Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Peter Sturm
Peter.Sturm@inrialpes.fr

Srikumar Ramalingam
ramalingam@merl.com

Jean-Philippe Tardif
tardifj@gmail.com

Simone Gasparini
Simone.Gasparini@inrialpes.fr

João Barreto
jpbar@deec.uc.pt

now
the essence of knowledge
Boston – Delft

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini and J. Barreto, Camera Models and Fundamental Concepts Used in Geometric Computer Vision, Foundations and Trends® in Computer Graphics and Vision, vol 6, nos 1–2, pp 1–183, 2010

ISBN: 978-1-60198-410-4
© 2011 P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini and J. Barreto

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Foundations and Trends® in Computer Graphics and Vision
Volume 6 Issues 1–2, 2010

Editorial Board

Editor-in-Chief:
Brian Curless
University of Washington
Luc Van Gool
KU Leuven/ETH Zurich
Richard Szeliski
Microsoft Research

Editors

Marc Alexa (TU Berlin)
Ronen Basri (Weizmann Inst)
Peter Belhumeur (Columbia)
Andrew Blake (Microsoft Research)
Chris Bregler (NYU)
Joachim Buhmann (ETH Zurich)
Michael Cohen (Microsoft Research)
Paul Debevec (USC, ICT)
Julie Dorsey (Yale)
Fredo Durand (MIT)
Olivier Faugeras (INRIA)
Mike Gleicher (U. of Wisconsin)
William Freeman (MIT)
Richard Hartley (ANU)
Aaron Hertzmann (U. of Toronto)
Hugues Hoppe (Microsoft Research)
David Lowe (U. British Columbia)
Jitendra Malik (UC. Berkeley)
Steve Marschner (Cornell U.)
Shree Nayar (Columbia)
James O'Brien (UC. Berkeley)
Tomas Pajdla (Czech Tech U)
Pietro Perona (Caltech)
Marc Pollefeys (U. North Carolina)
Jean Ponce (UIUC)
Long Quan (HKUST)
Cordelia Schmid (INRIA)
Steve Seitz (U. Washington)
Amnon Shashua (Hebrew Univ)
Peter Shirley (U. of Utah)
Stefano Soatto (UCLA)
Joachim Weickert (U. Saarland)
Song Chun Zhu (UCLA)
Andrew Zisserman (Oxford Univ)

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Editorial Scope

Foundations and Trends® in Computer Graphics and Vision will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

Information for Librarians
Foundations and Trends® in Computer Graphics and Vision, 2010, Volume 6, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Foundations and Trends® in
Computer Graphics and Vision
Vol. 6, Nos. 1–2 (2010) 1–183
© 2011 P. Sturm, S. Ramalingam, J.-P. Tardif,
S. Gasparini and J. Barreto
DOI: 10.1561/0600000023

now
the essence of knowledge

Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Peter Sturm[1], Srikumar Ramalingam[2], Jean-Philippe Tardif[3], Simone Gasparini[4], and João Barreto[5]

[1] *INRIA Grenoble — Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, Montbonnot, France, Peter.Sturm@inrialpes.fr*
[2] *MERL, Cambridge, MA, USA, ramalingam@merl.com*
[3] *NREC — Carnegie Mellon University, Pittsburgh, PA, USA, tardifj@gmail.com*
[4] *INRIA Grenoble — Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, Montbonnot, France, Simone.Gasparini@inrialpes.fr*
[5] *Coimbra University, Coimbra, Portugal, jpbar@deec.uc.pt*

Abstract

This survey is mainly motivated by the increased availability and use of panoramic image acquisition devices, in computer vision and various of its applications. Different technologies and different computational models thereof exist and algorithms and theoretical studies for geometric computer vision ("structure-from-motion") are often re-developed without highlighting common underlying principles. One of the goals of this survey is to give an overview of image acquisition methods used in computer vision and especially, of the vast number of camera models that have been proposed and investigated over the years,

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

where we try to point out similarities between different models. Results on epipolar and multi-view geometry for different camera models are reviewed as well as various calibration and self-calibration approaches, with an emphasis on non-perspective cameras. We finally describe what we consider are fundamental building blocks for geometric computer vision or structure-from-motion: epipolar geometry, pose and motion estimation, 3D scene modeling, and bundle adjustment. The main goal here is to highlight the main principles of these, which are independent of specific camera models.

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Contents

1 Introduction and Background Material **1**

1.1 Introduction 1
1.2 Background Material 4

2 Technologies **7**

2.1 Moving Cameras or Optical Elements 7
2.2 Fisheyes 13
2.3 Catadioptric Systems 14
2.4 Stereo and Multi-camera Systems 31
2.5 Others 33

3 Camera Models **35**

3.1 Global Camera Models 40
3.2 Local Camera Models 66
3.3 Discrete Camera Models 72
3.4 Models for the Distribution of Camera Rays 75
3.5 Overview of Some Models 83
3.6 So Many Models ... 85

4 Epipolar and Multi-view Geometry **91**

4.1 The Calibrated Case 92
4.2 The Uncalibrated Case 93

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

4.3 Images of Lines and the Link between Plumb-line Calibration and Self-calibration of Non-perspective Cameras 101

5 Calibration Approaches 105

5.1 Calibration Using Calibration Grids 105
5.2 Using Images of Individual Geometric Primitives 112
5.3 Self-calibration 116
5.4 Special Approaches Dedicated to Catadioptric Systems 126

6 Structure-from-Motion 129

6.1 Pose Estimation 130
6.2 Motion Estimation 132
6.3 Triangulation 135
6.4 Bundle Adjustment 136
6.5 Three-Dimensional Scene Modeling 138
6.6 Distortion Correction and Rectification 139

7 Concluding Remarks 145

Acknowledgements 147

References 149

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

1

Introduction and Background Material

1.1 Introduction

Many different image acquisition technologies have been investigated in computer vision and other areas, many of them aiming at providing a wide field of view. The main technologies consist of catadioptric and fisheye cameras as well as acquisition systems with moving parts, e.g., moving cameras or optical elements. In this monograph, we try to give an overview of the vast literature on these technologies and on computational models for cameras. Whenever possible, we try to point out links between different models. Simply put, a computational model for a camera, at least for its geometric part, tells how to project 3D entities (points, lines, etc.) onto the image, and vice versa, how to back-project from the image to 3D. Camera models may be classified according to different criteria, for example the assumption or not of a single viewpoint or their algebraic nature and complexity. Also, recently several approaches for calibrating and using "non-parametric" camera models have been proposed by various researchers, as opposed to classical, parametric models.

In this survey, we propose a different nomenclature as our main criterion for grouping camera models. The main reason is that even

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

so-called non-parametrics models do have parameters, e.g., the coordinates of camera rays. We thus prefer to speak of three categories: (i) A **global camera model** is defined by a set of parameters such that changing the value of any parameter affects the projection function all across the field of view. This is the case for example with the classical pinhole model and with most models proposed for fish-eye or catadioptric cameras. (ii) A **local camera model** is defined by a set of parameters, each of which influences the projection function only over a subset of the field of view. A hypothetical example, just for illustration, would be a model that is "piecewise-pinhole", defined over a tessellation of the image area or the field of view. Other examples are described in this monograph. (iii) A **discrete camera model** has sets of parameters for individual image points or pixels. To work with such a model, one usually needs some interpolation scheme since such parameter sets can only be considered for finitely many image points. Strictly speaking, discrete models plus an interpolation scheme are thus not different from the above local camera models, since model parameters effectively influence the projection function over regions as opposed to individual points. We nevertheless preserve the distinction between discrete and local models, since in the case of discrete models, the considered regions are extremely small and since the underlying philosophies are somewhat different for the two classes of models.

These three types of models are illustrated in Figure 1.1, where the camera is shown as a black box. As discussed in more detail later in the monograph, we mainly use back-projection to model cameras, i.e., the mapping from image points to camera rays. Figure 1.1 illustrates back-projection for global, discrete and local camera models.

After describing camera models, we review central concepts of geometric computer vision, including camera calibration, epipolar and multi-view geometry, and structure-from-motion tasks, such as pose and motion estimation. These concepts are exhaustively described for perspective cameras in recent textbooks [137, 213, 328, 336, 513]; our emphasis will thus be on non-perspective cameras. We try to describe the various different approaches that have been developed for camera calibration, including calibration using grids or from images of higher level primitives, like lines and spheres, and self-calibration. Throughout

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

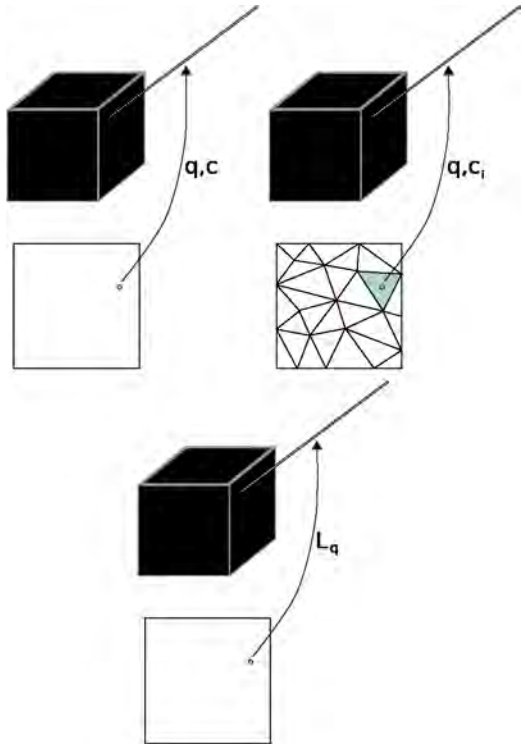


Fig. 1.1 Types of camera models. *Left*: For global models, the camera ray associated with an image point $\mathbf{q}$ is determined by the position of $\mathbf{q}$ and a set of global camera parameters contained in a vector $\mathbf{c}$. *Middle*: For discrete models, different image regions are endowed with different parameter sets. *Right*: For discrete models, the camera rays are directly given for sampled image points, e.g., by a look-up table containing Plücker coordinates, here the Plücker coordinates $\mathbf{L_q}$ of the ray associated with image point $\mathbf{q}$.

this monograph, we aim at describing concepts and ideas rather than all details, which may be found in the original references.

The monograph is structured as follows. In the following section, we give some background material that aims at making the mathematical treatment presented in this monograph, self-contained. In Section 2, we review image acquisition technologies, with an emphasis on omnidirectional systems. Section 3 gives a survey of computational camera models in the computer vision and photogrammetry literature, again emphasizing omnidirectional cameras. Results on epipolar and multi-view geometry for non-perspective cameras are summarized

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

in Section 4. Calibration approaches are explained in Section 5, followed by an overview of some fundamental modules for structure-from-motion in Section 6. The monograph ends with conclusions, in Section 7.

1.2 Background Material

Given the large scope of this monograph, we rather propose summaries of concepts and results than detailed descriptions, which would require an entire book. This allows us to keep the mathematical level at a minimum. In the following, we explain the few notations we use in this monograph. We assume that the reader is familiar with basic notions of projective geometry, such as homogeneous coordinates, homographies, etc. and of multi-view geometry for perspective cameras, such as the fundamental and essential matrices and projection matrices. Good overviews of these concepts are given in [137, 213, 328, 336, 513].

Fonts. We denote scalars by italics, e.g., s, vectors by bold characters, e.g., $\mathbf{t}$ and matrices in sans serif, e.g., A. Unless otherwise stated, we use homogeneous coordinates for points and other geometric entities. Equality between vectors and matrices, up to a scalar factor, is denoted by $\sim$. The cross-product of two 3-vectors $\mathbf{a}$ and $\mathbf{b}$ is written as $\mathbf{a} \times \mathbf{b}$.

Plücker coordinates for 3D lines. Three-dimensional lines are represented either by two distinct 3D points, or by 6-vectors of so-called Plücker coordinates. We use the following convention. Let $\mathbf{A}$ and $\mathbf{B}$ be two 3D points, in homogeneous coordinates. The Plücker coordinates of the line spanned by them, are then given as:

$$\begin{pmatrix} B_4\bar{\mathbf{A}} - A_4\bar{\mathbf{B}} \\ \bar{\mathbf{A}} \times \bar{\mathbf{B}} \end{pmatrix}, \tag{1.1}$$

where $\bar{\mathbf{A}}$ is the 3-vector consisting of the first three coordinates of $\mathbf{A}$ and likewise for $\bar{\mathbf{B}}$.

The action of displacements on Plücker coordinates is as follows. Let $\mathbf{t}$ and R be a translation vector and rotation matrix that map points according to:

$$\mathbf{Q} \mapsto \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{Q}.$$

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Plücker coordinates are then mapped according to:

$$\mathbf{L} \mapsto \begin{pmatrix} \mathsf{R} & \mathsf{0} \\ -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \end{pmatrix} \mathbf{L}, \tag{1.2}$$

where $\mathsf{0}$ is the 3×3 matrix composed of zeroes.

Two lines cut one another exactly if

$$\mathbf{L}_2^\mathsf{T} \begin{pmatrix} \mathsf{0}_{3\times3} & \mathsf{I}_{3\times3} \\ \mathsf{I}_{3\times3} & \mathsf{0}_{3\times3} \end{pmatrix} \mathbf{L}_1 = 0. \tag{1.3}$$

Lifted coordinates. It is common practice to linearize polynomial expressions by applying Veronese embeddings. We use the informal term "lifting" for this, for its shortness. Concretely, we apply lifting to coordinate vectors of points. We will call "n-order lifting" of a vector $\mathbf{a}$, the vector $\mathcal{L}^n(\mathbf{a})$ containing all n-degree monomials of the coefficients of $\mathbf{a}$. For example, second and third order liftings for homogeneous coordinates of 2D points, are as follows:

$$\mathcal{L}^2(\mathbf{q}) \sim \begin{pmatrix} q_1^2 \\ q_1 q_2 \\ q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix} \quad \mathcal{L}^3(\mathbf{q}) \sim \begin{pmatrix} q_1^3 \\ q_1^2 q_2 \\ q_1 q_2^2 \\ q_2^3 \\ q_1^2 q_3 \\ q_1 q_2 q_3 \\ q_2^2 q_3 \\ q_1 q_3^2 \\ q_2 q_3^2 \\ q_3^3 \end{pmatrix}. \tag{1.4}$$

Such lifting operations are useful to describe several camera models. Some camera models use "compacted" versions of lifted image point coordinates, for example:

$$\begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}.$$

We will denote these as $\bar{\mathcal{L}}^2(\mathbf{q})$, and use the same notation for other lifting orders.

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Documents
Document Sharing Hub

Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

2

Technologies

We briefly describe various image acquisition technologies, with an emphasis on omnidirectional ones. This section aims at describing the most commonly used technologies in computer vision and related areas, without any claim of being exhaustive. More information, including historical overviews, can be found in the following references, which include textbooks, articles, and webpages [50, 51, 54, 71, 103, 220, 227, 294, 326, 335, 528, 541].

2.1 Moving Cameras or Optical Elements

2.1.1 Slit Imaging

Slit imaging has been one of the first techniques to acquire panoramic images. Various prototypes existed already in the nineteenth century, usually based on a moving slit-shaped aperture. Historical overviews are given in the references in the first paragraph of this section. In the following, we only review some more recent, digital slit imaging systems, mainly those developed for robotic and computer vision; similar systems were also developed for photogrammetric applications [326].

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

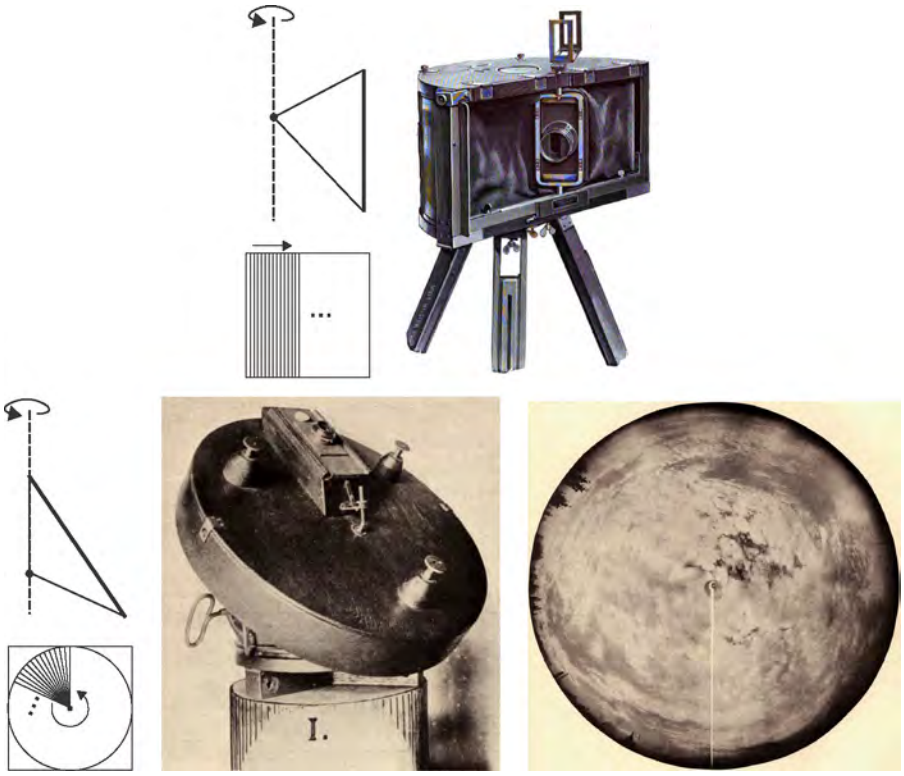


Fig. 2.1 Examples of generation of central slit images. *Top*: "Standard" slit imaging principle and an early realization of it, the cylindrograph of Moëssard [351]. *Bottom*: Slit imaging with a tilted 1D sensor, the so-called "cloud camera" and an image acquired with it (see text).

Most of these systems either use a 2D camera or a 1D camera (also called linear camera or pushbroom camera) which "scans" a scene while moving, generating a panoramic image (cf. Figure 2.1). In the 2D camera case, only one or several columns or rows of pixels are usually kept per acquired image, and stitched together to form a panoramic image. Note that pushbroom images are highly related to the so-called epipolar plane images, see for example [56].

Sarachik, Ishiguro et al., and Petty et al. acquired panoramas from a rotating perspective camera by glueing together pixel columns from each image, and used them for map building or 3D measurement [247, 400, 434]. Barth and Barrows developed a panoramic image acquisition

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

system based on a fast panning linear camera, thus significantly decreasing acquisition times [38]. A similar system was developed by Godber et al. [178]. Benosman et al. developed a panoramic stereo sensor based on the same principle, consisting of two linear cameras mounted vertically on top of each other [52]; they rotate together about their baseline and panoramic images are generated by stacking the acquired 1D images together. Issues of calibration, matching, and 3D reconstruction are addressed in [52]. Klette et al. reviewed several design principles and applications of such devices [281].

Usually, the rotation axis is parallel to the "slits" of pixels used to form the panoramic image. A historic example of rotating a tilted camera, is a design described by Fassig [136], called "cloud camera": it acquires a hemispheric field of view by revolving a tilted camera with a wedge-shaped opening around an axis (cf. Figure 2.1).

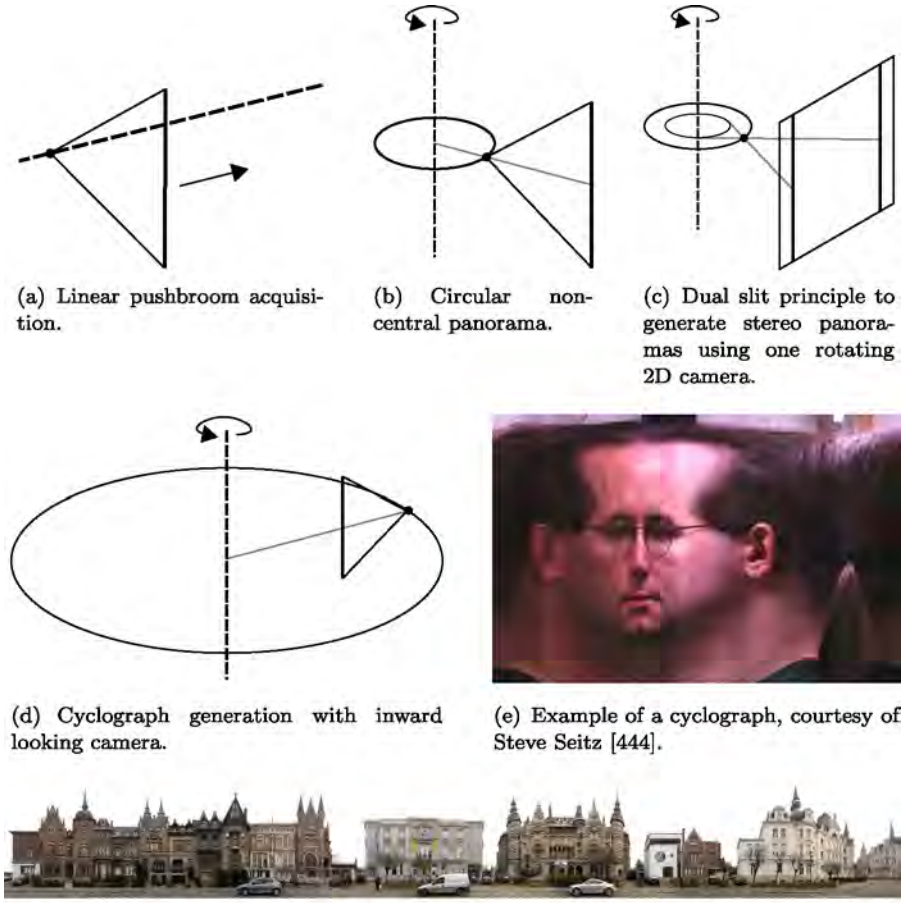
There are many more such systems; an exhaustive list is out of scope. The above systems are all designed to deliver **central panoramic images**: panoramic referring to omnidirectionality in one orientation, central referring to images having a single effective point of view. This is achieved by rotating the 1D or 2D camera about an axis containing its optical center.

In the following, we review several **non-central slit imaging systems**.

One-dimensional cameras, or pushbroom cameras, are routinely used in satellite imaging, since they provide a cheap way of obtaining high-resolution images and since they are well adapted to the way satellites scan planets. Pushbroom panoramas are non-central since each column is acquired at a different position of the satellite. The special case of linear pushbroom panoramas, where the camera motion is assumed to be a straight line (cf. Figure 2.2(a)), was extensively modeled by Gupta and Hartley [197], see Section 3.1.4. They also proposed a sensor model and a calibration method for the case of pushbroom cameras on an orbiting satellite, e.g., moving along an elliptical trajectory [198].

Concerning close-range applications, Zheng and Tsuji seem to be among the first researchers to have introduced and used non-central panoramic images [570, 571, 572], sometimes also called non-central mosaics, motion panoramas, or omnivergent images. Like in other

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision



(a) Linear pushbroom acquisition.

(b) Circular non-central panorama.

(c) Dual slit principle to generate stereo panoramas using one rotating 2D camera.

(d) Cyclograph generation with inward looking camera.

(e) Example of a cyclograph, courtesy of Steve Seitz [444].

(f) Example of a panorama obtained with the approach of [4], see text. Courtesy of Aseem Agarwala.

Fig. 2.2 Several ways of generating non-central slit or slit-like images.

systems, they proceeded by acquiring images using a moving camera, through a vertical slit, and stacking the acquired slit images one next to the other. If the camera is rotating about the vertical axis through its optical center, then the acquired image is a cylindrical mosaic, or panorama. If the camera is rotating about some other vertical axis, then we obtain non-central mosaics (cf. Figure 2.2(b)). Zheng and Tsuji have also proposed a generalization of this principle, for a camera moving on any smooth path. In the case of a straight line, we of course find the

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

above linear pushbroom panoramas. Zheng and Tusji used such motion panoramas for route following of robots. Besides explaining the generation of panoramas, they also analyzed the apparent distortions in the obtained images and other issues. They also used dual-slit panoramas (two panoramas acquired using the same camera and two vertical slits) for stereo computations (cf. Figure 2.2(c)), an idea that was later generalized by Peleg et al. [394, 395, 397] and Li et al.[314], Seitz et al. [444]. Here, 3D points can be reconstructed from point matches between the two non-central images by triangulation.

Ishiguro et al. used such panoramic views for map generation [246, 247]. They also used the dual-slit panoramas explained above for stereo computations, as well as stereo from panoramas acquired at different positions.

McMillan and Bishop used panoramic images for image-based rendering, inspired by the plenoptic function concept [338].

Krishnan and Ahuja studied how to obtain the sharpest panoramic images from a panning camera [288]. They showed that when using a regular camera, whose CCD is perpendicular to the optical axis, the camera should be rotated about an off-center point on the optical axis, together with acquiring images with a varying focus setting. This effectively yields non-central panoramas. Krishnan and Ahuja also proposed another sensor design, where the CCD plane is not perpendicular to the optical axis, and used it for panoramic image acquisition. The advantage of such a sensor is that the depth of field volume is skewed and so, while panning with a *fixed* focus setting, the union of the images' depth of field volumes, is larger than for a fronto-parallel CCD, hence avoiding to vary the focus setting while panning.

Usually, cameras are facing outward when acquiring slit images. Inward facing cameras were also considered, as early as in 1895 [105], leading to the principle of peripheral photography or images called cyclographs. A sample cyclograph obtained using this approach is shown in Figure 2.2(e). Let us mention another acquisition principle that does not produce slit panoramas but is considering the acquisition of images taken around an object: Jones et al. proposed a setup to acquire images as if taken all around an object, with a static camera and object [258], whereas the usual approach is to use multiple cameras or a

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

turntable. To do so, they placed a cylindrical mirror around the object and used a rotating planar mirror to select the effective viewpoint.

Peleg et al. showed how to acquire non-central panoramas without having to move the camera, by using a mirror of a particular shape or a lens with a particular profile [396, 395]. When combining a static central camera with the appropriate mirror or lens, the camera rays of the compound system are distributed the same way as in circular non-central panoramas, i.e., they are incident with a circle containing the camera's optical center (cf. Figure 2.2(b)).

Seitz et al. studied these ideas with the goal of obtaining the most accurate 3D reconstruction from possibly non-central stereo pairs [444]. Their problem definition is that only a 2D set of lines of sight may be acquired as an image, from viewpoints constrained within a sphere. In order to maximize the reconstruction accuracy of points outside the sphere, each point should be viewed along two lines of sight at least, with maximum vergence angle. The result is that the lines of sight to be stored are tangents of the sphere. How to best choose the appropriate tangents is described in [444], resulting in so-called omnivergent images. Interestingly, omnivergent stereo pairs, although being pairs of non-central images, do have a standard epipolar geometry, with horizontal epipolar lines, thus allow to directly apply standard stereo algorithms (more on this in Section 3.4.2). However, acquiring such spherical omnivergent images is not simple, although Nayar and Karmarkar proposed systems for this task, based on catadioptric or fisheye cameras [371]. An approximate omnivergent stereo pair can be acquired via the dual-slit principle introduced by Zheng et al., see above.

Agarwala et al. have extended the principle of using slits of images to compose panoramic images, toward the use of image regions, chosen in order to better conform to the scene's shape [4]. They apply this to generate multi-viewpoint panoramas of urban scenes, such as of the facades of buildings along streets (see Figure 2.2(f) for an example).

Ichimura and Nayar studied the problem of motion and structure recovery from freely moving 1D sensors and rigid rigs of two or three 1D sensors, as well as from special motions; their study subsumes several previously studied cases such as linear pushbroom panoramas [243].

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

2.1.2 Classical Mosaics

By "classical" mosaics, we refer to classical in terms of the computer vision community, i.e., mosaics generated by stitching together 2D images, but noting that slit imaging to generate mosaics has been done before, at least with analog cameras. A tutorial on image alignment and stitching has been published in the same journal as this survey, by Szeliski [487]. Due to this good reference and the fact, that classical mosaic generation is widely known, we do not describe this any further and simply give a few additional references.

An early and often overlooked work on digital mosaics is by Yelick and Lippman, who showed how to combine images obtained by a rotating camera to generate a mosaic [322, 547]. In his bachelor thesis [547], Yelick also discussed other omnidirectional image acquisition techniques, such as the fisheye lens and rotating slit cameras. Later works on digital mosaics include those by Teodosio and Mills [496], Teodosio and Bender [495] and Chen [92], to name but a few among the many existing ones.

2.1.3 Other Technologies

Murray analyzed the setup by Ishiguro et al. and others (see Section 2.1.1) and proposed an alternative solution where instead of rotating the camera about some axis, the camera looks at a planar mirror which rotates about an axis [362]. Such a system was already proposed for aerial imaging by Bouwers and van der Sande [58]. However, the issue of how to compensate for the camera displacement during rotations of the mirror, due to the airplane's motion, was not fully discussed. Other systems that use rotating mirrors or prisms are referenced by Yagi [541], see also Section 2.3.4.

2.2 Fisheyes

The concept of fisheye view and lens dates back to more than a century [48, 226, 350, 536]. Fisheye lenses or converters can achieve larger than hemispheric fields of view but are usually still relatively costly. A technical report suggesting inexpensive simple solutions to build

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

fisheye lenses was provided by Dietz [115]. Some more references on fisheyes are given in Section 3.1.7.

2.3 Catadioptric Systems

Using external mirrors together with cameras allows for a broad range of design possibilities, which is one of the reasons of the large number of catadioptric devices proposed alone in the computer vision community. Camera design may follow different goals, foremost in our context being a wide field of view, others being for example the compactness of a sensor, a single effective viewpoint, image quality, focusing properties, or a desired projection function. In the following, we exclusively describe geometric properties, in terms of projection function (focusing etc. of course also depend on mirror geometry).

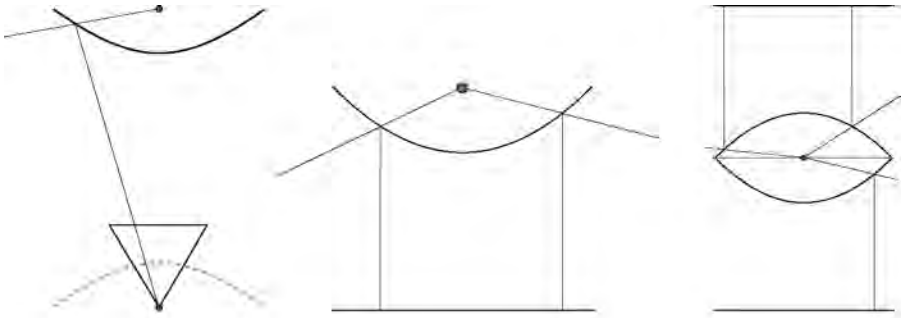
We may distinguish five types of catadioptric systems: (i) single-mirror central systems, having a single effective viewpoint, (ii) central systems using multiple mirrors, (iii) non-central systems, (iv) single-lens stereo systems, and (v) programmable devices.

The fourth category, single-lens stereo systems, is of course a subset of the third category, non-central systems, but is singled out here since the goal is to obtain a sufficiently non-central system in order to enable accurate 3D modeling whereas for other systems, "non-centrality" is usually not a design goal but an artifact or a consequence of other design goals.

2.3.1 Single-Mirror Central Catadioptric Systems

Baker and Nayar derived all single-mirror central catadioptric systems [22] (Bruckstein and Richardson obtained some of these results independently [66]). Essentially the same results were also known outside the scope of catadioptric vision, e.g., were reported in 1637 by Descartes [112] and later by Feynman et al. [142] and Drucker and Locke [121]; further, it is likely that they were already known in antiquity to Greek geometers. Single-mirror central catadioptric systems can only be constructed using mirrors whose surface is obtained by revolving a conic about a symmetry axis of the conic. In addition, the camera looking at the mirror must be central and its optical center

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision



(a) Hyper-catadioptric camera. The dashed curve is the second sheet of the hyperbola.

(b) Para-catadioptric camera.

(c) Two para-cata systems put back-to-back such that their effective viewpoints coincide.

Fig. 2.3 Illustration of (compositions of) single-mirror central catadioptric systems. Effective viewpoints are shown in green and the position of the true camera, in blue. The latter is only shown for the hyper-catadioptric case, for the para-catadioptric one, the camera is telecentric, thus has an optical center at infinity.

has to coincide with one of the conic's foci, otherwise the whole system is non-central.[1] The other focus is then the effective viewpoint of the catadioptric system: any (back-projection) line going out from the camera center goes, after being reflected in the mirror, through that second focus (cf. Figure 2.3).

The following special cases exist: hyperboloidal (cf. Figure 2.3(a)), paraboloidal (cf. Figure 2.3(b)), ellipsoidal, cone-shaped or planar mirrors. As for paraboloidal mirrors, one of the two real focus points is a point at infinity. In order to obtain a wide field of view, the only option is to "position" the camera at that point, which can be achieved using a telecentric lens. As for cone-shaped mirrors, the camera's optical center has to be located at the cone's tip; hence the only part of the mirror the camera sees corresponds to the rays that graze the mirror surface. Cone-shaped mirrors are thus theoretically excluded from the set of

[1] It may be possible to achieve a central catadioptric system using a *non-central* camera looking at a mirror which is not necessarily conic-based; all that matters is that the rays that are back-projected from the non-central camera converge to a single point after reflection in the mirror. For any non-central camera and a desired effective viewpoint, one may be able to design an appropriate mirror.

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

useful central catadioptric mirrors, although it has been shown that, with a more general modeling of optics than applied in [22], practical central viewing can still be achieved, see further below in this section. Ellipsoidal mirrors do not allow to increase the field of view and are thus not directly appropriate to build omnidirectional cameras; however, they are useful in designing so-called folded catadioptric systems, consisting of two or more mirrors (see Section 2.3.2). As for spherical mirrors, a special case of ellipsoidal ones, both real foci lie at the sphere center; a camera positioned there only sees the reflection of itself, which makes this case impractical. Overall, the only two systems deemed generally practical are the para-catadioptric and hyper-catadioptric ones, based on paraboloidal and hyperboloidal mirrors.

Central hyper-catadioptric systems seem to have been used first, cf. the patent by Rees in 1970 [424] and first applications in robotics and computer vision by Yagi and his co-workers [540, 546]. They highlighted the single viewpoint property if the camera is positioned at one of the mirror's foci and also discussed optical properties of the system such as blur. They also combined cameras of different types, such as in the MISS system which is composed of a cone-based catadioptric camera (non-central) and a standard stereo system [544]. This allows to combine mutual advantages of the sensors, such as good localization using the catadioptric camera and larger resolution with the stereo system. A full trinocular analysis of line images was also proposed.

Nayar introduced central para-catadioptric systems, consisting of an orthographic camera and a parabolic mirror, positioned such that the viewing direction of the camera is parallel to the mirror axis [367, 368]. He also showed that two para-catadioptric sensors with a field of view of 180° each, can be put back-to-back to achieve a full spherical field of view while still preserving a single effective optical center (cf. Figure 2.3(c)). This is possible since a 180° field of view is achieved with a para-catadioptric system if the mirror extends till the cross section containing its finite focus. Since the finite focus is exactly the effective optical center, putting two such mirrors back-to-back makes their finite foci coincide and thus also the two effective optical centers.

Such systems are rather widely used nowadays and commercialized by several companies, see e.g., [103].

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

As mentioned above, cone-shaped mirrors were predicted as being impractical to achieve central catadioptric systems. This issue was reconsidered by Lin and Bajcsy [319, 320]. Their starting point was that the geometric analysis of Baker and Nayar considered that each 3D point is imaged along a single light path, whereas due to the finite aperture of real lenses, light rays emitted by a point within a finite volume, hit the image area. Reciprocally, a camera whose main lens is located at the tip of a cone (the actual tip being cut off) actually sees the mirror surface and not only rays grazing it. Lin and Bajcsy showed that it is possible to obtain sharp single-viewpoint catadioptric images in this case.

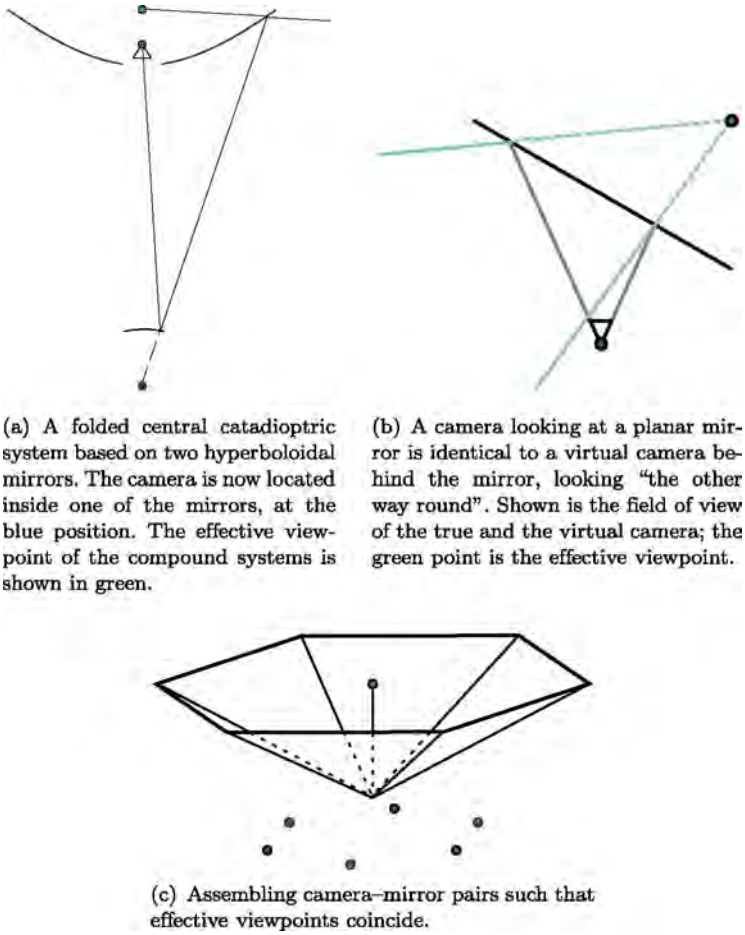
2.3.2 Central Catadioptric Cameras with Multiple Mirrors — Folded Catadioptric Cameras

Central catadioptric cameras can also be achieved when using more than one mirror. This allows more compact sensor designs and gives additional degrees of freedom to improve optical properties. Such designs are also termed *folded catadioptric cameras*; an excellent reference is [372], where Nayar and Peri presented several designs, design issues, and references to other works in this area.

Roughly speaking, when combining conic-shaped mirrors, and positioning them such that foci of successive mirrors in a sequence of reflections, coincide, then central image acquisition is possible: by placing a camera at the left-over focus of the "last" mirror, the compound system has a single effective viewpoint at the left-over focus of the "first" mirror (cf. Figure 2.4(a)).

Previously, Yagi and Yachida proposed such a system, consisting of two paraboloidal mirrors [542]. Nagahara et al. proposed a design for catadioptric optics for head mounted devices (HMDs), consisting of three mirrors, a planar, a hyperboloidal, and an ellipsoidal one, arranged such as to provide a single effective viewpoint, while achieving a wide field of view and avoiding occlusions of the field of view by the mirrors themselves [363]. Takeya et al. proposed another similar design [488]. Kim and Cho addressed the problem of calibrating a system composed of multiple successive mirrors, using a learning-based

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision



(a) A folded central catadioptric system based on two hyperboloidal mirrors. The camera is now located inside one of the mirrors, at the blue position. The effective viewpoint of the compound systems is shown in green.

(b) A camera looking at a planar mirror is identical to a virtual camera behind the mirror, looking "the other way round". Shown is the field of view of the true and the virtual camera; the green point is the effective viewpoint.

(c) Assembling camera–mirror pairs such that effective viewpoints coincide.

Fig. 2.4 Illustration of multi-mirror central catadioptric systems.

approach [280]. Nagahara et al. showed how to achieve uniform angular resolution for a folded catadioptric camera with two mirrors and a single effective viewpoint [364].

Another way of achieving central projection with multiple mirrors is what is often called the "Nalwa pyramid", although an earlier patent on essentially the same system is due to Iwerks [250]. Iwerks, Nalwa and independently, Kawanishi et al., proposed to use several regular cameras and as many planar mirrors [250, 271, 365]. A camera looking at a planar mirror produces the same image (up to side-swapping)

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

as a camera located behind the mirror, at the position obtained by reflecting the original center in the mirror, and "inversely" oriented (cf. Figure 2.4(b)). The reflected optical center position is the effective viewpoint here. The main idea consists in arranging camera–mirror pairs such that the effective viewpoints coincide. This is the easiest done in a pyramidal layout (cf. Figure 2.4(c)), but others are imaginable. Nalwa's original design consists of four camera–mirror pairs, Kawanishi et al. used six. Main advantages of such a system are omnidirectional viewing (the different camera images can be stitched together), a single effective viewpoint, and high resolution (as compared to mono-camera omnidirectional systems).

Gao et al. propose a similar design that has a hemispherical field of view [155]. Majumder et al. and Hua et al. placed two mirror–camera pyramids back-to-back such that the effective viewpoints of the two pyramids coincide, thus also enhancing the vertical field of view of the compound sensor [234, 235, 329].

Greguss developed the so-called panoramic annular lens (PAL) which combines reflective and refractive elements in a compact layout and achieves panoramic viewing with a full spherical field of view in horizontal direction and a vertical field of view of about 40° [188, 425]. Greguss' design has been used and/or improved for example by Powell [407] and Zhu et al. [574].

Yin and Boult built a sensor consisting of a tree of three different-sized coaxial paraboloidal mirrors and an orthographic camera looking at them [549]. Their motivation was not to achieve a single-lens stereo system, rather to obtain an image pyramid by optical means: the three mirrors are chosen such that they lead to omnidirectional images where each one doubles the resolution with respect to the previous one. Ideally, the mirrors' foci should coincide in order for the three pyramid images to correspond to the same viewpoint; this is not exactly the case in the proposed system but it was shown that the misalignment is negligible if the scene is sufficiently far away.

2.3.3 Non-central Catadioptric Cameras

Spherical mirrors. Hong et al. used a non-central catadioptric system with a spherical mirror for robot homing, a visual servoing

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

task [231]. They actually did not use the entire field of view that the camera offers, but just a region around the horizon plane; the horizon plane, i.e., the plane swept out by horizontal back-projection rays, corresponds to a circle in the image plane. The robot on which the camera is mounted is supposed to move on a horizontal ground plane. Hence, points in the horizon plane are always imaged on the above circle. In order to perform homing, the authors thus exploit a narrow region around that circle, perform image analysis and matching in it and use the result for a 2D visual servoing (translation and rotation in the ground plane). While a spherical mirror always leads to a non-central catadioptric system (unless the camera is placed at the sphere center), considering only the image portion corresponding to the horizon plane, actually corresponds to using an omnidirectional 1D camera with a single effective viewpoint.

Cone-shaped mirrors. Yagi and Kawato used a catadioptric system with a cone-shaped mirror (called COPIS) for robot motion estimation and map building [542]. The system was non-central since the camera was located at a certain distance from the tip of the cone. Yagi and Kawato described the forward projection model as well as 3D point triangulation and epipolar geometry for translational motion. They as well as other researchers have used cone-shaped mirrors mounted vertically on robots since vertical lines, which are omnipresent in man-made scenes, are imaged as (long) radial lines and are thus easy to extract, e.g., [61, 83, 321, 393, 456]. Mouaddib and his co-workers used this idea for pose estimation, motion estimation, map building, etc. with the sensor SYCLOP they developed [393]. They proposed a calibration method, where the sensor is immersed in a hollow cube with a calibration checkerboard painted on the inside of each face [61, 83]. They then use both, extracted points and vertical lines of the calibration object, for calibration. A method for pose estimation (or, absolute localization) is presented in [82].

Purpose-made mirrors optimizing resolution or satisfying other design goals. A drawback of "standard" catadioptric sensors is a significant variation of effective resolution, or spatial/angular resolution, across the image area. Several works aimed at removing

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

or alleviating this, see the nice overview by Hicks and Perline [225]. Different works aimed at achieving different properties for sensor resolution. Uniform angular resolution is achieved by what is usually called equidistant or equiangular projection: the angle spanned by the back-projection ray of an image point and the optical axis, is proportional to the distance between the image point and the principal point. This is approximately the case for most fisheye lenses. Ollis et al. showed how to compute mirror surfaces that provide an equiangular catadioptric sensor [386]. Previously, Chahl and Srinivasan obtained a similar result, a mirror shape where the angle between the optical axis and a ray back-projected from the camera, is proportional to the same angle after reflecting the ray in the mirror [84]. Conroy and Moore achieved a sensor with solid angle pixel density invariance, i.e., where the surface area of a circular image portion is proportional to the solid angle of the corresponding field of view [99]. A similar work is due to Hicks and Perline [225]. Another related work by Gächter et al. achieved a similar goal [153], but for the case where the camera looking at the mirror has a log-polar sensor arrangement, such as cameras studied by Tistarelli and Sandini [505] and Pardo et al. [392].

All these works lead to non-central cameras. Nagahara et al. proposed a catadioptric system that is central and has uniform angular resolution [364]; this was possible by using two appropriately shaped and placed curved mirrors, whereas the above works used a single mirror each.

Hicks and Bajcsy as well as Kweon et al. showed how to compute mirror shapes such that a particular scene plane that is "fronto-parallel" to the mirror is imaged without distortion, while still providing a very wide field of view [222, 223, 297]. This was done for both, perspective and orthographic cameras looking at the mirror. Srinivasan contributed a similar result, i.e., a mirror that gives a wide field of view while directly providing a rectified image: in [457], he showed how to compute the shape of a mirror that directly provides cylindrical panoramas, i.e., rectangular images where the two coordinates relate to azimuth respective elevation angles of points relative to the camera's optical axis. He found that there is no smooth mirror shape achieving this property, but that this can be approximated

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

by a piecewise planar mirror with a layout similar to a Fresnel lens array.

Kondo et al. proposed an anisotropic mirror shape, i.e., that is not a surface of revolution, with the aim of obtaining panoramic vision while allocating higher spatial resolution to a preferred azimuth range, e.g., corresponding to the driving direction of a robot [284, 286].

Nayar and Karmarkar showed how to acquire 360×360 mosaics by stitching together image slices acquired by a rotating *slice camera* [371]. The slice camera is designed such as to have a 360° field of view in one direction, while being orthographic in the other direction. This is achieved by a specially designed mirror; in case the camera looking at the mirror is orthographic itself, the mirror is a cone. Nayar and Karmarkar's design extends slit imaging, cf. Section 2.1.1, to the acquisition of a full spherical field of view, by being based on an omnidirectional slit camera that rotates.

Peleg et al. showed how to acquire circular non-central mosaics using a mirror of a special shape [395, 396], see Section 2.1.1.

A few more general approaches exist for mirror design, as follows. Gaspar et al. proposed a general approach allowing to derive several of the above mirror designs in a unified framework [156, 157]. Hicks as well as Menegatti formulated and provided a solution to the so-called prescribed projection problem [221, 341]: the input is a desired mapping between an object surface and the image plane of a camera in a given position. The goal is to compute a mirror and its location, that together realize this mapping, i.e., the image taken by the resulting catadioptric system is as specified. In many cases, there is no exact solution to the problem, but approximate solutions can be found. Swaminathan et al. addressed the same problem as Hicks in [486] and proposed a solution that minimizes reprojection errors in the image plane, i.e., where the desired and the actual scene-to-image mapping (the "inverse" of the prescribed projection problem) give image points as close to one another as possible. Kondo et al. proposed another approach for the same problem, akin to photometric stereo [285]. This approach allows to conceive discontinuous mirror shapes.

All the above sensors are necessarily non-central (with the exception of [364]), although for many applications, one may model them

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

sufficiently well using a central approximation, see also the discussion in Section 3.6.

Krishnan and Nayar recently proposed a catadioptric camera whose main optics is a fisheye, a camera they call cata-fisheye [289]. The motivation for this design is to achieve a spherical field of view in azimuth while not necessarily having a hemispherical one in elevation. Indeed, in many applications, such as videoconferencing and vehicle-mounted computer vision, a vertical field of view of a few dozens of degrees is often sufficient. Krishnan and Nayar proposed to achieve this by putting a convex mirror in front of a fisheye. The produced image consists of two annular regions: the outer one corresponds to the direct view at the scene through the fisheye, whereas the inner one shows the reflection in the mirror. Here, the outer image region would usually correspond to the part of the vertical field of view that is above the camera, while the inner region shows a part that is below. One advantage of this system is thus that the full image resolution is dedicated to the desired vertical field of view (as opposed to "wasting" resolution on a full hemispherical fisheye field of view if parts thereof are not relevant for a given application). Other advantages are a good image quality since a mirror of low curvature may be used, and a rather compact design. It is furthermore easy to regulate the desired vertical field of view by displacing the mirror or employing mirrors of different shapes. Strictly speaking, the system is non-central in general, but the working range in which the parallax is negligible, is explained in [289].

2.3.4 Single-Lens Stereo Systems

As mentioned at the beginning of Section 2.3, we consider here catadioptric systems that are intentionally non-central in order to enable stereovision, whereas the systems in the previous section are non-central "by accident" or due to conforming to design goals prohibiting a single effective viewpoint. Most catadioptric single-lens stereo systems are achieved by setting one or more planar or curved mirrors in front of the camera [73, 74, 99, 143, 183, 203, 255, 333, 366, 386, 455] or in addition by keeping a direct view on the scene for a part of the camera's field of view [15, 49, 261, 295, 343, 422, 569].

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Advantages of mirror-based stereo systems over multi-camera setups are that no camera synchronization and radiometric alignment of camera are required and that only one set of intrinsic parameters has to be calibrated, as opposed to calibrating each camera in a multi-camera setup.

Mouaddib et al. proposed a set of performance criteria to compare different single-lens catadioptric stereo systems [354, 355].

In the following, we describe some systems, first some using planar mirrors.

Planar mirrors. One of the earliest references suggesting a design for a mirror-based single-lens stereo system is probably a paper of 1899 by Finsterwalder [144], a seminal paper which summarized various results on camera calibration, pose and motion estimation, epipolar geometry, and even projective reconstruction from uncalibrated images. On pages 20–22 of this paper (written in German), Finsterwalder proposed to use a setup of three mutually perpendicular planar mirrors to photograph small objects put between the camera and the mirrors. He suggested that this gives up to eight perspective images (direct view up to multiple reflections) which can be used for stereo reconstruction. We do not know if this system has ever been built.

The idea of using a single planar mirror to acquire stereo images for 3D modeling, a direct and a reflected view of an object, has been considered by various researchers in the early twentieth century, see for example [39, 117, 409, 562, 563]. Some of these works were motivated by the modeling of coastal landscapes from images taken aboard a ship, the mirror surface being formed by a sea or lake. Note also that systems composed of planar mirrors were proposed for stereo *viewing* (as opposed to stereo imaging) as early as in 1838, by Wheatstone [526, 527] (see also [229]).

Kaneko and Honda showed that when acquiring an image of an object consisting of a part with a direct view of the object and another with its reflection in a planar mirror, this indeed corresponds to a stereo configuration and allows to reconstruct the object in 3D [261] (see also a similar approach by Zhang and Tsui [569]). This approach was extended by Mitsumoto et al. whose approach relaxed the requirement

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

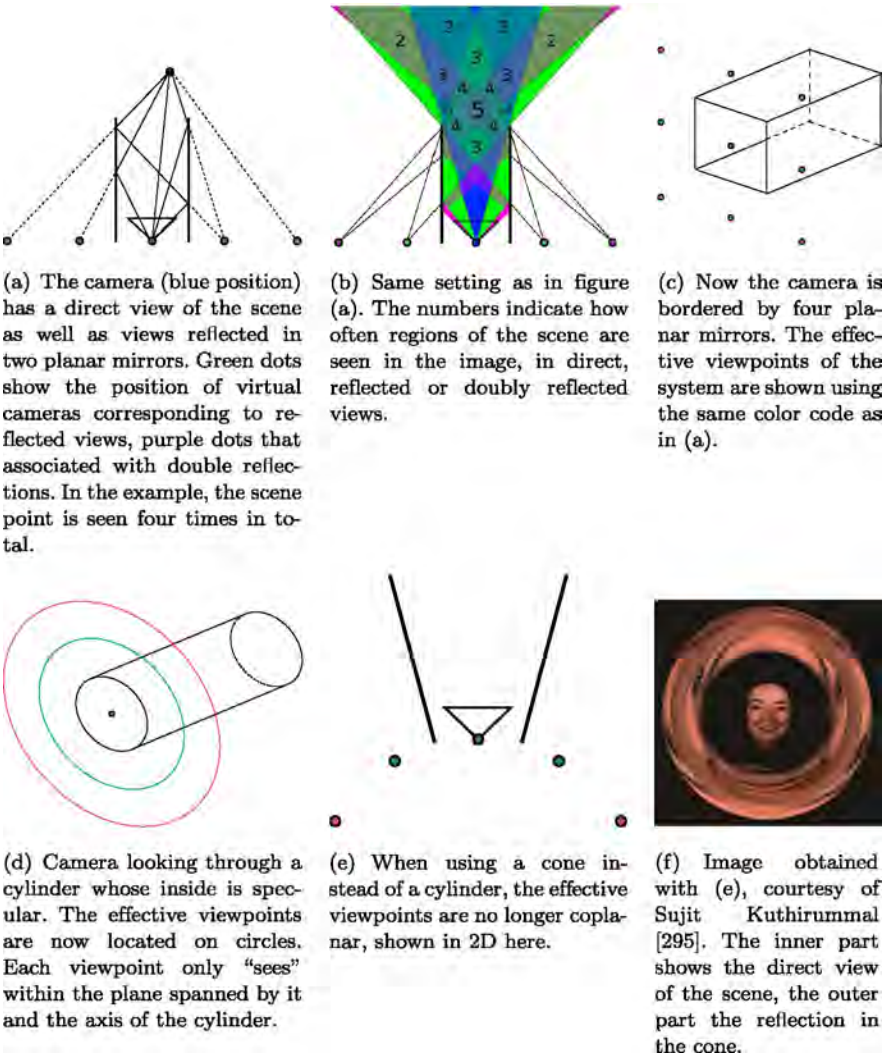
of knowing the mirror's position and who also proposed to use an additional planar mirror, to better recover occluded object parts [343]. Arnspang et al. also proposed to put planar mirrors in front of the camera [15, 422]. One of the setups they suggested consisted of two mirrors facing each other and which are parallel to the camera's optical axis (cf. Figure 2.5(a) and (b)). Here, a scene point may be seen multiple times: in the direct camera view, reflected in the mirrors or even after a double reflection, once in each mirror. Arnspang et al. formulated the triangulation problem for their setup and also proposed to extend the setup by arranging more than two mirrors in a cylindrical ring in front of the camera (cf. Figure 2.5(c)).

A similar idea was suggested by Han and Perlin who proposed a single-lens stereo system akin to kaleidoscopes [203]. The camera looks through a tapered tube whose inside is made of planar mirror facets. Each planar mirror, together with the actual camera, corresponds to a virtual camera. Hence, each scene point can be seen multiple times in what is effectively a multi-view stereo system. Furthermore, multiple reflections of light rays may happen inside the tube, thus multiplying the number of virtual cameras and viewpoints, much like what one can observe in a kaleidoscope. Han and Perlin's idea was to use this acquisition system for the acquisition of the Bidirectional Texture Reflectance of objects, based on the large number of viewpoints contained in a single snapshot. Kuthirummal and Nayar proposed a similar system, see further below.

Goshtasby and Gruver used two planar mirrors and no direct view of the scene [183]. Cafforio and Rocca, Inaba et al. as well as Mathieu and Devernay used four planar mirrors, where a pair each gives one virtual camera, via the successive reflections in two mirrors (see Figure 2.6); the system is thus basically equivalent to one with two planar mirrors, but allows for an easier setup by more easily avoiding self-reflections of the camera in the mirrors [74, 244, 333]. Gluckman and Nayar analyzed in detail the relative orientation and epipolar geometry of such systems as well as their self-calibration [174, 175].

In [177], Gluckman and Nayar studied the question of how to obtain an optically rectified stereo pair using a single-lens catadioptric system with planar mirrors. They found that an odd number of mirrors is

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision



(a) The camera (blue position) has a direct view of the scene as well as views reflected in two planar mirrors. Green dots show the position of virtual cameras corresponding to reflected views, purple dots that associated with double reflections. In the example, the scene point is seen four times in total.

(b) Same setting as in figure (a). The numbers indicate how often regions of the scene are seen in the image, in direct, reflected or doubly reflected views.

(c) Now the camera is bordered by four planar mirrors. The effective viewpoints of the system are shown using the same color code as in (a).

(d) Camera looking through a cylinder whose inside is specular. The effective viewpoints are now located on circles. Each viewpoint only "sees" within the plane spanned by it and the axis of the cylinder.

(e) When using a cone instead of a cylinder, the effective viewpoints are no longer coplanar, shown in 2D here.

(f) Image obtained with (e), courtesy of Sujit Kuthirummal [295]. The inner part shows the direct view of the scene, the outer part the reflection in the cone.

Fig. 2.5 Illustration of some single-lens catadioptric stereo systems.

required and for the cases of one and three mirrors, derived the constraints on mirror placement that ensure rectified images. They also showed how to optimize the mirror placement in order to minimize the overall size of the entire sensor.

Avni et al. built a system composed of two cameras and two planar mirrors: the cameras as well as an object to be modeled in 3D are

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

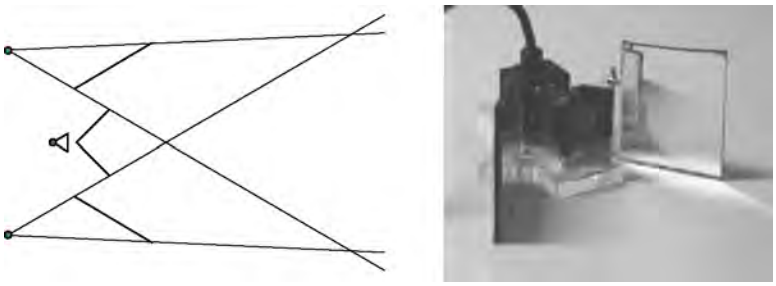


Fig. 2.6 *Left*: Single-lens stereo system using planar mirrors. Here, four mirrors are used for practical reasons, leading to two effective viewpoints, shown in green. *Right:* A practical realization of the system, image courtesy of Frédéric Devernay [333].

positioned above the mirrors such that each camera sees the object's reflections in both mirrors [20]. Hence, a total of four images of the object are obtained, allowing a multi-view 3D modeling.

Non-planar mirrors. Nayar used two specular spheres in the field of view of a camera to obtain stereo information [366]. Southwell et al. presented a design of a single-lens catadioptric stereo sensor with curved mirrors where one convex mirror is fixed on top of a second, larger one [143, 455]. Hence, points in the common field of view of the two individual catadioptric images are seen twice and can be reconstructed. Similar designs were proposed by Cabral et al. [73] and Jang et al. [255]. Jang et al. especially discussed how to maximize the effective stereo baseline.

Nene and Nayar described several single-lens catadioptric stereo configurations where both (virtual) cameras in a stereo system are central [374]. The first uses, like in other systems, two planar mirrors; Nene and Nayar described the epipolar geometry of this setup. The other proposed configurations use mirrors of revolution of conic-shape placed such that the camera is at a focus point of each of the mirrors, see Figure 2.7(a). Hence, each mirror corresponds to a central catadioptric system and by using two or more mirrors arranged this way, a stereo system is obtained where each (virtual) camera is central. Nene and Nayar described the cases of pairs of ellipsoidal, pairs of hyperboloidal and pairs of paraboloidal mirrors and derived the epipolar geometry for

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

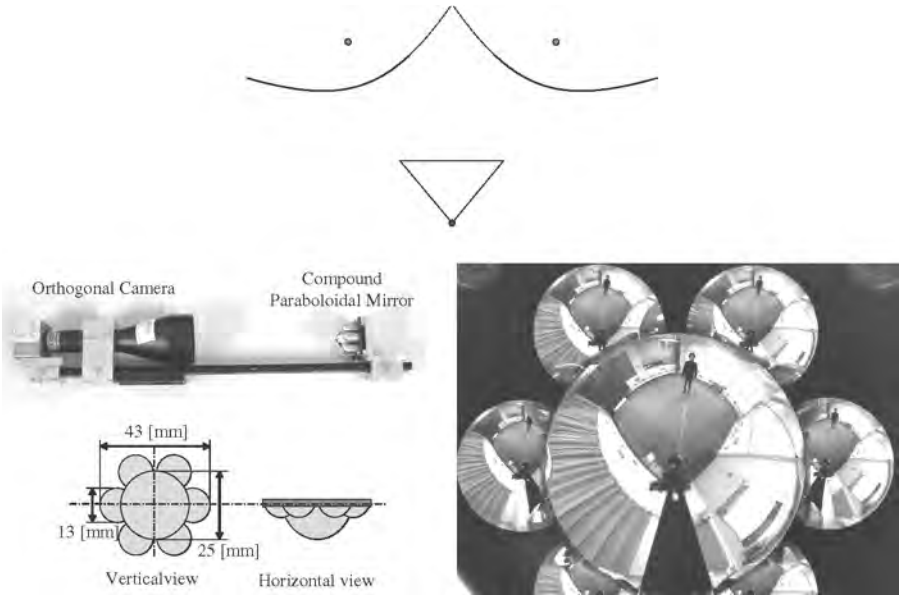


Fig. 2.7 Other single-lens catadioptric stereo systems. *Top*: System using two hyperbolic mirrors with a coinciding focus point, at which the camera is located. This is a single-lens stereo system where each view is central. *Bottom*: Single-lens stereo system by Sagawa et al. composed of paraboloidal mirrors and sample acquired image; the camera looks at the mirrors from below. Images courtesy of Ryusuke Sagawa.

each of those. Other cases are straightforward to imagine, e.g., combing ellipsoidal/hyperboloidal/planar mirrors in the appropriate manner.

Murphy presented a panoramic imaging system for planetary rovers where a single camera looks at a convex mirror with a hole in the middle, through which the camera sees through a fisheye lens [361]. Hence, a single image contains two panoramas, one for the lower part (the catadioptric view) and one for the upper one (fisheye). The system is similar to that of Krishnan and Nayar (cf. Section 2.3.3), although its aim is to obtain single-lens stereo while that of Krishnan and Nayar's design is omnidirectional viewing with a desired distribution of spatial resolution across the image. Benosman et al. proposed a sensor where a camera has both, a direct view of the scene and a view on a hyperbolic mirror [49]. The motivation is to have a high resolution in a dominant direction via the direct perspective image, e.g., the direction ahead of a robot, together with a lower-resolution panoramic view

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

of the surroundings. Yi and Ahuja proposed another single-lens stereo system, consisting of a hyperboloidal mirror and a concave lens [548]. The camera has a direct view of the mirror and an indirect one of it, through the concave lens, thus effectively producing a stereo pair.

Sagawa et al. built a single-lens catadioptric stereo system consisting of one camera looking at seven spherical or paraboloidal mirrors [432], see Figure 2.7(b) and (c). Although in principle such a system can be used for multi-baseline stereo, the prototype shown in [432] is of small dimension and thus has a small baseline. Its intended application is the detection of close-by objects, for which accurate 3D reconstruction is not necessary. The calibration of the sensor is discussed in [282], see also Section 5.3.2.

Lanman et al. built a similar catadioptric acquisition system consisting of a single high-resolution camera looking at a set of spherical mirrors arranged on a plate [298]. Their system is larger than Sagawa's since one intended application is 3D modeling. The system is obviously non-central since already a single spherical mirror leads to a non-central image. While a single spherical mirror is only "slightly non-central" (more on this in Section 3.6), the setup by Lanman et al. effectively generates a large "baseline", allowing for single image 3D reconstruction. A practical method for calibrating the system, including the properties of second surface mirrors, is proposed in [298], and the multi-mirror view geometry is analyzed, i.e., conditions that hold for points observed in different mirrors, to be images of the same scene point.

Kuthirummal and Nayar proposed a system similar to the one by Arnspang et al. and Han and Perlin (see above), where the piecewise planar tube is replaced by a cone or cylinder-shaped one [295], see Figure 2.5(d)–(f). They demonstrated the use of their system for single-image 3D scene and reflectance recovery.

Orghidan et al. designed a structured-light type depth sensor composed of a camera looking at two hyperboloidal mirrors and a laser emitter [388]. They showed how to model and calibrate this system and use it for 3D modeling.

Rotating mirrors. Other catadioptric single-lens stereo systems have been proposed, using mirrors that rotate between different image

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

acquisitions in order to produce stereo images, e.g., by Teoh and Zhang [497], Nishimoto and Shirai [376], Murray [362], and Gao and Ahuja [154].

2.3.5 Programmable Systems

Hicks and Nayar et al. proposed methods and system designs allowing to modify the shape of the mirror in a catadioptric system in order to adapt the sensor to a change in the scene or to acquire an image in a desired way [224, 296, 370].

Nayar et al. proposed to use a programmable array of micro-mirrors to achieve programmable, or purposive, imaging [370, 369]. They demonstrated the use of digital micro-mirror devices (DMDs), routinely used in projectors, for tasks such as high dynamic range image acquisition, optical appearance matching, or generally speaking, the change of imaging geometry. In a similar work, Hicks et al. also suggested to use micro-mirrors to control the way an image is acquired, e.g., by actively controlling the spatial resolution across an image [224].

The following two ideas are not strictly speaking programmable systems. Kuthirummal and Nayar studied the possibility of using a flexible mirror sheet as reflector in a catadioptric system [296]. They proposed to recover the current shape of the mirror from its outline in the image and additional assumptions; once the shape is determined, the catadioptric system is effectively calibrated. The system is not programmable in the same sense as those by Hicks and Nayar et al. but is described in this section since it operates by changing the mirror shape during image acquisition.

Fergus et al. proposed the concept of random lens imaging [140]. One practical instance of this concept is a camera looking at a collection of randomly positioned small mirrors or refractive elements. A calibration method is proposed and potential applications of the general concept for tasks, such as super-resolution and depth sensing, are described.

2.4 Stereo and Multi-camera Systems

Using two or more cameras to achieve omnidirectional viewing is a well-known principle and will not be covered in great detail here, besides a

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

few historical references and references to stereo systems based on omnidirectional cameras. A nice overview of omnidirectional stereo systems is due to Zhu [573].

Lin and Bajcsy proposed a catadioptric stereo system consisting of two cameras looking at one cone-shaped mirror from different distances [321]. The cameras are aligned with the cone's axis of revolution; to be precise, one camera lies on the axis, the other one looks at the mirror through a beam splitter such that it virtually comes to lie on the mirror axis. Hence, epipolar planes contain the axis, which simplifies the epipolar geometry and 3D reconstruction.

Spacek proposed to use two cone–camera pairs, one on top of the other and all cameras and mirrors axis-aligned [456]. He derives the equations for stereo computations and highlights that cone-based catadioptric cameras, although being non-central, can be used for stereo and may be advantageous over the more common central catadioptric systems due to their higher spatial resolution (lower vertical field of view). A similar system, consisting of two hyper-catadioptric cameras, was used in [57, 413].

Other stereo systems combine omnidirectional and traditional cameras, to combine their respective advantages, see e.g., [3, 79, 118, 268].

Multi-camera systems composed of several *perspective* cameras have been built at least as early as in 1884, initially mainly if not exclusively for aerial imaging. The earliest work known to us (no effort was made for an exhaustive bibliography research) is that of Triboulet, who, as reported in [504] experimented from 1884 on with a multi-camera system consisting of seven cameras attached to a balloon: one camera looked downward and six cameras were equally distributed around the balloon's circumference (the system thus resembles the popular Ladybug sensor by Point Grey, `http://www.ptgrey.com`). Similar other multi-camera systems from the early twentieth century include those by Scheimpflug [438] and Thiele [499], cf. Figure 2.8. Several other systems were developed throughout the first half of the twentieth century, see for example [87, 403, 423, 502].

Other systems used multiple lenses in the same camera body. Gasser as well as Aschenbrenner invented systems where multiple cameras share the same focal plane and film; due to a suitable arrangement

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

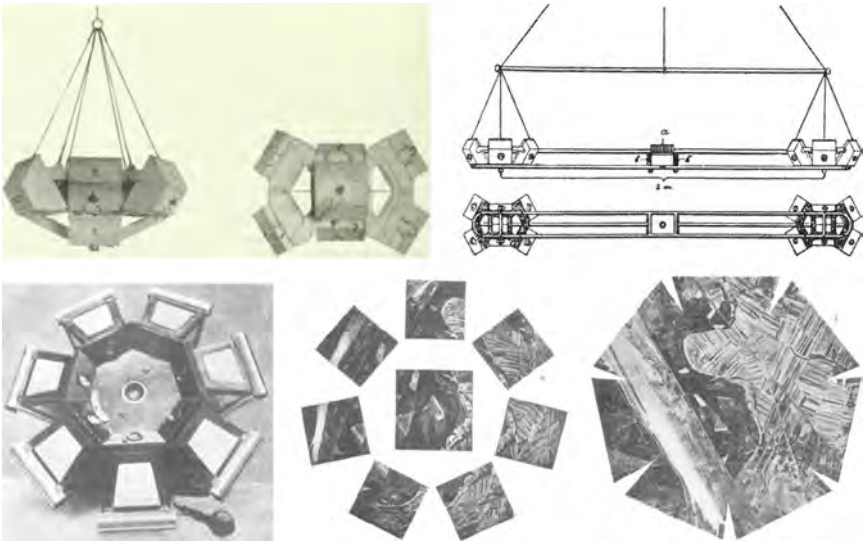


Fig. 2.8 Three multi-camera systems from the first decade of the twentieth century. *Top*: two systems developed by Thiele [499]. The first one, called auto-panoramograph, was attached at the bottom of a balloon. The second system, the stereo-panoramograph, is equivalent to an omnidirectional stereo rig with a baseline of 2 m. *Bottom*: A system developed by Scheimpflug [438], consisting of seven oblique cameras and a central one. Shown are eight original images acquired with the system and a mosaic composed from them using a special projection equipment called photo-perspectograph.

of prisms and mirrors, the different cameras directly generate a composite wide-angle image on the film [18, 160]. These systems were, like most of the above, intended for aerial imagery. Multi-lens cameras were used at least as early as in 1882: Londe used multi-lens cameras to study epileptic seizures, by capturing a succession of time-delayed images [71], similar to the famous motion series by Muybridge according to an acquisition principle he developed in 1872 [71].

2.5 Others

2.5.1 Plenoptic Camera

Adelson and Wang proposed the so-called plenoptic camera, a design that essentially gives multiple pinhole images while using a single main aperture and a single sensor array [2]. When placing a lenticular array in front of a camera's sensor plane, each lenslet together with a subset

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

of the sensor array's pixels forms a tiny pinhole camera. That pinhole camera effectively only captures rays entering the camera's main aperture within a subregion of the aperture, as opposed to regular cameras, where each pixel integrates light rays from all over the aperture. Conceptually, a plenoptic camera thus corresponds to a multi-camera system consisting of cameras arranged on a planar grid. Acquired images can be used to perform stereo computations or various computational photography tasks. We will not consider plenoptic cameras further in this article since they are not usually used for panoramic imaging, although in principle nothing would prevent to use them with fisheye lenses (besides the drop of spatial resolution for the already low-resolution individual pinhole sub-cameras).

More information on plenoptic cameras and similar designs can for example be found on `http://en.wikipedia.org/wiki/Integral_photography` and in [429, 162] The concept of plenoptic camera is highly related to that of integral imaging [323] and parallax stereogram [248] or panoramagram [249].

2.5.2 Biprism

A single-lens stereo system using a biprism rather than a mirror, that directly produces rectified images, has been developed by Lee et al. [303].

2.5.3 Spherical Lens and Spherical Image Area

Krishnan and Nayar proposed an omnidirectional sensor consisting of a transparent sphere, around which sensing elements are uniformly arranged on a spherical surface, with free space between neighboring elements [290]. Hence, individual sensing elements can see through the transparent sphere and other sensing elements, to perceive the scene. The design goals of the system are an omnidirectional field of view with a single center of projection and uniform spatial resolution.

2.5.4 Krill-Eye

Hiura et al. proposed an optical arrangement for wide-angle imaging system, termed krill-eye, that is inspired by compound animal eyes [228]

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

The system consists of gradient refractive index (GRIN) lenses aligned on a spherical surface. Hiura et al. proposed a theoretical study on image quality and focusing properties and built a proof-of-concept prototype.

2.5.5 Rolling Shutter Cameras

Meingast et al. and Ait-Aider et al. studied the case of rolling shutter CMOS cameras, where the image area is not exposed simultaneously, but in a rolling fashion across rows of pixels [10, 171, 340]. The consequence is that when taking a picture of a moving object, it will appear distorted. Meingast et al. developed projection models for the general and several special cases of object motion and showed how optical flow can be modeled [340]. They also addressed the calibration of the sensor. Ait-Aider et al. used the rolling shutter's apparent drawback to develop an approach that computes both, the pose of a known object and its instantaneous velocity, from a single image thereof [10]. There is a conceptual link to pushbroom cameras, as follows, highlighted in a similar fashion in [171, 340]. An alternative way of looking at the rolling shutter phenomenon is to consider a static object and a pushbroom camera moving around it with a velocity that combines the inverse velocity of the object and a motion that compensates the difference in pose associated with different rows of pixels in the camera.

The effect of rolling shutters was also taken into account by Wilburn et al. who built a multi-camera array with CMOS cameras [529, 530], with several applications such as synthetic aperture imaging or allowing to acquire videos with increased frame rate, dynamic range, resolution, etc. A method for calibrating such a camera array was proposed by Vaish et al. [516].

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

3

Camera Models

In the following, we review a certain number of camera models that were proposed in the literature. Some of them are based on physical models, others are of a more algebraic nature.

These models can be described and sorted according to various criteria. A first characteristic concerns the spatial distribution of the camera rays along which a camera samples light in the environment. Most models have a single optical center through which all camera rays pass. We also speak of **central camera models**. For these, the back-projection function (see below) delivers the direction of the camera ray. **Non-central camera models** do not possess a single optical center. In that case, the back-projection operation has to deliver not only the direction but also the position of a camera ray, e.g., some finite point on the ray. We will also use Plücker coordinates to represent camera rays. Special cases of non-central cameras are **oblique camera**, where no two camera rays meet [390] and **axial cameras** where there exists a line that cuts all camera rays. A special case of this are **x-slit** or **two-slit models** where there exist two lines that cut all camera rays. This is for example the case for linear pushbroom cameras [197].

A second property of camera models and one we would like to stress particularly, concerns how global or local a model is. The following

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

definitions have already been introduced in Section 1.1, cf. also Figure 1.1. Most models have a small set of parameters and in addition, each parameter influences the projection function all across the field of view. We call these, **global models**, since they hold for the entire field of view/image area. Second, there exist several more **local models**, e.g., models with different parameter sets for different portions of the image. These models have usually more parameters than global ones, but have a higher descriptive power. Finally, at the extreme, we have **discrete models**, where the projection or back-projection function is merely sampled at different points, e.g., sampling the back-projection function at every pixel. These models have many parameters, one set of them per sampled location. They are sometimes called non-parametric models, but we do not find this entirely appropriate, since they do have parameters; hence the proposed name of discrete models. Let us note that the transition from global to local to discrete models is not discontinuous: some local models have the same parametric form as global ones, and to use discrete models, one usually needs some interpolation scheme, for example to be able to back-project any image point, not only sampled ones. In that respect, discrete models plus, such as interpolation scheme, are actually local models in our language.

Many of the global and discrete models described in the following are well known. This is less true for the local models, although they may represent a good compromise between tractability (number of parameters, stability of calibration) and generality. We would like to point out the work of Martins et al. [331], which is rather often cited but often described only partially or not appropriately. They proposed three versions of the so-called two-plane model, see Sections 3.1.3 and 3.2.1 for more details. These foreshadowed several important contributions by others, mostly achieved independently. First, it contains one of the first proposals for a local camera model (the spline-based version of the developed two-plane model). Second, it sparked works on discrete camera models, where camera rays are calibrated individually from images of two or more planar calibration grids [85, 189], an approach rediscovered by [191, 477]. Third, the linear and quadratic versions of the two-plane model, when written in terms of back-projection matrices, are nothing else than particular

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

instances of rational polynomial models; they are thus closely related to the division and rational models explained by Fitzgibbon [145] and Claus and Fitzgibbon [98] or the general linear cameras of Yu and McMillan [559], as well as others, which found a lot of interest recently. Martins et al. may also be among the first to explicitly use lifted coordinates (cf. Section 1.2) in camera models.

Finally, a third main property of camera models we consider is that some models have a direct expression for forward projection, others for back-projections, some work easily both ways. This is important since forward and back-projections are convenient for different tasks: forward projection for distortion correction of images and bundle adjustment, back-projection for minimal methods for various structure-from-motion tasks, such as pose and motion estimation.

A few more general notes and explanations of notations follow.

Back-projection versus forward projection. One defines the other but one or the other may be more difficult to formulate algebraically. For example, the classical radial distortion model is written using polynomials; even with only one distortion coefficient, the polynomials are cubic, making it cumbersome to write the inverse model. This model is traditionally used for back-projection although some researchers used the same expression for forward projection. Other polynomial models are also used for either direction.

Generally speaking though, back-projection is often easier to formulate. This is especially true for catadioptric systems, where back-projection comes down to a deterministic "closed-form" ray-tracing whereas forward projection entails a search for which light ray(s) emitted by a 3D point gets reflected by the mirror(s) into the camera. Also, when using rational polynomial functions in point coordinates to express camera models, this is straightforward for back-projection; using such a model for forward projection, results in general in curved "camera rays", i.e., the set of 3D points mapped onto an image point, form a curve, not a straight line (cf. Section 3.1.8 on the cubic camera model by Hartley and Saxena [210]).

For these and other reasons, we will emphasize back-projection formulations in the following. Another reason for this is that back-projection is less often used in the literature although it enables to

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

clearly formulate several structure-from-motion tasks and epipolar geometry.

Back-projection will be formulated in two different ways, either by giving expressions allowing to compute two points on a camera ray, or by directly giving the Plücker coordinates of the ray. As for the first case, one of the two points on the ray will be its point at infinity and the other, a finite point, denoted by

$$\mathbf{B}^i \quad \text{and} \quad \mathbf{B}^f,$$

respectively. The point at infinity will be given as a 3-vector, the finite point as a 4-vector, both in homogeneous coordinates. Note that in this survey we do not consider the possibility of camera rays that lie completely on the plane at infinity; such a case is of small practical interest and if necessary, all formulas are easy to adapt. Plücker coordinates of camera rays will be written as

$$\mathbf{B}^l.$$

They can of course be easily computed from the two points $\mathbf{B}^i$ and $\mathbf{B}^f$, cf. Section 1.2.

The reason for using both expressions for back-projection is that the two-point expression is better suited to formulate pose estimation, whereas the Plücker-based expression allows to formulate motion estimation and epipolar geometry in a straightforward manner. When we deal with central camera models, we will only give the point at infinity $\mathbf{B}^i$ of camera rays, i.e., their direction, always assuming that cameras are in canonical position, i.e., with the optical center at the origin.

Whenever possible, we try to write the back-projection operation using back-projection matrices, operating on (lifted) image point coordinates. We will use full back-projection matrices

$$\mathsf{B}^l$$

of size $6 \times n$, that map image points to camera rays in Plücker coordinates, and partial back-projection matrices

$$\mathsf{B}^i$$

of size $3 \times n$, that map image points to camera ray directions. The values of n for different models depend on which liftings of image coordinates are required.

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Finally, let us note that the back-projection for a true camera usually gives rise to a half-line in 3D; many camera models do not take this fully into account and only model back-projection via infinite lines. We do the same in this monograph and, for ease of language, use the term ray or camera ray to denote a line of sight of a camera.

Radial distortion. In this section, models are presented in different forms, depending if they are backward or forward ones, or only model radial distortion or more general distortions. Radial distortion models are given in different forms; they are defined by a 1D (un)distortion profile or (un)distortion function, mapping radial distances in the distorted image (distances of image point to distortion center) to either radial distances in the undistorted image or the incidence angle between camera rays and the optical axis.

We use the following notations. The radial distance is denoted by r_d (in distorted images) and r_u (in undistorted images). The angle between a camera ray and the optical axis is denoted by θ. Radial distortion models are usually given as a mapping between any of these three entities. For simplicity, we always assume in the following that the radial distortion center is equal to the principal point and that it is known and located at the origin. When considering a distortion center different from the origin, the given back-projection equations can be modified in a straightforward manner, by preceding them with the translation bringing the distortion center to the origin. When considering a distortion center different from the principal point, the modifications of back-projection equations are equally straightforward. Whereas in the following, we suppose that the perspective back-projection consists merely in "undoing" the focal length, considering a distortion center that is different from the principal point, requires a full perspective back-projection.

Extrinsic parameters. Camera pose, or extrinsic parameters, is straightforward to take into account. As for back-projection, translations and rotations can be applied as follows (cf. Section 1.2):

$$\mathbf{B}^f \mapsto \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{B}^f$$

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision

$$\mathbf{B}^i \mapsto \mathsf{R}\mathbf{B}^i$$
$$\mathbf{B}^l \mapsto \begin{pmatrix} \mathsf{R} & 0 \\ -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \end{pmatrix} \mathbf{B}^l.$$

3.1 Global Camera Models

3.1.1 Classical Models

By "classical models", we mean those used most often in applications and academic research, i.e., the pinhole model, affine camera models, and pinhole models enhanced with classical terms for radial and tangential distortions.

Pinhole model. The pinhole model, or perspective projection, assumes that all camera rays pass through a single point, the optical center and that there is a linear relationship between image point position and the direction of the associated camera ray. That relationship can be expressed via a so-called calibration matrix which depends on up to five intrinsic parameters:

$$\mathsf{K} = \begin{pmatrix} f_u & s & x_0 \\ 0 & f_v & y_0 \\ 0 & 0 & 1 \end{pmatrix},$$

where f_u respectively f_v is the focal length measured in pixel dimensions, horizontally respectively vertically, s is a so-called skew term which may for example model non-rectangular pixels or synchronization errors in the image read-out, and (x_0, y_0) are the coordinates of the principal point, the orthogonal projection of the optical center onto the image plane. The linear relation between image points and ray directions can be written as:

$$\mathbf{q} \sim \mathsf{K}\mathbf{B}^i.$$

The model is equally easy to use for forward and back-projections. In what follows, we will assume square pixels, i.e., $s = 0$ and $f_u = f_v = f$:

$$\mathsf{K} = \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.1}$$

Documents
Document Sharing Hub
Full document details: IEEE Camera Models and Fundamental Concepts Used in Geometric Computer Vision