obtain the results given in the last column of Table 3.4. Although the reprojection error is not lower than

the unrestricted case, it is slightly higher indeed, these ξ and f values are closer to the theoretical ones. Thus, we suggest to use $\xi = 1$ and corresponding f value. We will see in Chapter 5 that epipolar constraint can be expressed linearly for para-catadioptric cameras assuming $\xi = 1$. Therefore, calibration with $\xi = 1$ restriction brings flexibility in epipolar geometry and pose estimation steps.

3.5 Conclusions

We presented a calibration technique based on the sphere camera model which is able to represent every single-viewpoint catadioptric system. When compared to the only previous work on calibration of sphere camera model [26], our approach has the advantages of not requiring input for parameter initialization and being able to calibrate perspective camera as well. On the other hand, our algorithm needs a 3D calibration object.

Another way for parameter initialization could be directly using the values in the product specification. This approach provides a good approximation for perspective cameras. However, for catadioptric cameras it is rarely possible to obtain system parameters from the manufacturer since they are considered as intellectual property. Moreoever, conversion of actual parameters to the parameters of the sphere camera model is not always straight-forward. Thus, an automatic parameter initialization method is valuable in our case.

We tested our method both with simulations and real images of catadioptric cameras. Although we left it as a future work, it is also possible to use the proposed technique for fisheye lenses since it was shown that the sphere model can approximate fisheye projections [53].

CHAPTER 4

Feature Matching

In this chapter, we present our method of matching feature points between mixed camera images. In Section 4.1, we describe the proposed algorithm to increase the matching performance of Scale Invariant Feature Transform (SIFT) and to eliminate the false matches in the SIFT output. In Section 4.2, we briefly decribe how to create virtual camera plane (VCP) images for matching. Finally, in Section 4.3 we present the results of experiments conducted with catadioptric and fish-eye cameras, showing that matching performance significantly increases with the proposed approaches.

4.1 Improving the Initial SIFT Output

To match features in hybrid image pairs automatically, we employ SIFT and propose an algorithm to obtain better feature matches between catadioptric and perspective images.

SIFT detects features in the so-called *scale space* comprising levels and octaves which are obtained by low-pass filtering and downsampling the original image systematically [21]. This enables the detection of features at different scales. In our case of matching points between images of different resolutions, a feature in the perspective image can be matched to a feature in the catadioptric image. However, we observed that, together with the correct matches, a considerable number of false matches occur in the SIFT output due to matching a high-resolution feature in perspective image to a feature in catadioptric image which is lower resolution.

Table 4.1 shows the number of extracted features in different octaves for images Pers2 and Omni2 (Fig. 4.7). There is an approximate ratio between scales of true correspondences (SR=$\sigma_{pers}/\sigma_{omni}$), which is $\approx$3.6 for the given images. Corresponding octaves of correct matches are indicated in the table with arrows.

Table 4.1: Number of SIFT features detected in a catadioptric and perspective image pair (Pers2-Omni2).

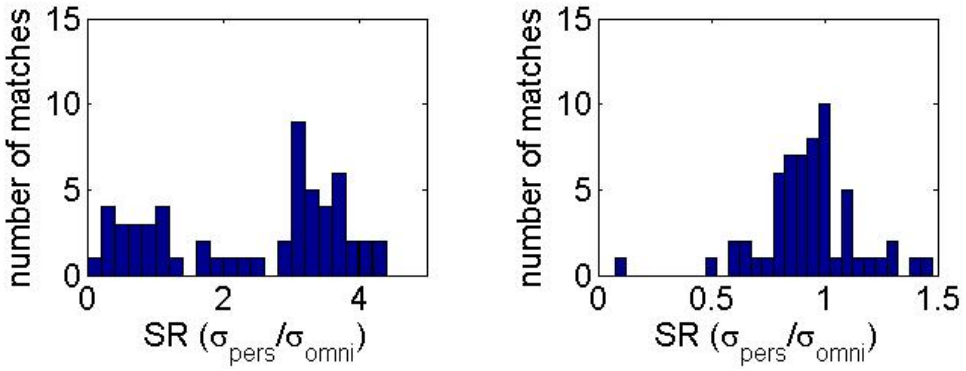
Octave	Approximate scale(σ) in SIFT scale space	Omni 2 (1024x960 pixels)	Pers 2 (1100x800 pixels)	Pers 2, blurred and downsampled
-1	1	1365	2459	288
0	2	489	463	97
1	4	202	174	23
2	8	68	76	5
3	16	23	20	0
4	32	4	5	0

SIFT extracts many features (nearly 3000) at the first two (high-resolution) octaves as can be observed. When there is no scale restriction, it is quite likely that some of the many candidates from first two octaves of the perspective image are incorrectly selected as the best match of features in the omnidirectional image. For the given image pair, there are 25 false matches out of 60 and 23 of these false matches have a scale ratio (SR) less than 2.0, whereas average SR of true matches is 3.57. It is possible to eliminate most of these false matches by simply discarding the matches with improper SR. This idea has been recently used for perspective camera images in two independent studies [54, 55].

Yi *et al.* [54] form the histogram of scale differences and define a window around the peak of this histogram. The matches with scale differences outside this window are rejected. They only considered image pairs with approximately same scale. However, for larger scale differences, the ratio of scales rather than scale difference is meaningful.

Alhwarin *et al.* [55] divide the SIFT features according to the octaves they are extracted from. They detect the octave pair between which the number of found matches is maximum and they assign the ratio between these octaves as the correct scale factor. All matches from other octave pairs are rejected. Since only the matches between octaves are analyzed, the scale ratio can be obtained only in the form of 2^k and ratios in between are not considered.

The method we propose here also uses the dominant scale ratio between images but it is not an eliminate-after-matching method as the above mentioned approaches. As will be described in Section 4.1.1, we preprocess the perspective image in the hybrid pair to adjust its scale and we observe a significant



(a) (b)

Figure 4.1: Histogram of SR for the matches in the catadioptric-perspective image pair given in Table 4.1. (a) SIFT applied on original image pair, (b) SIFT applied on downsampled perspective image and original catadioptric image.

improvement in SIFT matching. In Section 4.1.2, we discuss the advantages of our method compared to the existing approaches [54, 55].

4.1.1 Preprocessing Perspective Image

The histogram of the example hybrid image pair is shown in Fig. 4.1a. The accumulation on the left (matches with SR<2.0) is explained by false matches due to matching features in the first octaves of perspective image SIFT. We found out that it is possible to improve the SIFT matching by blurring and downsampling the perspective image.

Blurring is achieved by low-pass filtering with a Gaussian filter. With this preprocessing, the scale ratio of matching features becomes close to 1 and the possibility of matching valuable features in omnidirectional image with features in the correct octaves of perspective image considerably increases. For the given example, last column of Table 4.1 shows the number of detected features where the perspective image is downsampled by 3.6 (both in horizontal and vertical axis) following a blurring operation. Fig. 4.1b shows the SR histogram when the perspective image is low-pass filtered and downsampled as explained. This matching resulted in a true/total ratio of 56/60.

Correct and false matches for the example Pers2-Omni2 hybrid image pair, with direct SIFT matching and with the proposed preprocessing method are given in Fig. 4.2.

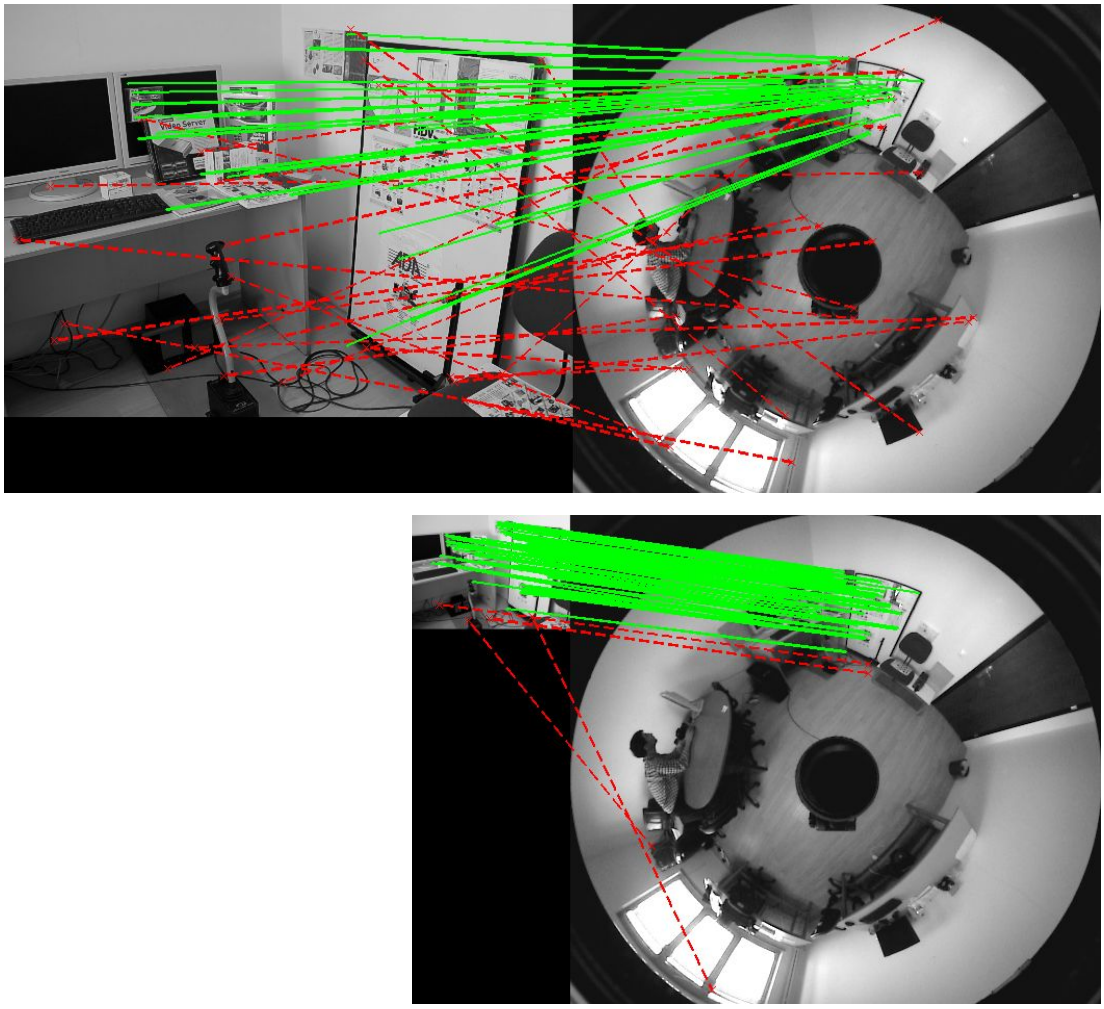
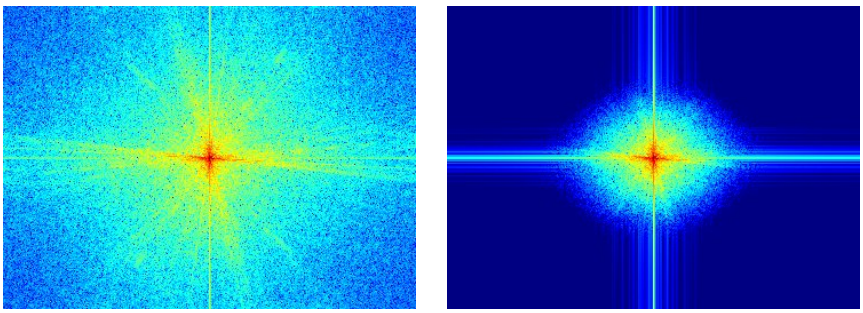


Figure 4.2: Matching results for the Pers2-Omni2 image pair with direct matching resulted in 25/60 false/total match ratio (at the top) and with the proposed preprocessing method resulted in 4/60 false/total match ratio (at the bottom). Red dashed lines indicate false matches, whereas green lines indicate correct ones.

The parameters of downsampling and low-pass filtering. We selected the downsampling factor from the histogram as the mean of the most dominant Gaussian in the mixture, because the SIFT scale space ratio also reveals the scale ratio of features in the images. To avoid aliasing, we need to low-pass filter the perspective image before downsampling. We selected the cut-off frequency as $2.5/\sigma$ in the frequency domain and the standard deviation of the Gaussian filter for the blurring becomes $\sigma = 2.5d/\pi$ where d is the downsampling factor. Figure 4.3 shows the discrete time FFT graphs before and after the described low-pass



(a) (b)

Figure 4.3: Discrete time FFT before low-pass filtering (a) and after low-pass filtering with $\sigma = 2.5d/\pi$ where $d = 3.6$ (b).

filtering scheme. While downsampling, we did not directly remove the columns and rows in between, but we employed resampling by Lanczos filter.

Comparison with Lowe's elimination method with affine parameters. Lowe proposes a verification method for the detected SIFT matches aiming to increase the object detection performance [21]. The matches are grouped according to their scale ratios, rotations and translations. Afterwards, an affine transformation is assigned to these groups and matches which do not conform to any of these transformations are eliminated.

We checked whether this technique is able to eliminate false matches for our case. We observed that, although it is able to eliminate a few false matches, its performance is far from our proposed preprocessing approach. Main reason is that, Lowe's approach seeks for only three matches to create a group of transformation and some matches with incorrect scale ratios but with similar transformations can form such groups. However, there is a window of true scale ratio outside of which indicates false matches. Let us explain this phenomenon with the scale ratio histogram (Fig. 4.4) when Lowe's elimination is applied to the hybrid image pair of Fig. 4.1a. The approximate window of 3<SR<4 belongs to true matches, however matches with other SR values are not eliminated.

Furthermore, even we assume that all false matches are eliminated, the number of true matches is very low when compared to the proposed approach of preprocessing the perspective image. This is due to the fact that the proposed approach improves the performance of initial SIFT matching rather than eliminating the false matches afterwards. To explain this with figures, let us examine

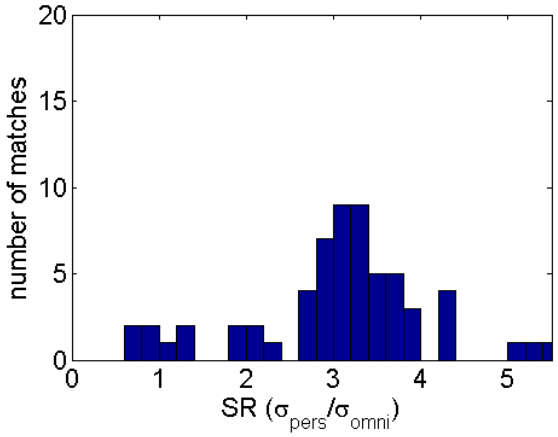


Figure 4.4: Histogram of SR for the matches of the catadioptric-perspective image pair given in Table 4.1 when Lowe's elimination is applied.

the entries in Table 4.2 for the same example hybrid image pair (Pers2-Omni2). The table also shows the results of VCP approach (Section 4.2). True/False ratios for Pers2-Omni2 and Pers2-VCP2 pairs are not satisfactory even with the elimination proposed by Lowe, however when the perspective image is preprocessed as proposed (bold-face lines), results are significantly improved. Therefore, preprocessing the perspective image increases the performance of SIFT matching with or without employing the elimination with affine parameters.

We also observe that, the results are slightly better with Lowe's elimination. Thus, elimination with the assumption of an affine transformation works up to an extent although it is not a good approximation especially when we work on hybrid pairs. Moreover, the elimination of matches that do not conform to a transformation is not crucial at this step since we do eliminate the matches that do not conform to the epipolar geometry during the estimation of fundamental matrix with RANSAC (cf. Chapter 5) which imposes a more accurate geometrical constraint.

Sensitivity to varying preprocessing parameters. So far, we described the procedure of preprocessing the perspective image and selection of low-pass filtering and downsampling parameters. Fig. 4.5 shows the number of correct matches for varying low-pass filering (σ) and downsampling (d) parameters for the example mixed image pair. We observe that the performance with the selected parameters is quite close to the peak and all the parameter pairs represented in

Table 4.2: Comparison with Lowe's proposal of elimination. Number of total and true/false matches for the example hybrid image pair (Pers2 - Omni2).

Image pairs	**No. of matches**	**True/False**
Pers2 - Omni2	60	35/25
Pers2 - Omni2 (Lowe)	62	48/14
Pers2 σ2.5 d3.6 - Omni2	**71**	**64/7**
Pers2 σ2.5 d3.6 - Omni2 (Lowe)	**68**	**67/1**
Pers2 - VCP2	73	43/30
Pers2 - VCP2 (Lowe)	74	43/31
Pers2 σ2.5 - VCP2	**76**	**70/6**
Pers2 σ2.5 - VCP2 (Lowe)	**76**	**76/0**

Pers$N$ σA dB indicates that Pers$N$ image was blurred with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction. N in VCPN indicates the index of omnidirectional image that VCP image is generated from.

the graph performed much better than matching without preprocessing where the number of correct matches is only 35. We infer that slight variations of parameters do not cause a significant performance difference.

4.1.2 Final Elimination

We can further restrict the scale ratio (SR) to remove a few false matches with improper SR similar to the scale restriction approaches [54, 55]. To do this, we define a window around the mean SR and discard the matches outside the window. For our experiments we chose the bounds of the elimination window as 0.6SR<SR<1.4SR. After this final elimination, true/total ratio becomes 54/55 for the given example.

If we directly apply this scale restriction without preprocessing, true/total ratio is 32/34. Since it is important to keep as many true matches as possible in most computer vision applications, especially for structure-from-motion, we preferred the preprocessing approach to the elliminate-after-matching approaches [54, 55].

Results for all image pairs of our experiment are given in Section 4.3. In the following, we outline the proposed SIFT matching algorithm including the preprocessing steps to improve the SIFT performance.

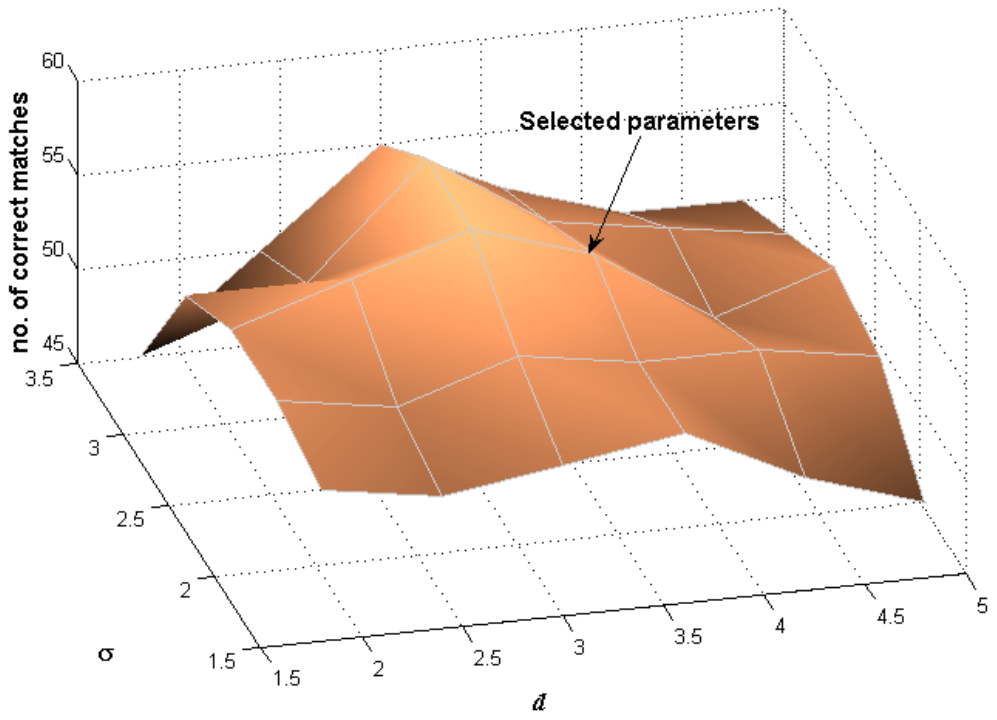


Figure 4.5: Number of correct matches out of 60 matches for varying low-pass filtering (σ) and downsampling (d) parameters for the example mixed image pair (Table 4.1).

Algorithm 4.1: Hybrid SIFT matching with the proposed approach

1. **Extract downsampling factor.** Apply SIFT matching between the given perspective and catadioptric image pair and plot the histogram of scale ratios (eg. Fig. 4.1a). Select the downsampling factor (d) from the histogram as the mean of the most dominant Gaussian in the mixture.

2. **Preprocess the perspective image.** Low-pass filter the perspective image with a Gaussian filter having a $\sigma = 2.5d/\pi$ and downsample the low-pass filtered image by d both in horizontal and vertical directions.

3. **Perform SIFT.** Apply SIFT matching between the preprocessed perspective image and catadioptric image.

4. **Perform final elimination.** Plot the histogram of scale ratios (SR) for the final matching (eg. Fig. 4.1b) and select only the matches with 0.6SR<SR<1.4SR.

4.2 Creating Virtual Camera Plane Images

Another approach that can be employed in conjuction with preprocessing method is to use so-called *virtual camera planes* (VCP) to create virtual perspective images from omnidirectional images and perform matching between VCP and perspective images. If the omnidirectional camera is calibrated, i.e. if the intrinsic parameters of the camera are known, catadioptric-to-VCP or fisheye-to-VCP conversion can easily be performed.

To generate virtual perspective image, a virtual camera plane with a certain viewing direction (azimuth), a vertical angle and a distance from the mirror focal point (origin) have to be defined (Fig. 4.6). To find the intensity value of a pixel in the virtual perspective image (x_v, y_v), corresponding pixel coordinates in the paraboloidal catadioptric image (x_i, y_i) are given by [56]:

$$x_i = \frac{h}{Z_S+\sqrt{d^2}} X_S \quad y_i = \frac{h}{Z_S+\sqrt{d^2}} Y_S \tag{4.1}$$

where, $d = \sqrt{{X_S}^2 + {Y_S}^2 + {Z_S}^2}$ and (X_S, Y_S, Z_S) are the 3D coordinates of the virtual image pixel w.r.t. the origin. These coordinates are computed using the assigned azimuth, vertical angle and distance value for the virtual plane.

If the calibration information is not available for a para-catadioptric image, we can recover the mirror parameter h by assigning a position in the image which corresponds to an object with the same height with mirror focal point. The distance between the image center and that point gives us the parameter h in terms of pixels, for example the radius of white circle in the top-left image of Fig. 4.8. The principal point can be assumed to be at the center of omnidirectional image circle.

VCP image can be created using the sphere model parameters in the same manner. A VCP with an azimuth, a vertical angle and a distance from the viewpoint (center of sphere) is defined and for each pixel in the VCP, its 3D ray and corresponding pixel coordinates in the catadioptric image are determined with sphere model parameters.

Created virtual images are matched with the perspective camera images viewing the same scene. When we create VCP images with sizes close to the size of perspective images, SR of the true matches is close to 1.0 and no downsampling

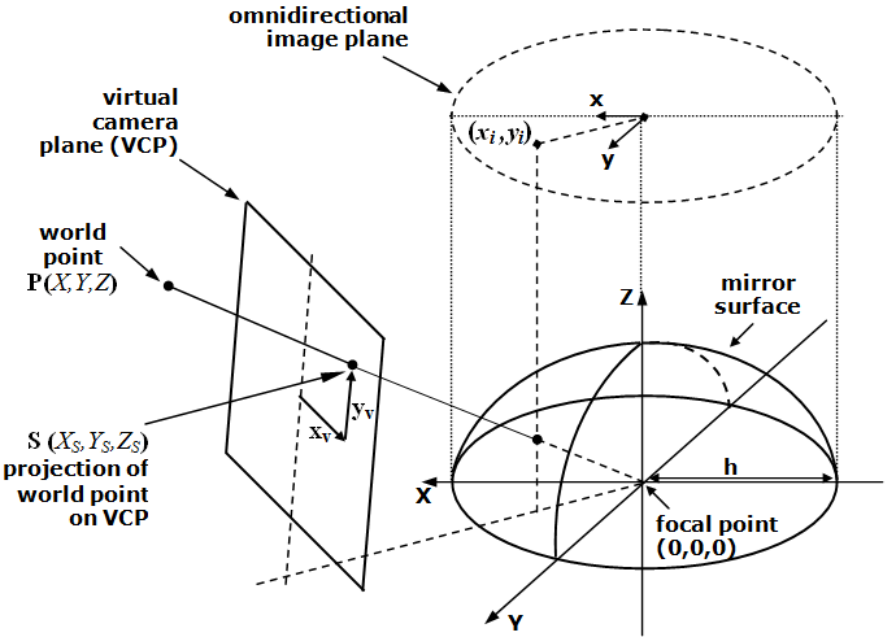


Figure 4.6: Generation of a virtual perspective image in a paraboloidal omnidirectional camera.

is needed. However, since the VCP image is created by interpolating the limited number of pixels in the omnidirectional image the resolution difference still exist between perspective and VCP image. Therefore, we again low-pass filter the perspective image to balance its resolution, i.e. we perform the second step of Algorithm 4.1 (Section 4.1.2, p.39) without the downsampling part. Alternatively we could have produced a small-size VCP and downsampled the perspective image after low-pass filtering. In our experiments, we observed that it produces similar results. The results of our experiments are given in Section 4.3.

4.3 Experiments

4.3.1 Experiments with Catadioptric Cameras

We conducted tests using two types of catadioptric apparatus, namely Remote Reality S80 Optic[1] and 0-360 Panoramic Optic[2] and captured images in indoor and outdoor environments, respectively.

[1]http://www.remotereality.com
[2]http://www.0-360.com

To detect and match feature points, we used SIFT implementation of Andrea Vedaldi[3] with a modification to provide one-to-one matching. In the original algorithm it is possible for many points in the first image to match to the same point in the second image. This causes inconsistencies depending on which image is defined as "first". To eliminate this problem, we run the SIFT algorithm in both ways changing the order of images and declare a match only if it is found in both runs.

As the output of the catadioptric system is viewed through a mirror, the objects are *flipped* in the image. This is corrected by flipping the image w.r.t. a line passing through the center of catadioptric image circle.

For different hybrid camera pairs and varying baseline length, performances of direct perspective-omnidirectional matching, preprocessed perspective-omnidirectional matching and preprocessed perspective-VCP matching approaches are compared. To keep the number of matched points same for different trials of an image pair, we adjusted the matching threshold of SIFT, which defines the strength of the matched point w.r.t. the second candidate match. Let D1 and D2 be the SIFT descriptor vectors of two points in the first and second image respectively and t is the matching threshold. D1 is matched to D2 only if the distance d(D1,D2) multiplied by t is not greater than the distance of D1 to all other descriptors in the second image. Typically t is chosen as 1.5.

Indoor Experiment. The images of this first experiment were captured with the Remote Reality S80 Optic in a controlled indoor environment. The locations and orientations of the cameras and objects in our scene are given in Fig. 4.7 and corresponding images are given in Fig. 4.8. Omnidirectional images have a size of 1024x960 pixels, whereas perspective and VCP images are 1100x800 pixels. With this setup, we are able to investigate the effect of increasing baseline both in the direction towards the scene and perpendicular to the scene.

Matching results are shown in Table 4.3. As described in Section 4.1, to increase the performance, we downsampled the perspective images after low-pass filtering with a Gaussian filter. In Table 4.3, Pers$N$ σA dB indicates that Pers$N$ image was low-pass filtered with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction. The extraction of these parameters are explained

[3]http://vision.ucla.edu/∼vedaldi/code/sift/sift.html

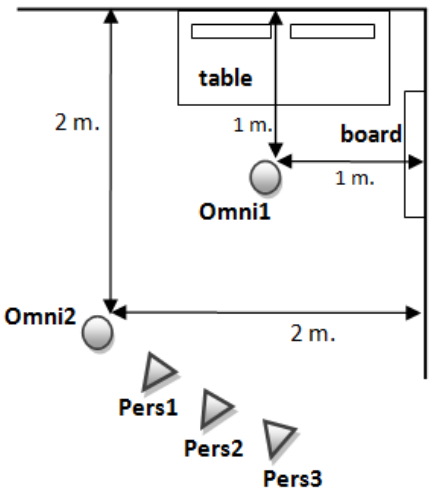


Figure 4.7: Locations and orientations of cameras in the indoor environment catadioptric-perspective point matching experiment.

in Section 4.1.1. Results of matching with VCP approach are also given in Table 4.3. N in VCPN indicates the index of omnidirectional image from which it is generated. As explained in Section 4.2 we match the VCP images with the low-pass filtered perspective images.

We plot the ratio of true/total matches in Fig. 4.9 for different approaches. The table also shows, in the last column, the final true/false ratios obtained after applying final scale ratio elimination. We do not plot these ratios because the number of true matches is also important along with the ratio (cf. Section 4.1.2). Please note that the success rates of both downsampling and VCP approaches in Fig. 4.9 (solid lines) are further increased by this final elimination.

Pers1, Pers2 and Pers3 have approximately 30cm., 60 cm. and 90 cm. lateral baselines with omnidirectional cameras respectively. We observe that the matching performance decreases with increasing baseline for all approaches and the number of false matches decreases significantly for both downsampling and VCP approaches. VCP approach is more robust to increase in baseline at least for Pers$N$-Omni1 pairs. We also observe that ratios of Pers$N$-Omni1 pairs are higher. Since the scene is represented with a larger area in Omni1, this resulted in an increased number of correct matches.

Outdoor Experiment. The images of this second experiment were captured with the 0-360 Panoramic Optic and shown in Fig. 4.11. The locations and

Figure 4.8: Images of point matching experiment, cameras are shown in Fig. 4.7. From top to bottom, first line is Omni1 and VCP image generated from it (VCP1), second line is Omni2 and its virtual image (VCP2), third line is two of the perspective images, Pers1 and Pers3, with small and wide lateral baseline respectively.

orientations of the cameras and objects in the scene are given in Fig. 4.10.

The matching results are shown in Table 4.4 and true/total match ratios are plotted in Fig. 4.12 for the three different approaches. Varying baseline scenario of Pers$N$-Omni2 pairs is similar to the previous experiment, i.e. increasing form Pers1 to Pers3. However, for Pers$N$-Omni1 pairs, there is not a significant difference between baselines since Omni1 is not especially located close to the viewing direction of any of the perspective cameras. This is why we do not observe any significant success increase or decrease for Pers$N$-Omni1 pairs in Fig. 4.12.

Table 4.3: Matching results for the image pairs of indoor matching experiment. True/false match ratios (T/F) after the initial and scale restricted matching.

Image pairs	**no. of matches**	**T/F (T/total %)**	**T/F final**
Pers1 - Omni1	100	97/3 (97%)	84/1
Pers1 σ1.5 d1.65 - Omni1	100	99/1 (99%)	94/0
Pers1 σ1.5 - VCP1	100	99/1 (99%)	99/0
Pers2 - Omni1	75	56/19 (75%)	51/7
Pers2 σ1.5 d1.65 - Omni1	75	70/5 (93%)	67/3
Pers2 σ1.5 - VCP1	75	73/2 (97%)	73/1
Pers3 - Omni1	60	42/18 (70%)	39/9
Pers3 σ1.5 d1.65 - Omni1	60	50/10 (83%)	50/6
Pers3 σ1.5 - VCP1	60	57/3 (95%)	57/1
Pers1 - Omni2	80	63/17 (79%)	62/1
Pers1 σ2.5 d3.6 - Omni2	80	80/0 (100%)	80/0
Pers1 σ2.5 - VCP2	80	80/0 (100%)	80/0
Pers2 - Omni2	60	35/25 (58%)	32/2
Pers2 σ2.5 d3.6 - Omni2	60	56/4 (93%)	54/1
Pers2 σ2.5 - VCP2	60	56/4 (93%)	56/3
Pers3 - Omni2	45	15/30 (33%)	15/1
Pers3 σ2.5 d3.3 - Omni2	45	35/10 (78%)	30/8
Pers3 σ2.5 - VCP2	45	37/8 (82%)	35/3

Pers$N$ σA dB indicates that Pers$N$ image was low-pass filtered with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction.

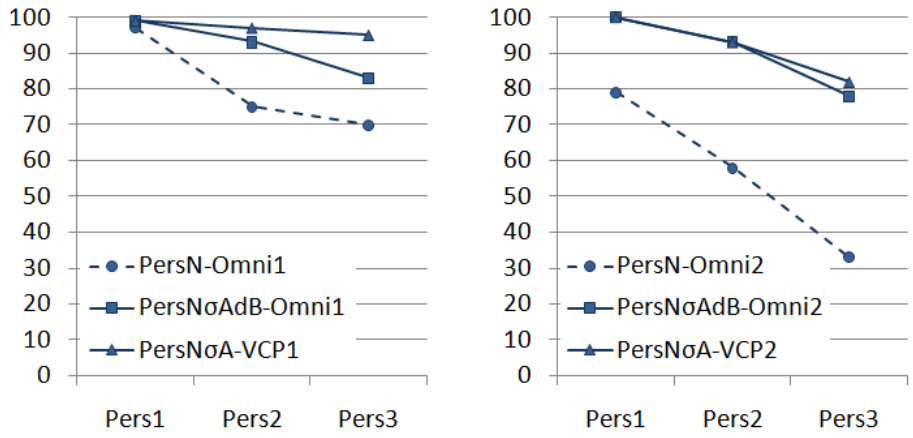


Figure 4.9: The true/total match ratios (in percentage) for Omni1-Pers$N$ (on the left) and Omni2-Pers$N$ (on the right).

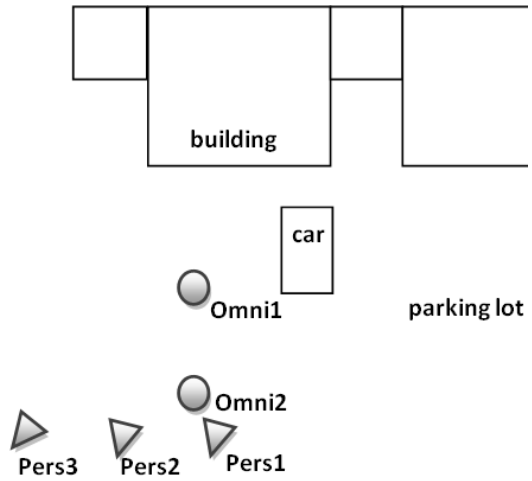


Figure 4.10: Locations and orientations of the cameras in the outdoor environment catadioptric-perspective point matching experiment.

We again observe that the true/total match ratio increases for both downsampling and VCP approaches. However, this time the difference between these two approaches is significant. VCP approach produces the best results for all pairs. We are also able to confirm that the proposed method to extract low-pass filtering and downsampling parameters, previously described in Section 4.1.1, is valid for this new experiment which was conducted with a different camera in a different environment.

Fig. 4.13 shows correct and false matches of Pers1-Omni1 pair for the three different matching alternatives.

4.3.2 Experiments with Fish-eye Camera

To investigate if the proposed approach is also valid for cameras with fish-eye lenses, we conducted an experiment similar to the catadioptric-perspective matching experiments, with Fujinon FE185C046HA-1 185° fish-eye lens. We used the same scene and cameras depicted in Fig. 4.7 and perspective images shown in Fig. 4.8, with the difference that we put fish-eye cameras instead of catadioptric ones. We refer to these new images as Fish1 and Fish2 which are shown in Fig. 4.14.

To create VCP images from fish-eye camera images, we calibrated our camera with the method (and Matlab Toolbox) of Scaramuzza *et al.* [47] which estimates

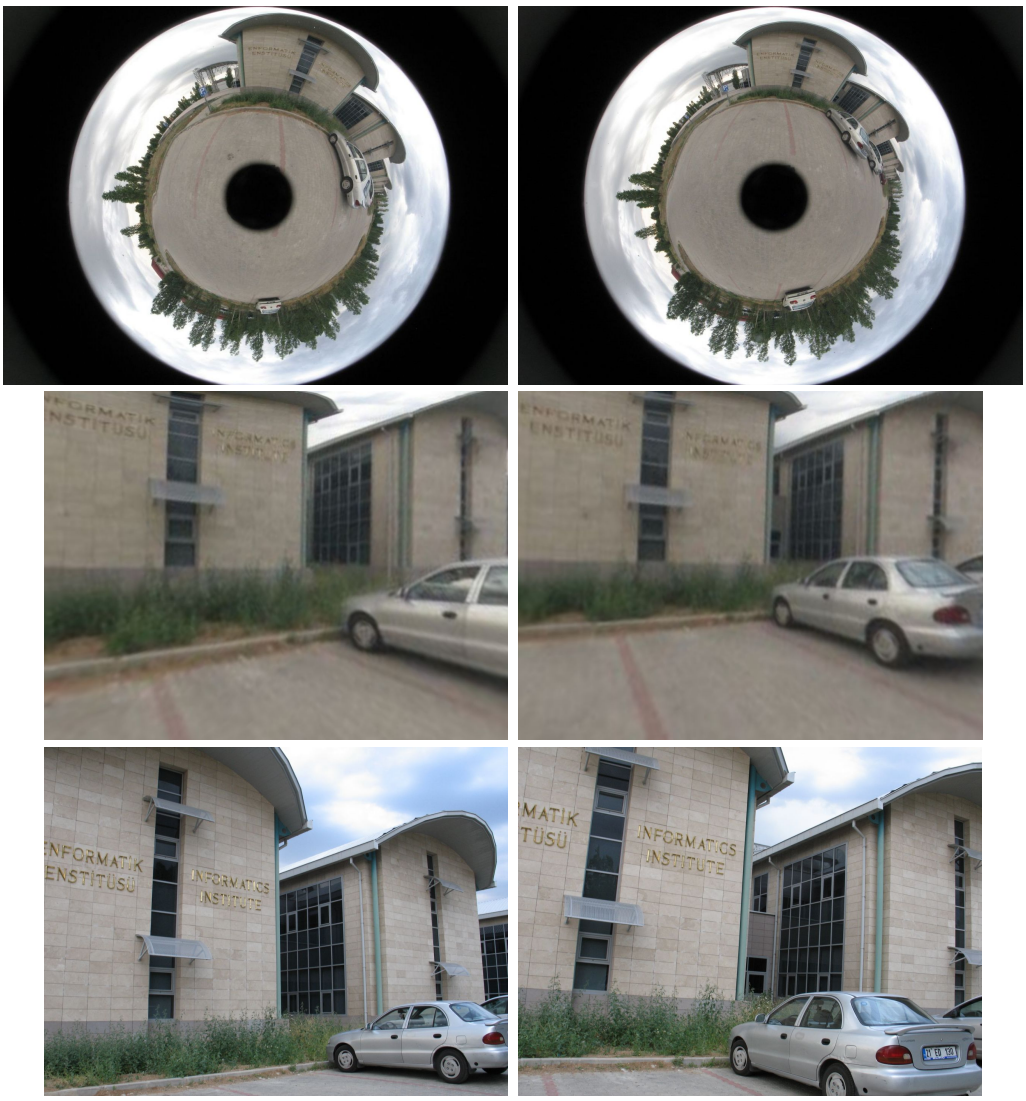


Figure 4.11: Images of 2^{nd} point matching experiment, scene is shown in Fig.4.10. Top row shows catadioptric omnidirectional images, Omni1 and Omni2. Middle row shows corresponding VCP images generated from catadioptric images. Bottom row shows two of the perspective images, Pers3 and Pers1.

the radial distortion model for 3D rays corresponding to image points. Calibrating fish-eye camera with sphere model as proposed in [53] and used by [57] can also be considered as an alternative.

Fig. 4.16 shows correct and false matches of Pers2 - Fish1 image pair. The improvement with the proposed approach can be easily observed. Matching results are shown in Table 4.5 and ratios of true/total matches are plotted in Fig. 4.17 for the three different approaches similar to previous experiments. We observe

Table 4.4: Matching results for the image pairs of outdoor matching experiment. True/false match ratios (T/F) after the initial and scale restricted matching.

Image pairs	**no. of matches**	**T/F (T/total %)**	**T/F final**
Pers1 - Omni1	60	28/32 (47%)	26/6
Pers1 σ3.0 d3.5 - Omni1	60	44/16 (73%)	40/10
Pers1 σ3.0 - VCP1	60	57/3 (95%)	53/2
Pers2 - Omni1	60	30/30 (50%)	27/5
Pers2 σ2.5 d3.0 - Omni1	60	45/15 (75%)	41/7
Pers2 σ2.5 - VCP1	60	56/4 (93%)	51/2
Pers3 - Omni1	60	31/29 (52%)	23/9
Pers3 σ2.5 d3.0 - Omni1	60	40/20 (67%)	36/14
Pers3 σ2.5 - VCP1	60	57/3 (95%)	55/2
Pers1 - Omni2	75	47/28 (63%)	47/1
Pers1 σ3.5 d4.8 - Omni2	75	64/11 (85%)	61/5
Pers1 σ3.5 - VCP2	75	75/0 (100%)	73/0
Pers2 - Omni2	60	33/27 (55%)	32/3
Pers2 σ3.0 d4.2 - Omni2	60	49/11 (82%)	46/9
Pers2 σ3.0 - VCP2	60	60/0 (100%)	59/0
Pers3 - Omni2	60	29/31 (48%)	27/4
Pers3 σ2.5 d3.3 - Omni2	60	34/26 (57%)	30/19
Pers3 σ2.5 - VCP2	60	54/6 (90%)	52/4

Pers$N$ σA dB indicates that Pers$N$ image was low-pass filtered with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction.

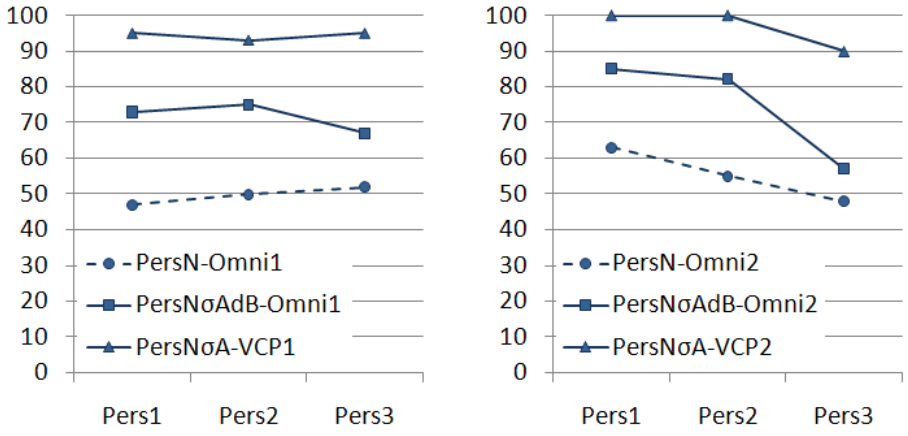


Figure 4.12: The true/total match ratios (in percentage) of outdoor environment catadioptric-perspective point matching experiment (cf. Fig. 4.11 and Fig. 4.10).

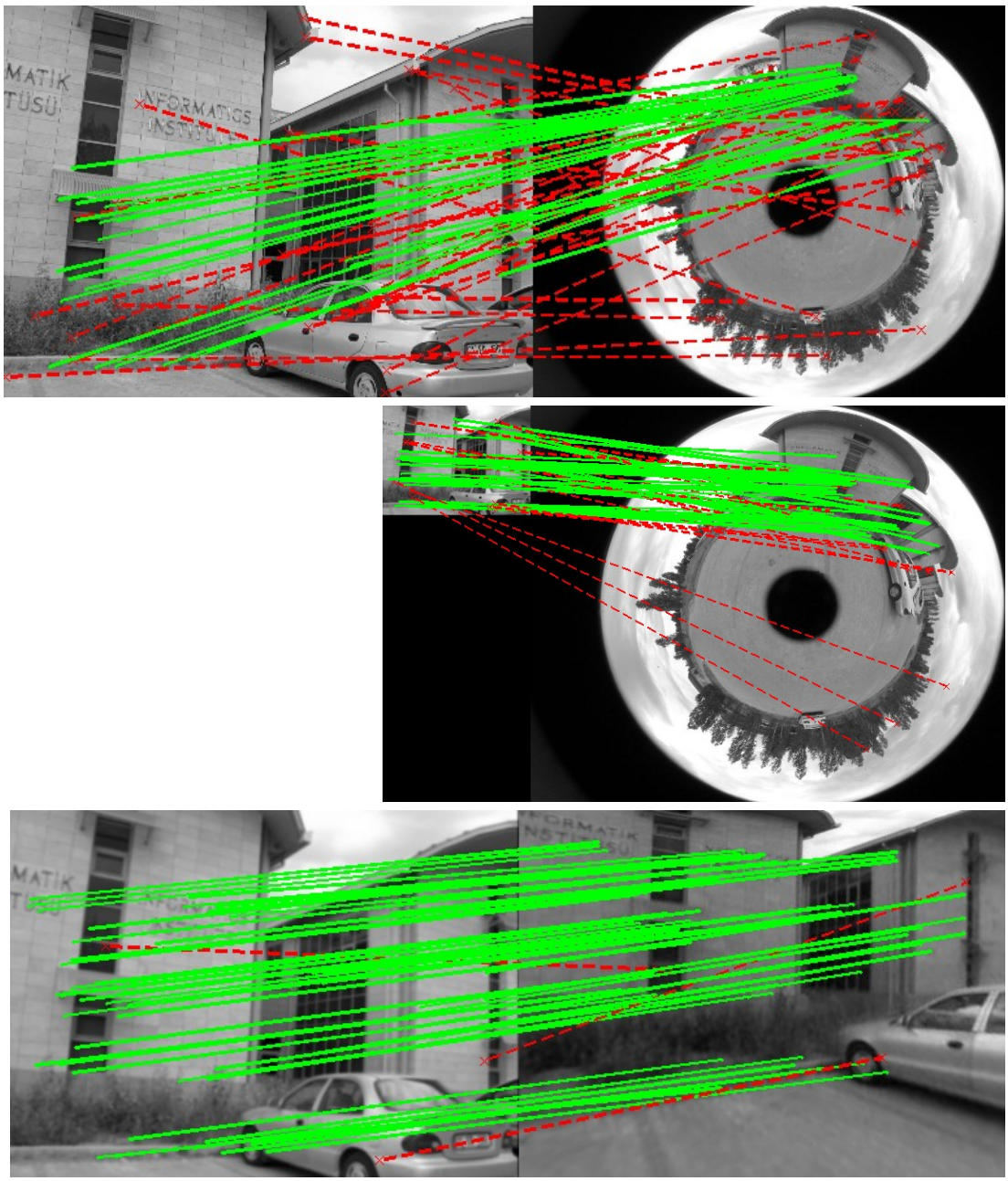


Figure 4.13: Matching results for the Pers1-Omni1 pair in the outdoor experiment (Table 4.4) with direct matching (top), preprocessed perspective - omnidirectional matching (middle) and preprocessed perspective - VCP matching (bottom). Red dashed lines indicate false matches, whereas green lines indicate correct ones.

that the performance significantly increases for both downsampling and VCP approaches, and the performances of these two proposed approaches are same for both sets and for increasing baseline.

Figure 4.14: Fish-eye camera images of the point matching experiment: Fish1 (left) and Fish2 (right). The scene is given in Fig. 4.10. Fish1 and Fish2 are at the same location with Omni1 and Omni2, respectively.

Figure 4.15: A perspective fish-eye hybrid image pair, where the common features are not located at the central part of the fish-eye image but closer to periphery.

When compared to the results of experiments with catadioptric cameras, we are able to say that performance of VCP approach did not change for fish-eye images, however performance of low-pass filtering and downsampling approach increased. A possible reason is that the scene represented in fish-eye cameras is closer to the perspective images when compared to catadioptric cameras (please compare Fig. 4.14 with Figures 4.8 and 4.11). We also investigated if this result is valid when the features are not located at the central part of the fish-eye image but closer to periphery, such as the hybrid image pair given in Fig. 4.15. We again observed that downsampling approach is as successful as VCP approach.

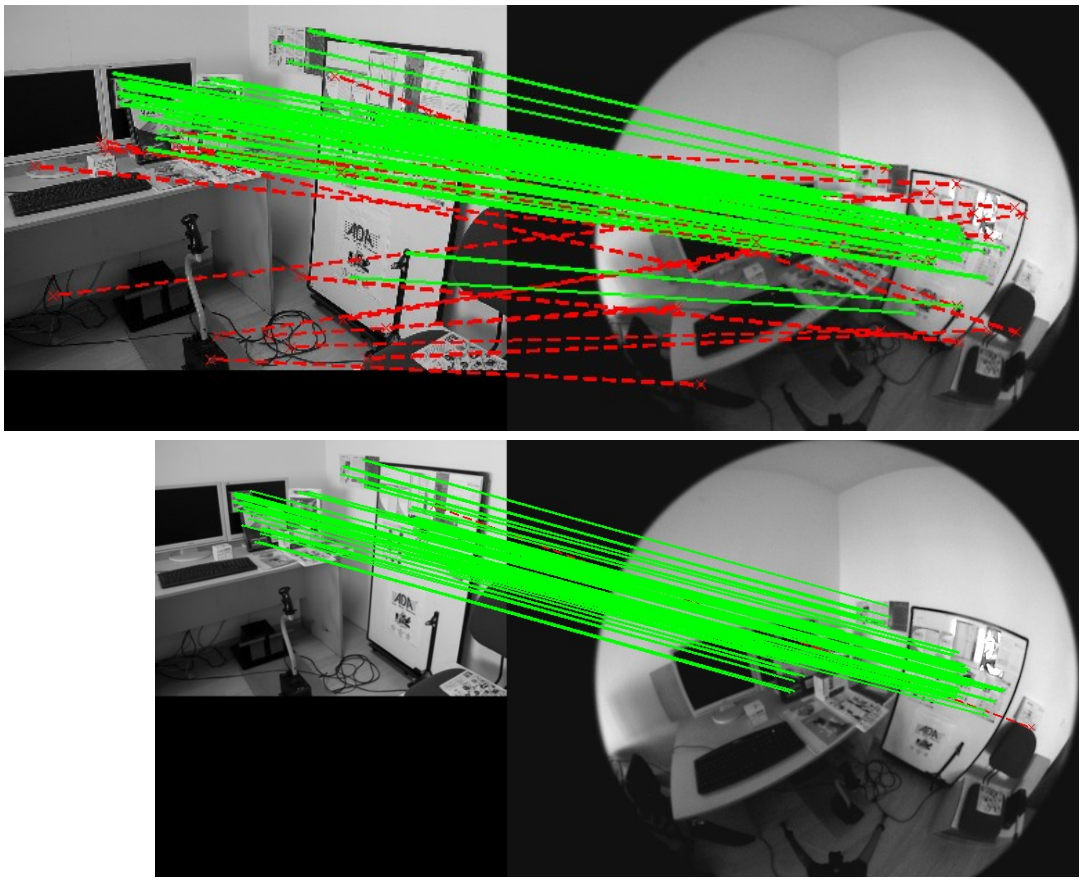


Figure 4.16: Matching results for the Pers2-Fish1 pair in the fish-eye experiment (Table 4.5) with direct matching (top) and downsampling approaches (bottom). Red dashed lines indicate false matches, whereas green lines indicate correct ones.

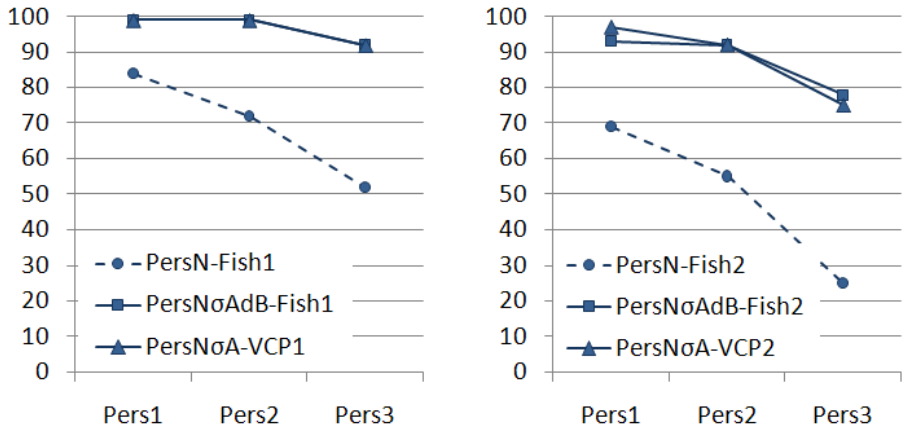


Figure 4.17: The true/total match ratios (in percentage) of fisheye-perspective point matching experiment (cf. Figs. 4.7 and 4.14).

Table 4.5: Matching results for the fisheye-perspective matching experiment. True/false match ratios (T/F) after the initial and scale restricted matching.

Image pairs	**no. of matches**	**T/F (T/total %)**	**T/F final**
Pers1 - Fish1	100	84/16 (84%)	82/5
Pers1 σ1.5 d1.43 - Fish1	100	99/1 (99%)	98/1
Pers1 σ1.5 - VCP1	100	99/1 (99%)	98/1
Pers2 - Fish1	75	54/21 (72%)	53/4
Pers2 σ1.5 d1.43 - Fish1	75	74/1 (99%)	74/1
Pers2 σ1.5 - VCP1	75	74/1 (99%)	74/1
Pers3 - Fish1	50	26/24 (52%)	25/5
Pers3 σ1.5 d1.25 - Fish1	50	46/4 (92%)	45/1
Pers3 σ1.5 - VCP1	50	46/4 (92%)	45/1
Pers1 - Fish2	70	48/22 (69%)	46/1
Pers1 σ3.0 d3.6 - Fish2	70	65/5 (93%)	64/4
Pers1 σ3.0 - VCP2	70	68/2 (97%)	68/0
Pers2 - Fish2	60	33/27 (55%)	31/1
Pers2 σ3.0 d3.3 - Fish2	60	55/5 (92%)	52/1
Pers2 σ3.0 - VCP2	60	55/5 (92%)	54/4
Pers3 - Fish2	40	10/30 (25%)	10/3
Pers3 σ3.0 d3.3 - Fish2	40	31/9 (78%)	28/5
Pers3 σ3.0 - VCP2	40	30/10 (75%)	25/9

Pers$N$ σA dB indicates that Pers$N$ image was low-pass filtered with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction.

4.4 Conclusions

It had been stated that directly applying SIFT is not sufficient to obtain good results for hybrid image pairs. In our study, we showed that the performance of SIFT considerably increases with the proposed algorithm of preprocessing the perspective image in the hybrid pair. It brings the advantage of automatic point matching between catadioptric omnidirectional and perspective images. Another approach we proposed is generating VCP image from omnidirectional image first and then applying SIFT.

We conducted three sets of experiments, having six pairs of hybrid images in each, with two different types of catadioptric cameras and a fish-eye camera. We observed that VCP approach performed best for catadioptric images and is more robust to increasing baseline. For fish-eye images, downsampling approach

performed as well as the VCP approach due to the fact that a fish-eye camera acts as a perspective camera with a large lens distortion and serves as a good candidate for matching with perspective cameras.

The proposed algorithm of preprocessing the perspective images works with different hybrid camera types as shown by experiments. Thus, we are able to say that proposed technique of extracting parameters of low-pass filtering and downsampling (cf. Section 4.1.1) is versatile for omnidirectional cameras up to a large extent.

We also investigated whether Lowe's false match elimination method with affine parameters solves the problem of matching incorrect scales. Experiment results show that, although it is able to eliminate a few false matches, it is not comparable to our algorithm because when all the false matches are eliminated, the number of remaining correct matches is very low. The reason is that the proposed approach improves the performance of initial SIFT matching rather than just eliminating the false matches afterwards. Therefore, we conclude that proposed algorithm is useful with or without employing an extra elimination. Moreover, the elimination of matches that do not conform to a transformation is not crucial at this step since we do eliminate the matches that do not conform to the epipolar geometry during the estimation of fundamental matrix with RANSAC (cf. Chapter 5) which is a more accurate geometrical constraint.

We should also mention that the optimal SR detection and preprocessing the high-resolution image approach proposed here can be used for other applications employing SIFT where an approximate scale ratio exists between the objects in the given images. Otherwise, if scale ratios considerably vary for objects in the scene, this approach is not suitable. In such a case, elimination method with affine or projective parameters as proposed by Lowe and explained in Section 4.1 is more meaningful.

Another approach to detect and match scale-invariant features for omnidirectional cameras was proposed by Hansen *et al.* [57]. Rather than extracting features by convolving the image, image is projected onto a sphere first and scale-space images are obtained as the solution of the heat (diffusion) equation on the sphere which is implemented in the frequency domain using spherical harmonics. They compare the performance of this spherical processing approach with the conventional SIFT. They performed experiments on a set of synthetic

parabolic and fish-eye images, where features are matched between the images of rotated and zoomed views of the same camera type and robustness to varying rotation and scaling is tested. The approach they proposed improved the results for fish-eye lenses for all cases, however conventional SIFT performed well enough for para-catadioptric cameras when there are both rotation and scaling. Also, the number of features detected by conventional approach is higher. On the other hand, it worths to mention that more work on this new approach may lead to better results.

CHAPTER 5

Robust Epipolar Geometry and Pose Estimation

This chapter focuses on the epipolar geometry and pose estimation steps of the SfM pipeline. Using the point correspondences obtained in the previous step, epipolar geometry between the camera views is extracted. This is used for both eliminating the false matches that do not obey the epipolar constraint and obtaining the motion parameters between views.

At the beginning of the chapter, the hybrid epipolar geometry is explained and related literature is given. Then, the details of our random sample consensus (RANSAC) implementation for fundamental matrix estimation and experiment results presented in Sections 5.2 and 5.3 respectively. Finally, the experimental comparison of the options for pose estimation (extraction of motion parameters) is given in Section 5.4.

5.1 Hybrid Epipolar Geometry

Epipolar geometry for omnidirectional cameras has been studied in the last decade. Svoboda and Pajdla [58] derived epipolar geometry constraints for catadioptric cameras with different mirror types where cameras are assumed to be calibrated. Geyer and Daniilidis [59] defined fundamental matrix for catadioptric cameras with paraboloidal mirrors employing lifted coordinates. They also presented their SfM work with uncalibrated cameras. Claus and Fitzgibbon [60] worked on epipolar geometry between cameras with large lens distortions such as fish-eye cameras. They proposed to use rational functional model and another lifting scheme to estimate fundamental matrix between images. Micusik and Pajdla [13] aimed to perform metric SfM with uncalibrated omnidirectional cameras and presented a method to estimate intrinsic and extrinsic parameters

at the same time by generalizing the technique given in [61].

Epipolar geometry between hybrid camera views was first explained by Sturm [17] for mixtures of para-catadioptric (catadioptric camera with a parabolic mirror) and perspective cameras. Barreto and Daniilidis showed that the framework can also be extended to cameras with lens distortion due to the similarities between the para-catadioptric and division models [18]. According to these studies, a 3x4 fundamental matrix describes the relationship between a perspective and a paracatadioptric image.

To summarize this relationship, let us denote the corresponding image points in perspective and catadioptric images as $\mathbf{q}_p$ and $\mathbf{q}_c$ respectively. They are represented as 3-vectors in homogeneous 2D coordinates. To linearize the equations between catadioptric and perspective images, *lifted coordinates* are used for the points in omnidirectional images. Lifting for para-catadioptric cameras can be performed by $\hat{\mathbf{q}}_c = (x^2 + y^2, x, y, 1)^\mathsf{T}$.

A point in the perspective image is related to a point in the catadioptric image by F_{pc}, which is the 3x4 hybrid fundamental matrix:

$$\mathbf{q}_p^\mathsf{T} \mathsf{F}_{pc} \hat{\mathbf{q}}_c = 0 \tag{5.1}$$

Using F_{pc}, geometric entity relations are:

$$\mathbf{l}_p = \mathsf{F}_{pc} \hat{\mathbf{q}}_c \ , \quad \mathbf{c}_c = \mathsf{F}_{pc}^\mathsf{T} \mathbf{q}_p \ , \quad \hat{\mathbf{q}}_c^\mathsf{T} \mathbf{c}_c = 0 \ , \quad \mathbf{q}_p^\mathsf{T} \mathbf{l}_p = 0 \tag{5.2}$$

where $\mathbf{l}_p$ is the epipolar line in the perspective image and $\mathbf{c}_c$ is the epipolar curve in the catadioptric image. Actually $\mathbf{c}_c$ is a 4-vector containing the four distinctive elements of the conic matrix encoding a circle:

$$\mathsf{C} = \begin{pmatrix} 2c_1 & 0 & c_2 \\ 0 & 2c_1 & c_3 \\ c_2 & c_3 & c_4 \end{pmatrix}$$

Hybrid epipolar geometry can be visualized in Fig. 5.1. An example of solved epipolar geometry and corresponding epipolar lines/conics are given in Fig. 5.2.

In the same manner, the relation between two para-catadioptric views can be represented by a 4x4 fundamental matrix. Lifted coordinates for hyperbolic

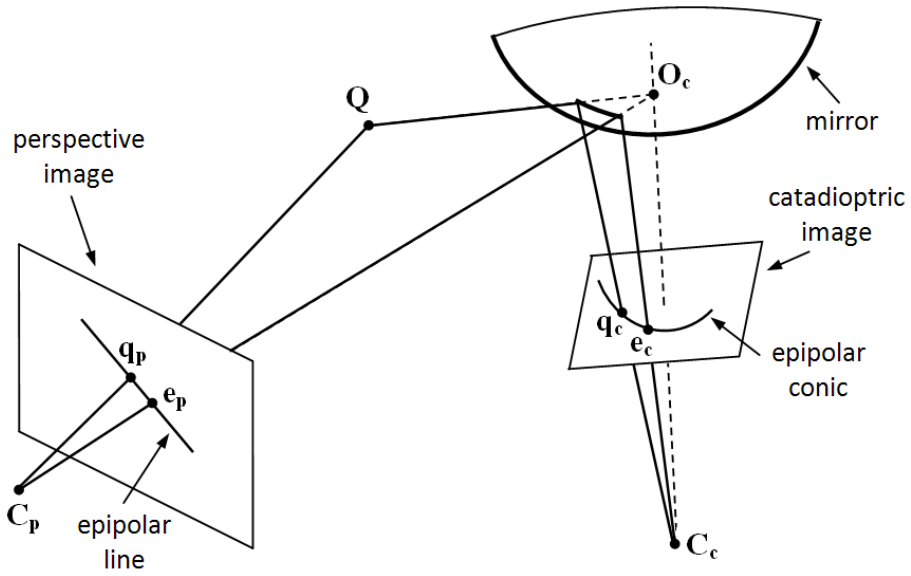


Figure 5.1: Epipolar geometry between a perspective and a catadioptric image. $\mathbf{q}_p$ and $\mathbf{q}_c$ are the corresponding points, $\mathbf{e}_p$ and $\mathbf{e}_c$ are the epipoles in the perspective and catadioptric images respectively.

mirrors are represented by 6-vectors since the corresponding conic does not have to be a circle. However, hyper-catadioptric images fail to satisfy a linear form of epipolar constraint with this schema [18]. It has been shown that a linear relation exists with a 15x15 fundamental matrix [19], on the other hand, if the mirror shape is close to a parabolic, 3x4 and 4x4 matrices can be used for hyper-catadioptric cameras as well since they are able to satify the relation to some extent [20].

5.2 Robust Epipolar Geometry Estimation

After initial detection of matches, RANSAC [22] algorithm based on hybrid epipolar relation can be used to eliminate false matches and estimated accurate F_{pc}. This step will be followed by extracting camera motion parameters as described in Section 5.4. Now, we will discuss the details of our robust epipolar geometry estimation approach.

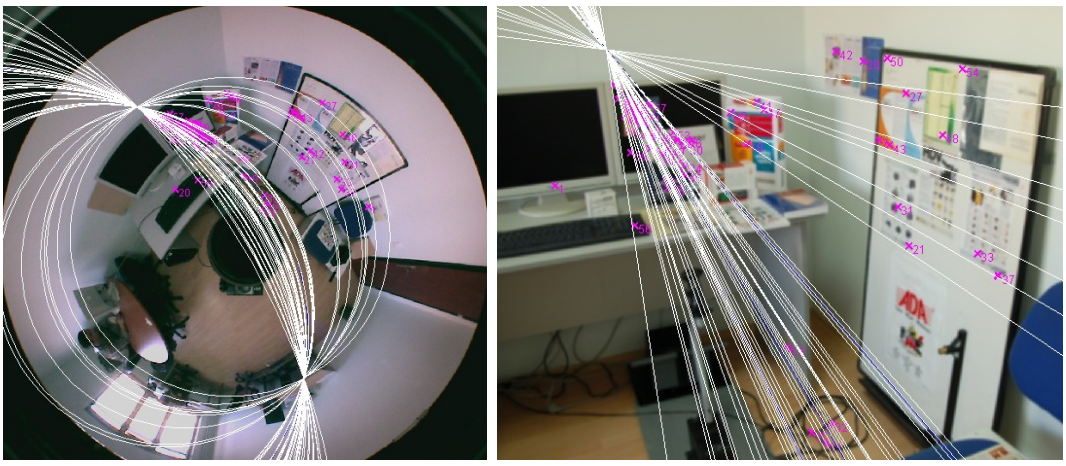


Figure 5.2: Example catadioptric-perspective pair and epipolar conics/lines of point correspondences.

5.2.1 Linear Estimation of Fundamental Matrix

Analogous to the perspective case, to obtain the elements of the hybrid fundamental matrix, we compose the equations representing epipolar constraint (Eq. 5.1). Since we employ lifted coordinates for catadioptric images, points and F_{pc} are in the following form:

$$\mathbf{q}_p = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad \hat{\mathbf{q}}_c = \begin{pmatrix} x'^2 + y'^2 \\ x' \\ y' \\ 1 \end{pmatrix} \quad \mathsf{F}_{pc} = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \end{pmatrix}$$

A pair of point correspondences results in one equation as

$$\begin{aligned}(x'^2 + y'^2)xf_{11} + (x'^2 + y'^2)yf_{21} + (x'^2 + y'^2)f_{31} + x'xf_{12} + x'yf_{22} + x'f_{32} \ \ldots \\ +y'xf_{13} + y'yf_{23} + y'f_{33} + xf_{14} + yf_{24} + f_{34} = 0\end{aligned} \tag{5.3}$$

When we stack up equations for all correspondences we obtain the linear system of equations $\mathsf{A}\mathbf{f} = 0$, where $\mathbf{f}$ is the column vector containing the elements of the fundamental matrix and matrix A is filled with point coordinates (a row per correspondence) using Eq. 5.3. Actually, 11 point correspondences are enough to find a unique solution for F_{pc} since it is defined up to a scale factor. Again due to

the fact that F_{pc} is defined up to a scale, adding an additional constraint $\|\mathbf{f}\| = 1$ enables us to avoid trivial solution. For overdetermined systems (more than 11 correspondences) and in the case of noise, the least-squares solution minimizing $\|\mathsf{A}\mathbf{f}\|$ subject to $\|\mathbf{f}\| = 1$ is selected.

It is known that solution of this problem is the unit eigenvector corresponding to the smallest eigenvalue of $\mathsf{A}^\mathsf{T}\mathsf{A}$. Namely, $\mathbf{f}$ is the singular vector corresponding to the smallest singular value of A, that is, the last column of V in the singular value decomposition (SVD): $\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{V}^\mathsf{T}$.

5.2.2 Normalization of Point Coordinates

Normalization of coordinates comprises carrying the origin to the centroid of points and scaling the coordinate values. Let n be the amount of scale normalization and (u_x, u_y) be the centroid of the points in the image, normalized form of 3-vector homogeneous coordinates is as follows [10]:

$$\mathbf{q}_{norm} = \begin{pmatrix} \frac{x-u_x}{n}, & \frac{y-u_y}{n}, & 1 \end{pmatrix} \tag{5.4}$$

One way to perform normalization for lifted coordinates is normalizing point coordinates before lifting them. Then, F_{pc} is computed with lifted coordinates. Lastly, we are supposed to denormalize the corresponding points, lines and conics to be able to calculate the distance error for purposes of outlier elimination or non-linear optimization of fundamental matrix.

Normalization can be performed after lifting as well. We define 4x4 T matrices for normalization of lifted coordinates ($\hat{\mathbf{q}}_{1norm} = \mathsf{T}_1\hat{\mathbf{q}}_1$ and $\hat{\mathbf{q}}_{2norm} = \mathsf{T}_2\hat{\mathbf{q}}_2$), so that normalized coordinates still suit to the lifted form, i.e. $(x^2 + y^2, x, y, 1)$.

$$\hat{\mathbf{q}}_{norm} = \begin{pmatrix} \frac{(x-u_x)^2}{n^2} + \frac{(y-u_y)^2}{n^2}, & \frac{x-u_x}{n}, & \frac{y-u_y}{n}, & 1 \end{pmatrix} \tag{5.5}$$

The transformation T defined as in Eq. 5.6 yields $\hat{\mathbf{q}}_{norm}$ when multiplied with unnormalized lifted coordinates ($\hat{\mathbf{q}}$).

$$\hat{\mathbf{q}}_{norm} = \mathsf{T}\hat{\mathbf{q}} = \begin{pmatrix} \frac{1}{n^2} & \frac{-2u_x}{n^2} & \frac{-2u_y}{n^2} & \frac{u_x^2+u_y^2}{n^2} \\ 0 & \frac{1}{n} & 0 & \frac{-u_x}{n} \\ 0 & 0 & \frac{1}{n} & \frac{-u_y}{n} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^2+y^2 \\ x \\ y \\ 1 \end{pmatrix} \tag{5.6}$$

With these T matrices, denormalization of F can be performed linearly by $\mathsf{F} = \mathsf{T}_2^\mathsf{T}\mathsf{F}_{norm}\mathsf{T}_1$, which produces correct epipolar conics/lines on which we can calculate distance error directly contrary to the *normalize before lifting* approach.

Normalization is crucial for fundamental matrix estimation. As discussed in [62], using unnormalized coordinates causes inhomogeneous weighting for the elements of the F matrix during the linear estimation. Regarding F estimation for perspective cameras, consider an image point with homogeneous 2D coordinates (100,100,1) and assume that its correspondence has the same coordinates as well. A row of the matrix A becomes

$$\begin{aligned} \mathbf{r}^\mathsf{T} &= \begin{pmatrix} x'x & x'y & x' & y'x & y'y & y' & x' & y' & 1 \end{pmatrix} \\ &= \begin{pmatrix} 10^4 & 10^4 & 10^2 & 10^4 & 10^4 & 10^2 & 10^2 & 10^2 & 1 \end{pmatrix} \end{aligned}$$

The contribution to the matrix $\mathsf{A}^\mathsf{T}\mathsf{A}$ is of the form $\mathbf{r}\mathbf{r}^\mathsf{T}$, which contains entries ranging between 10^8 and 1. This results in a very large condition number for $\mathsf{A}^\mathsf{T}\mathsf{A}$ and negatively affects the computation. The ideal condition would be reached by normalizing the point coordinates to (1,1,1) since it would bring the situation that $\mathbf{r}^\mathsf{T}$=(1 1 1 1 1 1 1 1 1). Instead of choosing different scale factors for each of the two axes, an isotropic scaling factor is chosen so that the average distance of a point to the origin is equal to $\sqrt{2}$. This means that the *average point* is equal to (1,1,1).

For hybrid case, due to the lifted coordinates, normalizing to (1,1,1) with $n = \sqrt{2}$ does not makes all entires of $\mathbf{r}^\mathsf{T}$ equal to one:

$$\mathbf{r} = \left((x'^2+y'^2)x, (x'^2+y'^2)y, (x'^2+y'^2), x'x, x'y, x', y'x, y'y, y', x', y', 1\right)^\mathsf{T}$$

Table 5.1: Entries of $\mathbf{r}$ for varying scale normalization factors, (n_{omni},n_{pers}).

(n_{omni}, n_{pers})	$\mathbf{r}^\mathsf{T}$
$(\sqrt{2}, \sqrt{2})$	(2 2 2 1 1 1 1 1 1 1 1 1)
$(1, \sqrt{2})$	(1 1 1 0.707 0.707 0.707 0.707 0.707 0.707 1 1 1)
$(\sqrt{2}/2, \sqrt{2})$	(0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 1 1 1)

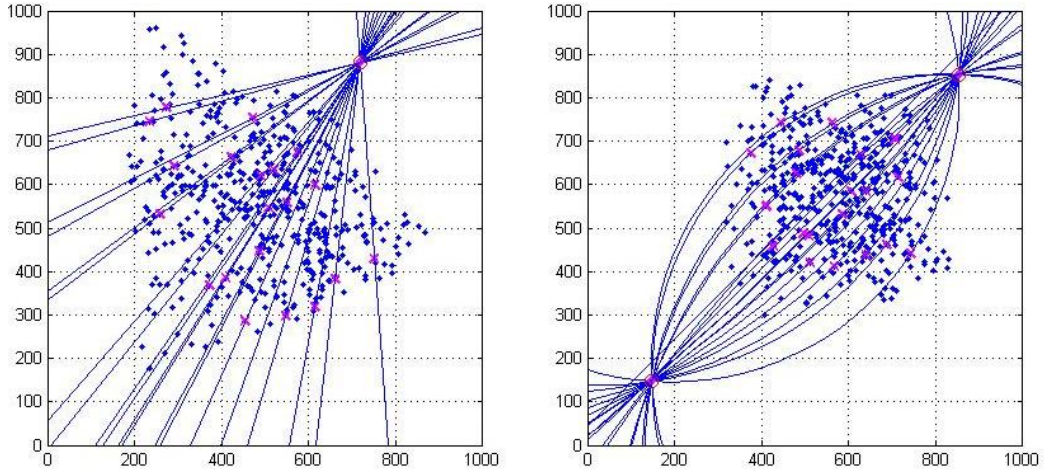


Figure 5.3: Simulation hybrid images for the normalization experiment.

In fact, it does not seem possible to find an n value for the omnidirectional image to produce constant values for entries of $\mathbf{r}$. Table 5.1 shows a few possible n_{omni} values and corresponding values of $\mathbf{r}$. In the following, we present our experimental analysis on different n values.

Experiment. We conducted an experiment for a hybrid pair of simulated images to evaluate the effectiveness of coordinate normalization and to choose the best values of (n_{omni}, n_{pers}). Simulated images are shown in Fig. 5.3. Points are created randomly in a volume of 1.0x1.0x1.5 m^3. Image size is 1000x1000 pixels and Gaussian location noise with $\sigma = 1.0$ added to both images. (ξ, f, c_x, c_y) values of the perspective and omnidirectional cameras are (0,600,500,500) and (1,600,500,500) respectively.

Table 5.2 shows the results of the experiment for varying (n_{omni}, n_{pers}) values. F_{pc} is estimated with 12, 24, 36 and 48 points. Errors in the table are, for a given correspondence, the sum of distances from points to corresponding epipolar

Table 5.2: Median distance errors (in pixels) for different fundamental matrices computed with varying number of points and scale normalization values. Experiment was repeated 1000 times, values shown are the mean of median distance errors over 1000 trials. Image size is 1000x1000 pixels and Gaussian location noise with $\sigma = 1.0$ added to both images.

n_{omni}, n_{pers}	**12 points**	**24 points**	**36 points**	**48 points**
$2, \sqrt{2}$	5.451	1.955	1.932	1.934
$1.8, \sqrt{2}$	5.141	1.934	1.937	1.908
$1.6, \sqrt{2}$	4.923	1.973	1.917	1.923
$\sqrt{2}, \sqrt{2}$	4.871	1.938	1.914	1.919
$1.2, \sqrt{2}$	4.992	1.888	1.919	1.919
$1, \sqrt{2}$	5.325	1.932	1.923	1.909
$0.85, \sqrt{2}$	5.146	1.950	1.923	1.892
$\sqrt{2}/2, \sqrt{2}$	5.366	1.998	1.934	1.928
$1, 1$	4.985	1.987	1.945	1.938

curves/lines in both images.

We infer from the table that the performance difference between different n_{omni} values is quite small and $\sqrt{2}$ is the best performer for most of the cases. Therefore, we continued our work by using (n_{omni},n_{pers})=($\sqrt{2}$,$\sqrt{2}$).

5.2.3 Distance Threshold

As in the perspective camera case, we define a distance (d) to distinguish outliers and inliers, where the points closer to its corresponding epipolar line/curve than d are called inliers. In our experiments, we use $d = d_l + d_c$, where d_l is the point-to-line distance in the perspective image and d_c is the point-to-conic distance in the catadioptric image.

We aimed to employ a varying threshold which would be small for points close to epipoles and larger for points far away from the epipoles but estimated location of epipoles greatly varies between trials of RANSAC which made such a varying threshold schema unstable.

We also performed experiments to evaluate the use of Sampson distance, which is the first order approximation to the geometric error [10]. Results were slightly worse when compared to the case of using geometric distance error (d).

5.2.4 Applying Rank-2 Constraint

F is a singular rank-2 matrix. Enforcing this constraint is important to obtain epipolar lines intersecting each other at the same points, i.e. epipoles. We need to correct the F estimated linearly by point correspondences. A fast way of doing this is *direct imposition* (DI) of rank-2 constraint, a slower but more successful way is non-linear optimization of F using its orthonormal representation (OR).

Direct imposition: F is replaced by F' that minimizes the Frobenius norm $\|\mathsf{F} - \mathsf{F}'\|_F$ subject to the condition $\det(\mathsf{F}') = 0$. Here, $\|.\|_F$ represents Frobenius norm, i.e. the square root of the sum of the squares of all entries of the matrix. This operation can be performed using SVD. Let $\mathsf{F} = \mathsf{U}\mathsf{D}\mathsf{V}^\mathsf{T}$, where D is a diagonal matrix $\mathsf{D} = \mathrm{diag}(r, s, t)$ satisfying $r \geq s \geq t$. Then, $\mathsf{F}' = \mathsf{U}\ \mathrm{diag}(r, s, 0)\ \mathsf{V}^\mathsf{T}$.

Non-linear optimization of orthonormal representation: F is refined by performing a minimization of distances from points to their corresponding epipolar conic/line. A way to guarantee rank-2 is using the matrix parameterization proposed in [63] which is called the orthonormal representation of the fundamental matrix. To describe this representation we again employ SVD. Let $\mathsf{F} = \mathsf{U}\mathsf{D}\mathsf{V}^\mathsf{T}$ where D is supposed to be $\mathrm{diag}(r, s, 0)$ satisfying $r \geq s$. If we define $\mathsf{D} = \mathrm{diag}(1, \sigma, 0)$ such that $\sigma = s/r$, fundamental matrix can be represented as

$$\mathsf{F} \sim \mathbf{u}_1 {\mathbf{v}_1}^\mathsf{T} + \sigma \mathbf{u}_2 {\mathbf{v}_2}^\mathsf{T} \tag{5.7}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} columns of U and V respectively. For hybrid case, $\mathbf{u}_i$ is a 4-vector whereas $\mathbf{v}_i$ is a 3-vector. Together with σ, we have 15 parameters to be optimized in Eq. 5.7. We use Levenberg-Marquardt (LM) method provided by the function **lsqnonlin** in Matlab.

Experiment. To investigate the amount of improvement gained by non-linear optimization we perfomed a test on simulated images similar to the one given in Section 5.2.2. We infer from Table 5.3 that performance increase with non-linear optimization is significant. However, we also know that optimization requires considerable time. Therefore, we suggest to employ *direct imposition* during the RANSAC algorithm where F is computed hundreds of times and to apply non-linear optimization for the final estimation with the inlier points.

Table 5.3: Median distance errors to compare the two rank-2 imposition methods. Experiment was repeated 50 times and the mean of these 30 trials are given. Image size is 1000x1000 pixels and Gaussian location noise with $\sigma = 1.0$ added to both images.

method	**12 points**	**24 points**	**36 points**	**48 points**
Direct imposition	4.871	1.938	1.914	1.919
Non-linear optim. of OR	2.034	1.352	1.593	1.583

The reader should also note that rank-2 imposition methods should be applied to the F just after its computation with normalized coordinates and before denormalizing F with the transformation matrices. The reason is that these methods treat all entries of the matrix equally, regardless of their magnitude and the magnitude ratios between the elements of F when it satisfies the epipolar constraint with unnormalized/denormalized coordinates are different.

5.3 Experiments of Outlier Elimination

To eliminate the false matches in the SIFT output, we applied RANSAC based on hybrid epipolar geometry. For the wide baseline image pairs of Fig. 4.7 and Table 4.3, the number of matches and successful match ratios before and after RANSAC elimination are given in Table 5.4. We repeated RANSAC 30 times for each pair and recorded the mean values.

We employed the general scheme of RANSAC which is given in [10, p. 291]. We need to decide on the number of point correspondences to be used for linear estimation of F during RANSAC iterations. Let k be the minimum number of correspondences (for F_{pc} $k = 11$). Using more than k points improves the results provided that the number of correct correspondences permits. In Table 5.4, we used $2k$ because we are comfortable that a number of random selection of $2k$ will be free of outliers. Only for Pers3-Omni2 pair we used k correspondences due to fewer number of correspondences.

Please note that, we run RANSAC always on the original images. If we use VCP approach, we first calculate the coordinates of matched points in omnidirectional image by backward mapping and use them in RANSAC. In this way, same distance threshold value (d) corresponds to same error for VCP approach.

Table 5.4: Matching results after RANSAC for hybrid image pairs (cf. Table 4.3). Distance threshold of RANSAC, d, was set to 15 pixels. Pers$N$ σA dB indicates that Pers$N$ image was blurred with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction.

	initial matching		**T/F after**
Image pairs	**total**	**T/F restricted**	**RANSAC**
Pers2 - Omni1	75	51/7	49.3/2.1
Pers2 σ1.5 d1.65 - Omni1	75	67/3	66.8/0.3
Pers2 σ1.5 - VCP1	75	73/1	70.7/0.0
Pers3 - Omni1	60	39/9	36.9/3.6
Pers3 σ1.5 d1.65 - Omni1	60	50/6	47.9/2.5
Pers3 σ1.5 - VCP1	60	57/1	54.5/0.0
Pers2 - Omni2	60	30/8	29.1/0.0
Pers2 σ2.5 d3.6 - Omni2	60	54/1	53.8/0.5
Pers2 σ2.5 - VCP2	60	56/3	54.6/0.0
Pers3 - Omni2	45	15/1	12.9/0.4
Pers3 σ2.5 d3.3 - Omni2	45	32/6	26.3/6.7
Pers3 σ2.5 - VCP2	45	35/3	33.4/1.0

As a result, we can say that remaining false matches can be eliminated by RANSAC to a great extent. If there are still a few false matches it means these false matched points are very close to the corresponding epipolar line/conic by coincidence.

5.4 Pose Estimation

Pose estimation is the step of extracting motion parameters of the cameras w.r.t each other. To do this we first obtain the essential matrix (E). Then the motion parameters (R,$\mathbf{t}$) can be extracted from it using the technique given in [10, p. 258].

We analyze two methods for the estimation of E. We also compare the effectiveness of these methods by experiment as will be seen shortly. First option is directly estimating E with the calibrated 3D rays of the correspondences in the RANSAC output. The other option is estimating F_{pc} with RANSAC and then extracting E from F_{pc} using the relation [18]:

$$\underbrace{\mathbf{q}_p^{\mathsf{T}}\mathsf{K}_p^{-\mathsf{T}}}_{\bar{\mathbf{q}}_p^{\mathsf{T}}}\,\mathsf{E}\,\underbrace{\Theta^{\mathsf{T}}\hat{\mathsf{K}}_c^{\mathsf{T}}\hat{\mathbf{q}}_c}_{\bar{\mathbf{q}}_c} = 0 \tag{5.8}$$

where $\bar{\mathbf{q}}_p$ and $\bar{\mathbf{q}}_c$ are the normalized 3D rays for perspective and catadioptric cameras respectively. K_p is the calibration matrix of perspective camera, $\hat{\mathsf{K}}_c$ is the lifted calibration matrix of catadioptric camera, expressed in the sphere model. Finally, Θ^{T}, given by

$$\Theta^{\mathsf{T}} = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

carries the origin of coordinate system to the center of sphere (cf. Fig. 2.2 and Eq. 2.1) linearly with lifted coordinates for paracatadioptric cameras. A rather complicated solution for hypercatadioptric cameras is given in [18].

We conducted an experiment on simulated data to compare these two options. In this experiment, point correspondences set does not contain outliers and 3D points are selected from a volume of 4 m^3 which is one meter away from the two simulated cameras. Table 5.5 shows 2D and 3D errors for direct-E and E-from-F methods for varying number of point correspondences and varying amount of Gaussian location noise added to both images.

We observe from the table that direct-E approach is more successful for all cases. E-from-F method is more susceptible to noise. The top two rows in the table shows the results for noise-free case. Results of E-from-F method are still slightly worse even for such small errors, which indicates the susceptibility of the computation of E-from-F method. We also conducted experiments with two perspective cameras to compare direct-E and E-from-F methods. They performed same in that case. Thus, we infer that E-from-F becomes disadvantageous in hybrid case. Possible reasons are: hybrid F has 12 elements to estimate whereas E has 9, lifted point coordinates are used in hybrid case and extracting E from hybrid F is relatively complicated.

In the experiment presented above, true camera parameters are used for both direct-E and E-from-F methods. To investigate the case when the calibration is erroneous and camera parameters are not perfectly estimated, we repeated the experiment with calibration noise in addition to the location noise on pixels. We added Gaussian noise with $\sigma = 2 - 5\%$ to the parameter values. For instance,

Table 5.5: Comparison of the two E-estimation methods: direct-E and E-from-F. Table shows median 2D reprojection errors (in pixels) and 3D estimation errors (in meters) after linear-Eigen triangulation. Values are the mean of 50 trials. σ indicates the standard deviation of Gaussian location noise added to both images.

Method, # of points, noise	**Pers. 2D err.**	**Omni. 2D err.**	**3D err.**
direct-E, 20 points, $\sigma = 0$	1.9e-13	1.0e-13	2.5e-15
E-from-F, 20 points, $\sigma = 0$	5.4e-13	2.7e-13	3.5e-15
direct-E, 20 points, $\sigma = 1$	1.66	0.802	0.0263
direct-E, 40 points, $\sigma = 1$	1.19	0.580	0.0206
direct-E, 80 points, $\sigma = 1$	1.04	0.511	0.0182
E-from-F, 20 points, $\sigma = 1$	4.07	1.99	0.0395
E-from-F, 40 points, $\sigma = 1$	2.89	1.43	0.0308
E-from-F, 80 points, $\sigma = 1$	1.85	0.936	0.0216
direct-E, 20 points, $\sigma = 2$	3.34	1.64	0.0607
direct-E, 40 points, $\sigma = 2$	2.38	1.15	0.0443
direct-E, 80 points, $\sigma = 2$	2.10	1.02	0.0370
E-from-F, 20 points, $\sigma = 2$	9.11	4.61	0.0957
E-from-F, 40 points, $\sigma = 2$	5.02	2.50	0.0527
E-from-F, 80 points, $\sigma = 2$	4.14	2.09	0.0464

focal length (f) with a value of 500 pixels was perturbed with a Gaussian noise with $\sigma = 10$ to $\sigma = 25$. The results are parallel to the ones given in Table 5.5 with the difference that the errors are increased due to the calibration noise. E-from-F performed worse than direct-E even in the case when there is no location noise but only calibration noise.

We should also keep in mind that direct-E approach is independent from camera type and can be used for all type of cameras as long as calibration is performed, whereas E-from-F approach practically possible only for perspective and para-catadioptric cameras.

5.5 Conclusions

We robustly estimated the hybrid epipolar geometry using RANSAC and used it to eliminate a few false matches. We discussed the important aspects of fundamental matrix estimation such as coordinate normalization, rank-2 imposition and distance threshold.

We defined normalization matrices for lifted coordinates and performed analysis on the effectiveness of coordinate normalization and to choose the best values of scale normalization factor for hybrid case. We concluded that, as suggested for perspective cameras, carrying the origin to the centroid of points and scaling coordinates so that the average distance of a point to the origin is equal to $\sqrt{2}$ is proper for normalizing point coordinates in omnidirectional camera images.

We also evaluated the alternatives for pose estimation and decided on estimating the essential matrix with the calibrated 3D rays of point correspondences rather then extracting the essential matrix from fundamental matrix. This conclusion is based on both the performance difference observed in the experimental analysis and the limitations of obtaining E-from-F for hybrid case.

CHAPTER 6

TRIANGULATION

Triangulation is the step of estimating 3D coordinates of the matched 2D points using camera poses. In this chapter, we present the proposed weighting strategy for *iterative linear-Eigen* triangulation method and show its effectiveness in increasing the 3D structure estimation performance with simulated images. Also, a two-view hybrid SfM experiment is presented in Section 6.3 in order to evaluate the proposed triangulation approach for real images.

6.1 Weighted Triangulation for Mixed Camera Images

We generalized *iterative linear-Eigen* triangulation method for effective use in a mixed structure-from-motion pipeline. According to the comprehensive study by Hartley and Sturm [28], *iterative linear-Eigen* is one of best triangulation methods for Euclidean reconstruction. It is superior to midpoint method and non-iterative linear methods especially when 2D error is considered. For projective reconstruction, polynomial triangulation method performs better, however it was also mentioned that, although *iterative linear-Eigen* is not projective invariant its performance under projective reconstruction is close to that of polynomial triangulation method. Moreover, polynomial method requires a considerable amount of computation time and not easily generalizable to more than two images.

Let us briefly go over the *iterative linear-Eigen* method. Let the two corresponding point coordinates be $\mathbf{q} = (x, y, 1)$, $\mathbf{q}' = (x', y', 1)$ and their projections are represented by $\mathbf{q} = \mathsf{P}\mathbf{Q}$, $\mathbf{q}' = \mathsf{P}'\mathbf{Q}$. Letting $\mathbf{p}_i$ denote the i^{th} row of P, $\mathbf{Q}$ satisfies:

$$\mathsf{A}\mathbf{Q} = 0 \quad \text{where} \quad \mathsf{A} = \begin{pmatrix} x\mathbf{p}_3 - \mathbf{p}_1 \\ y\mathbf{p}_3 - \mathbf{p}_2 \\ x'\mathbf{p}'_3 - \mathbf{p}'_1 \\ y'\mathbf{p}'_3 - \mathbf{p}'_2 \end{pmatrix} \tag{6.1}$$

For multi-view triangulation, two rows are added to A for each view. Least square solution is the last column of V in the singular value decomposition $\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{V}^\mathsf{T}$, which is the unit eigenvector corresponding to the smallest eigenvalue of the matrix $\mathsf{A}^\mathsf{T}\mathsf{A}$. This method is extended by adjusting the weights of rows iteratively such that reprojection error will be decreased. The weights for the first and second views are $\frac{1}{\mathbf{p}_3\mathbf{Q}}$ and $\frac{1}{\mathbf{p}'_3\mathbf{Q}}$ respectively [28].

Please note that we employ this method with calibrated 3D rays instead of raw pixels. Since the projection in omnidirectional cameras can not be expressed linearly as in perspective cameras, hybrid triangulation can be performed with the 3D rays outgoing from the effective viewpoints of the cameras.

The perspective cameras in mixed systems tend to have higher resolution than the omnidirectional ones. To benefit from their resolution, we increased the weight of rows coming from perspective images. With the mentioned weighting strategy, we observed improvement in the accuracy of estimated 3D coordinates. We relate the amount of weighting to three factors: the scale ratio, the distance to the scene and the position of the object in the omnidirectional image. Let us examine these one by one.

Since we perform triangulation with the 3D rays, an object in two images with different focal lengths cover different amount of areas in the images but have same field of view in the scene. One pixel noise in the zoomed image corresponds to lower angular error and distance error on the object. In Fig. 6.1a, the case of doubling the focal length is depicted. If one pixel noise in the left image causes δ distance error, then same noise in the right image causes 2δ error. In this case, we need to increase the weight of the zoomed camera, otherwise triangulation gives equal weight to both of them.

Distance to the scene affects the triangulation with 3D rays. Rays diverge as the camera comes closer to the scene. Object in the image gets larger but this should be distinguished from the zoom effect. We depict this case in Fig. 6.1b. Both cameras have equal focal length values but distance between the camera on

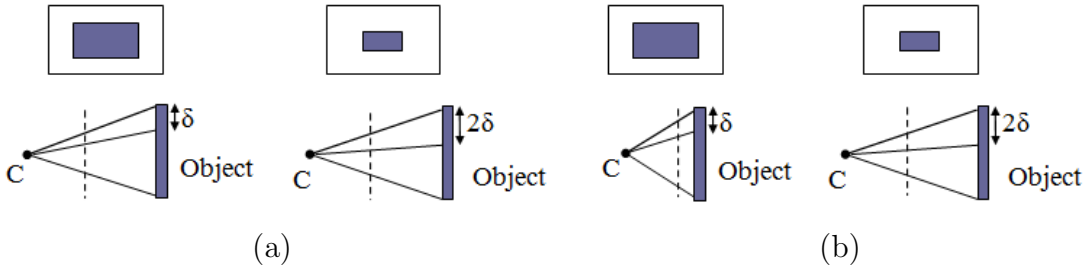


Figure 6.1: Depiction of doubling the focal length and decreasing the camera-scene distance for triangulation on normalized rays, i.e. normalized image plane. On the top row, images are shown with the object observed. Bottom row shows the field of view of the camera (top and bottom 3D rays) corresponding to the object and the distance error on the object (δ or 2δ) corresponding to one pixel error in the images. C represents the camera center (i.e. pinhole). Dashed line represents the normalized image plane. Iterative linear triangulation on 3D rays minimizes the error on that plane. (a) Camera on the left has twice the focal length of the camera on the right. Distance between the camera and the scene is equal for both cameras. (b) Both cameras have equal focal length values but distance between the camera on the left and the scene is half the distance between the camera on the right and the scene.

the left and the scene is half the distance between the camera on the right and the scene. Triangulation already gives support to left camera because it minimizes the reprojection error on the normalized 3D rays. We do not need to increase the weight.

The third and the last factor is the position of the points in the catadioptric image. When the objects in the scene have approximately the same height with the camera, x and y values in the $(x, y, 1)$ form of normalized 3D rays have quite high values compared to the values in the perspective images. These high values cause an unwanted support for the rows coming from catadioptric image. When the objects are below the camera this does not occur. We observed that, for such elevated objects, i.e. for the points near the periphery of the catadioptric image, increasing weight of perspective camera improves the results.

To combine the effects in a single weight coefficient, we propose to multiply the rows of perspective image by s

$$s = r_s \cdot r_d \cdot p \tag{6.2}$$

where r_s is assigned as the ratio of the scales of the object in perspective and omnidirectional images repsectively. Please note that this value is not supposed to be the ratio of focal length values of the cameras. When camera type differs, focal length ratio is not the ratio of the sizes of object in the images. Distance ratio is represented by r_d and assigned as the ratio of distances to the scene from perspective and omnidirectional cameras respectively. Approximate information of camera positions can be used for less controlled image capturing. Finally p represents the position factor. If objects/points are high and represented at the periphery of the catadioptric image, we increase p i.e. weight of the perspective camera. The value of p was chosen emprically from our experiments. Detailed results of the experiments analyzing various cases are given in Section 6.2.

Please note that for two perspective cameras, triangulation can be performed with pixels (not with 3D rays) and since the reprojection error in the image is minimized, the iterative linear triangulation supports the zoomed image without requiring an extra weighting for the zoomed images. For the distance effect, becoming closer to the scene is similar to the zoom effect and position of the point in the image does not have a significant effect for perspective cameras.

6.2 Experiments

In this section, we analyze the improvement for the proposed weighted triangulation approach (cf. Section 6.1) on simulated data. We generated a total of 1000 points that are regularly distributed on a planar grid. We added Gaussian location noise to all simulated images with σ=2.0 pixels. We define two main scenarios to distinguish between the cases when observed points are below the catadioptric camera or at the same horizontal level with the camera. We depicted these two cases and camera positions in Fig. 6.2. Two of the camera positions are selected each time also with varying focal length values to create analyzed scale ratios.

For the case that the grid is below the catadioptric camera, we have four experiments, results of which are given in Table 6.1. Camera positions, scale ratios (r_s) and distance ratios w.r.t. the grid (r_d) are indicated in the table. Error is expressed by the median 3D estimation error. Fig. 6.3 shows the images of the experiment in the first row of Table 6.1 as an example. Applied weight

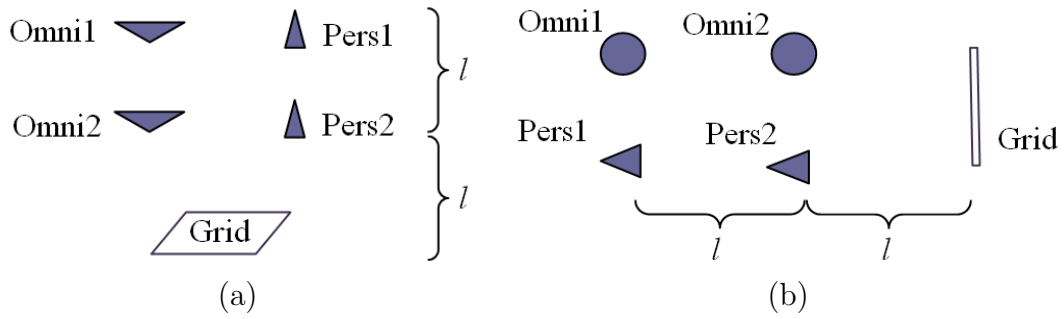


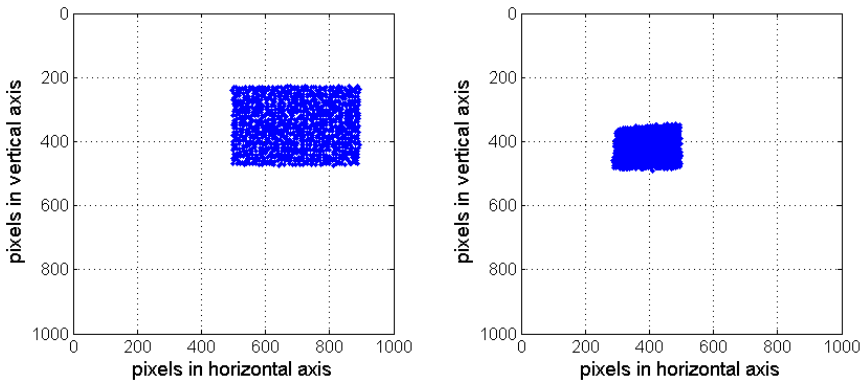
Figure 6.2: Camera and grid positions in the scene of triangulation experiments. (a) The grid is below the catadioptric camera, side view. (b) The grid is at the same horizontal level with the cameras, top view. Distance between the cameras is 2 meters. Maximum distance between the cameras and the scene ($2l$) is 2.5 meters.

Table 6.1: Results of triangulation experiments for the scene given in Fig. 6.2a. Error is expressed by 3D coordinate estimation error in meters, median of 1000 points in the grid. Experiments were repeated 30 times and the values in the table are the mean of these 30 experiments. Gaussian location noise with σ=2.0 pixels was added to both images.

pair	r_s	r_d	$w = 1$	$w = s$	**improvement**
Omni1-Pers1	2	1	0.0250	0.0242	3.2%
Omni1-Pers1	4	1	0.0334	0.0302	9.6%
Omni1-Pers2	2	0.5	0.0200	0.0200	-
Omni2-Pers1	1	2	0.0170	0.0166	2.4%

value is represented with w. Errors for $w = 1$ and $w = s$ are compared in the table, where s is calculated by Eq. 6.2 and $p = 1$ for the current case. We observe that proposed weighting (s) produces better results. We expressed the improvement as percentage of decrease in error. Improvement becomes significant when r_s increases which is quite likely for hybrid pairs. For Omni1-Pers2 case, the effects of r_s and r_d cancel each other and $s = 1$ already. We put this experiment to indicate the importance of r_d because we tested with our experiments that $w = 1$ is better than $w = 2$.

When the grid is close to the same horizontal level with the camera (Fig. 6.2b) we again conducted the same experiments, results of which are given in Table 6.2. This time $p \neq 1$ due to the position of grid w.r.t. the catadioptric camera. We observed in our experiments that $p = 2$ is the best performer for



(a) (b)

Figure 6.3: Simulated hybrid image pair of the experiment in the first row of Table 6.1. (a) perspective image, (b) omnidirectional image.

Table 6.2: Results of triangulation experiments for the scene given in Fig. 6.2b. Error is expressed by 3D coordinate estimation error in meters, median of 1000 points in the grid. Experiments were repeated 30 times and the values in the table are the mean of these 30 experiments. Gaussian location noise with σ=2.0 pixels was added to both images.

pair	r_s	r_d	p	$w = 1$	$w = s$	**improvement**
Omni1-Pers1	2	1	2	0.0266	0.0254	4.5%
Omni1-Pers1	4	1	2	0.0317	0.0288	9.2%
Omni1-Pers2	2	0.5	2	0.0221	0.0216	2.3%
Omni2-Pers1	1	2	2	0.0171	0.0162	5.3%
Omni1-Pers1	2	1	1	0.0266	0.0261	1.9%
Omni1-Pers1	4	1	1	0.0317	0.0298	6.0%
Omni1-Pers2	2	0.5	1	0.0221	0.0221	-
Omni2-Pers1	1	2	1	0.0171	0.0167	2.3%

this case and increasing p further does not improve the results. The table also shows the results for $p = 1$ which is the case neglecting the position factor. Please compare the improvements in the top four and bottom four rows to observe the difference gained by including position factor.

When the observed scene points are below the horizontal level of the catadioptric camera but not directly below (a case between Fig. 6.2a and Fig. 6.2b), it is appropriate to increase p from 1 to 2 gradually as the 3D points get higher. The value of p can be assigned according to the vertical angles of the 3D rays corresponding to the points as demonstrated in the experiment in Section 6.3.

6.3 Structure-from-Motion with Real Images

In this section, with the experience gained by the experiments performed for steps of point matching, pose estimation and triangulation, we choose the best performing methods and complete a SfM experiment with a real hybrid image pair.

We estimate the essential matrix with the calibrated rays of point correspondences selected by RANSAC. We use the RANSAC output for Pers2 σ1.5 - VCP1 pair (refer to Table 5.4), which has 70 correspondences. We employ the proposed weighted iterative linear-Eigen triangulation. In Fig. 6.4, at the top row, we see the correspondences selected by RANSAC, at the bottom row we observe 2D side-view (left) and 2D top-view (right) of the reconstructed scene, where O_p and O_c shows perspective and catadioptric camera centers, $(\mathsf{X}_p, \mathsf{Y}_p, \mathsf{Z}_p)$ and $(\mathsf{X}_c, \mathsf{Y}_c, \mathsf{Z}_c)$ shows perspective and catadioptric camera axes, respectively.

For triangulation, we employ weighted approach and compute w with $s = r_s \cdot r_d \cdot p$ consulting to the results of Section 6.2. We take r_s=1.65 which was already extracted in point matching step for the current image pair (Section 4.3, Table 5.4). We know that $r_d \approx 2$ since we set up the experiment environment, however one can also use the result of an initial triangulation (like Fig. 6.4 bottom row) to obtain an approximate ratio of distances to the scene. We employ varying p values for the points according to the vertical angles of their corresponding 3D rays. The vertical angles change between 50°-90° (0° indicates directly below the omnidirectional camera) and we take p gradually increasing from 1.5 to 2 with the increasing angle.

To estimate the improvement gained by the proposed weighting scheme, we compare a number of real world distances with the ones in the estimated 3D structure. We measured 30 distances at the real scene corresponding to the distances between estimated 3D points. They are not in the same scale, thus we equalized the scale of the distances using the ratio between the averages of 30 distances. We measure the accuracy with the absolute difference of the distances at the real scene and at the reconstructed scene. Table 6.3 shows the median of these 30 distance errors (in centimeters) for $w = 1$ and $w = s$. One can see the improvement brought by employing the proposed weighting scheme. For reference, these 30 measured distances vary between 11.2 cm. and 31.8 cm. having

Table 6.3: Distance estimate errors after triangulation for hybrid real image pair. Error is expressed with the absolute difference between the measured real-world distances and the estimated distances after triangulation (in centimeters). Values are the median of 30 distance errors.

r_s	r_d	p	$w = 1$	$w = s$	improvement
1.65	2	1.5-2	0.88	0.83	5.7%

a median value of 16.5 cm. and standard deviation of 5.0 cm. The distance error obtained by the proposed approach has a median value of 0.83 cm. (indicated in the table) and standard deviation of 0.69 cm. varying between 0.42 and 3.2 cm.

6.4 Conclusions

The work presented in this chapter aimed to achieve more accurate triangulation for hybrid camera images and based on the fact that the significant resolution difference between omnidirectional and perspective images should be considered.

We chose the *iterative linear-Eigen* triangulation method since it is superior to *midpoint* method and non-iterative linear methods and one of the best alternatives for Euclidean reconstruction. We developed a weighting strategy by analyzing the factors affecting the 3D estimation accuracy and showed its effectiveness with simulated images. We also presented a two-view hybrid SfM experiment and demonstrated the improvement gained by the weighted triangulation approach with an analysis on distance estimation errors.

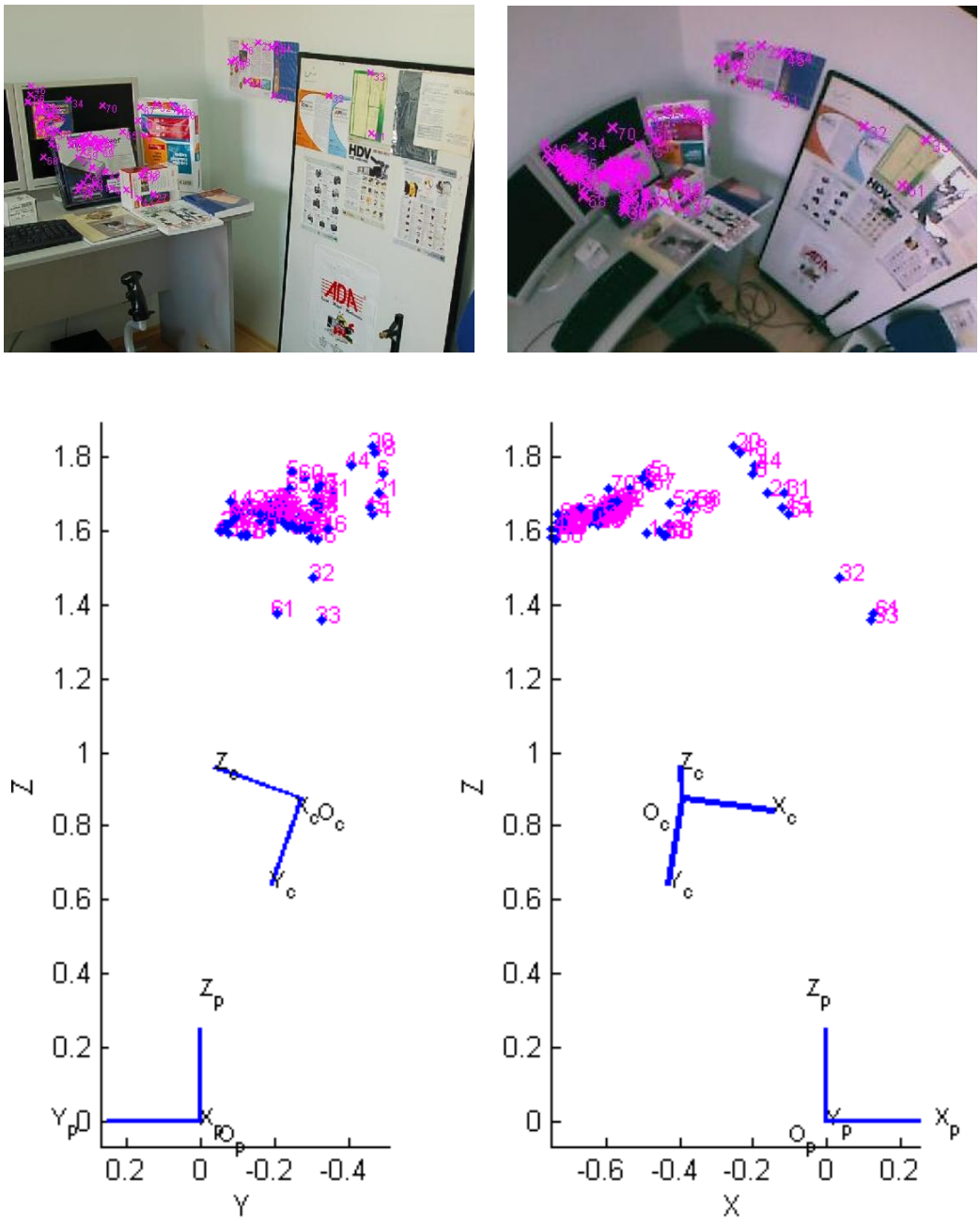


Figure 6.4: Reconstruction with hybrid real image pair. Selected correspondences on images are viewed on top. Images are cropped to make points distinguishable. At the bottom row, 2D side-view (left) and 2D top-view (right) of the reconstructed scene can be observed.

CHAPTER 7

Multi-view SfM

This chapter presents the work on multi-view Structure-from-Motion (SfM) with hybrid images. We start by describing the approach we employ to integrate additional views for multi-view SfM (Section 7.1). The implementation of sparse bundle adjustment and the improvement it provides are discussed in Section 7.2. One of the main motivations in this thesis is presented with the theory and experiment given in Section 7.3 where the structures estimated with two perspective cameras with no view in common are combined by pairing them with an omnidirectional camera.

7.1 Computing the Projection Matrices of Additional Views

To integrate additional views for multi-view SfM, we employed the approach proposed by Beardsley *et al.* [29]. In this approach, when a sequence of views is available, initially SfM is applied for the first two views. Then, for each new view i, feature detection and matching is applied to establish 2D correspondences with the previous view $i-1$, which are then matched with the already constructed 3D points. The projection matrix of the new view is computed using these final 2D-3D matches as explained below.

Let the 2D-3D matches are represented by $\mathbf{x}_i$ and $\mathbf{X}_i$, respectively. Projection can be written as $\mathbf{x}_i = \mathsf{P}\mathbf{X}_i$ where

$$\mathbf{x}_i = \begin{pmatrix} w_i x_i \\ w_i y_i \\ w_i \end{pmatrix}$$

and

$$\mathbf{X}_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

This leads to three equations per 2D-3D match to solve for P:

$$\begin{aligned} w_i x_i &= \mathbf{p}_1^\mathsf{T} \mathbf{X}_i \\ w_i y_i &= \mathbf{p}_2^\mathsf{T} \mathbf{X}_i \\ w_i &= \mathbf{p}_3^\mathsf{T} \mathbf{X}_i \end{aligned}$$

where $\mathbf{p}_k^\mathsf{T}$ represents the k^{th} row of P. However, we need to eliminate the unknown scale factor w. Thus, we write the two equations below:

$$\begin{aligned} \mathbf{p}_1^\mathsf{T} \mathbf{X}_i - x_i \mathbf{p}_3^\mathsf{T} \mathbf{X}_i = 0 \\ \mathbf{p}_2^\mathsf{T} \mathbf{X}_i - y_i \mathbf{p}_3^\mathsf{T} \mathbf{X}_i = 0 \end{aligned}$$

When represented as a matrix multiplication:

$$\begin{pmatrix} \mathbf{X}_i{}^\mathsf{T} & \mathbf{0}^\mathsf{T} & \mathbf{x}_i \mathbf{X}_i{}^\mathsf{T} \\ \mathbf{0}^\mathsf{T} & \mathbf{X}_i{}^\mathsf{T} & \mathbf{y}_i \mathbf{X}_i{}^\mathsf{T} \end{pmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{pmatrix} = 0.$$

We stack the equations for n points (2 per correspondence) and obtain a $2n$x12 matrix A which holds $\mathsf{A}\mathbf{p} = 0$ where $\mathbf{p}$ is the column vector containing the elements of P. We compute the least-squares solution of $\mathbf{p}$ by singular value decomposition (SVD).

Lastly, 3D coordinates of the newly matched 2D points are computed with triangulation and they are added to the structure.

Multi-view SfM with Projective Factorization. We would like to briefly describe an alternative approach used for multi-view structure computation and explain why it is not suitable to be used for our problem. This approach is widely referred as *projective factorization* and constructs a measurement matrix which

contains the projections of all 3D points in all cameras:

$$\begin{pmatrix} w_1^1\mathbf{x}_1^1 & w_2^1\mathbf{x}_2^1 & \cdots & w_n^1\mathbf{x}_n^1 \\ \vdots & & \ddots & \vdots \\ w_1^m\mathbf{x}_1^m & w_2^m\mathbf{x}_2^m & \cdots & w_n^m\mathbf{x}_n^m \end{pmatrix} = \begin{pmatrix} \mathsf{P}^1 \\ \vdots \\ \mathsf{P}^m \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_n \end{pmatrix} \tag{7.1}$$

where $i = 1, .., m$ denotes images and $p = 1, .., n$ denotes points. $\mathbf{X}_p$ are the homogeneous coordinate vectors of the 3D points, P^i are the unknown 3x4 projection matrices, $\mathbf{x}_p^i$ are the measured homogeneous coordinate vectors of the image points and w_p^i are the unknown scale factors since projection of $\mathbf{X}_p$ is defined up to scale:

$$w_p^i\mathbf{x}_p^i = \mathsf{P}^i\mathbf{X}_p \tag{7.2}$$

The measurement matrix can be approximated to its nearest rank-4 form by using SVD and when decomposed into UDV^T, $(\mathsf{P}^{1\mathsf{T}}, \mathsf{P}^{2\mathsf{T}}, .., \mathsf{P}^{m\mathsf{T}})^\mathsf{T} = \mathsf{UD}$ and $(\mathbf{X}_1, \mathbf{X}_2, .., \mathbf{X}_n) = \mathsf{V}^\mathsf{T}$. Please note that this factorization is not unique and this makes the result of the method always a projective reconstruction even if the cameras are calibrated. Thus, conversion to an Euclidean reconstruction should be performed using additional methods.

The original algorithm proposed by Sturm and Triggs [64] requires all points are available in all frames. Later on, this approach was improved to handle occlusions up to an extent [65]. The proposed improvement uses the available measurements to fill the missing entries of the measurement matrix and image sequence should have an amount of coverage in terms of tracked points. The degradation of reconstruction with decreasing amount of coverage is discussed in [66] where upgrading the projective reconstruction to an Euclidean one is also demonstrated.

In addition to these two reasons, projective reconstruction and the sparseness of the measurement matrix, factorization based multi-view approach is not suitable for our problem of merging hybrid reconstructions due to the fact that projection of omnidirectional images can not be written in a linear fashion together with the perspective images as given in Eq. 7.2.

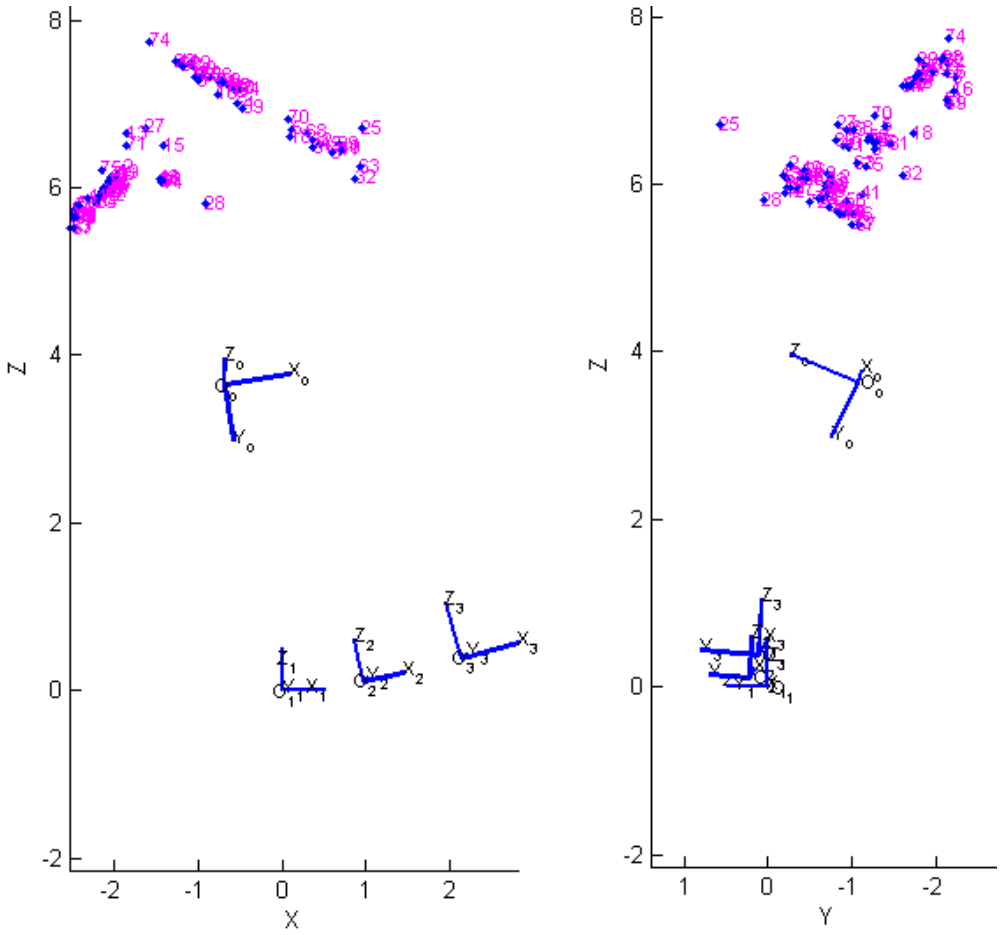
7.2 Sparse Bundle Adjustment

Sparse bundle adjustment method proposed by Lourakis and Argyros [30] has become popular in the community due to its capability of solving enormous minimization problems (with many cameras and 3D points) in a reasonable time. We employed this method for our system of mixed cameras. We modified the projection function with the sphere model projection and intrinsic parameters with sphere model parameters to encompass the mixed camera types. The details of sphere model projection was presented in Section 2.2.

7.2.1 Experiment

We conducted a real image hybrid multi-view experiment and refined the results with sparse bundle adjustment (SBA). Thus we completed the entire pipeline of hybrid SfM which was shown in Fig. 1.1. Views Pers1-Pers2-Pers3 and Omni1 in Fig. 4.7 were used. Initial structure estimation was performed with Pers1-Pers2 pair, then Pers3 and Omni1 views were added. 75 feature points are common in all the images. Estimated coordinates of these points and estimated camera positions are shown in Fig. 7.1.

We performed SBA on this structure (scene point coordinates) and camera parameters. The reprojection errors before and after SBA (in pixels) are given in Table 7.1. We infer from the table that the reprojection errors are considerably decreased after SBA. The error before SBA for the omnidirectional image is higher than the perspective images. This is mainly due to the fact that the number of common points between omnidirectional and perspective images is less when compared to the number of common feature points between two perspective images, which decreases the accuracy of pose estimation. When adding a perspective camera to the structure, the projection matrix (extrinsic parameters) of this new camera is computed by available 2D-3D correspondences (usually more than 500 points), however when an omnidirectional camera is added, there are less (usually less than 100 points) number of 2D-3D correspondences to compute its extrinsic parameters.



(a) (b)

Figure 7.1: Estimated camera positions, orientations and scene points for the hybrid multi-view SfM experiment. (a) top-view (b) side-view.

Table 7.1: The mean values of reprojection errors before and after SBA (in pixels).

	Pers1	Pers2	Pers3	Omni1
Before SBA	0.49	0.48	0.53	0.97
After SBA	0.28	0.23	0.29	0.39

7.3 Merging 3D Structures of Different Hybrid Image Pairs

In this section, we discuss the theoretical and practical aspects of how an omnidirectional camera can combine the 3D structures viewed by two or more perspective cameras which do not have a scene in common (Fig. 7.2). By pairing

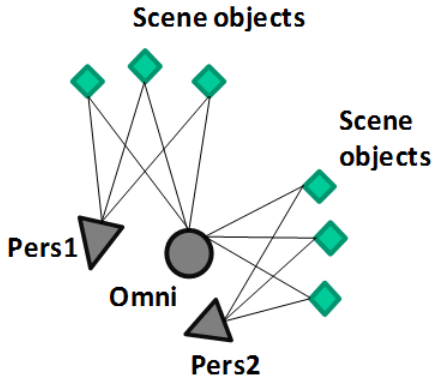


Figure 7.2: Depiction of merging 3D structures estimated with different hybrid image pairs.

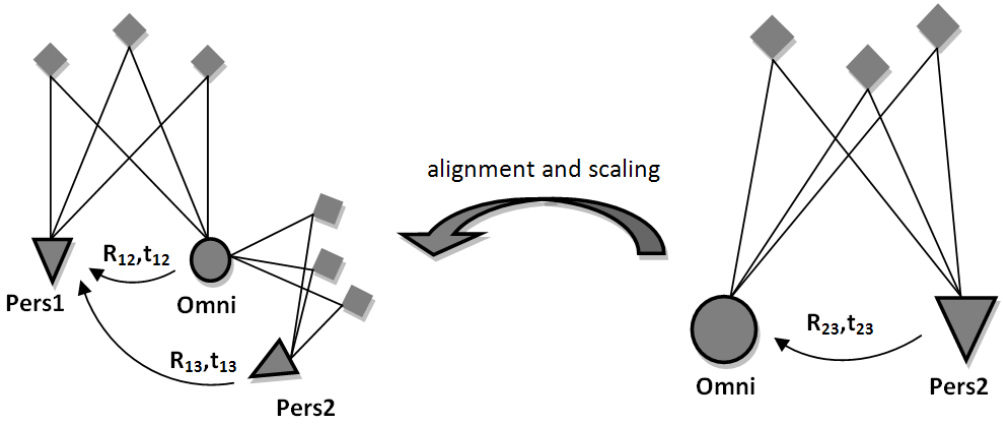


Figure 7.3: Depiction of aligning and scaling the 2^{nd} 3D structure w.r.t. the first one to obtain a combined structure.

the perspective views with the same omnidirectional view, it is possible to combine the 3D structures estimated with different hybrid image pairs. This is one of the main motivations in this thesis since it is not possible to combine different perspective cameras if they do not view the same scene. It is even very difficult if they share a small portion of the view because epipolar geometry estimation becomes inaccurate and if the correspondence points lie on a planar surface it becomes totally impossible.

The multi-view approach described in Section 7.1 does not allow us to obtain a combined structure from these two hybrid pairs directly. When we perform SfM for two hybrid pairs, we obtain two 3D structures which are in different scales because the generated structures are true up to a scale factor. Therefore, first we

have to align the two structures using the rotation and translation of the common view, second we have to adjust the scale by estimating or defining the true ratio of scales. These alignment and scaling processes are depicted in Fig. 7.3 and briefly described in the following.

Let (R_{12},$\mathbf{t}_{12}$) be the rotation and translation between the first perspective image and the omnidirectional image, and (R_{23},$\mathbf{t}_{23}$) be the ones between the omnidirectional image and the other perspective image. When the center of the first perspective camera is taken as the origin of coordinate system, the external calibration matrix of the second camera (omnidirectional camera) becomes $\mathsf{K}_{ext,2} = [\mathsf{R}_{12} \,|\, \mathbf{t}_{12}]$.

The rotation of the third camera (2^{nd} perspective image) in the same coordinate system can easily be defined as the multiplication of rotations:

$$\mathsf{R}_{13} = \mathsf{R}_{23} \cdot \mathsf{R}_{12}$$

The translation vector of the third camera can be formulated as the translation between second and third cameras in addition to the translation between first two cameras when rotated with R_{23} to be represented w.r.t. the third camera:

$$\mathbf{t}_{13} = \mathbf{t}_{23} + \mathsf{R}_{23} \cdot \mathbf{t}_{12}$$

As a result, the external calibration matrix of the second perspective image is as below. The origin is still the center of the first camera.

$$\mathsf{K}_{ext,3} = [\mathsf{R}_{13} \,|\, \mathbf{t}_{13}]$$

However, as we have mentioned earlier, the scales of these two motions are not necessarily the same. To obtain the structure as a whole, we need to estimate the ratio of scales and adjust the scale of the second structure by multiplying the translation vector by this ratio. In the following experiment, benefiting from the small overlap between two perspective images, to estimate the scale ratio we used the 3D points which are available in both reconstructions. We minimized the distance between these points in the two reconstructions.

If there is no overlap, knowledge of real world distances in the scene or the distance between the camera pair can be used to obtain the true scale ratio.

7.3.1 Experiment

In Fig. 7.4, on the top row, we observe two perspective images, which have little intersection in their field of view. At the bottom-left, the omnidirectional view is seen in which 2D points matched with perspective images are indicated with different colors. Finally, at the bottom-left, we see the top-view of the estimated structure. Along with the reconstructed 3D points we also observe the positions, orientations and field of views of the cameras. The circle around the middle camera indicates the 360° field of view. 71 points are common between left perspective view and omnidirectional view, whereas 124 points are common in the right hybrid pair. Only four of these feature points are common in all three images.

As mentioned, the scales of the two structures estimated by the hybrid pairs are not the same and to align these two reconstructed sections we used the four 3D points which are available in both reconstructions. In ideal case, when the structures are perfectly aligned, the distances from these points to the origin should be equal in both structures. When imperfections exist, however, the difference of distances is not zero. We seek for the best scale ratio that minimizes this distance error. With the scale ratio corresponding to the minimum error, it is assumed that the best possible alignment is obtained.

The estimated scale ratio is 0.334 for our experiment. In the 3D structure given in Fig. 7.4 (bottom-left), the structure estimated by the right hand side perspective camera and omnidirectional camera is downscaled with this ratio.

7.4 Conclusions

In this chapter, we explained our work on multi-view hybrid Structure-from-Motion. We employed the approach proposed by Beardsley *et al.* [29] in which the initial structure is estimated with the first two views and additional views are added to the structure via 3D-2D correspondences between structure and new view. This approach seems to be the only option for multi-view hybrid SfM since the other option used for perspective cameras, *projective factorization*, can not be applied to our case due to the fact that the projection of omnidirectional images can not be written linearly together with the perspective images. It may be

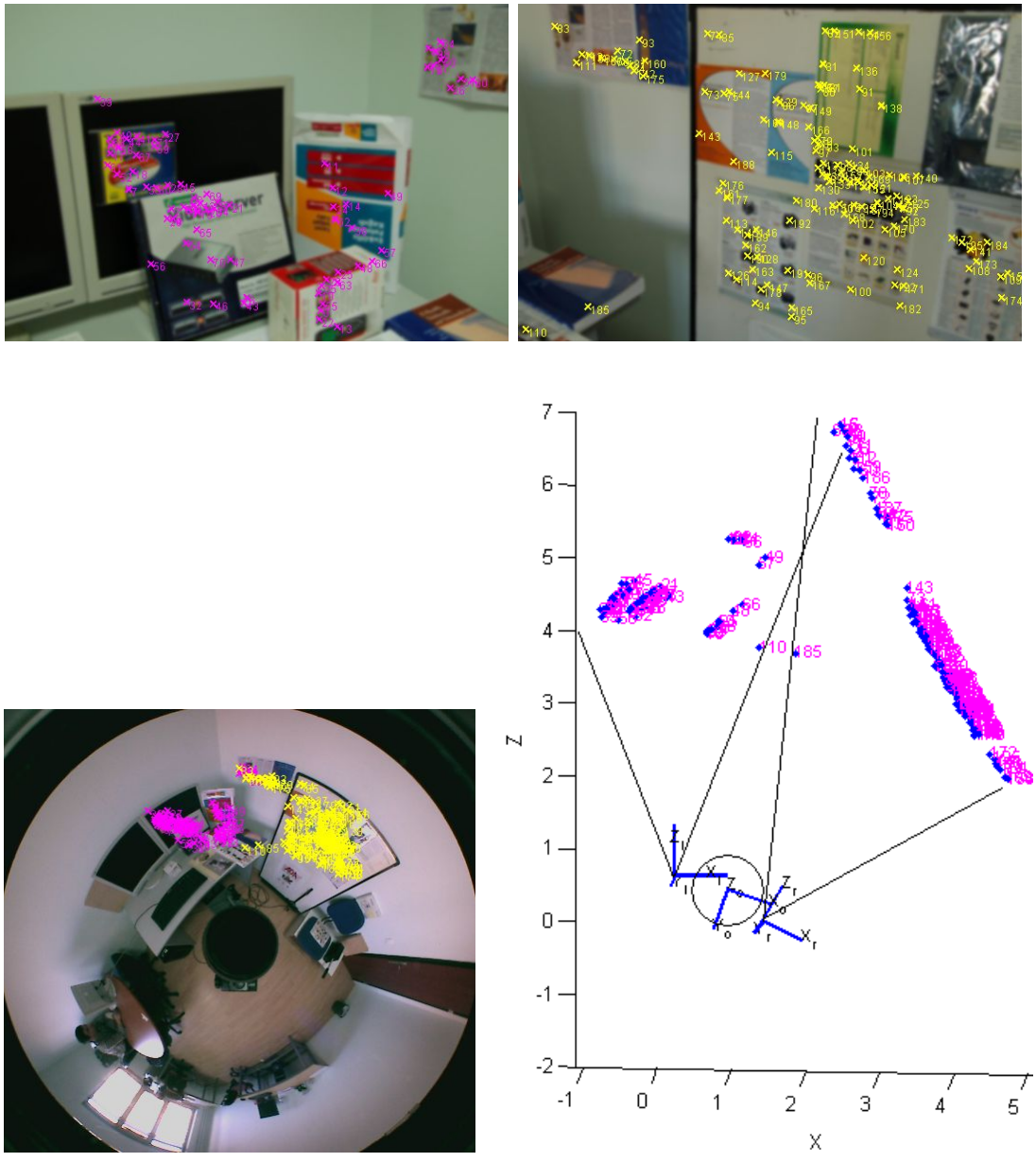


Figure 7.4: Matched points between the perspective images (top row) and omnidirectional image (bottom-left) and the estimated structure (bottom-right) for the experiment of merging 3D structures.

possible to employ lifted coordinates and write 6x10 generic projection matrices for all points as described in Section 3.2.2 to represent the projections linearly and form Eq. 7.1. However it would be impractical since the number of unknowns increase five-fold for camera parameters and three-fold for point coordinates.

We employed sparse bundle adjustment method by modifying the projection function with sphere model projection and intrinsic parameters with sphere model parameters. The improvement gained by bundle adjustment is demonstrated with an experiment of multi-view hybrid SfM. Thus, the complete pipeline of hybrid SfM, given in Fig. 1.1, is realized.

One of the main motivations in this thesis is presented with the theory and experiment given in Section 7.3 where the structures estimated with two perspective cameras which do not view the same part of the scene can be combined by pairing them with an omnidirectional camera.

CHAPTER 8

Conclusions

In this thesis, theoretical and practical issues concerning structure-from-motion (SfM) with mixed camera images are investigated in detail. The work is motivated by the possible advantages of employing omnidirectional and perspective cameras together. An example scenario is the usage of an omnidirectional camera with several perspective cameras (or several views of a perspective camera) which do not necessarily have a view in common. Such a scenario occurs when the cameras are located at the central region and view the perimeter. Omnidirectional view is able to combine partial 3D structures obtained by different omnidirectional-perspective image pairs. Otherwise, 3D reconstruction with only perspective cameras requires a large number of cameras.

Surveillance is another potential application area. Hybrid systems were proposed where slave pan-tilt-zoom cameras are directed according to the information obtained by an omnidirectional camera which performs event detection. Such systems can be enhanced by adding 3D structure and location estimation algorithms without increasing the number of cameras.

We described an end-to-end pipeline for hybrid multi-view structure-from-motion and proposed new approaches or modified existing methods for the steps of this pipeline which can be summarized as camera calibration, feature point matching, epipolar geometry and pose estimation, triangulation and bundle adjustment.

We employed the sphere camera model [25] to represent mixed types of cameras. We did not need to change the camera model for different camera pairs or between SfM steps.

We presented a calibration technique based on the sphere camera model which is able to represent every single-viewpoint catadioptric system. We tested this method both with simulations and real images. When compared to previous techniques of sphere model calibration, our method has the advantages of not

requiring input for parameter initialization and being able to calibrate perspective cameras as well. On the other hand, it needs a 3D calibration object.

It had been stated that directly applying SIFT [21] is not sufficient to obtain good results for hybrid image pairs. In our study, we showed that the performance of SIFT considerably increases with the proposed algorithm of low-pass filtering and downsampling the perspective image in the hybrid pair. Another approach we evaluate is generating virtual camera plane (VCP) image from the omnidirectional image first and then applying SIFT. We conducted experiments with two different catadioptric cameras and a fish-eye camera. We observed that VCP approach performed best for catadioptric images and is more robust to increasing baseline. For fish-eye images, preprocessing without VCP approach performed as well as the VCP approach due to the fact that a fish-eye camera looking towards the scene acts as a perspective camera with a large lens distortion and serves as a good candidate for matching with perspective cameras.

The proposed algorithm for feature point matching brings the advantage of automatic point matching between omnidirectional and perspective images. Moreover, we are able to say that the proposed technique of extracting parameters of low-pass filtering and downsampling is versatile up to a large extent since it worked in our experiments with three different omnidirectional cameras.

The proposed point matching algorithm should not be considered as a false match elimination technique since it increases the performance at the matching stage. Thus, it is not compensated by elimination techniques such as the one proposed by Lowe [21] which eliminates matches that do not conform to an affine transformation. On the other hand, employing such an elimination together with the proposed technique has the potential of improving the result especially for VCP-perspective matching. We did not integrate such a step since next step in our pipeline is the estimation of hybrid epipolar geometry and elimination of matches that do not conform to the epipolar geometry which is a more accurate geometrical constraint.

We robustly estimated the hybrid epipolar geometry using RANSAC. We discussed the important aspects of fundamental matrix estimation such as coordinate normalization, rank-2 imposition and distance threshold. We defined normalization matrices for lifted coordinates and performed detailed analysis on the effectiveness of coordinate normalization and choosing the best values of scale

normalization factor for hybrid case. We concluded that, as suggested for perspective cameras, carrying the origin to the centroid of points and scaling coordinates so that the average distance of a point to the origin is equal to $\sqrt{2}$ is proper for normalizing point coordinates in omnidirectional camera images.

We also evaluated the alternatives for pose estimation and decided on estimating the essential matrix with the calibrated 3D rays of point correspondences rather than extracting the essential matrix from fundamental matrix. This decision is based on both the performance difference observed in the experimental analysis and the limitations of obtaining E from F for hybrid systems with catadioptric cameras with hyperbolic mirrors.

Aiming to achieve more accurate triangulation for hybrid camera images and based on the fact that the significant resolution difference between omnidirectional and perspective images should be considered, we proposed a weighting strategy for *iterative linear-Eigen* triangulation method. We showed its effectiveness with simulated images and real image hybrid SfM experiment.

For multi-view hybrid SfM, we implemented the approach proposed by Beardsley *et al.* [29] in which the initial structure is estimated with the first two views and additional views are added to the structure via 3D-2D correspondences between the structure and the new view. This approach seems to be the only option for multi-view hybrid SfM since the other option used for perspective cameras, *projective factorization*, can not be applied to our case due to the fact that the projection of omnidirectional images can not be written linearly together with the perspective images.

We employed the sparse bundle adjustment method [30] by modifying the projection function with the sphere model projection and intrinsic parameters with the sphere model parameters. The improvement gained by bundle adjustment was demonstrated with an experiment of hybrid multi-view SfM.

In conclusion, with the suggested techniques and proposed improvements in this thesis, it is possible to perform hybrid multi-view structure-from-motion in an effective and automatic way.

8.1 Limitations and Future Work

The presented calibration method was tested with simulated and real images of catadioptric cameras. Although we left it as a future work, it may be possible to generalize the algorithm to cover the fish-eye cameras as well. Ying and Hu [53] showed that the sphere model can approximate fisheye projections and Hansen *et al.* [57] employed this approach. The hybrid epipolar geometry and pose estimation steps can also be performed for fish-eye cameras as the theory was given in [18]. Sparse bundle adjustment can be directly applied for fish-eye cameras by modifying its projection function accordingly.

The optimal scale ratio detection and preprocessing the high-resolution image approach proposed in this work can be used for other applications employing SIFT where an approximate scale ratio exists between the objects in the given images. Otherwise, if scale ratios considerably vary for objects in the scene, this approach is not suitable.

Another approach to detect and match scale-invariant features for omnidirectional cameras was proposed by Hansen *et al.* [57] where features are detected in the spherical domain. When compared to the conventional SIFT applied on 2D image plane, it brought improvement for some camera types and in certain motion scenarios. This is a new approach and may lead to better results in the near future. Therefore, integrating or comparing the proposed preprocessing method with this "SIFT on Sphere" approach may be beneficial.

REFERENCES

[1] **Chahl, J. and Srinivasan, M.**, 2000. A Complete Panoramic Vision System, Incorporating Imaging, Ranging, and Three Dimensional Navigation, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 104–111.

[2] **Gaspar, J.**, **Winters, N. and Santos-Victor, J.**, 2000. Vision-Based Navigation and Environmental Representations with an Omnidirectional Camera, *IEEE Transactions on Robotics and Automation*, **16(6)**, 890–898.

[3] **Ng, K.**, **Ishiguro, H.**, **Trivedi, M. and Sogo, T.**, 2004. An Integrated Surveillance System: Human Tracking and View Synthesis using Multiple Omni-directional Vision Sensors, *Image and Vision Computing*, **22**, 551–561.

[4] **Cui, Y.**, **Samarasekera, S.**, **Huang, Q. and Greiffenhagen, M.**, 1998. Indoor Monitoring via the Collaboration between a Peripheral Sensor and a Foveal Sensor, IEEE Workshop on Visual Surveillance, pp. 2–9.

[5] **Yao, Y.**, **Abidi, B. and Abidi, M.**, 2006. Fusion of Omnidirectional and PTZ Cameras for Accurate Cooperative Tracking, IEEE Int. Conference on Video and Signal Based Surveillance (AVSS).

[6] **Scotti, G.**, **Marcenaro, L.**, **Coelho, C.**, **Selvaggi, F. and Regazzoni, C.**, 2005. Dual Camera Intelligent Sensor for High Definition 360 Degrees Surveillance, *IEE Proc.-Vision Image Signal Processing*, **152(2)**, 250–257.

[7] **Fleck, S.**, **Busch, F.**, **Biber, P.**, **Andreasson, H. and Strasser, W.**, 2005. Omnidirectional 3D Modeling on a Mobile Robot using Graph Cuts, Proc. of IEEE International Conference on Robotics and Automation (ICRA), pp. 1748–1754.

[8] **Lhuillier, M.**, 2007. Toward Flexible 3D Modeling using a Catadioptric Camera, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).

[9] **Chang, P. and Hebert, M.**, 2000. Omni-directional Structure from Motion, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 127–133.

[10] **Hartley, R. and Zisserman, A.**, 2004. Multiple View Geometry in Computer Vision, Cambridge Univ. Press, 2nd edition.

[11] **Lhuillier, M.**, 2005. Automatic Structure and Motion using a Catadioptric Camera, Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS).

[12] **Bunschoten, R. and Krose, B.**, 2003. Robust Scene Reconstruction from an Omnidirectional Vision System, *IEEE Transactions on Robotics and Automation*, **19(2)**, 351–357.

[13] **Micusik, B. and Pajdla, T.**, 2006. Structure from Motion with Wide Circular Field of View Cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28(7)**, 1135–1149.

[14] **Chen, X.**, **Yang, J. and Waibel, A.**, 2003. Calibration of a Hybrid Camera Network, Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 150–155.

[15] **Adorni, G.**, **Cagnoni, S.**, **Mordonini, M. and Sgorbissa, A.**, 2003. Omnidirectional Stereo Systems for Robot Navigation, Proc. of Workshop on Omnidirectional Vision and Camera Networks (OMNIVIS).

[16] **Chen, D. and Yang, J.**, 2005. Image Registration with Uncalibrated Cameras in Hybrid Vision Systems, IEEE Workshop on Application of Computer Vision, pp. 427–432.

[17] **Sturm, P.**, 2002. Mixing Catadioptric and Perspective Cameras, Proc. of Workshop on Omnidirectional Vision (OMNIVIS), pp. 37–44.

[18] **Barreto, J. and Daniilidis, K.**, 2006. Epipolar Geometry of Central Projection Systems using Veronese Maps, Int. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1258–1265.

[19] **Sturm, P. and Barreto, J.**, 2008. General Imaging Geometry for Central Catadioptric Cameras, European Conference on Computer Vision (ECCV), pp. 609–622.

[20] **Puig, L.**, **Guerrero, J. and Sturm, P.**, 2008. Matching of Omnidirectional and Perspective Images using the Hydrid Fundamental Matrix, Proc. of Workshop on Omnidirectional Vision (OMNIVIS).

[21] **Lowe, D.**, 2004. Distinctive Image Features from Scale Invariant Keypoints, *International Journal of Computer Vision*, **60**, 91–110.

[22] **Fischler, M. and Bolles, R.**, 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, **24(6)**, 381–395.

[23] **Ramalingam, S.**, **Lodha, S. and Sturm, P.**, 2004. A Generic Structure-from-motion Algorithm for Cross-camera Scenarios, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 175–186.

[24] **Sturm, P. and Ramalingam, S.**, 2004. A Generic Concept for Camera Calibration, European Conference on Computer Vision (ECCV), pp. 1–13.

[25] **Geyer, C. and Daniilidis, K.**, 2000. A Unifying Theory for Central Panoramic Systems, European Conference on Computer Vision (ECCV), pp. 445–461.

[26] **Mei, C. and Rives, P.**, 2007. Single Viewpoint Omnidirectional Camera Calibration from Planar Grids, IEEE International Conference on Robotics and Automation (ICRA), pp. 3945–3950.

[27] **Matas, J.**, **Chum, O.**, **Urban, M. and Pajdla, T.**, 2002. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions, Proc. of British Machine Vision Conference (BMVC), pp. 384–393.

[28] **Hartley, R. and Sturm, P.**, 1997. Triangulation, *Computer Vision and Image Understanding*, **68(2)**, 146–157.

[29] **Beardsley, P.**, **Zisserman, A. and Murray, D.**, 1997. Sequential Updating of Projective and Affine Structure from Motion, *IJCV*, **23(3)**, 235–259.

[30] **Lourakis, M. and Argyros, A.**, 2004, The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package based on the LM Algorithm, FORTH-ICS Technical Report, TR-340.

[31] **Rees, D.**, 1970. Panoramic Television Viewing System, US Patent No. 3505465.

[32] **Nayar, S. and Peri, V.**, 1999, Folded Catadioptric Cameras, Technical Report, Dept. of Computer Science, Columbia University.

[33] **Conroy, T. and Moore, J.**, 1999. Resolution Invariant Surfaces for Panoramic Vision Systems, Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 392–397.

[34] **Hicks, R. and Bajcsy, R.**, 2000. Catadioptric Sensors that Approximate Wide-Angle Perspective Projections, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 97–103.

[35] **Gaspar, J.**, **Decc, C.**, **Okamoto, J. and Santos-Victor, J.**, 2002. Constant Resolution Omnidirectional Cameras, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 27–36.

[36] **Swaminathan, R.**, **Grossberg, M. and Nayar, S.**, 2004. Designing Mirrors for Catadioptric Systems that Minimize Image Errors, Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS), pp. 91–102.

[37] **Baker, S. and Nayar, S.**, 1999. A Theory of Single-viewpoint Catadioptric Image Formation, *IJCV*, **35(2)**, 175–196.

[38] **Swaminathan, R.**, **Grossberg, M. and Nayar, S.**, 2001, Non Single-viewpoint Catadioptric Cameras, Technical Report, Department of Computer Science, Columbia University.

[39] **Derrien, S. and Konolige, K.**, 2000. Approximating a Single-viewpoint in Panoramic Imaging Devices, Proc. of Workshop on Omnidirectional Vision (OMNIVIS), pp. 85–90.

[40] **Ho, T.**, **Davis, C. and Milner, S.**, 2005. Using Geometric Constraints for Fisheye Camera Calibration, Proc. of Workshop on Omnidirectional Vision (OMNIVIS).

[41] **Barreto, J. and Araujo, H.**, 2005. Geometric Properties of Central Catadioptric Line Images and Their Application in Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27(8)**, 1327–1333.

[42] **Geyer, C. and Daniilidis, K.**, 2002. Paracatadioptric Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24(5)**, 687–695.

[43] **Kang, S.**, 2000. Catadioptric Self-calibration, Int. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1201–1207.

[44] **Orghidan, R.**, **Salvi, J. and Mouaddib, M.**, 2003. Calibration of a Structured Light-based Stereo Catadioptric Sensor, Proc. of Workshop on Omnidirectional Vision and Camera Networks (OMNIVIS).

[45] **Cauchois, C.**, **Brassart, E. and Drocourt, C.**, 1999. Calibration of the Omnidirectional Vision Sensor: SYCLOP, IEEE International Conference on Robotics and Automation (ICRA), pp. 1287–1292.

[46] **Kannala, J. and Brandt, S.**, 2004. A Generic Camera Calibration Method for Fish-eye Lenses, Proc. of International Conference on Pattern Recognition (ICPR), pp. 10–13.

[47] **Scaramuzza, D.**, **Martinelli, A. and Siegwart, R.**, 2006. A Toolbox for Easily Calibrating Omnidirectional Cameras, Int. Conference on Intelligent Robots and Systems (IROS), pp. 5695–5701.

[48] **Abdel-Aziz, Y. and Karara, H.**, 1971. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-range Photogrammetry, Symposium on Close-Range Photogrammetry, pp. 1–18.

[49] **Horn, R. and Johnson, C.**, 1991. Topics in Matrix Analysis, Cambridge Univ. Press, 2nd edition.

[50] **Horn, R. and Johnson, C.**, 1985. Matrix Analysis, Cambridge Univ. Press.

[51] **Bastanlar, Y.**, **Puig, L.**, **Sturm, P.**, **Guerrero, J. and Barreto, J.**, 2008. DLT-like Calibration of Central Catadioptric Cameras, Proc. of Workshop on Omnidirectional Vision (OMNIVIS).

[52] **Heikilla, J. and Silven, O.**, 1997. A Four-step Camera Calibration Procedure with Implicit Image Correction, Proc. of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1106–1112.

[53] **Ying, X. and Hu, Z.**, 2004. Can We Consider Central Catadioptric Cameras and Fisheye Cameras Within a Unified Imaging Model?, European Conference on Computer Vision (ECCV), pp. 442–455.

[54] **Yi, Z.**, **Zhiguo, C. and Yang, X.**, 2008. Multi-spectral Remote Image Registration based on SIFT, *Electronic Letters*, **44(2)**, 107–108.

[55] **Alhwarin, F.**, **Wang, C.**, **Ristic-Durrant, D. and Gräser, A.**, 2008. Improved SIFT-Features Matching for Object Recongnition, Visions of Computer Science - BCS International Academic Conference.

[56] **Peri, V. and Nayar, S.**, 1997. Generation of Perspective and Panoramic Video from Omnidirectional Video, Proc. of DARPA Image Understanding Workshop, pp. 243–246.

[57] **Hansen, P.**, **Corke, P.**, **Wageeh, B. and Daniilidis, K.**, 2007. Scale-Invariant Features on the Sphere, IEEE International Conference on Computer Vision (ICCV).

[58] **Svoboda, T. and Pajdla, T.**, 2002. Epipolar Geometry for Central Catadioptric Cameras, *International Journal of Computer Vision*, **49**, 23–37.

[59] **Geyer, C. and Daniilidis, K.**, 2001. Structure and Motion from Uncalibrated Catadioptric Views, Int. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 279–286.

[60] **Claus, D. and Fitzgibbon, A.**, 2005. A Rational Function Lens Distortion Model for Generic Cameras, Int. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 213–219.

[61] **Fitzgibbon, A.**, 2001. Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pp. 125–132.

[62] **Hartley, R.**, 1997. In Defense of the Eight-Point Algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19(6)**, 580–593.

[63] **Bartoli, A. and Sturm, P.**, 2004. Non-linear Estimation of the Fundamental Matrix with Minimal Paremeters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26(3)**, 426–432.

[64] **Sturm, P. and Triggs, B.**, 1996. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion, Proc. of European Conference on Computer Vision (ECCV), pp. 709–720.

[65] **Martinec, D. and Pajdla, T.**, 2002. Structure from Many Perspective Images with Occlusions, Prof. of European Conference on Computer Vision (ECCV), volume 2, pp. 355–369.

[66] **Martinec, D. and Pajdla, T.**, 2005. 3D Reconstruction by Fitting Low-rank Matrices with Missing Data, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pp. 198–205.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Başţanlar, Yalın
Nationality: Turkish
Date of Birth: 21.09.1978
Place of Birth: Ankara, Turkey
Marital Status: Single
E-mail: yalinb@yahoo.com

EDUCATION

Degree	Institution	Year
Ph.D. in Information Systems	Middle East Technical University	2009
M.Sc. in Information Systems	Middle East Technical University	2005
B.Sc. in Civil Engineering	Middle East Technical University	2001

WORK EXPERIENCE

Year	Place	Enrollment
Jan.'06 - Sept.'09	Middle East Technical University	Research Assistant
Mar.'08 - Aug.'08	INRIA Rhône-Alpes (France)	Visiting Researcher
Sept.'04 - Dec.'05	Middle East Technical University	Project Researcher

RESEARCH INTERESTS

Catadioptric omnidirectional cameras and image formation processes;
Calibration and epipolar geometry of omnidirectional cameras;
Structure-from-motion and 3D reconstruction;
Corner detection;
Usage of 360° in Web-based virtual tours;
Human factors and usability in virtual reality applications.

FOREIGN LANGUAGES

English: Very Good
French: Intermediate

PUBLICATIONS

Journals:

• Bastanlar, Y., Yardimci, Y., Temizel, A., Sturm, P., "Automatic and Robust Structure-from-Motion for Mixed Camera Systems", ***submitted to*** IJCV Special Issue on Omnidirectional Vision, Camera Networks and Non-Classical Cameras.
• Puig, L., Bastanlar, Y., Sturm, P., Guerrero Campo, J., Barreto, J., "Calibration of Central Catadioptric Cameras Using a DLT-Like Approach", ***submitted to*** IJCV Special Issue on Omnidirectional Vision, Camera Networks and Non-Classical Cameras.
• Bastanlar, Y., Yardimci, Y. (2008), "Corner Validation based on Extracted Corner Properties", Computer Vision and Image Understanding (CVIU), vol.112, p.243-261.

International Conferences:

• Bastanlar, Y., Puig, L., Sturm, P., Guerrero Campo, J., Barreto, J. (2008), "DLT-like Calibration of Central Catadioptric Cameras", Proc. of Workshop on Omnidirectional Vision (OmniVis 2008), 12-18 October, Marseille, France.
• Bastanlar, Y., Grammalidis, N., Zabulis, X., Yilmaz, E., Yardimci, Y., Triantafyllidis, G. (2008), "3D Reconstruction for a Cultural Heritage Virtual Tour System", XXI Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS 2008), 3-11 July, Beijing, China.
• Zabulis, X., Grammalidis, N., Bastanlar, Y., Yilmaz, E., Yardimci, Y. (2008), "3D Scene Reconstruction Based on Robust Camera Motion Estimation and Space Sweeping for a Cultural Heritage Virtual Tour System", 3DTV-Conference, 28-30 May, Istanbul, Turkey.
• Bastanlar, Y. (2007), "User Behaviour in Web-Based Interactive Virtual Tours", 29th International Conference on Information Technology Interfaces (ITI 2007), 25-28 June, Dubrovnik, Croatia.
• Bastanlar, Y., Canturk, D., Uke, H. (2007), "Effects of Color-multiplex Stereoscopic View on Memory and Navigation", 3DTV-Conference: The True Vision: Capture, Transmission and Display of 3D Video, 7-9 May, Cos, Greece.
• Bastanlar, Y., Cetin, A.E., Yardimci, Y. (2005), "Enhancement of Panoramas Generated from Omnidirectional Images", International Workshop on Sampling Theory and Applications (SampTA 2005), 10-15 July, Samsun, Turkey.

National Conferences:

• Bastanlar, Y., Temizel, A., Yardimci, Y. (2009), "Automatic Point Matching and Robust Fundamental Matrix Estimation for Hybrid Camera Scenarios", (in Turkish), IEEE National Conference on Signal Processing and Applications (SIU 2009), 9-11 April, Antalya, Turkey.

• Serce, H., Bastanlar, Y., Temizel, A., Yardimci, Y. (2008), "On Detection of Edges and Interest Points for Omnidirectional Images in Spherial Domain" (in Turkish), IEEE National Conference on Signal Processing and Applications (SIU 2008), 20-22 April, Didim, Turkey.
• Bastanlar, Y., Yardimci, Y. (2007), "Selecting Image Corner Points Using Their Corner Properties" (in Turkish), IEEE National Conference on Signal Processing and Applications (SIU 2007), 11-13 June, Eskisehir, Turkey.
• Bastanlar, Y., Altingovde, I.S., Aksay, A., Alav, O., Cavus, O., Yardimci, Y., Ulusoy, O., Gudukbay, U., Cetin, A.E., Bozdagi Akar, G., Aksoy, S. (2006), "E-Museum: Web-based Tour and Information System for Museums" (in Turkish), IEEE National Conference on Signal Processing and Applications (SIU 2006), 17-19 April, Antalya, Turkey.
• Bastanlar, Y., Yardimci, Y. (2005), "Parameter Extraction for Hyperboloidal Catadioptric Omnidirectional Cameras" (in Turkish), IEEE National Conference on Signal Processing and Applications (SIU 2005), 16-18 May, Kayseri, Turkey.