

OMNIDIRECTIONAL OPTICAL FLOW AND VISUAL MOTION DETECTION FOR AUTONOMOUS ROBOT NAVIGATION

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften

eingereicht an der
Universität Osnabrück
Fachbereich Mathematik/Informatik

von
Irem Stratmann

Gutachter:

1. Prof. Dr. J. Hertzberg, Institut für Informatik, Universität Osnabrück
2. Prof. Dr. R. Schwarte, Institut für Nachrichtenverarbeitung, Technische Universität Siegen

Tag der mündlichen Prüfung: 16 Oktober 2007

"I wish you wouldn't keep appearing and vanishing
so suddenly; you make one quite giddy!"
"All right," said the Cat;
and this time it vanished quite slowly, beginning
with the end of the tail, and ending with the
grin, which remained some time after the rest
of it had gone.
"Well! I've often seen a cat without a grin,"
thought Alice;
"but a grin without a cat! It's the most curious
thing I ever saw in all my life!"

'Alice in Wonderland', Lewis Carroll

Acknowledgments

I am grateful for the guidance and support of my supervisors Prof. Joachim Hertzberg (University of Osnabrück), Prof. Rudolf Schwarte (Technical University of Siegen), and Dr.-Ing. Erich Rome (FhG-AIS).

The major part of this research has been performed at the Fraunhofer Institute for Autonomous Intelligent Systems at Sankt-Augustin, Germany. I would like to thank Prof. Thomas Christaller, the Head of the Institute, for giving me this opportunity. Many members of the FhG-AIS have supported my work in many ways. Special thanks to Ralph Breithaupt, Hartmut Surmann, Andreas Nüchter and Erik Solda for their help in providing the image sequences. The interesting discussions with Hans-Ulrich Kobialka and Vlatko Becanovic were without doubt very helpful. I thank all the members of the ARC-Team for the friendly atmosphere.

I would also like to thank Thorsten Ringbeck from PMDTechnologies GmbH, Siegen, Germany for providing the 3D camera setup, which I used in some of the experiments.

Finally, I am indebted to my family for their emotional and intellectual support and patience. Without them, this work wouldn't be complete.

Contents

1	Introduction	9
1.1	On Motion Seeing in Living Beings	11
1.2	Robotics and Visual Motion Detection	12
1.2.1	On the advantages of a wide-angle field of view	13
1.3	Objectives of the thesis	15
1.3.1	Contributions	16
1.3.2	Structure of the Thesis	18
2	State of the Art	19
2.1	Basics of Optical Flow Computation	20
2.2	Methods for Computing Optical Flow	23
2.2.1	Matching Techniques	23
2.2.2	Differential Techniques	24
2.2.3	Frequency-based Techniques	27
2.3	Comparative Evaluations	29
2.4	Recent Advances in OFC	30
2.5	Motion Recovery	32
2.6	Catadioptrical Vision	34
2.6.1	Examples from Arthropod Vision	35
2.6.2	Catadioptric Sensors and Mirror Design	37
3	OFC on Omnidirectional Images	43
3.1	Spatial Filter Design	44
3.2	Homogeneous Optical Flow on Omnidirectional View	50
4	Visual Motion Detection	55
4.1	Egomotion Estimation	57
4.2	Visual Clues	62
4.3	Independent Motion Estimation	65
4.3.1	Outlier Interpretation	66

4.3.2	Particle Filter	70
4.4	3D-Omnidirectional View	76
4.4.1	Recent Technology for 3D Acquisition	78
4.4.2	Integrating 3D Measurement into the IME Scheme	81
5	Experimental Results	85
5.1	Experimental Setup	85
5.2	Visual Egomotion Detection	88
5.2.1	Estimation of translation direction in synthetical images	89
5.2.2	Egomotion Detection: Experiments with real images	90
5.2.3	Summary of the egomotion detection experiments	100
5.3	Independent Motion Detection	101
5.3.1	Segmentation of independently moving scene objects	101
5.3.2	Tracking independent motion	103
5.3.3	Summary of the independent motion detection experiments	108
6	Conclusion and Outlook	111
6.1	Summary	111
6.1.1	Strengths and Limitations	113
6.2	Outlook	114
A	Egomotion Estimation	121
B	Nonlinear Least-Squares Fitting	125

List of Figures

1.1	Optical flow field with focus of expansion.	10
1.2	Ambiguous optical flow	14
1.3	Task Description	17
2.1	Correspondence and Aperture Problem	22
2.2	OF Calculation scheme: Local Smoothness Assumption	27
2.3	Energy based methods	28
2.4	Dragonfly Speed-control	36
2.5	The compound eye	37
2.6	The mirror shape	41
2.7	The mapping of an omnidirectional image onto polar coordinates	42
3.1	A standard and spherically distorted picture	45
3.2	Gauss kernel	46
3.3	Variable spatial filters	47
3.4	The view sphere	48
3.5	Projections of a spatial filter patch	49
3.6	Variable spatial filtering	51
3.7	Comparison of resulting optical flow fields	52
3.8	Calculation Scheme	53
4.1	Environment and the active agent	56
4.2	Egomotion Detection	58
4.3	Optical flow and sphere projection	59
4.4	Qualitative analysis of spherical flow fields	60
4.5	Flow Field on the Sphere	61
4.6	Egomotion Calculation	64
4.7	Pixel displacement during the warping operation	67
4.8	Independent motion segmentation	69
4.9	Temporal Analysis of Residual Images	70
4.10	Resulting particle image	74

4.11	Resulting particle image 2	75
4.12	Resulting particle image 3	75
4.13	Binocular Stereo	77
4.14	3D-from-Motion	79
4.15	PMD Principle	81
4.16	Integration of 3D information	82
5.1	Experimental Setup	86
5.2	Software Structure	87
5.3	Omnidirectional flow field analysis	88
5.4	Flow field analysis in synthetical image sequences	89
5.5	Image from the pure translational motion sequence	91
5.6	Components of the pure translation	92
5.7	2D path of the purely translational motion	93
5.8	Purely rotational motion and flow distribution	94
5.9	The components of the purely rotational motion	95
5.10	Purely rotational motion path	95
5.11	The robot follows a curvilinear path in a corridor	97
5.12	Complex navigation path	98
5.13	Rotational component of the curvilinear motion	99
5.14	Experimental results of IM segmentation	102
5.15	Independent motion segmentation	103
5.16	Tracking Independent Motion 1	104
5.17	Tracking Independent Motion 2	105
5.18	Tracking Independent Motion Path	106
5.19	Tracking IM and Non-IM regions	107
5.20	Comparison of IM and non-IM Region Tracking	108
6.1	3D Omnivision: Pictures of the experimental setup	115
6.2	3D Omnivision: Experimental setup	116
6.3	Images captured from the 3D Omnicamera	118
6.4	Omnidirectional view from a rotated 3D camera	119
A.1	World and Camera Coordinate Systems	122

Chapter 1

Introduction

Many living organisms have senses to perceive **motion**, helping them to survive in and to cope with a dynamical, ever changing environment. Evolution provided beings with different senses for motion perception. With tactile senses, beings can perceive motions of objects that they touch. Using their hearing alone, barn owls, for instance, may locate moving, noise emitting prey in total darkness. But for many species, the main sense for perceiving motion is vision. With vision, it is possible to perceive objects or events that are too far away to be touched or heard. Not constrained to motion perception alone, vision yields an almost instantaneous, rich impression of the environment and its possible changes, allowing for fast reactions in critical situations.

Clearly, the ability to react fast in critical situations is also a desirable property of autonomous mobile robots (AMR). It is a prerequisite for the safe interaction of AMRs with a dynamic environment, that is, with living beings, mobile machines including other robots, or valuable artefacts. While, for instance, avoiding collisions between AMRs and stationary objects seems feasible in many situations, it is much harder to achieve when moving objects are involved. Thus, the AMR design should, among many other issues, consider the necessity of fast and reliable motion detection mechanisms. An AMR should be able to robustly estimate its own motion (egomotion) and the independent motion of objects in the environment.

One approach for retrieving the motion information needed for AMRs is *motion field analysis* or, to use a more popular, but somewhat too general term, the computation of *optical flow*. The latter notion was introduced in 1950 by psychologist James J. Gibson [52]. He modeled a class of perceptual phenomena that we experience everyday. One such phenomenon is *motion parallax*, named and investigated by von Helmholtz in the 1860s [125], but first mentioned as early as 300 B.C. in Euclid's *Optics* [37]. Motion parallax is the name of a perceptual effect caused by the lateral movement of an observer: stationary objects seem

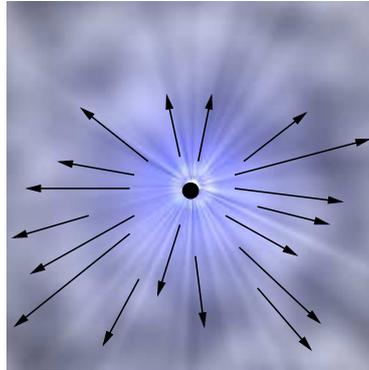


Figure 1.1: *Example of an optical flow field with focus of expansion.*

to move in the opposite direction of the observer's heading direction, but with different relative speeds. Distant objects seem to move slower than close objects. Apparently, the induced relative speed may be used as a depth cue.

Motion parallax is one aspect of a more general class of phenomena that are perceived by a moving observer and that have been investigated by Gibson. Moving about in an environment creates quite different visual impressions in the moving observer, depending on the relation between the heading direction and the view direction. If both coincide, then surfaces of stationary parts of the environment seem to move towards or to *flow* past the observer. Some surfaces flow out of the observer's field of view, and new, expanding surfaces seem to emerge from a certain imaginary point, the flow's *pole* or *focus of expansion* (Fig.(1.1)). When moving backwards and looking into the opposite of the heading direction, surfaces seem to flow into the observer's field of view, to shrink and to vanish finally in another imaginary point, the *focus of contraction*. The observed relative directional velocities of points on moving surfaces constitute the *optical flow field*.

The sensed 2D motion is a geometric projection of the actual 3D motion. Solely on the basis of the observed motion, it is impossible to always recover full information of the respective 3D motion. But it *is* possible to derive or estimate certain informations from the cues that the motion field provides: time-to-contact, egomotion direction, motion of independent objects, and more.

As we will see in the next section, there is strong support for the assumption that motion seeing is the most important visual perception mechanism, and that the cues provided by optical flow are indeed being used by living beings. Thus, respective results and models provided by the life sciences are a valid source of inspiration for the development of methods for handling similar tasks in robotics. There is a rich body of work describing how information gained from optical flow

computation (OFC) can be used in AMRs for solving tasks like collision avoidance, self localization, navigation, and others. However, there are inherent problems and principle limitations of the approach.

In the following sections, the importance of motion seeing for living beings and artificial systems will be discussed. The inherent problems of artificial motion seeing mechanisms will be explained in Section 1.2, and a possible solution to the problem of reliable motion estimation will be introduced.

1.1 On Motion Seeing in Living Beings

Do living beings equipped with visual senses really see motion, and, if so, do they make use of the mentioned cues? These questions have been extensively investigated in experimental psychology, neurobiology and related disciplines for several decades. A part of this research was – and is – devoted to the neural basis of vision. In 1961, Reichardt [94] created the first mathematical model of the local motion detection mechanism of flies, the *elementary motion detectors*, that gained wide acceptance. A year later, Hubel et al. [61] were the first to discover neurons in the visual cortex of cats that respond selectively to oriented patterns of light. Similar neurons and directionally selective motion sensitive neurons have been discovered in other species, too [3]. There is evidence that several areas of the visual cortex are involved in motion perception (e.g., V1: orientation, direction and disparity, MT (V5): direction and disparity, MST: optic flow [35; 17] and navigation; cf.[103] for a short overview). The complicated interplay between these areas and their forward and backward connections is not yet fully understood. But motion perception seems to be the most important visual ability, because – as Nakayama [81](in p. 627) pointed out – “No animals have been found that lack mechanisms for motion processing.”

Experimental psychology provides evidence that living beings make use of the cues provided by a motion field. An excellent overview of respective research may be found in [103]. Here, we will give just a very brief summary. Warren and colleagues [38; 128] showed that, if the environment provides sufficient structure, human locomotion control is dominated by optical flow, and that optical flow supports the perception of heading [127]. This is consistent with the neurobiological findings of Bradley et al. [17]. Humans use several different cues to estimate the time to collision [100]. One such cue is the ratio between the retinal image size and the rate of change in that image size. Lee [71] introduced the variable *tau* to name this ratio. *Tau* can be derived from optical flow, and its reciprocal is the time to collision. In a recent investigation, Redlick et al. [93] stated that humans use optical flow to estimate the distance of travel. Last but not least,

there is evidence that optical flow supports the perception of object structure, including shape and spatial arrangement. The respective phenomena are known as *kinetic form perception*, including perception of non-rigid *biological motion*, and *kinetic depth effect* [126]. In modern works, both are subsumed under the notion of *structure from motion*.

1.2 Robotics and Visual Motion Detection

Robotics can be a biomimetic way of understanding biological systems. The other way around, biological systems can be an inspiration source for technical systems, like robots. Specifically, the question if and to which extent can the visual competency of animates and other beings be understood to realize technical systems, which can “see”, has been an interesting one for the researchers. In [34], the authors claim that the behavior based robotics could profit in many ways from the ecological approach to vision, introduced by Gibson.

Following Reichardt’s pioneering work, several models of motion detectors have been proposed. In the last decade, some of them have been implemented in hardware for experimental evaluation. A special issue of “Robotics and Autonomous Systems” is devoted to this subject (cf. [26; 47]). But, for reasons of flexibility, the majority of algorithms for motion analysis for use in AMR CPUs is implemented in software.

In visual servoing tasks, computer vision systems are used to extract sufficient information to control the position of the robot’s end-effector relative to the environment. Visual servoing may include various task components like the detection of moving objects [48], target tracking [67] and egomotion detection [75]. These tasks require sufficient motion sensing mechanisms. The design of a visual system is beneficial for robot systems like the autonomous driving robots, too, as was proposed by Giachetti et al.[51]. Tracking systems and specific tasks like hand-eye control, path finding and object avoidance are only some examples of the whole field of research considering robotics and visual perception.

Algorithms for visual motion detection generally consider sequences of images and evaluate the brightness changes in the consequent pixels of two or more images of a sequence. Basically, it is aimed at determining the direction and the magnitude of the relative velocities between the camera and the scene objects. In the respective literature, it is stated that the motion field can be approximately recovered by means of optical flow computation techniques if the image sequence fulfills some constraints that assure the spatial variations of the brightness correspond to physical features on the 3D surfaces [124]. A contrary example to this is the projection of a smooth surfaced ball illuminated with a changing light intensity

and direction, where one may detect a flow field, which represents the changes in the illumination rather than a motion.

As far as robot vision is concerned, motion detection algorithms have to consider additional constraints, which were also highlighted by Camus [20]. Beyond the accuracy of the computed motion field, there has to be the computation time and load considered. Some optical flow methods fail to give satisfactory results on real robotic applications due to their high computing time requirements. Also the methods have to cope with miscellaneous perturbations of the images like sudden changes in the illumination conditions.

In an attempt to speed up the real time calculations in motion detection tasks, alternative approaches using specific hardware have been proposed. Röwekamp used a special flow sensor for real-time optical flow calculation for depth and time-to-contact estimation on a mobile robot [96]. Also Stocker et al. [112] report a hardware realization of optical flow calculations. Special hardware implementations of the desired visual system (in this case, optical flow calculation), reduce the high computational demand, but introduce some dependency on the given task or on the limitations of the hardware realization.

In some applications, additional information from the task or the environment in which the robot operates, is combined to achieve more reliable motion field information. For example Neven et al. [87] refer to a sensor fusion system, which combines other available sensor data with the visual measurements. Ohya et al. [88] have used vision and ultrasonic sensors for static and mobile object detection.

1.2.1 On the advantages of a wide-angle field of view

Obtaining exact quantitative solutions to the problem of determining structure and 3D motion from optical flow under a planar perspective projection requires precise knowledge of the projective transformation (camera) parameters, non-trivial mathematical methods and nonlinear transformations. The inherent noise in the optical flow field makes this task even harder to accomplish.

Considering this problem in a more qualitative way, one can make use of the structural properties of optical flow (OF). Basically, OF can be decomposed into two independent components, translational and rotational. Pradzny [91] pointed out that only the translational component of the motion field contains information about the local surface orientation, or relative depth. The effects of the rotational motion are equivalent to geometric transformations of the translational motion field. Thus it is possible to derotate the projection of the motion field without any additional knowledge about the 3D structure of the scene. Determining the translational component of the motion, however, requires additional knowledge about the 3D structure or the magnitude of the egomotion.

The projected paths generated by rotational motion of the scene elements are qualitatively different from the straight paths generated by the translation. Thereafter, this fact can be exploited to determine the egomotion direction of a camera-robot system. The singular points of the derotated optical flow field yield this information.

The major problem with this proposal is that, in general, there are at least two possible interpretations of a perspective projected motion field. For example, a perspective camera on a robot looking along the x-axis, and rotating about the y-axis (ω_y) produces a flow field which is similar to that by translation along the z-axis (t_z) in a homogeneously structured environment (Fig.(1.2)). Intuitively, although the optical flow fields produced by different camera motions may match closely over restricted fields of view, they cannot do so globally. Thus an efficient way to solve the ambiguity of the interpretation of the optical flow field to achieve a 3D motion field is to widen the topological constraints of the OF by using wide-angle cameras.

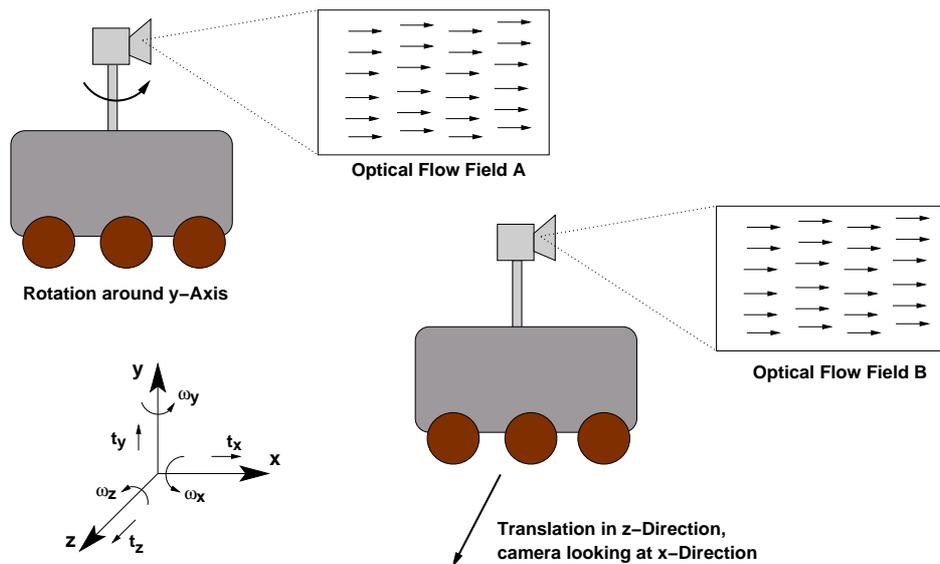


Figure 1.2: *Ambiguous motion field induced by two different types of robot ego-motion. In the first situation (above), the robot rotates around its y axis (ω_y). The limited field of view of the camera captures a motion field which is very similar to the second (below) situation, where the robot translates in z direction of a homogeneously structured environment (t_z), capturing images as it looks in the x direction.*

1.3 Objectives of the thesis

Gibson [52] points out that the perceptive structure of a changing visual scene is feasible for specifying the invariant structure of an environment. Following this statement, this thesis seeks answers to the questions:

- To which extent can the dynamical image measurements of a mobile camera-robot system and the optical flow be useful for gaining navigational information, which is relevant for a secure interaction of the robot with its environment?
- Can the inherent structural features in omnidirectional motion fields allow for efficient motion detection, specifically for tracking independent motion?

Several solutions exist to solve the egomotion detection problem; for example using sparse point matchings, tracking and subspace methods (e.g. [121], [64], further, see [55] for a comprehensive review), but the dynamic environments, in which a robot typically navigates, complicates the application of such methods and ambiguates the motion estimations. Point occlusions due to other moving objects in the scene or projective deformations, are the major sources of errors. On the other hand, the topological structure of a dense optical flow field reveals essential information for the egomotion detection. The optical flow field induced by a translational motion, for example, includes singular points, indicating the heading direction of the camera-robot system. As mentioned in Section 1.2.1, the interpretation of the OF field is only unambiguous and distinct, if the scene is projected with a wide-angle sensor.

In the thesis, the omnidirectional images are captured by a catadioptric sensor, which approximates spherical projections of the real indoor scenes. The egomotion direction of the robot, with translational and rotational components, is detected by analyzing the induced patterns of the omnidirectional optical flow field. Although the fact that the egomotion causes specific patterns on the omnidirectional motion fields has been mentioned before (see e.g. [85]), this thesis introduces, to our knowledge, the results of the first applications which employ this topological information for real indoor robot navigation.

The major goal of the thesis is to determine the independent motion, that is frequently present in dynamical scenes. This problem has been tackled in the past by several researchers, but it still remains unsolved so far. Most of the proposed methods makes use of the background homogeneity of the specific scenes to segment the independent motion from the global motion field [89], [62]. This thesis shows that the omnidirectional imaging allows for a more general solution to the problem. In omnidirectional projections, the background motion forms specific

topological structures on the optical flow field, which allows for the constitution of a relevant background motion model. The detection of the flow outliers or the residue of the background model may then yield to the independent motion segmentation, directly. Also, in omnidirectional images, it is possible to track the moving objects of the scene for a longer time, which facilitates the detection of outliers in the temporal domain.

The inherent depth parallax problem in independent motion detection tasks stems from the fact that the 3D structure of the scenes induces similar changes on the optical flow field as the independent motion does. Depth parallax problem cannot be solved from the direct analysis of optical flow fields, since it is defined only on the 2D plane. The temporal analysis proposed above to solve this problem, can be enhanced by using additional sensory equipment, which measures the 3D structure of the scene directly. In the thesis, we introduce the first experiments with a fast 3D camera¹, which we have modified to capture a wide field of view.

Summarizing the objectives of the thesis, we aim at solving the problem of visual independent motion detection in interdependent subtasks, each of which is solved in a novel approach. These subtasks are depicted in Fig.(1.3). The task of *Independent Motion Detection* can only be accomplished if the egomotion detection yields a relevant background motion model. Therefore, the problem of *Egomotion Detection* should be solved first. As mentioned above, the thesis exploits the topological structures of the global omnidirectional optical flow fields to detect egomotion. The estimation of the optical flow field is, therefore, the prerequisite of this subtask. Since the omnidirectional projection introduces non-affine deformations on the image plane, the known optical flow calculation methods have to be modified to yield accurate results. This modification is introduced here as a subtask, *Omnidirectional Optical Flow Estimation*. The experiments concerning the 3D omnidirectional scene capturing are grouped under the subtask *3D Omni-Image Processing*.

1.3.1 Contributions

The thesis introduces a novel solution to the problem of visual independent motion detection by interpreting the topological features of omnidirectional dense optical flow field and determining the background - egomotion direction. Several interdependent tasks have been solved using novel techniques or modified standard methods, which make up the contributions of the thesis:

¹PMD-Camera: a novel camera setup for the parallel 3D-scanning of the scene in real time, - developed by Prof. Schwarte, University of Siegen- was kindly provided by the company PMD-Tech, Siegen-Germany, and used for the experiments.

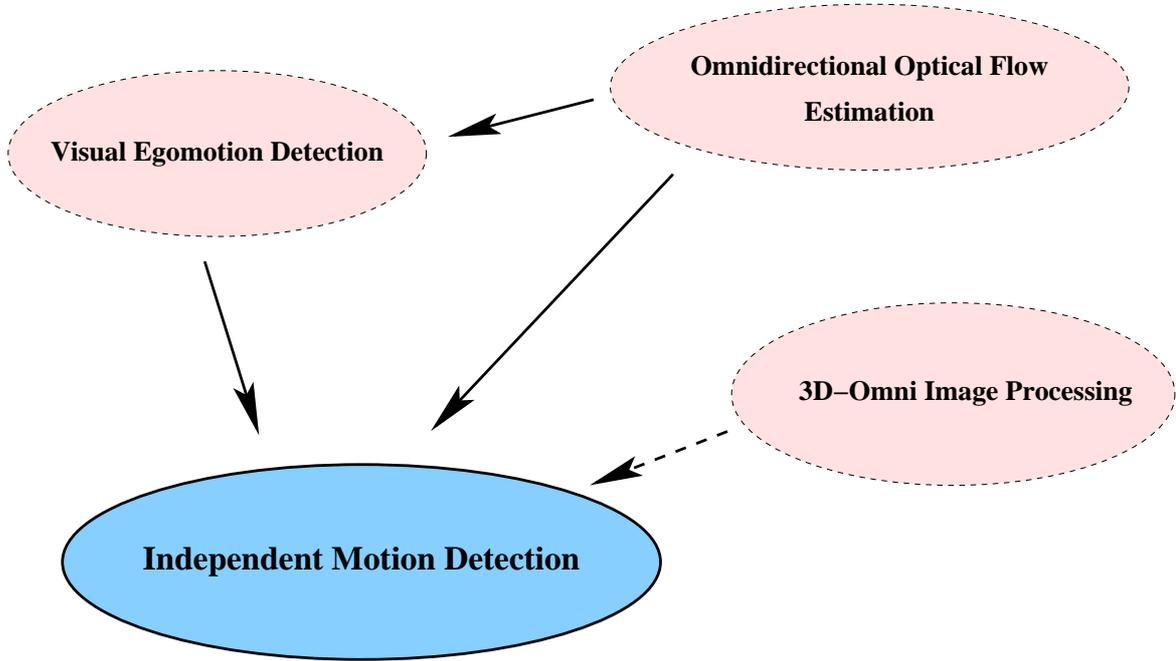


Figure 1.3: *The subtasks of the thesis with their interdependencies*

1. Our comparative evaluation of the known optical flow calculation methods showed that the local gradient method, known as the Lucas - Kanade Method, is appropriate for time efficient omnidirectional flow estimation but requires modifications to cope with the specific image distortion. The thesis introduces a time efficient modification to the estimation method using separable filter kernels that can be mapped onto the view sphere. The evaluation of the known methods has been published in [113].
2. The results of egomotion detection by interpreting the global omnidirectional flow field have shown that the egomotion direction of the robot can be determined visually as published in [114]. Additionally, it is possible to enhance this method by including other gyrodometric measurements in a Kalman-Scheme as has been theoretically elaborated and published in [116] and [115].
3. The thesis introduces an independent motion segmentation method using warping and elimination of visually estimated background model. This has been published in [114], also.
4. A novel interpretation of the independent motion segmentation problem

combines the segmentation with a probability based tracking, namely using a particle filter. The thesis, for the first time, makes use of the spatial widening of the field of view that is introduced by omnidirectional imaging, to track the independently moving scene objects. This allows for a more reliable and robust tracking than the standard tracking systems since the objects remain longer in the scene and since it is possible to model the background motion using the topological features of the catadioptric projections.

1.3.2 Structure of the Thesis

The thesis is organized as follows:

The next chapter (Chapter 2) reviews the basics of optical flow calculation, and points to the inherent problems of motion estimation in general. Also, a brief review of motion recovery methods and application examples in robotics will be given. In Section 2.6, the omnidirectional image processing in natural and artificial systems will be introduced. An efficient way of achieving all-round view, namely using catadioptric sensors, will be elucidated. Since the omnidirectional view captured with the catadioptrical sensors bears projective deformation, the image processing, specifically, the calculation of optical flow, has to be modified in order to yield reliable and accurate results. The proposed modifications and the results will be introduced in Chapter 3.

The visual egomotion detection and independent motion tracking will be reviewed in Chapter 4. The proposed optical flow field analysis to obtain egomotion information of a mobile camera-robot system will be described in Section 4.1. Then, a detailed explanation of the proposed independent motion segmentation technique and the probabilistic tracking based on the particle filter scheme will follow in Section 4.3. The methods for gaining the 3D information will be summarized in Section 4.4 and a novel technique of capturing the depth scene using a 3D camera will be introduced.

In Chapter 5, the experimental results of indoor robot navigation employing the methods will be described and the results will be discussed.

Chapter 6 will summarize the thesis and discuss the open issues.

Additionally, the Appendices A and B describe the known egomotion estimation scheme and the nonlinear minimization technique "Levenberg-Marquardt" in detail to support the discussions in Chapter 4.

Chapter 2

State of the Art: Estimating Optical Flow for Motion Recovery and Basics of Catadioptrical Vision

Robotic systems may use the visual data of sequences of 2D raster images supplied by a camera setup to estimate motion information qualitatively or quantitatively. Technically, this can be achieved by tracing certain features across a number of successive images and analyze their spatio-temporal changes at discrete points in time and space. In principle, spatially extended features like edges, labelled object surfaces or lines are suited for this task. But most of the optical flow calculation (OFC) algorithms utilize a more elementary feature, the pixel intensity value, which does not require preprocessing in order to be retrieved.

The basic steps of OFC can be summarized as follows: The projection of the relative motions of a camera system and the objects in a 3D scene onto the image plane introduces a 2D vector field, called the motion field. This field is estimated by measuring the changes in the images features, e.g. the pixel intensity variations, and by using one of the estimation techniques reviewed below. The result of this computation step is a directional vector field in the image domain, called the optical flow field. The accuracy of the estimation influences the quality of the recovered motion information.

Several problems make it difficult to reliably and robustly estimate the motion field. These problems may be divided into three categories: problems due to the chosen features, problems inherent to OFC, and problems due to the assumptions made by various estimation techniques. This chapter will briefly describe these problem categories, introduce the most important OFC techniques, and describe

how they deal with specific problems. For many applications, the recovered 2D motion contains sufficient information for the task at hand. However, applications that need to estimate the original 3D motion require additional procedures, e.g. depth estimation, self-localization, and/or 3D reconstruction. Examples thereof will be given in Section 2.5. Section 2.6 reviews the advantages of omnidirectional vision, especially for mobile robotics. The section also introduces the techniques for capturing all-round view using catadioptrical sensors.

2.1 Basics of Optical Flow Computation

A basic assumption of OFC techniques is that a change in the intensity of one and the same pixel between two consecutive frames is only caused by motion, i.e. illumination does not change significantly across the image sequence under consideration. This assumption is called the *brightness conservation constraint (BCC)* [59]. Clearly, in real-world situations illumination may often change, hence the presented OFC techniques account for such changes in different ways. Methods that deal with brightness variations include the equalization of the pixel intensities of given features across image sequences, e.g. by successively adapting the histograms [27], employment of suitable temporal filters, and explicit modeling of brightness variations [56; 110; 15].

Assuming ideal contrast conditions, and if $I(x, y, t)$ represents the image intensity function of a 2D image at a pixel position (x, y) at a time t , the BCC can be written as:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t), \quad (2.1)$$

where δx and δy are the spatial displacement of the local image region after time δt . Expanding the right-hand side of this equation in a Taylor series yields

$$I(x, y, t) = I(x, y, t) + \frac{\partial I(x, y, t)}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial I(x, y, t)}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial I(x, y, t)}{\partial t} + \dots, \quad (2.2)$$

where \dots represents the higher order derivatives, which can be ignored for the simplicity of the calculations. The terms $\frac{\partial x}{\partial t}$ and $\frac{\partial y}{\partial t}$ represent the image velocity in x and y directions, and will be represented with \mathbf{v}_x and \mathbf{v}_y in the following equations. The image velocity is then written in vector form: $\vec{\mathbf{v}} = (\mathbf{v}_x, \mathbf{v}_y)$.

Introducing the representation $\nabla I = \left(\frac{\partial I(x, y, t)}{\partial x}, \frac{\partial I(x, y, t)}{\partial y} \right)$ for the spatial gradients and ignoring the higher order terms in Eqn.(2.2), we can rewrite the

BCC:

$$\nabla I \cdot \vec{v} + I_t = 0, \quad (2.3)$$

where I_t represents the temporal gradient $\frac{\partial I(x, y, t)}{\partial t}$.

Eqn.(2.3) defines a single local constraint on image motion. Thereafter, it is not sufficient to compute both components of $\vec{v} = (\mathbf{v}_x, \mathbf{v}_y)$. Only the velocity vector in the perpendicular direction of the gradient, represented here with \vec{v}_n , referred as the *Normal Flow*, can be calculated by solving the following equation:

$$\vec{v}_n = -\frac{I_t \nabla I}{\|\nabla I\|^2} \quad (2.4)$$

Normal flow can provide qualitative information about the motion field. Aloimonos et al. ([4]) propose to use this normal flow field to constrain the search area for the actual velocity vectors of the images. Santos-Victor et al. [99] applied the information supplied by the normal flow field to construct a model of a motion field for obstacle detection by a moving robot system.

Problems inherent to OFC may arise both from partial occlusion or exposition of objects and from objects disappearing from or appearing in the scene, making it difficult to establish a one-to-one correspondence between the optical flow field and the actual 2D motion field.

Generally, matching two locations in successive image frames requires specifying distinct features which distinguish these locations from others. If the chosen features are not unique, i.e. if more than one location in the second frame is sufficiently similar to the location in the first frame, then there is a *correspondence problem*, as depicted in Fig.(2.1(a)). Images with little structure and smooth flat surfaces with constant intensity may cause this kind of ambiguity in real image sequences.

A special case of the correspondence problem is the *aperture problem*, which is caused by the limitation of the field of view. Object surfaces that are larger than the dimensions of the aperture can only be partially perceived. If, in addition, the features of the observed surfaces are not unique, then ambiguous matches may be obtained again. Fig.(2.1(b)) shows a line that changes from t_0 to t_1 in two successive image frames. The normal velocity vector, labeled with \mathbf{v}_n , can be derived from the intensity gradient of the signal, as shown above, but the actual 2D motion vectors of the line remain ambiguous because the points on the line instances t_0 and t_1 have indistinguishable features (same color and intensity). Eqn.(2.4) describes this problem mathematically and allows for a unique solution, although the correct point correspondences may not be known, i.e., \vec{v}_n is unique and known, whereas \vec{v} is not.

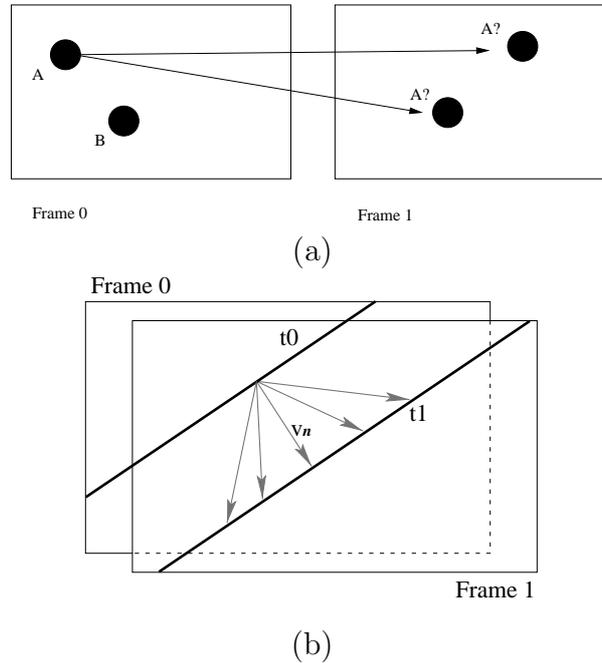


Figure 2.1: (a) *The correspondence problem* (b) *The aperture problem*

Additionally, it should be noted that numerical differentiation by convolving the image with a small derivative filter kernel, as is suggested in differential techniques, causes a special local aperture problem. This type of differentiation process restricts the calculation to a local neighborhood that is defined by the kernel. The boundaries of the neighborhood form an aperture that is even smaller than the global aperture formed by the image boundaries.

Summarizing the above considerations, the actual 2D motion components \mathbf{v}_x and \mathbf{v}_y can be calculated by either using a one-to-one pixel matching method that defines the pixel features uniquely and solves the correspondence problem, or by using additional assumptions derived from the supplementary knowledge of the task and the environment.

Except for the matching techniques, all other OFC classes require additional assumptions for calculating the actual velocity vectors. Matching methods have to cope with the correspondence problem, which presumes finding the best feature that allows for a distinct description of the pixels to be matched.

The modeling of the expected motion may provide additional constraints. For example, if the environment is static and the task is to estimate motion in images taken with a camera mounted on a robot arm that performs only one-dimensional

translations, then it is valid to assume that the observable motion is globally smooth. Therefore, the neighboring pixels in the images are likely to expose similar velocities. Here, a global smoothness assumption can be used as an additional constraint for the solution of the data conservation assumption. This means that two unknown velocity components \mathbf{v}_x and \mathbf{v}_y can be computed using two constraint equations.

The OFC methods described in the literature, that we have evaluated, generally differ in their approaches of handling the correspondence and aperture problems. Some authors suggest alternative calculation techniques for robust or fast calculations of the optical flow field from the given image sequences. In the following section, a review is given to cover these methods by grouping them according to their different 2D velocity estimation strategies.

2.2 Methods for Computing Optical Flow

2.2.1 Matching Techniques

In these approaches, corresponding image points (pixels) of consecutive frames with similar features are matched to get a displacement measure. Usually, the matching criteria are the similarity of the intensity values of the pixels and the proximity of their locations. If the time lags between the image frames are known, the relative velocity of the objects and the image plane can be estimated using the computed displacement measure. Finding the best match between the corresponding pixels in the image sequences leads to the problem of minimizing a difference measure such as a cross-correlation of the image intensity functions of successive image frames of pixel dimensions $N \times N$. Cross-correlation of the brightness functions of the first image $I_1(x, y)$ and the second image $I_2(x + d_x, y + d_y)$, where the pixels are moved with a distance (d_x, d_y) , can be written as follows:

$$\sum_{i,j=-N/2}^{N/2} I_1(x+i, y+j)I_2(x+i+d_x, y+j+d_y) \quad (2.5)$$

Note that the above equation assumes similar pixel displacements (d_x, d_y) over the whole image. It is possible to assume a regional similarity by applying the cross correlation function over the small regions of the image plane consecutively. Those similarity regions can also assumed to be of varying dimensions.

Other distance similarity functions such as *sum of squared differences* (SSD) can also be used for the minimization as demonstrated by [107]. For example, Konrad et al. [68] propose a probabilistic method, maximizing a similarity function

based on the *a posteriori* probabilities of displacement vectors, given the image intensity changes. A review of matching techniques using other distance similarity functions is given by Giachetti [50].

As indicated above, matching techniques do not require additional assumptions to cope with the special aperture problem introduced by numerical differentiation, but still they have to cope with the correspondence problem in general. Mathematically, this can be described as finding a global minimum (or maximum, if correlations are used) of the similarity function. The computational complexity of such a search is $O(N^4)$ for images with dimensions $N \times N$.

Some approaches estimate the matching points in resolution prisms of different levels (e.g. [5]), and these tend to be suitable for parallel processing. One should note that a strict localization of the search window may result in ambiguous correspondences.

2.2.2 Differential Techniques

High computational complexity of the matching techniques leads to considering alternative techniques. The differential (or gradient) methods avoid the difficult task of finding distinct features to match. Instead they rely on the calculation of gradients of the image intensity, which provide clues about the relative motion projected onto the images. These methods assume that the time lags between the frames are small enough to capture the expected motion in the scene, which assures the differentiability of the image intensity signal. The velocity vector components minimize the *gradient constraint* as suggested by Horn and Schunck [59].

$$E_{gc} = \frac{\partial I}{\partial x} \underbrace{\frac{dx}{dt}}_{\mathbf{v}_x} + \frac{\partial I}{\partial y} \underbrace{\frac{dy}{dt}}_{\mathbf{v}_y} + \frac{\partial I}{\partial t} = 0 \quad (2.6)$$

Note that the error function E_{gc} rewrites the BCC (Eqn.(2.3)), representing the main assumption that the image intensity of a particular point in a pattern is constant over time and the intensity changes are only related to the motion, here indicated by the underbraced \mathbf{v}_x and \mathbf{v}_y components. This assumption constrains the estimation such that only the image velocity component in the direction of the image gradient (*normal velocity*), $\tilde{\mathbf{v}}_n$, can be calculated as follows:

$$\tilde{\mathbf{v}}_n = -\frac{\partial I}{\partial t} \cdot \frac{1}{\sqrt{\frac{\partial I^2}{\partial x} + \frac{\partial I^2}{\partial y}}} \quad (2.7)$$

As we have mentioned above, spatial differentiation (mostly based on four point differences of a particular pixel on the image) is a strictly local analysis, also

limited by the size of the filter kernel used for the calculation. This introduces an aperture problem that requires making additional assumptions to calculate the actual velocity components. Different such assumptions correspond to different estimation techniques, such as:

Global smoothness. Assuming opaque objects undergoing rigid motion, these techniques minimize an error function based upon the gradient constraint introduced above (Eqn.(2.6)). In this case the method integrates a smoothness constraint assuming that the neighboring points on the objects have similar velocities and the optical flow field varies smoothly almost everywhere on the image. This global smoothness assumption postulates smoothness for the spatial variation of the optical flow field and the local relationship of the pixel intensity variation, together. The latter one can be formulated as the minimization of the Laplacian of the optical flow vectors $\vec{v}(\mathbf{v}_x, \mathbf{v}_y)$, which is written as follows:

$$\nabla^2 \mathbf{v}_x = \frac{\partial^2 \mathbf{v}_x}{\partial x^2} + \frac{\partial^2 \mathbf{v}_x}{\partial y^2} \quad (2.8)$$

$$\nabla^2 \mathbf{v}_y = \frac{\partial^2 \mathbf{v}_y}{\partial x^2} + \frac{\partial^2 \mathbf{v}_y}{\partial y^2} \quad (2.9)$$

Then the total error function writes from combining BCC and the new smoothness constraint using the L2 -Norm:

$$E_{total}^2 = \int \int \left(\frac{\partial I}{\partial x} \mathbf{v}_x + \frac{\partial I}{\partial y} \mathbf{v}_y + \frac{\partial I}{\partial t} \right)^2 + \alpha^2 (\nabla^2 \mathbf{v}_x + \nabla^2 \mathbf{v}_y) \quad (2.10)$$

where the term α^2 reflects the influence of the local smoothness term in the global minimization task. The higher values of α^2 takes the local smoothness of the intensity variation into account, whereas the smaller values give more importance to the global smoothness of the image velocity. In [59], the minimization of this total error function is estimated iteratively and over the whole image, to yield the components of the optical flow vector, \mathbf{v}_x and \mathbf{v}_y .

Oriented smoothness. Nagel [80] has proposed a *directional smoothness assumption* in an attempt to eliminate the drawbacks of the global smoothness method in the cases of object occlusions and spatial discontinuities in the images. The corresponding constraint equation uses higher order derivatives of the BCC, including the partial derivatives of the velocity estimates. A solution is computed by an iterative minimization of the constraint equation.

Local smoothness. Some methods assume only local constancy or linearity of the image intensity variations. This local smoothness assumption constrains the BCC for a particular pixel position and its local region and allows for constructing a system of linear equations. This is achieved by fitting the gradient measurements in each chosen neighborhood to a local model for 2D velocity, mostly using least-squares minimization. The error function, E , is constituted by modifying and spatially limiting the BCC with a local windowing function $\omega^2(\mathbf{x})$:

$$E = \sum_{\mathbf{x} \in \Omega} \omega^2(\mathbf{x}) [\nabla I(\mathbf{x}) \cdot \mathbf{v} + I_t(\mathbf{x})]^2 \quad (2.11)$$

where the weights $\omega(\mathbf{x})$ in Ω are set to give more importance to the center of the window than to its periphery. ∇I represents the spatial gradient operation on the image I . I_t represents the temporal gradient and $\mathbf{v} = [\mathbf{v}_x \mathbf{v}_y]^T$ is the velocity vector. The solution to this minimization task,

$$\frac{\partial E}{\partial \mathbf{v}} = 0 \quad (2.12)$$

leads to a linear system of equations, which can be formulated with $W\mathbf{A}\mathbf{v} = W\mathbf{b}$, where:

$$\begin{aligned} \mathbf{A} &= [\nabla I] \\ W &= \text{diag}[w(\mathbf{x})] \\ \mathbf{b} &= -[I_t] \end{aligned}$$

The closed-form solution for the velocity \mathbf{v} of one image point writes:

$$\mathbf{v} = [\mathbf{A}^T W^2 \mathbf{A}]^{-1} \mathbf{A}^T W^2 \mathbf{b} \quad (2.13)$$

when the matrix $[\mathbf{A}^T W^2 \mathbf{A}]$ is nonsingular. This can be written explicitly:

$$[\mathbf{A}^T W^2 \mathbf{A}] = \begin{bmatrix} \sum_{\Omega} \omega^2 I_x^2 & \sum_{\Omega} \omega^2 I_x I_y \\ \sum_{\Omega} \omega^2 I_x I_y & \sum_{\Omega} \omega^2 I_y^2 \end{bmatrix} \quad (2.14)$$

where I_x and I_y represent the spatial gradients of the image point. The sums are taken over the neighboring region of the image point (over the spatial support Ω), for which the velocity vector is calculated. The spatial weighting parameter ω can be represented as a function of space, $\omega(x, y)$ and be defined by a 2D Gauss function. This assures the maximum contribution of the actual image pixel intensity for the estimation process and the neighboring pixels are considered with a spatially decreasing importance.

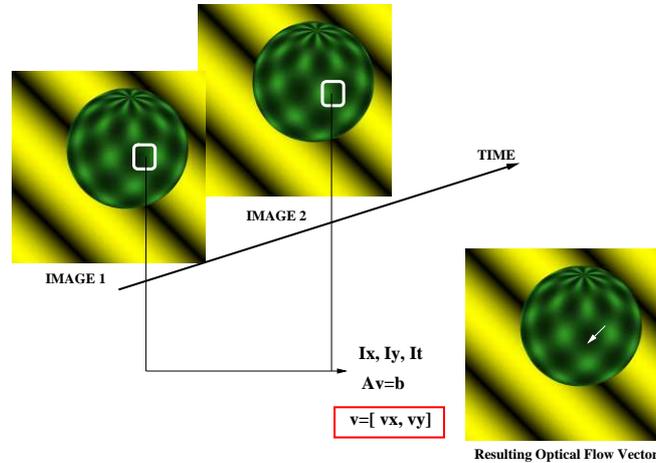


Figure 2.2: Calculation scheme of one optical flow vector using the local smoothness assumption. The sequential images of a circulating sphere are captured. The small neighborhood of one pixel on the sphere pattern has been assumed to have almost linear motion. The spatial and temporal gradients are estimated for this region, and the linear system of equations is solved to yield the optical flow vector.

In [65], it has been noted that the differential OFC methods yield inaccurate results when the images include motion boundaries. This is more severe for global methods, since they propagate the estimation errors throughout the image, whereas by the local methods the error propagation remains limited to the chosen estimation window.

2.2.3 Frequency-based Techniques

Energy-Based Methods

Energy-based methods view the 3D motion projected onto the image sequences in a space-time corridor. Treating the image sequence as a space-time signal, the velocity of the motion can be defined as orientation in this 3D volume. Adelson et al. [1] propose to determine the image motion from the output energy of velocity-tuned filters in the Fourier domain. An example is shown in Fig.(2.3). A static world point P is projected onto three consecutive images taken from a camera while it moves horizontally. The projections of P are seen at different locations A , B and C in three consecutive images. Viewing these images in a time-space cube, P will be mapped onto a slanted line in the $x-t$ space slice. The slope of this line yields the velocity of the relative motion. If the motion of the camera or

the points in the 3D scene is complex, then the contents of the time-space slices becomes complex, too. Filtering out a particular slant in the Fourier domain means filtering for the corresponding specific velocity of the projected 3D motion.

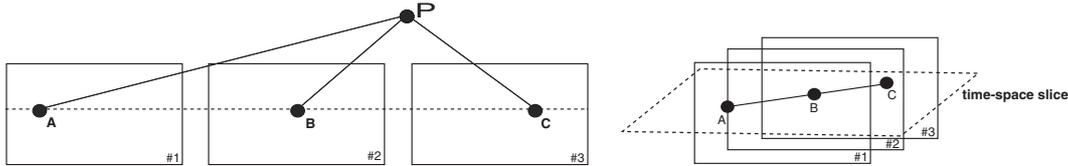


Figure 2.3: *Projections of the point P onto three consecutive images at locations A , B , and C (a) and their view in a time-space slice (b)*

The responses of quadrature pairs of filters¹, which detect the energy of signals in different directions in the time-space slices, are squared and summed up to extract an energy response in the search directions. Heeger et al. [57] have used Gabor energy filters to extract the local energy of the edge motion. These filters can be tuned to return a maximum amplitude response for a given spatial and temporal velocity. The authors proposed an algorithm that can be performed using parallel distributed processing. Simoncelli [106] points that this technique is equivalent to the gradient-based method proposed by Lucas and Kanade [74].

Phase-Based Methods

Phase-based methods define the velocity in terms of the phase response of velocity-tuned filters. The image velocity gets determined in two steps; calculation of velocity in the gradient's direction and interpretation of this field using assumptions about smoothness. In [42], the normal velocity is calculated using the phase responses of various Gabor filters. These filters are orientation selective, yielding information on local line and edge positions and directions. It should be noted that this extensive filtering depending on the chosen sensitivity of the detection (velocity resolution) yields high computational loads in most of the applications.

¹A quadrature filter is zero over one half of the Fourier space. In other words two filter functions are in quadrature when they are each others Hilbert transform. In the spatial domain, the filter is complex: an even real part and an odd imaginary part. In two dimensions, the even part can be seen as a line filter, and the odd part as an edge filter [131]

2.3 Comparative Evaluations and Analyses of OFC Techniques

A number of papers provide more detailed introductions to OFC methods [12; 8; 111] and comparative performance analyses of various OFC techniques [49]. Comprehensive in-depth analyses and comparisons of OFC methods are given in [9] and [10], the most prominent respective works. These authors found out that the phase-based methods are superior as far as accuracy is concerned.

A basic problem of the comparison of OFC techniques is the fact that for real image sequences the ideal optical flow field is unknown, i.e. there is no *ground truth* in these image sequences, and hence there is no real basis for absolute error and performance measurements. As a consequence, some authors have proposed to use synthetic image sequences as a basis for comparisons and performance analyses [9]. In synthetic image sequences, the crucial illumination can be controlled, problematic occlusions can be avoided, and the ideal optical flow field is known by construction, enabling precise error measurements for the techniques under consideration.

Liu et al. [72] compared the efficiency in terms of speed, flexibility and robustness of different OFC methods and reported that energy- and phase-based methods have a major drawback when speed is concerned. Clearly, for real-time robotic applications the choice of efficient optical flow methods is crucial, so energy- and phase-based methods might not be the first choice. Fleury et al. [43] assess the suitability of OFC techniques for implementation on parallel computers. They have implemented correlation (matching), gradient-based, and phase-based algorithms on a modestly parallel machine. All algorithms exposed speed-ups, with matching techniques in a clear lead.

Accurate and efficient calculation of an optical flow field remains to be a problem despite of the many methods proposed so far. The reason is that the calculations depend on measurements which contain almost always some kind of noise and that there are always cases where the assumptions of the respective methods are violated. Verri et al. [124] argue that meaningful information on actual motion can only be gained from the qualitative analysis of the flow field. Fermueller et al. [40] review OFC methods with respect to accuracy and conclude that the noise in the data causes a bias, i.e. a tendency of flow calculations to underestimate not only the lengths, but also the directions of the flow vectors. They also report that this kind of inaccuracy can be experienced by the human visual system in the context of visual illusions like the Ouchi pattern. They propose that the motion detection algorithms should consider other visual clues additional to optical flow.

2.4 Recent Advances in OFC Methods

In recent years, a number of authors investigated theoretical aspects of OFC and motion recovery problems. Beauchemin [11] analyzed the aperture problem and showed that the information content of an occlusion almost always disambiguates the velocity of an occluding signal suffering from the aperture problem.

Negahdaripour [83] introduces a revised and more general definition of OF by integrating both geometric and radiometric transformations of image brightness patterns, resulting in the generalized dynamic imaging model (GDIM) and a generalized aperture problem. The author showed that the GDIM provides a framework for describing several OFC methods. Modifications of the given classical methods for computationally efficient and robust calculations have been proposed by different authors. The term robustness in these cases means reduced sensitivity to the input noise and flexibility in different applications with varying environmental conditions. Such varying condition might be occluding edges, as well as fast intensity changes due to sudden appearance or disappearance of objects. The solution to those problems remains an open issue, but usually the approach is to use implementation specific methods and supplementary visual information, that will robustify the calculation.

Log-polar pixel representation: A space variant pixel representation that has a maximum density in the center, and a logarithmically decreasing resolution out from this point, is referred to as a log-polar pixel representation. Some approaches for calculating the OF field in log-polar coordinates, specially using foveated vision, have been given by Tistarelli et al. [120] and Daniilidis and Krüger [32]. The representation of the image field in log polar coordinates and calculation of the OF field from these images introduce a compromise between computational cost and image resolution. Experiments on attentive navigation and fixation using such flow fields have been described by Daniilidis [29]. He also claims that the space variant sensing facilitates the active vision tasks like real-time or attentive tracking [30]. Log-polar approaches have also been applied to feature tracking and ego-motion detection tasks as proposed by Bernardino et al. [13] and Silva et al. [104]. They both make use of the data reduction property of space variant sensors. Additionally, Silva et al. utilize the geometric features of the log-polar normal flow vectors to determine the ego-motion.

Clustering, regression and outlier detection: Black [14] uses a hierarchical approach in a robust incremental algorithm, warping in time via bilinear interpolation and eliminating the outliers. The method is robust, especially in image regions where motion boundaries exist but the calculational speed remains to be improved

if real-time applications are considered. Nesi et al.[86] propose an algorithm based on gradient methods. They introduce a clustering approach using combinatorial Hough transform instead of least-squares method (LSM) for the solution of the constraint equations. The authors argue that LSMs tend to produce averaged solutions instead of most probable ones. Their approach increases the accuracy in cases of occlusions and input noise, but suffers from high computational load. Bab-Hadiashar et al. [7] introduce a robust method that identifies and removes the outliers in the flow field. They also propose to use a robust regression method for the solution of the overdetermined system of constraint equations. The authors propose that this method produces more accurate optical flow vectors compared to the other published methods so far. However, they do not characterize their method in terms of computational load. Weber et al. [129] convolve the images with a set of spatiotemporal filter kernels and combine the response of those using the total least squares method. This method is similar to the phase-based method of Fleet and Jepson, but can be implemented with a parallel algorithm, suited to gain significant acceleration on parallel computing hardware.

The Camus model: Camus [20] described a linear-time algorithm for OFC, based on a patch correlation method. He proposed a matching method that keeps the spatial search field for a pixel constant and varies the temporal search field. This technique reduces the computational complexity, spreading the calculation onto the temporal domain. The method implicitly assumes rigid object motion. Although this method allows for faster solution of "time-to-contact" problems and some other typical robotic tasks that use OFC, the accuracy of the resulting velocity fields remains to be improved.

Robustification of Lucas and Kanade's Model: While structure from motion algorithms must rely on accurate flow estimations, real-time robotics stresses the speed and the flexibility of the calculations. We close this section with a review of several modifications to robustify the differential method based on the local smoothness constraint proposed by Lucas and Kanade, which we described in Section 2.2.2. This method, together with the modifications of Simoncelli [105], has been used in many of the robotic applications we have surveyed, e.g. [95]. It provides an acceptable trade-off between accuracy and computational load. Local spatiotemporal filtering using separable kernels allows time-efficient applications of OFC. Simoncelli [106] modified the method introducing a probabilistic approach and corresponding confidence measures. These measures help to improve the accuracy and the reliability of the calculated OF field. Fleet and Langley [41] proposed to use IIR (infinite impulse response) filters for a faster implementation of the tem-

poral derivatives. The robustness of this method results from the assumption of a local smoothness of the motion in a small area of the image rather than a global smoothness. This makes it more appropriate for the applications that have to cope with object deformations. One drawback of the method results from the general problem of digital differentiation which is always error prone.

2.5 Motion Recovery

Analysis of visual motion for the reconstruction of the 3D motion field is made up of two main stages. The first one is the measurement of the motion field. In this stage, temporal image sequences are processed to yield a representation of a velocity field (e.g. optical flow field measurements). In the second stage, these representations are interpreted to extract comprehensive information about the 3D structure of the motion field.

Generally, the 3D motion field results from one of the three cases; (1) relative motion of the static camera and the moving objects in the scene, (2) mobile camera and the static scene, (3) relative motion of the mobile camera and the dynamic scene objects. In the first case, a local flow field, which is induced by the moving object, will be measured. The second and third cases introduce global motion fields of the images depending on the type of the sensor motion and object motion.

The motion recovery problem aims at constituting the relationship between the 2D velocity measurements and the actual 3D motion. This task is in essence ill posed, unless one has some additional knowledge about the scene or obtains some proper assumptions. It is mostly assumed that the objects in a static scene move rigidly or the camera moves in a static environment. In these cases, all points on a rigid object share the same six motion parameters (one depth, two translational, three rotational parameter).

Starting with the above considerations, Ullman [122] states a solution to the problem of motion recovery and the estimation of 3D structure of the environment from motion (*structure-from-motion* problem) as follows: If one knows the correspondence of points from each image frame to the next and if one assumes the moving object is rigid then it is possible to recover both the 3D location and the motion of the object from four non-coplanar points in three distinct orthographic projections. This also provides an internal test to tell if the points are moving rigidly or not. In another approach, Ullman [123] introduces the concept of incremental rigidity which assumes that the system incrementally constructs an internal model of the 3D scene. The scene is initially estimated as being flat and the perception of its 3D structure develops over time as more information becomes

available.

This kind of iterative solutions to the problem introduces large computational load. Alternatively, Tsai et al. [121] introduce a discrete approach. A set of parameters are determined by solving a set of linear equations and these parameters are used to recover the motion and structure. They showed that seven measurement points are sufficient to determine the motion uniquely. Bruss et al. [18] proposed a global approach that combines information in the entire visual field to choose the 3D motion and structure that fits the given flow field best in the least square sense. Jepson et al. [64] suggested a subspace method, which decomposes the general problem in three steps and obtains linear constraints for the translational parameter by eliminating the rotational components locally. They extracted information about the rotational parameters and depth based on the translational parameter. Gupta et al. [55] proposed a similar approach calculating some basic parameters by solving linear systems of equations, which are related to the motion parameters one to one, so that the 3D motion recovery is achieved in closed form.

Estimating the independent motion, which may be induced by a moving camera platform *and* moving scene objects, is not straightforward while applying the above mentioned methods, since they mostly assume the rigidity of the expected motion field. The non-homogeneous measured velocity field due to multiple motion and complex background motion, complicates the interpretation and the determination of the actual 3D motion of the scene and the objects. A few studies have been performed [63; 84; 2]. Adiv et al. [2] identified regions in the image whose motion was consistent with the movement of a planar surface and grouped those to their mutual consistency for various 3D motions. Nelson et al. [84] proposed another approach where the observer motion is given. The author compares the expected motion field against measured component velocities; wherever significant deviation is found, independent object motion is assumed. Maclean et al. [75] applied motion segmentation based on clustering constraints of 3D translational velocity and bilinear constraints on 3D translational and rotational motion, which was derived from OFC. Irani and Anandan [62] proposed a general approach to the problem of independent motion detection, combining the model based 2D and 3D approaches. The authors introduced a layered detection and tracking concept, which might be computationally complex, specially when the images are captured with a narrow field of view. Argyros et al. [6] used the normal flow components and stereoscopic measurements to segment independent motion. In the approach by Boluda [16], the log-polar representations of image sequences are processed to detect the regions of motion boundaries.

Generally, the algorithms proposed for solving the problem of independent motion detection rely on the assumption of a planar background, smooth egomotion

and standard camera view. The given methods are mostly based on the detection from two frames (with the exceptions of [62], and [89]), and do not consider the temporal and probabilistic domain of the problem. The reliable and efficient detection of independent motion from a moving camera setup, therefore, remains an open issue in visual motion estimation tasks.

2.6 Catadioptrical Vision: Towards Omnidirectional Motion Segmentation

A mobile robot equipped with a standard vision system has typically more degrees of freedom to move around than it can actually visually capture. For example, moving forward, a standard camera-robot system can normally capture a field of view of about $40-50^\circ$ in horizontal and vertical directions, which is essentially too narrow to detect obstacles lying behind, below or beside itself. This may result in collisions with such obstacles. An efficient way to solve this problem is to increase the field of view using omnidirectional image sensors.

In recent years, cameras with a large field of view, often referred to as *Omnidirectional Cameras*, quickly gained popularity in robotics. Expansive sources of information, as provided by wide angle and omnidirectional visual sensors facilitate visual motion detection mechanisms as observed in some biological prototypes like insects. Typically *Catadioptric Omnidirectional Cameras*, which are a combination of a standard camera with a lens pointing towards a curved mirror, are used. This setup provides a horizontal 360° view. The larger field of view comes at the cost of lower resolution in the outermost image regions. The geometrical properties and the distortion introduced by the reflection of the mirror are dependent on its shape and should thus be designed carefully. The mirror geometry can be shaped such that it approximates a perspective projection or a spherical projection. It is also possible to design mirrors that eliminate some reflectional distortions (e.g. [25]).

Catadioptric sensors introduce non-affine distortions on the captured images, therefore assumptions about the global smoothness and rigid motion cannot be exploited for the OFC. Local smoothness can be assumed as long as the temporal intensity changes are small in the chosen evaluation window (see Eqn.(2.11)). Using single centered catadioptric cameras, one achieves omnidirectional images, which can be mapped onto spherical coordinates. Daniilidis et al. [33] referred to this property and proposed to apply the method of Lucas and Kanade in spherical coordinates. In the case of non-single centered cameras, accurate OFC remains to be a current research field.

Some of the recent research on the estimation of navigational preliminaries

of the mobile robotics consider the use of omnidirectional imaging. Chahl and Srinivasan [23] described a method for range estimation based on the estimation of image deformations, Yamazawa et al. [134] used a catadioptric sensor for obstacle detection from a mobile platform, and Yagi et al. [133] introduce a method to determine the rolling motion of a mobile robot using radial flow estimations. They used a spatiotemporal differential method for the optical flow estimation. Chahl and Srinivasan [22] applied an image interpolation technique for range estimation, localization and ego-motion estimation. Gluckman and Nayar [53] adapted the known egomotion estimation techniques to a motion estimation method using spherical projections of the optical flow field.

In the following subsections, we refer to the relevance of omnidirectional vision in mobile robotics. The inspiration source hereby is nature itself. In insect vision omnidirectional sensing and optical flow play important roles, as will be referred in Subsection 2.6.1. Technical aspects of the field-of-view enhancement will be reviewed in Subsection 2.6.2.

2.6.1 The Role of Omnidirectional Vision and Optical Flow in Nature: Arthropod Vision

Insects, arachnids (spiders, scorpions), and crustaceans (lobsters, shrimp, crabs) compose the species called arthropods. The compound eye structure found in arthropod vision, allows a very fast capturing of a wide (panoramic) field of view in a space variant resolution. Physically, each unit of the compound eye is an independent optical system, which is called ommatidium. The compound eye structure has a larger relative size and lower spatial resolution compared to the vertebrate visual system. The visual system of the insects is also far less complex than that of a vertebrate. Nevertheless, insects have an outstanding ability to navigate in their environment. Srinivasan et al. [22; 109] conducted experiments showing that honeybees estimate object distances in terms of apparent speeds of motion of the object images. The bees also distinguish objects from backgrounds by sensing the apparent relative motion at the boundary between objects and background [78], [79]. This mechanism of object - background separation does not require fast computing power with a sophisticated memory to distinguish the angular direction of a relative motion and to identify the discontinuities in the flow field as the outline of the objects [60]. There is also evidence that in the insect visual system, there are neurons which immediately interpret the relative motion against the background as the separateness of an object, whether or not the background appears to move.

Additionally, by exploiting the spectral information from the various light

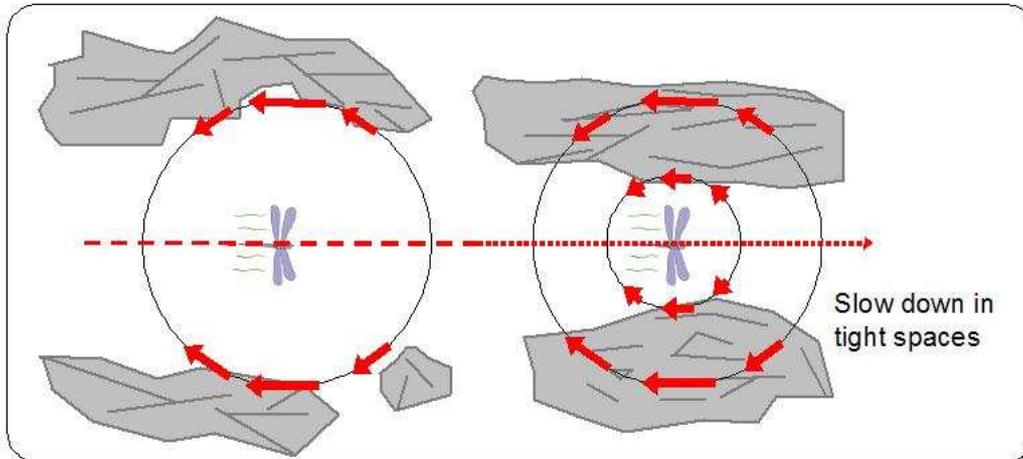


Figure 2.4: *Speed-control of a dragonfly, (used with the kind permission of G.Barrows). The detection of omnidirectional optical flow allows the dragonfly to control its speed in dense environments. The insect slows down (here symbolized with the changing magnitudes of the arrows) in tight spaces, where the optical flow field becomes dense and large in magnitude, avoiding a collision with the objects.*

sources present in nature, reliable orientation behavior can be achieved with less neural circuitry. Bees and ants use a skylight compass. The upper portion of the eyes of these insects are equipped to detect the polarization pattern of light in the sky and to determine the orientation from this information [130]. But neither the bees nor the ants can compete with the spectral range that is encompassed by the compound eye of the mantis shrimp, a marine crustacean, whose eye is known to include more photoreceptor classes than any other known visual system [28]. This remarkable compound eye structure found in mantis shrimp (Stomatopoda) (see Fig.(2.5)) have the ability to sense a single location in space by optical elements in three different parts of the same eye - the upper hemisphere, midband, and lower hemisphere. Thus, each individual eye of the mantis shrimp has depth perception and trinocular vision. In contrast, we humans require both of our eyes for depth perception and binocular vision.

The neural circuits known as elementary motion detectors [94] calculate the motion vectors that build the optical flow field needed for insect navigation [36; 70]. Optical flow does also play an important role for the odometry and route validation in flying insects (e.g. speed control behavior of a dragonfly as can be seen in Fig.(2.4)) . It has been shown in experiments with bees that edges are important for the recognition of visual landmarks, as well as higher level information, such as



Figure 2.5: *The compound eye of the mantis shrimp*, Copyright Takata (<http://www.blueboard.com/mantis/bio/vision.htm>)

the relative position of landmarks [109]. A model that describes this phenomenon well is the *neural snapshot model* [21]. This model, as the name implies, suggests that for certain pivotal locations on the bee's route, all visual lower level cues are simultaneously stored in its memory. This information is recalled at a later stage, and thus the bee can align its position such that its current visual information corresponds to the recalled one. With such methods, insects are able to achieve quite complex behavioral tasks with far less computational effort than vertebrates do.

Insect vision could inspire the robot designs that make an efficient use of their sensory information and processing power. This is true in particular for smaller robots that have a higher demand on power efficiency and where the weight of a visual system has considerable impact on the dynamics of the robot. Several robotic applications have followed the so far gained insights (e.g. [46], [119] and [98]) already and this inspiration is still an objective of ongoing research as referred in [19].

2.6.2 Catadioptric Sensors and Mirror Design

In technical systems, there are several ways to enhance the field of view of an imaging system. Rotating a standard imaging system around one of its axes and stitching the snapshots to a panoramic view is just one possibility, which requires additional mechanical and computational processing. A more practical solution is to incorporate reflective surfaces into traditional imaging systems. Such a camera setup that is complemented with a reflective surface for a wide angle field of view is called a *catadioptric sensor*. Often, the reflective surface is curved and enlarges the field of view in a chosen direction.

Several examples of catadioptric sensor design can be found in the respective

literature. In [132], a conical mirror and in [58] a spherical mirror was used in conjunction with a standard perspective sensor. [134] used a hyperboloid mirror and showed that this combination allows for a single viewpoint. [82] reports another design technique that consists of four planar mirrors, which form the faces of a pyramid and four separate imaging systems, each placed above one of the faces. The optical axis of the imaging systems and the angles made by the four planar surfaces are adjusted so that the four viewpoints produced by the four planar mirrors coincide. The resulting sensor achieves a panoramic field of view of appr. $360^\circ \times 50^\circ$.

The extent of the the field-of-view expansion, the geometrical properties of the images and the distortion caused by the refraction can be affected by the mirror design. The geometrical form of a mirror that allows an approximation of the perspective projection or a spherical projection, should have a single center of projection. On the other hand, one can design mirrors which eliminate or at least control the refractive distortion of the scene. The two different types of sensor systems can be grouped as follows:

Single centered cameras Cameras with single effective view point allow perspective imaging on a plane at the effective focal distance. Every pixel in the sensed images measures the irradiance of the light passing through the viewpoint in one particular direction. Since one knows the geometry of the catadioptric system, one can precompute this direction for each pixel. Therefore the irradiance value measured by each pixel can be mapped onto a plane at any distance from the viewpoint to form a planar perspective image. This facilitates the use of known image processing algorithms for perspective images.

Equi-areal or equi-angular cameras For applications like robot navigation, range finding, etc., having a single center of projection is not the restricting criterion for the mirror design. Other design issues like keeping a constant resolution over the image or the ease of mapping are more important. For example, the mirror designed by Chahl and Srinivasan [25] keeps the angular resolution constant so that the incidence ray angle is proportional to the reflectance ray angle (see Fig.(2.6)). This allows for a camera setup that can capture a global image of the environment projecting it as a polar image on the plane. Thereby, the elevation and azimuth in world coordinates are mapped linearly to the radius and the angle on the imaging array.

The angle of elevation is defined as the angle between a ray of light impinging on the surface and the downward vertical (ϕ). The radial angle is the angle between the optical axis of the camera and the incoming ray of light after reflection by the

mirror surface(θ). Keeping the linearity of these two angles constant with α , one can write the following mathematical relationship:

$$\alpha = \frac{\Delta\phi}{\Delta\theta} \quad (2.15)$$

This relationship ensures that the change of elevation of the incoming light rays is mapped onto a proportional change in the radial distance from the center of the 2D - image captured by the camera. The constant α represents the angular magnification. The profile of a mirror surface which satisfies the relationship (2.15) has been derived by [25]. Thereafter the surface profiles, which satisfy the following relationship (2.16) allow for an omnidirectional imaging, which can preserve the range proportionality of the 3D points and their projections onto the 2D image plane.

$$\cos[\theta(1 + \alpha)/2] = (r_0/r)^{(1+\alpha)/2} \quad (2.16)$$

In Eqn.(2.16), r is the distance of a point on the mirror surface to the nodal point of the camera lens, the point inside a lens where light paths cross before being focused onto a photosensor, θ is the radial angle of the reflected light ray, and r_0 is the distance of the point on the mirror surface to the camera nodal point where $\theta = 0$, which is the optical axis (see Fig.(2.6)). The relationship (2.16) describes a family of surface profiles, varying on the choice of the angular magnification α . For example $\alpha = 1$ describes a plane mirror surface profile, while $\alpha = 3$ describes a rectangular hyperbola.

The mirror surface we use to conduct the experiments throughout this work has the angular magnification constant $\alpha = 3.6$. This means that the surface profile was designed according to Eqn.(2.17):

$$r^{2.3} = \frac{r_0^{2.3}}{\cos(2.3\theta)} \quad (2.17)$$

In cartesian coordinates, r and θ can be reformulated (Note that the surface profile is estimated on the $X - Y$ plane.):

$$r^2 = X^2 + Y^2 \quad (2.18)$$

$$\tan(\theta) = X/Y \quad (2.19)$$

Using the following approximation for the Eqn.(2.17):

$$r^2 = \frac{r_0^2}{\cos(2\theta)} \quad (2.20)$$

which it can be reformulated in cartesian coordinates as:

$$X^2 - Y^2 = r_0^2 \quad (2.21)$$

This new equation characterizes a rectangular hyperbola on the $X - Y$ plane. Rotating this profile curve along its Y axis results a solid mirror surface, which preserves range and expands the field of view with a factor of $\alpha = 3.6$. The omnidirectional images captured with such a setup can be easily mapped onto polar coordinates, too. The mapping of an omnidirectional image onto polar coordinates is depicted in Fig.(2.7). The coordinates of a pixel on the polar image r and θ are estimated from its coordinates in the original image (x_i, y_i) using the following relationships:

$$r = \sqrt{x_i^2 + y_i^2} \quad (2.22)$$

$$\theta = \text{atan}\left(\frac{y_i}{x_i}\right) \quad (2.23)$$

where the coordinate origin is the center of the omnidirectional image and the angular term θ is measured counterclockwise.

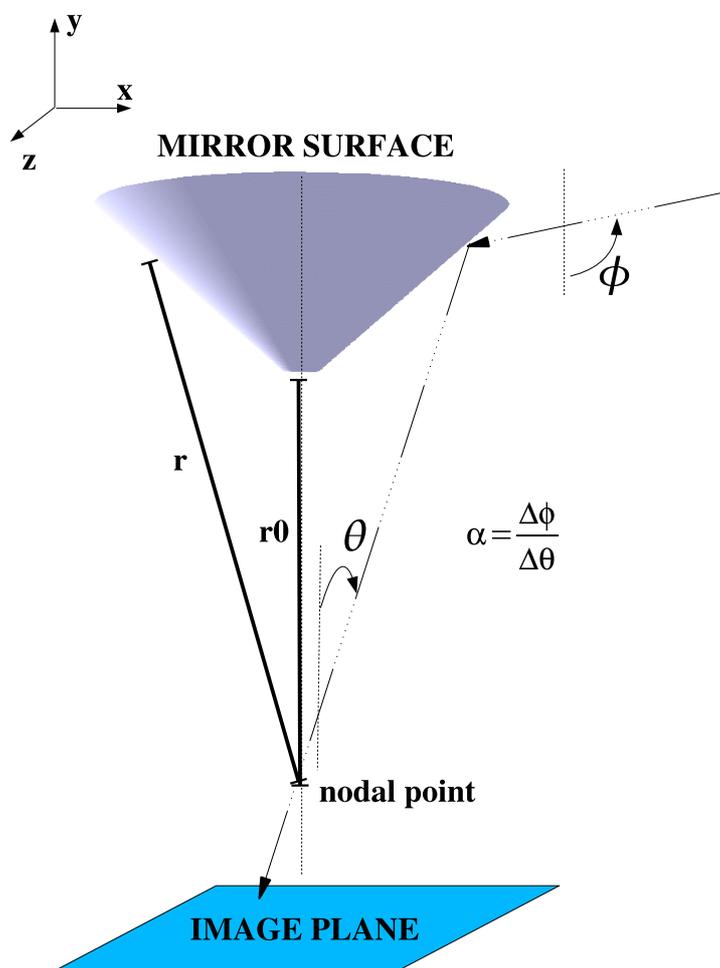


Figure 2.6: *The mirror shape designed by Chahl and Srinivasan. The form of the mirror is estimated by considering the desired expansion of the field of view, which is specified by the constant α . ϕ is called the angle of elevation and represents the angle between the incoming ray and the surface normal of the image plane. θ is the radial angle and represents the angle between the reflected and the vertical normal. The mirror is assumed to be exactly positioned such that it is vertically aligned with the camera nodal point. The distances r_0 and r are measured from the mirror surface to this nodal point.*

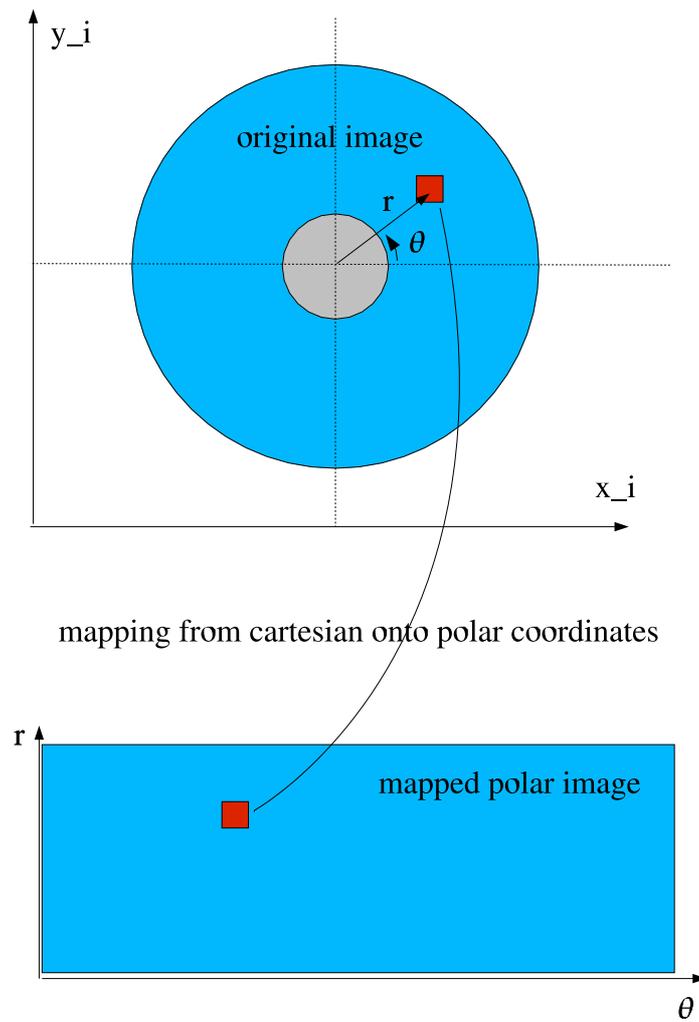


Figure 2.7: *The mapping of an omnidirectional image onto polar coordinates.*

Chapter 3

Optical Flow Calculation on Omnidirectional Images

Optical flow estimation techniques, which have been reviewed in Chapter 2, consider images captured by standard perspective cameras. Usually, these cameras project the scenery onto a rectangular plane represented by the cartesian coordinates. The majority of the image processing and optical flow estimation algorithms are designed to be applied to such images. Distortions due to the camera optics are mostly ignored or assumed to be constant. In [113], we showed that most of the standard OFC methods, with the exception of the local differential method (Lucas and Kanade method reviewed in Section 2.2.2), do not yield optimal results for the omnidirectional optical flow estimation.

Specifically, catadioptric sensors introduce non-affine distortions on the captured images. The extent and the type of the catadioptrical distortions depend on the geometry of the refracting surface and the capturing sensor. They may also be spatially varying. For example, some of the catadioptric sensors supply images where the amount of deformation changes with the radius of the circular image plane. Others are optimized to yield images without distortion on the near regions of the camera view center and induce significant deformation on the distant regions of the captured scene (e.g. [76]). These examples imply that the standard optical flow calculation methods assuming rigid scene variations cannot be applied onto omnidirectional images without modifications. The assumptions on global smoothness of the image motion or the rigidity do not hold any longer. The local smoothness assumption should also be modified considering that the constant flow regions are deformed as the 2D image itself.

Moreover, the smoothing and gradient operations, which are required by most of the OFC methods, are spatial operations. This means that they are estimated by the convolution of spatial kernels with the image pixels. Usually, these filters

are constructed in the sense of neighborhood windowing, and the spatial support is chosen to be a square of some pixel number, e.g. 5×5 smoothing filter operating on 25 neighboring image pixels. The spatial operations should be modified to consider the projective distortions on the omnidirectional images. For example, Fig.(3.1) depicts a picture of a cameraman, captured with a standard perspective camera and the spherical projection of the same image, which includes projective distortions. The red squares show the support of the spatial filter, which can be applied to smooth the image or to estimate spatial derivatives. It can be seen that the support of the filters is only on the undistorted image homogenous. On the distorted image, on the right, the spatial support of the filter varies radially. For example, the face of the man can be filtered completely using the shown filter support on the undistorted image, whereas on the distorted image, a larger filter support is required to fulfill the same task. Also, the distortion causes the originally linear regions, like the pillar of the camera appear curved. Such distortions cause non-linear effects that should be handled by an appropriate spatial filtering on the distorted image. Generally, the modification of the spatial filters for homogenous filtering on the omnidirectional images, reduces the effects of the aperture problem and is also required for a more accurate optical flow estimation.

It has been proven that the single centered catadioptric cameras capture omnidirectional images, which can be projected onto spherical coordinates ([31],[118]). In [33], the authors propose to map the omnidirectional images back onto the spherical coordinates and calculate the optical flow field in this coordinate system. Although this method considers the shift variance of the derivative filters on the spherical images, it needs intensive calculation. The mapping and the derivative filtering in spherical coordinates cannot be calculated in time-efficient manner which is a handicap in robot navigation tasks. In addition to this, one can apply this method solely on the images captured by the single centered catadioptrical cameras.

As a trade-off solution to the problem of time efficiency vs. accuracy in the OFC for omnidirectional images, we propose to use separable smoothing and derivative filters, which are spatially variable according to the distortion type caused by the refractive surface. As an example, we apply the method to calculate optical flow for the image sequences captured by the range preserving, non-single centered catadioptric sensor.

3.1 Spatial Filter Design

The spatial filter functions are applied in the image space domain (x, y) to enhance the images like smoothing or sharpening. Also, spatial derivative filters can be

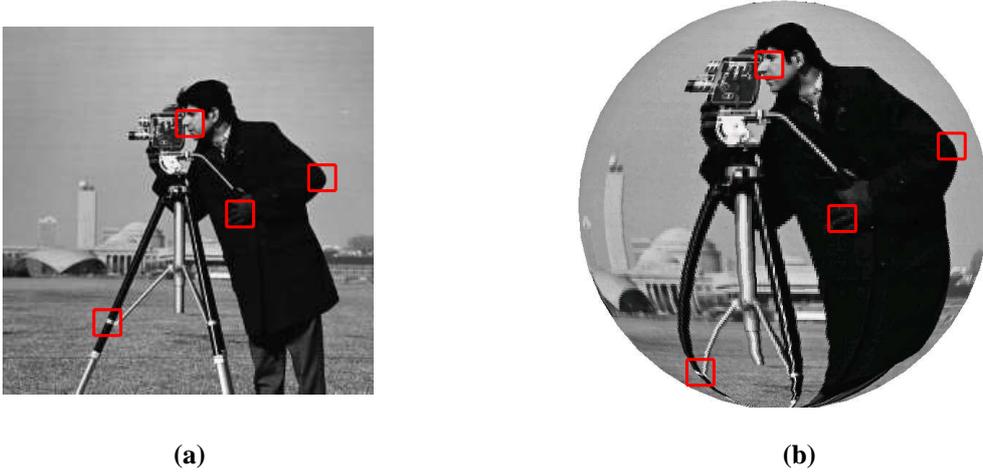


Figure 3.1: A picture of a cameraman, (a) captured with a standard perspective camera and (b) the spherical projection of the same image, which includes projective distortions. The red squares show the support of the spatial filter, which can be applied to smooth the image or to estimate spatial derivatives.

used to estimate the gradient of the images intensity function in various directions. In the case of digital image data, spatial filtering in the domain of image space is usually achieved by local convolution with an $\mathbf{n} \times \mathbf{n}$ matrix operator as follows:

$$G(i, j) = \sum_{k=i-w}^{i+w} \sum_{l=j-w}^{j+w} I(k, l)h(i - k, j - l), \quad (3.1)$$

where I represents the input image, h the filter function, G the output image, i, j the image column and row indices, and w the windowing index. The convolution is created by a series of shift-multiply-sum operators with a $\mathbf{n} \times \mathbf{n}$ matrix (\mathbf{n} : odd number). The dimensions of this matrix determine the spatial support of the convolution filter. It is mostly chosen to be a square region of 3×3 or 5×5 pixels.

The 2D convolution can be performed fairly quickly if the kernel (e.g. the filter function h in Eqn.(3.1)) is a separable function. For example, the 2D Gaussian kernel, shown in Fig.(3.2(a)) is separable into x and y components. Thus the 2D convolution can be performed by first convolving with a 1D Gaussian in the x direction, and then convolving with another 1D Gaussian in the y direction.

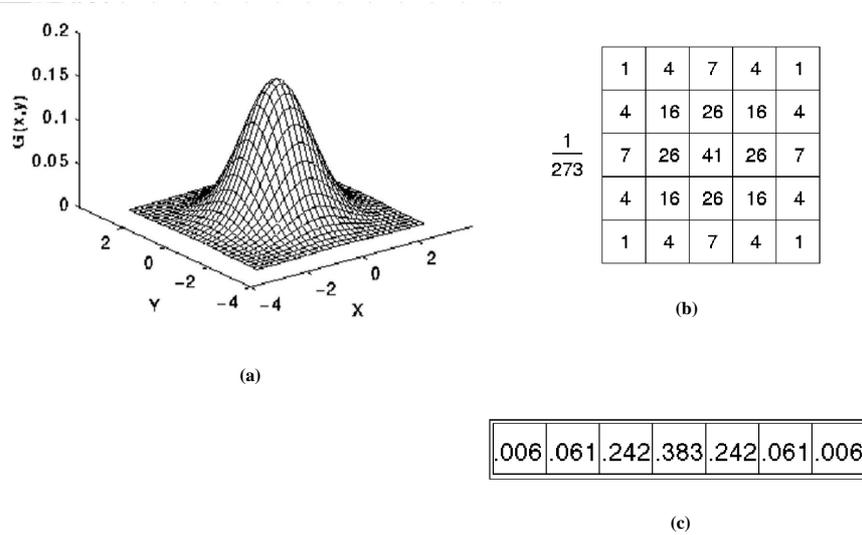


Figure 3.2: (a) 2D Gaussian distribution with mean $(0,0)$ and $\sigma = 1$ (b) Discrete approximation to Gaussian function with $\sigma = 1.0$ (c) One of the pair of 1D convolution kernels used to calculate the full kernel shown in (b) more quickly.

Fig.(3.2(c)) depicts the 1D x component of the Gauss kernel that would be used to produce the full Gauss kernel shown in Fig.(3.2(b)) (after scaling by 273, rounding and truncating one row of pixels around the boundary because they mostly have the value 0. This reduces the 7×7 matrix to the 5×5 shown above). The y component is exactly the same but is oriented vertically.

The calculation efficiency introduced by the separable spatial filters can be appreciated if one considers that the complexity of the separable convolution reduces from $O(N^2 * n^2)$ to $O(N^2 * n)$, where N is the image dimension of a squared image, and n is the kernel dimension.

Separable spatial filters for omnidirectional images The spatial distortion due to the mirror surface, affects the spatial relationship of the neighboring image pixels. Therefore, new neighborhood forms should be defined, that are similarly deformed as the reflected image. These new spatial regions should conserve the neighborhood relationships of the image pixels as if they were non-deformed by the mirror. This means that the homogenous quadratical filter functions defined on a spherical 3D world should be mapped onto a 2D image plane. This mapping causes radially variant deformation of the quadratic filter form on the 2D image plane, as depicted in Fig.(3.3).

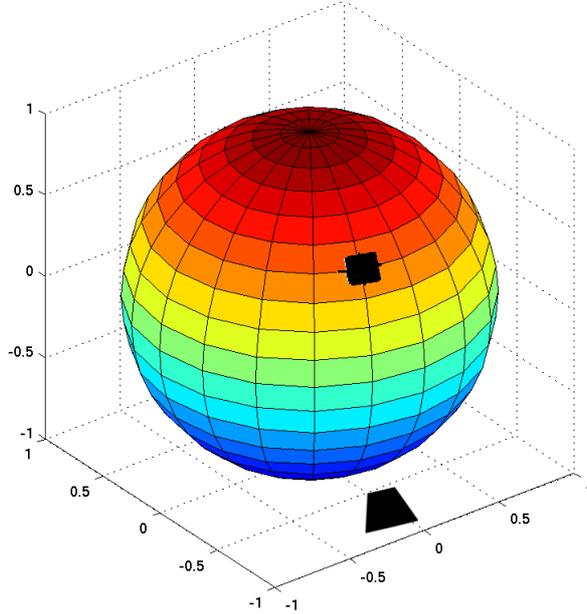


Figure 3.3: *Projection of a quadratical region of a sphere onto the image plane*

Homogenous filtering of the omnidirectional images on the view sphere is achieved by ensuring that each filter patch symmetrically straddles the longitudinal and elevational great circles passing through the center of the patch, as shown in Fig.(3.4). This technique was first introduced by Chahl et al. [24] in the context of separable smoothing filters for omnidirectional images. Conformal to the notation given in [24], we define \mathbf{V} as the directional vector on the principal axis of the local longitude of the filter patch, and \mathbf{U} as the directional vector on the orthogonal great circle passing through the middle of the local patch, where we constrain the spherical image. The boundaries of a square filter patch pass through the great circles of the view sphere. If the square filter patch center is at the Φ elevational, and Θ azimuthal coordinates as shown in Fig.(3.4), the projection of this patch onto the 2D image plane can be achieved by projecting the principal axes \mathbf{U} and \mathbf{V} .

Therefore, the projection of the \mathbf{U} axis can be written as:

$$\theta = \Theta + \arctan\left[\frac{\tan(\mu)}{\sin(\Theta)}\right] \quad (3.2)$$

$$r = \arccos[\cos(\mu)\cos(\Phi)] \quad (3.3)$$

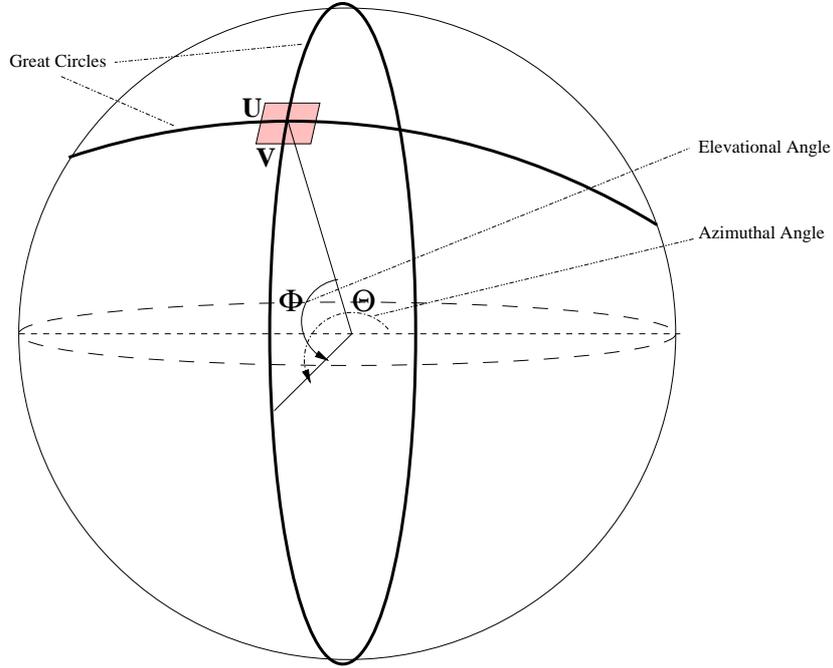


Figure 3.4: *The view sphere with a homogenous filter patch of dimensions \mathbf{U} and \mathbf{V} , defined on the corresponding great circles.*

And the projection of the \mathbf{V} axis:

$$\theta = \Theta \quad (3.4)$$

$$r = \Phi + v \quad (3.5)$$

where θ and r refer to the pixel locations on the 2D image in polar coordinates (formally $\theta = \text{atan}(\frac{y}{x})$ and $r = \sqrt{x^2 + y^2}$). μ and v represent the angular deviation from the center to the boundaries of the filter patch on the 3D view sphere. They represent the spatial support of the filter in angular means.

The projections of these squared filters can be depicted as in Fig.(3.5). For the un-warped image on the θ - r plane, these projections can be employed to construct separable filters. The 1D convolutions can then be first applied on the \mathbf{U} axis and consequently on the \mathbf{V} axis.

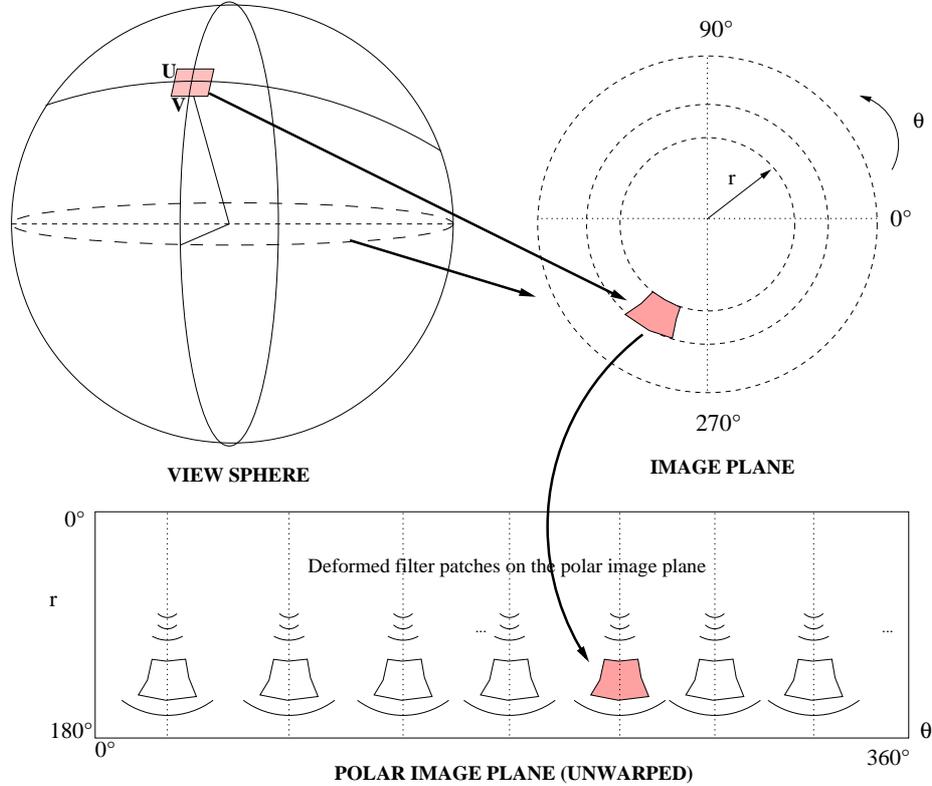


Figure 3.5: Projections of a spatial filter patch on the view sphere, image plane and on the unwarped polar image. The homogenous filters must be of constant size and shape on the view sphere, which will be measured with respect to the great circles passing through it. The resulting spatial filter is mapped onto the image plane as shown in the right upper part of the figure. The azimuthal angle (θ) is measured counterclockwise, starting from the right horizontal axis. r represents the elevational angle (also called radial angle) on the image plane. The filter patch can also be mapped onto the unwarped polar image as shown on the third diagram (below) of the figure. The shape of the spatial patch on the image plane varies for different elevations but it is invariant to the changes in the azimuthal axis. This is a consequence from the fact that the mappings of the great circles of the sphere change with the elevation but not with the azimuth. Considering that all the spatial filters have the same elevational expanse, one can precompute the filter coefficients for each dimension (θ and r) and apply the separable filtering scheme for the spatial filtering of the images.

3.2 Homogeneous Optical Flow Calculation on the Omnidirectional View

The OFC techniques reviewed in Chapter 2 require additional assumptions to estimate the actual optical flow vectors from the normal flow vectors. Some of the gradient techniques, that are also efficient in calculation time, such as the method proposed by Lucas and Kanade (Subsection 2.2.2), assume the local rigidity of the 2D motion. For the omnidirectional images achieved by catadioptrical sensors, this local rigidity assumption should be modified and the local regions where this assumption should hold, must be redefined on the spherical image.

Also, the gradient OFC methods require the application of spatial smoothing and derivative filters, which can be combined to one spatial filter kernel. As indicated above, images can be filtered using horizontal and vertical separable filters for a time efficient processing. However, these filters should be modified for the catadioptric images to yield a meaningful spatial support and to allow for homogeneous processing. For omnidirectional images, this means that the outermost radii of the 2D image should be filtered with a larger support than the innermost regions. Two sample spatial filters are depicted in Fig.(3.6). The spatial support of the filter components varies in the azimuthal direction depending on the elevation of the image pixel.

The method of Lucas and Kanade solves the local gradient minimization problem formalized with Eqn.(2.11) in Section 2.2.2. Our implementation of the local gradient optical flow method (Lucas and Kanade) solves the following alternative equation:

$$\mathbf{A} \vec{\mathbf{v}} = \mathbf{b} \quad (3.6)$$

where

$$\mathbf{A} = \begin{bmatrix} \sum_{\Omega} \omega(\Theta, r) I_{\Theta}^2 & \sum_{\Omega} \omega(\Theta, r) I_{\Theta} I_r \\ \sum_{\Omega} \omega(\Theta, r) I_{\Theta} I_r & \sum_{\Omega} \omega(\Theta, r) I_r^2 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \sum_{\Omega} \omega(\Theta, r) I_{\Theta} I_t \\ \sum_{\Omega} \omega(\Theta, r) I_r I_t \end{bmatrix}$$

I_{Θ} and I_r represent the gradient in the azimuthal and radial directions, respectively. They are estimated using the separable filters mentioned above. The spatial support function $\omega(\Theta, r)$ varies radially with the azimuthal coordinates, as depicted in Fig.(3.5). I_t is the temporal derivative, estimated by the relative gray value differences in two sequential omnidirectional images.

Our calculation scheme can be outlined as follows:

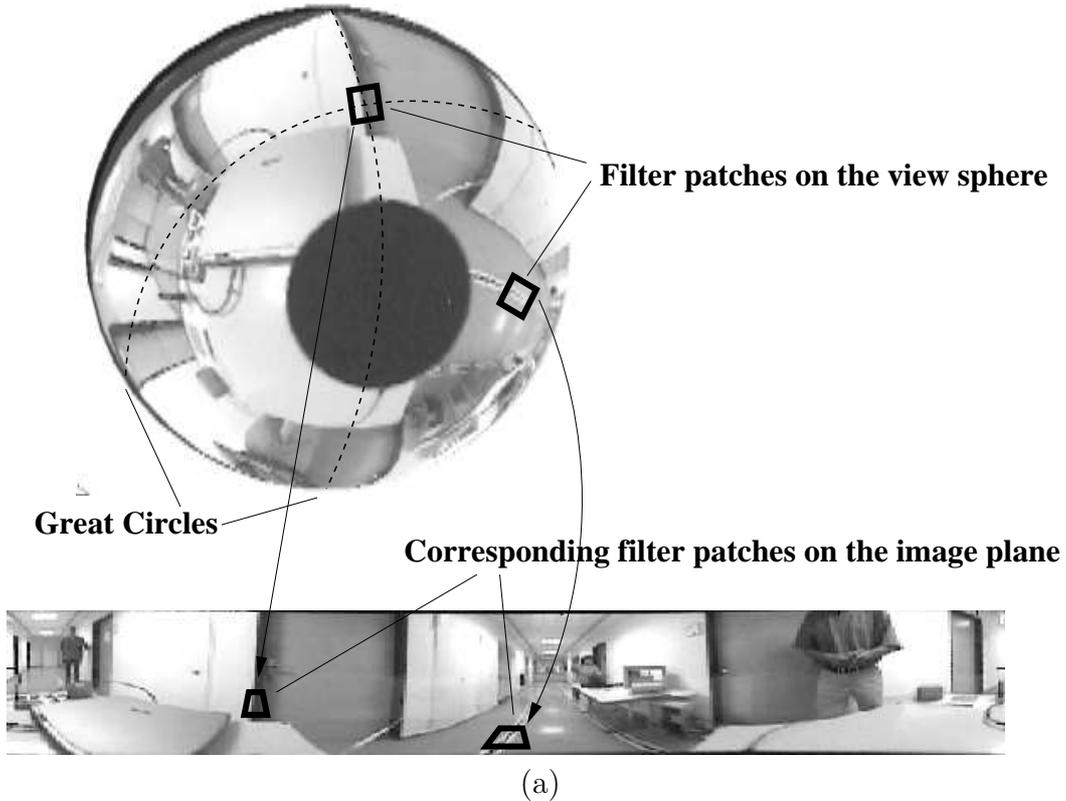
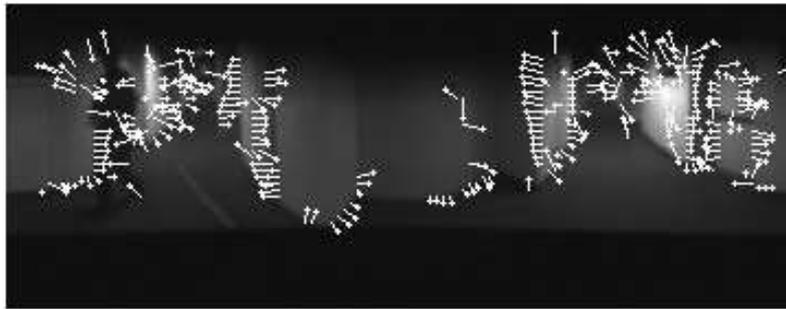


Figure 3.6: *Homogenous spatial filters defined on the view sphere with respect to the great circles crossing across them. The filter supports are projected onto the polar image plane, where they are spatially deformed depending on their elevational coordinates.*

- Capture two of the omnidirectional images in a sequence and unwarped them (representation in polar coordinates)
- Smooth the images and calculate the spatial and the temporal gradients using a homogenous spatial filter kernel as defined in Section 3.1
- Build the matrix \mathbf{A} , and vary ω , the extent of the local neighborhood, in the azimuthal and elevational directions
- Solve the local linear system (Eqn.(3.6)) for each neighborhood using the least squares estimation

Solving the above mentioned system of equations for each pixel of the image

yields a dense optical flow field. The estimation time for a gray valued VGA Image Resolution (640x480 pixels) is about 2 millisecon on a Pentium III, 933MHz PC. The angular support of the smoothing and the derivative filters is 5° . The resulting flow field represents the actual 2D motion rather than the distortional changes of the grayvalues, as seen in Fig.(3.7(a) and (b)). The original calculation technique suffers mostly on the outermost regions where the smoothness can only in larger spatial neighborhoods be assumed.



(a)



(b)

Figure 3.7: (a) *modified Lucas and Kanade Technique* (b) *standard Lucas and Kanade Technique applied to calculate optical flow on omnidirectional images. Note that the image distortion due to the refraction are interpreted as motion and the system estimates flow vectors, which do not correspond to the actual motion field.*

In Fig.(3.8), the individual calculation steps are depicted to show the resulting gradient images as well as the resulting optical flow field. The flow field is represented with the classical vector format and in intensity code.

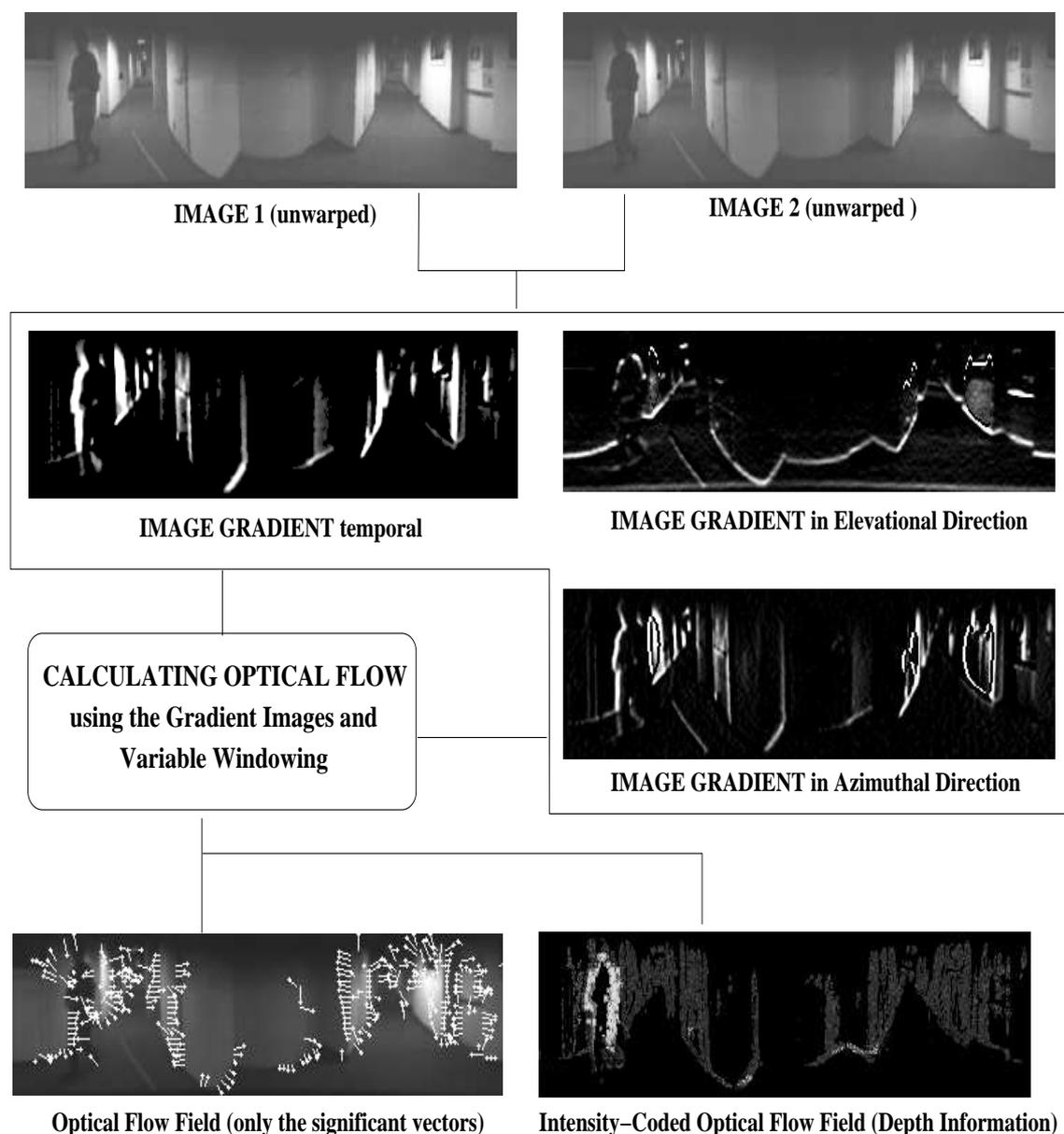


Figure 3.8: *Omnidirectional OFC Scheme: The two sequential images are unwarped to facilitate the following spatial processing. The temporal and spatial gradients are calculated using the scheme explained in Section 3.2. The derivative filters and the windowing functions are spatially variable and account for the projective distortions on the images. The resulting optical flow field and the intensity coded depth image are shown at the bottom line of the diagram. (The lighter the color, the higher the vector magnitude.)*

Chapter 4

Visual Motion Detection using Omnidirectional Optical Flow

Mobile robots that operate in an arbitrary environment should be able to explore this environment, decide on the security of their path, and locate themselves at different points in time. These tasks, particularly, and autonomous navigation, generally, require reliable and robust motion detection mechanisms. For many species, the most intuitive way to detect motion is using vision, since the correct interpretation of the visual data yields the structural and the navigational information directly (Fig.(4.1)). Our everyday experience of combining visual clues to explore an unknown environment or walk through the world without thinking much about it, supports the idea of exploiting vision to detect motion for the artificial systems like the robots, too.

The features inherent in the visual motion field of a mobile robot indicate important clues about its navigation and environment. For example, during straight translation, there is a focus of expansion within the visual field of the camera where image velocity is zero. Optical flow seems to radiate from this point. Closer objects move faster across a camera image field than those farther away. Simultaneous rotation and translation create optical flow fields that are more difficult to interpret. Independently moving objects also devise significant changes in the optical flow field. This chapter describes a novel methodology, which uses adequate measurement techniques and the principles of pattern extraction, to detect egomotion and to segment independent motion in a dynamic scenery

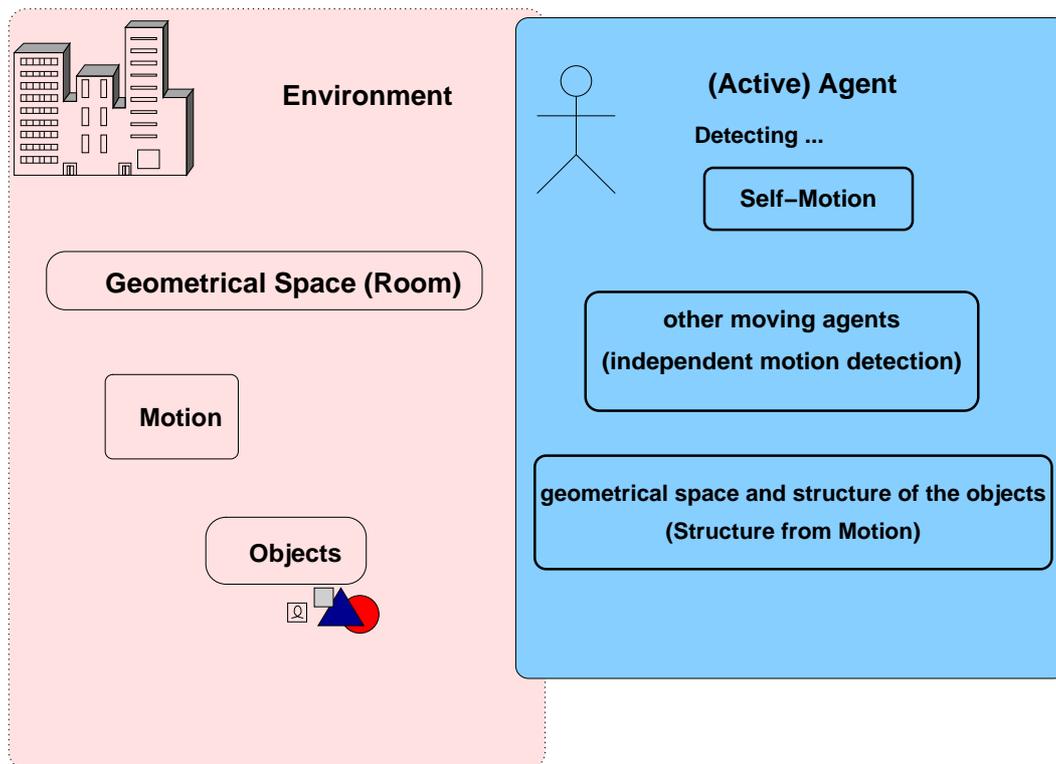


Figure 4.1: *The importance of motion seeing for many species motivates the idea of dynamic machine vision. The visual detection of the egomotion, structure from motion and the independent motion are only some of the challenging subtasks.*

4.1 Egomotion Estimation

The egomotion estimation problem is the problem of detecting the self-motion of a mobile camera-robot system. We assume that the robot-camera system moves in an environment where some structure is visible. The self motion of the robot (here referred to as the vector \vec{V}) can be measured visually, exploiting the structural correspondences in a sequence of images (see Fig.(4.2)).

Taking the camera system as the center of projection, the general motion of the camera system can be written as:

$$\vec{V} = -\vec{T} - \vec{\omega} \times \vec{P} \quad (4.1)$$

where \vec{T} is the translational component of the motion, \vec{P} is the position vector of the system ($\vec{P} = [X, Y, Z]^T$) and $\vec{\omega}$ represents the rotation of the robot-camera system around any axis, that originates from the center of projection..

If we assume that the environment is static besides the camera motion, then the self motion of the robot-camera system induces a projective motion field. It can be represented with $\vec{v} = [\dot{X}, \dot{Y}, \dot{Z}]^T$, which is a vector of spatial changes in each cartesian direction. Assuming perspective projection, in camera coordinates x and y , we write:

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}$$

If we assume that the 3D Motion vector (\vec{V}) can be approximated by the optical flow measurements (here referred as \vec{v}) of the projected image pixels ($\vec{p}(x, y, 1)$), we can rewrite Eqn.(4.1) as follows:

$$\vec{v}(\vec{p}) = \left(-\frac{1}{Z}\right) \begin{bmatrix} 1 & 0 & (-x) \\ 0 & 1 & (-y) \end{bmatrix} \left(\vec{T} + \vec{\omega} \times \vec{p}\right) \quad (4.2)$$

(see Appendix A for a more detailed derivation of this equation)

The goal of the visual egomotion detection problem is to estimate the translational velocity, \vec{T} and the rotational velocity, $\vec{\omega}$ given that the optical flow, $\vec{v}(\vec{p})$ measurements.

Most of the existing algorithms assume an exact correlation between the 3D motion field and the optical flow measurements to solve the egomotion detection problem or employ sparse optical flow calculations, which estimate the spatio-temporal changes of significant scene points, like landmarks (see Fig.(4.2)). Then they use the bilinear constraint to solve the problem, which can be derived by algebraic manipulations and removal of the depth (Z) from the Eqn.(4.2).

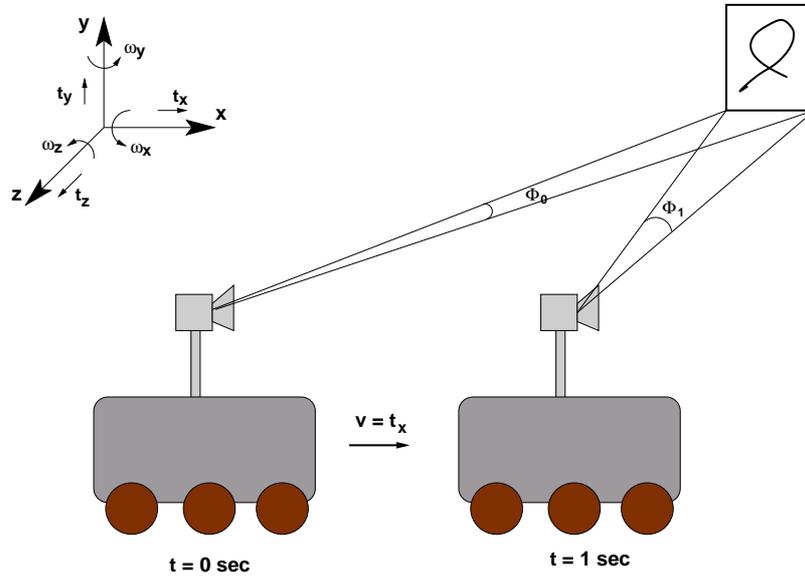


Figure 4.2: **Visual Egomotion Detection Problem:** *From the theoretical point of view, imaging and measuring the self-distance to a landmark at two different times allows the determination of the self motion of the camera-robot system. But the major practical problem is to find the correspondences of the same landmark from two different field-of-view.*

The major disadvantage of these strategies stems from the correspondence problem that is inherent in almost all visual tracking tasks. Since the estimation of motion parameters depends only on the flow measurement of several pixels on the images, it is flawed when the correspondence between the pixels in an image sequence fails. Considering Fig.(4.2), the detected shape of the landmark differs as the camera takes pictures from different view angles of Φ_0 and Φ_1 . In this situation, the correspondence of these two images of the same landmark is not given readily.

Note that the estimated visual motion (optical flow) consists of translational and rotational components, which can be processed separately (see Eqn.(4.2)). Rotation is a geometrical property of a moving body and does not vary from point to point and does not depend on the depth structure of the scene. Translation causes optical flow vectors to vary not only with the egomotion but also with depth structure of the scene. Therefore, the translation component of a general motion can be estimated up to a constant factor.

In [92], Pradzny proposed an algorithm that first estimates the rotation component of the motion and interprets the translational component as a rotation

about the axis parallel to the projection plane. The algorithm has a drawback when applied to standard perspective images; the interpretation of the estimated optical flow field may be ambiguous. There are examples where an optical flow field due to the non-zero rotation is identical over the entire image plane to one induced by translation alone (see Fig.(1.2)).

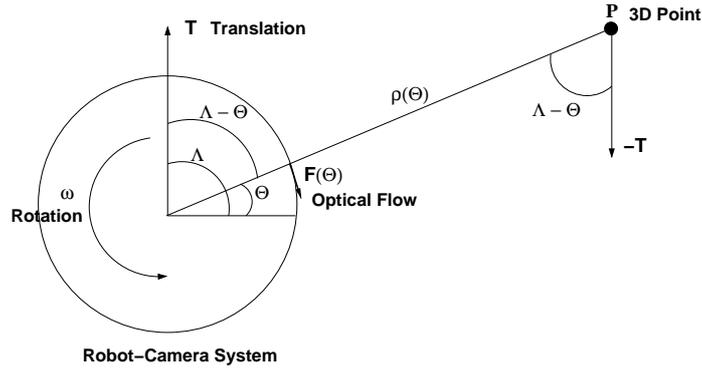


Figure 4.3: *Considering the robot camera system as a spherical projection system, one can interpret the optical flow vectors as angular velocities projected onto the sphere.*

The limited field of view and the perspective projection as provided by standard cameras causes the ambiguity mentioned above. When different types of camera motion yield similar optical flow fields, the interpretation of those fields is not always distinct and the estimated egomotion is not unique. This is not the case while interpreting global optical flow fields obtained by spherical projections. Considering Fig.(4.3), which depicts a general motion of a robot-camera system in a static environment, one optical flow vector on the global spherical view can be interpreted as an angular relative velocity, $F(\Theta)$. It is caused by the translation of the robot in the Λ direction, measured relative to the static 3D-Point P and the rotation, ω , around the vertical axis of the robot. Using the geometrical relationship shown in the figure, the angular velocity vector, $F(\Theta)$, can be written as follows:

$$F(\Theta) = \frac{v}{\rho(\Theta)} \sin(\Lambda - \Theta) + \omega \quad (4.3)$$

Pattern-Matching Approach There is inherent information, that can be gained from the topological structure of the dense omnidirectional optical flow fields. Pure translational motion produces a field, which is characterized by the singular points, focus-of-expansion (FOE) and the focus-of-contraction (FOC), located on

the opposite poles of the spherical field. Any rotation about the axis of translation disturbs this arrangement. The rotation around an axis generates a motion field that whirls around this axis. The vector passing through these two singular points indicates the translation direction. As the rotational velocity increases, the FOE and FOC move around the circle towards each other (see Fig.(4.4)).

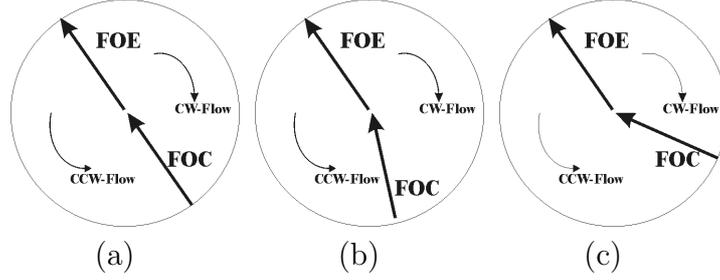


Figure 4.4: (a) FOE and FOC lines by pure translational motion of the camera system. (b) relative positions of both lines change as the camera system follows a curvilinear path with a rotation to the left side. (c) rotational movement to the right side.

Fig.(4.5) visualizes the spherical flow fields of translational and rotational motion. A spherical sensor undergoing rigid motion in a static environment induces a motion field that is measurable as a vector field on the image sphere. The motion of the sensor can be written in terms of three translational (t_x, t_y, t_z) and three rotational components (w_x, w_y, w_z). The two dimensional flow field, measured around the orthogonal equator of the spherical motion field, has a sine-like distribution. Formally, the circumferential motion field ($F(\Theta)$) due to the sensor translating in direction Λ with velocity \vec{v} and rotating with angular velocity $\vec{\omega}$ can be written as follows:

$$F(\Theta) = A \cdot \sin(\Theta - \Lambda) + \vec{\omega} \quad (4.4)$$

$$A = \frac{-\|\vec{v}\|}{r}$$

Note that this equation is analogous to Eqn.(4.3), which was obtained analytically. The amplitude (A) of this sine distribution is proportional to magnitude of the translational velocity and the distance (r) of the 3D scene point to the center of projection. Any rotation around the vertical axis adds a constant term ($\vec{\omega}$) to this distribution. The phase deviation of the sine, Λ , indicate the translation direction and the zero-crossings of the distribution are the FOE and FOC.

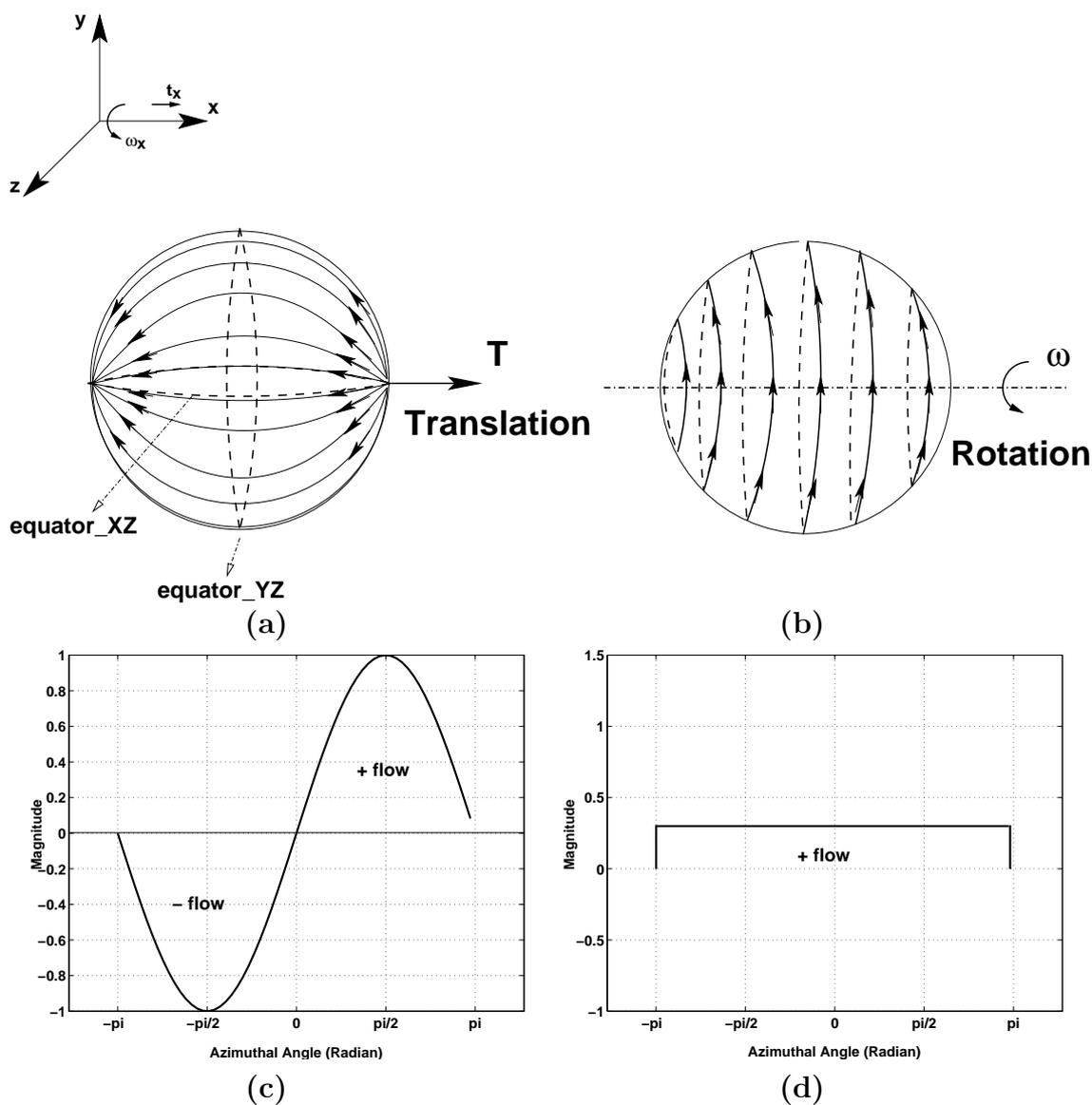


Figure 4.5: *The motion field on a sphere. (a) The sphere translates in x direction of a 3D space. It is assumed that the scene is homogeneously structured. (c) The flow distribution on the equator_{XZ} of the sphere renders a measurable sine form. (b) The rotation of the sphere around one of its axes produces uniform flow distribution on the sphere. (d) The flow distribution on the equator_{XZ} is uniform and proportional to the rotational velocity.*

4.2 Visual Clues inherent in the Catadioptrical Image Sequences

The structural analysis of spherically projected motion fields have been first introduced in the description of qualitative flow invariants by Koenderink [66]. Nelson and Aloimonos [85] have proposed to use the spherical projections of a scene to calculate the translational and rotational components of the motion fields with a qualitative analysis of flow fields. Both of the mentioned papers are theoretical in nature and their analyses were done by projecting standard synthetical images onto a sphere.

Our analysis, clearly inspired by the former work, depends on real image sequences produced by a camera-robot system equipped with a catadioptrical omnidirectional camera, moving in an indoor environment. The catadioptrical camera captures omnidirectional images, which can be mapped onto a sphere directly. The spherical image sequence can then be analyzed by estimating the optical flow field and determining its structural parameters. Our analysis differs from the previous work also in that we consider the dense flow field on the whole rather than just the equatorial field of the spherical images as proposed in [85]. One of the significant differences to the previous proposals is that our method implements a non-linear fitting algorithm to determine the best fitting sine function to the radial flow distribution. This allows for an exact description of the rotational parameters as well as the translational velocity for a chosen point with a known radial distance. It also allows for a thorough residual analysis for the task of detecting independent motion, to which Section 4.3 will be devoted.

The omnidirectional optical flow field consists of angular and radial components with respect to the center of projection. The radial component is induced by the pitch and roll motions of the camera on an uneven horizontal plane. The angular component represents the egomotion on the horizontal plane. In our approach we assume that the camera-robot system moves on an even horizontal plane and the radial component of the flow is negligible compared to the angular component.

The catadioptric sensor capable of capturing a 360° view of the horizontal plane allows for a global analysis of the angular flow field on this plane. As noted in the discussion about the spherical motion field (see Section 4.1), in such an omnidirectional optical flow field, the Focus of Expansion (FOE), the points on the motion field where the flow vectors seem to be emerging, and the Focus of Contraction (FOC), the points where the flow vectors are vanishing, are always in the field of view and span an angle of 180° , if the camera motion is purely translatory. By pure rotational motion neither FOE nor FOC are in the field of view. In real robotic applications, though, there is mostly an additional translatory

component in the motion field. A rotational component in a translational motion field causes the relative positions of FOE and FOC to vary (see Fig.(4.4)).

Virtually, a camera-robot system moving in a real indoor environment views enough structure so that the optical flow field patterns imply the type of the egomotion. Translational or curvilinear motion induces global flow fields with singular points indicating the translational direction. This fact enables determining the navigation direction, which can be interpreted as a pattern recognition problem, that relies on the estimation of the similarity of the flow field to an ideal sine-like distribution. This distribution may be disturbed by the rotational component. As the rotation becomes significant, the distribution deviates on the amplitude axis. Then the pure rotation causes an approximately constant distribution with an amplitude proportional to the magnitude of the rotational velocity.

Our egomotion estimation scheme can be summarized as follows:

- **The global motion field is estimated by the calculation of optical flow:** The algorithm proposed in Chapter 3 is used for the accurate and time efficient OFC. This method considers the geometric distortion of the projected images.
- **The omnidirectional flow field is analyzed and an angular flow histogram is formed:** The distribution of the flow data on different angular directions is determined by averaging the clockwise and counterclockwise flow distribution on each radius.
- **Estimation of the best fitting sine-distribution on this flow data determined by nonlinear optimization:** The algorithm minimizes the error function:

$$E = \sum_{\Theta} \|F(\Theta) - (A \cdot \sin(\Theta - \Lambda) + \omega)\| \quad (4.5)$$

for a discrete set of circumferential angles Θ . The algorithm employs a known nonlinear optimization scheme (see Appendix B) to minimize this error function. The optimization parameters are; A , the amplitude, Λ , the deviation from the pure translation and ω , the rotational velocity.

- After the **residual analysis**, the algorithm compares the translational component (A) and the rotational component (ω) and decides about the type of the egomotion (translational and/or rotational). The pure translation has a significantly larger amplitude component (A) compared to the rotational component (ω). The decision boundary between translation and rotation is measured by using the ratio $\frac{|A|}{|\omega|}$.

- If the flow pattern implies a significant translational motion, **the algorithm estimates the singular points of the distribution**. These points indicate the translation direction of the robot-camera system.

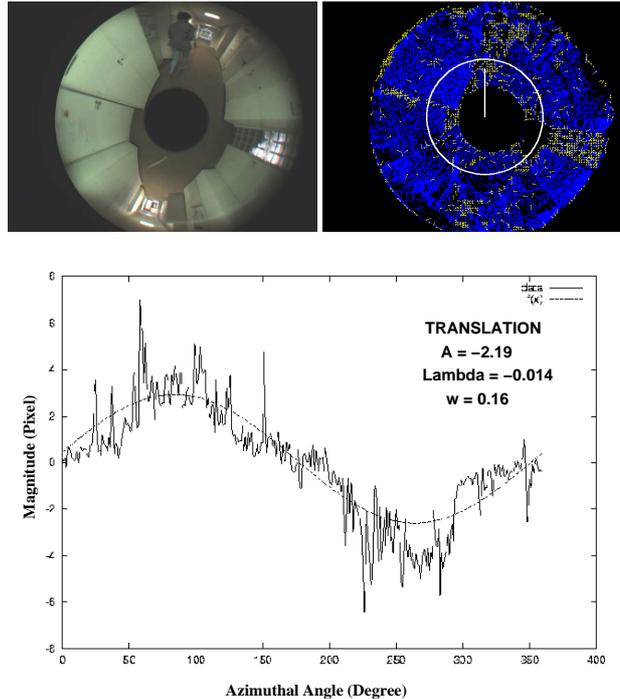


Figure 4.6: *The estimation of egomotion direction begins with calculating the omnidirectional flow distribution. If this distribution is similar to a perfect sine-form, the system decides that the motion is translational. The component **Lambda** represents the singular point of the distribution (here it is the deviation from the pure translation in direction 90°), which is the actual translation direction. The component **w** is the rotational component in radian.*

Since this analysis is based on the recognition of global structural motion patterns, it is far less sensitive to error or noise in the estimation of the motion field. Local occlusions of features or locally dynamic scene changes do not result in major errors in the global motion detection task. These local changes occur as residuals of the global fitting function. Moreover, the analysis of this residual field allows accomplishing further tasks like the independent motion estimation.

4.3 Independent Motion Estimation

Detecting moving scene objects is an important task for a mobile autonomous robot. Several other navigational tasks, e.g. object tracking, collision avoidance, coordinated navigation, docking depend upon the success of this detection. Moving objects can be detected from a static camera-robot system visually by frame differencing and change detection. But the task is more complicated if the camera-robot system moves, too. Even if the egomotion of the camera system is known, the structure of the scene and the 3D parallax caused by the stationary objects appearing in the scene as the robot moves, cause ambiguities.

The task of detecting moving scene objects from a moving camera-robot system is called the independent motion estimation (IME). Several attempts have been made to solve the IME problem: [6] proposed a method which combines stereoscopic depth measurements from a pair of cameras, and the visual motion measurements obtained by estimating the sparse normal flow field, to segment the independent motion. In [73], this stereoscopic measurement is replaced by a monocular stereoscopy. The method assumes that at least some part of the moving scene is composed of planar objects. In [62], the authors introduced a framework for the estimation of independent motion in 2D and 3D scenes with dense or sparse 3D parallax. Generally, the methods proposed so far utilize perspective projection with a standard camera to capture the image scenery. Also, they assume specific egomotion of the camera or they employ a 2D registration technique to estimate the egomotion. In most cases, a rigid semi-planar background motion is assumed (e.g. ground image taken from a flying aeroplane camera), that facilitates the background registration task.

Our approach differs from the known methods in several ways. We analyze a larger part of the scene, which we capture with an omnidirectional camera. Therefore, we can interpret and segment the moving parts of the scene and follow them for a longer period of time. This allows for a thorough analysis of the image motion regions in the spatial and temporal dimensions. The induced changes in the global flow field as the omnidirectional camera undergoes a specific egomotion, reveal the background motion model, which we estimate by a statistical analysis and a numerical fitting function, as described in Section 4.2.

The residual analysis detects the outliers of the omnidirectional flow field. Since all depth changes in the scene cause residuals, we analyze the flow distribution further to distinguish between the depth changes induced by independent motion and nearing static objects. Temporal changes of the depth give clues about their characteristics. The depth changes caused by the static objects have a temporal development, which is conformal with the background motion, whereas independent object movement deviates strongly from the assumed background

spatio-temporal change. This fact will be exploited to construct a probabilistic model for segmenting and tracking these objects.

4.3.1 Interpretation of Outliers in the Independent Motion Segmentation Problem

While the smooth background motion occupies the larger part of the omnidirectional motion field of a moving camera-robot system, independent motion arises as regions of disturbances in this global pattern. These regions can be detected by the IME - Analysis. But the stationary objects, which appear closer to the camera as the system moves towards them, produce also optical flow vectors with higher magnitudes, which arise as disturbances in the global flow pattern, too. This is the *depth parallax*, addressed earlier.

The stationary objects cause greater flow vectors varying in their magnitude but not in their direction, which should be consistent with the egomotion direction. In other words, the objects at different depths display significant differences in the flow field, due to the depth parallax, but their motion is constrained to lie on a one dimensional locus in a two dimensional space, that agrees with the egomotion direction. Independently moving objects, on the contrary, are not constrained to lie on this locus. Exploiting this fact, we build a probability based detection system. The outlier detection yields hypothetical locations of the independent motion, and a subsequent probabilistic tracking scheme verifies them. The proposed method can be outlined as follows:

- The omnidirectional optical flow field is estimated using two sequential frames of an image sequence.
- The global pattern of the optical flow field is analyzed to determine the clockwise and counterclockwise flow vectors. The distribution of these vectors may indicate the FOE and FOC if the egomotion has translational components, as described in Section 4.2.
- The mean value of the flow vector magnitudes ($\|V_r(\vec{\Theta})\|$) in each angular direction is estimated. This yields an angular distribution function F_{ang} , which can be written as follows:

$$F_{ang} = \frac{1}{r_{max}} \sum_{r=0}^{r=r_{max}} V_c(\Theta) \quad (4.6)$$

where

$$V_c(\Theta) = \begin{cases} \|\vec{V}_r(\Theta)\| & \text{if clockwise} \\ -\|\vec{V}_r(\Theta)\| & \text{else} \end{cases} \quad (4.7)$$

- The distribution F_{ang} resembles a sine form if the egomotion has translational components. Otherwise, it is a negative or a positive constant function, depending on the rotation direction.
- An iterative non-linear numerical fitting method (Levenberg-Marquardt algorithm) is used to approximate the F_{fit} . The estimated sine function represents the background motion model $F(\Theta)$ given in Section 4.1 with Eqn.(4.4)

$$\hat{F}_{ang} = F_{fit} = F(\Theta) \quad (4.8)$$

Using the background motion model given with Eqn.(4.8), sequential images can be transformed into a common coordinate system, and can be analyzed further to find the regions with high residual fields. In our system, the warping function (\mathcal{W}) transforms the first image (I_1) into a synthetic second image using the transform function \hat{F}_{ang} . This synthetic image (here represented with I_{warped}) can then be compared with the real second image directly, which yields a residual image (I_R). Significantly higher residual values on I_R indicate independent motion regions.

$$I_{warped} = \mathcal{W}(I_1, \hat{F}_{ang}) \quad (4.9)$$

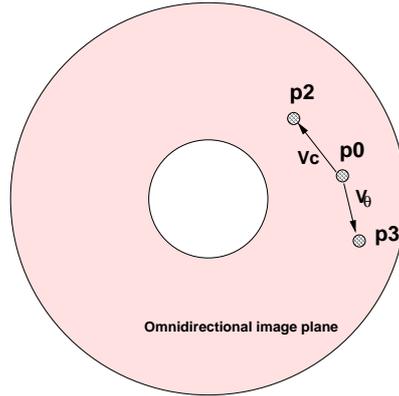


Figure 4.7: *Pixel displacement during the warping operation.*

Warping The warping (\mathcal{W}) of the first image onto a synthetic second image is a pixelwise operation. Let $p_0(r, \theta)$ be the pixel position of one image point in the

first image and $p_2(r, \theta)$ the position in the synthetic second image (I_{warped}). We can estimate this position as follows:

$$p_2(r, \theta) = p_0(r, \theta) + V_c \delta t \quad (4.10)$$

where V_c represents the angular velocity estimated using the background motion model for any radius, r , and δt the time between successive frames. The image point $p_0(r, \theta)$ which has the actual angular velocity V_θ , that is significantly different from V_c , will move to a point $p_3(r, \theta)$ in the real second image.

$$p_3(r, \theta) = p_0(r, \theta) + V_\theta \delta t \quad (4.11)$$

If the direction of $|p_3(r, \theta) - p_2(r, \theta)|$ is conformal with the egomotion direction, the residue on this image point indicates most probably a depth change. Otherwise we can hypothesize that this image point undergoes an independent motion. We use this hypothesis for the residual image construction. Considering an image I_R , the brightness value of each pixel should represent the probability of having independent motion or significant depth parallax at that pixel position. The following algorithm is used to generate such a residual image, I_R :

```

for every  $p_0$  in image  $I_R$ 
if  $|p_3(r, \theta) - p_2(r, \theta)| < Thresholdvalue$ 
    then  $I_R(p_0) = 0$ 
else if Direction of  $|p_3(r, \theta) - p_2(r, \theta)|$  is conformal to egomotion
direction
    then  $I_R(p_0) = |p_3(r, \theta) - p_2(r, \theta)|$ 
else if Direction of  $|p_3(r, \theta) - p_2(r, \theta)|$  is NOT conformal to
egomotion direction
    then  $I_R(p_0) = CONSTANT * |p_3(r, \theta) - p_2(r, \theta)|$ 

```

This algorithm assigns higher brightness values to the pixels with independent motion and sudden depth changes. A constant factor, which should be greater than 1.0, increases the brightness value for the region with a negative residual value, which indicates IM. The sign of the residue indicates the direction of the actual image point displacement. Since we assume that such regions may indicate independent motion with a higher probability, the brightness value is increased additionally. Independent motion detection results then in a task of thresholding higher brightness regions of the residual image.

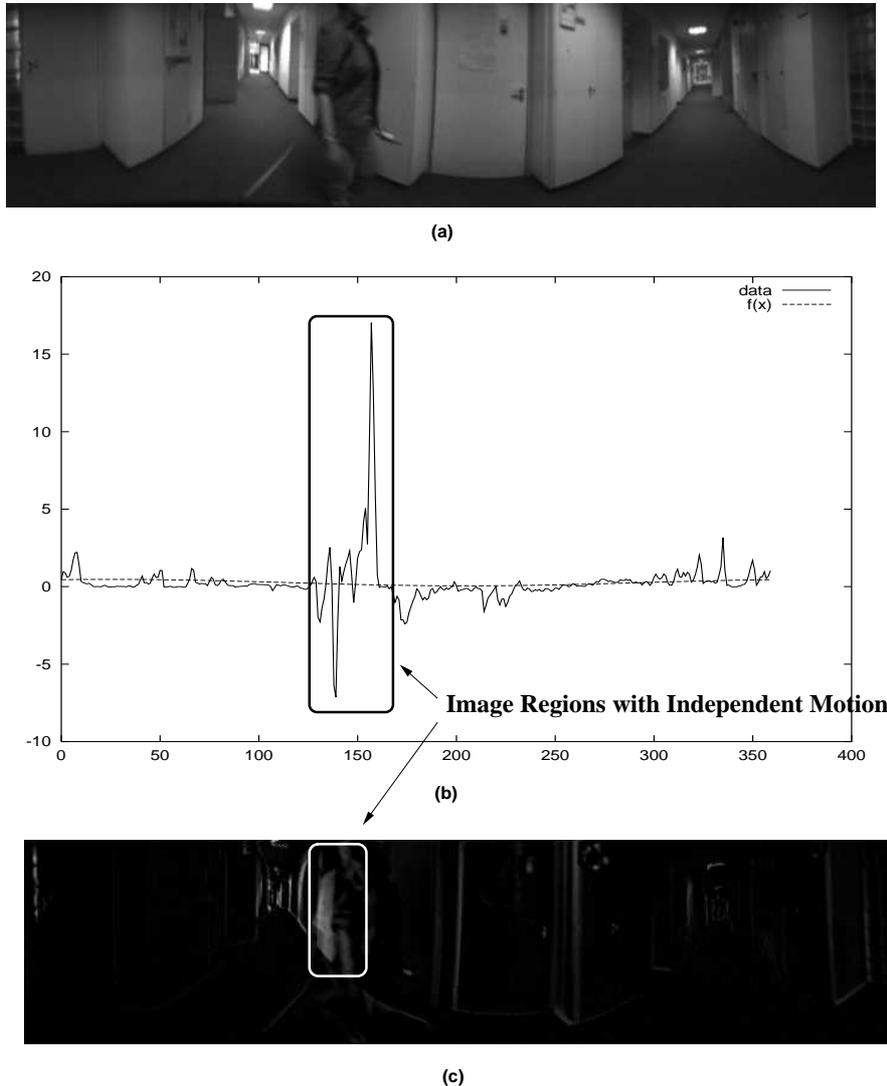


Figure 4.8: **Independent motion segmentation:** *The first picture shows one frame of a sequence captured from an omnidirectional camera system moving in a corridor. The image is represented in polar coordinates to allow for an easier comparison with the angular flow distribution function given in (b). The person that can be seen in the picture moves in the opposite direction and with a significantly different velocity. The second diagram (b) shows the angular flow distribution and the fitted sine function. The fitting was successful because the camera motion had a significant translational component. The third image (c) shows the residual regions of the image after warping and background suppression. Note that the region with the independent motion has higher brightness values. Depth parallax due to the nearing door frames, cause residual regions, too, but they are represented with lower brightness values in the resulting residual image.*

4.3.2 Temporal Tracking using the Probability: Particle Filters for Independent Motion Detection

The ambiguity in IME caused by the sudden depth changes due to the nearing static objects in the scene and by the independent object motion, can be resolved by analyzing the image sequences in time. The residual images, which are produced with the IME algorithm as described in the previous section, can be normalized to represent the probability distribution of having an independent motion. The higher the brightness value at some region of the normalized residual image, the higher the probability that it is a 2D projection of an independent motion in the scene. The temporal evolution of such image regions shows significant deviations from the background. A relevant tracking algorithm, like particle filter tracking, which can cope with varying probabilities and cluttered measurements, can then be employed to track such regions.

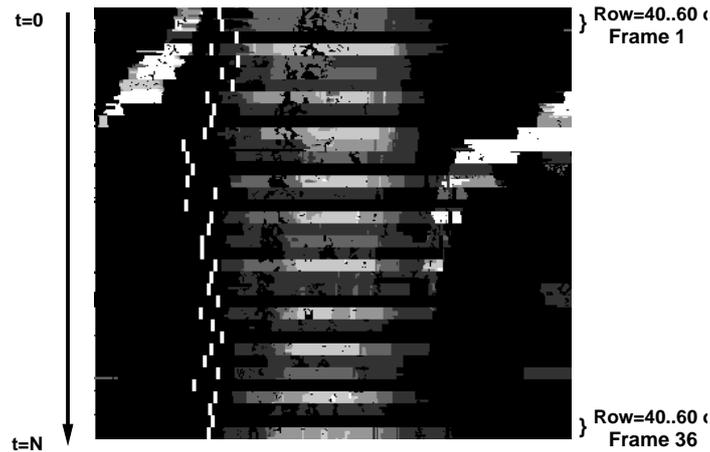


Figure 4.9: An image of 36 sequential residual image slices concatenated in the vertical dimension. The slices are extracted from the middle rows (rows 40 to 60) of each residual image, which were formed using the procedure described before. The slices are concatenated in the vertical dimension as new frames are provided in time. The brighter values indicate higher depth change and independent motion. One can notice that the independent motion regions evolve significantly differently than the background scene object motion.

In the following, we will first give a brief review of particle filtering. Then we describe our probabilistic model of tracking independent motion. More detailed information on particle filters can be found in [97].

A Brief Introduction to Particle Filter Tracking A probability density function describing the likely state of the objects is propagated over time using a dynamic model. The measurements influence the probability function and allow new objects to be incorporated into the tracking scheme.

Notation and Methodology

An object is characterized by a state vector $x \in X$. Assuming that we are unable to know the exact state, we describe the knowledge about an object by a probability function $P(x)$.

Dynamics: As the observed scene changes over time, the probability function evolves to represent the altered object states. The dynamics of the evolution is described by a stochastic differential equation, where the deterministic part of this equation models the system knowledge. The stochastic part allows us to model the uncertainties. Applying the Markov assumption, the density function $p(x_t)$ depends only on the immediately preceding distribution $p(x_{t-1})$. The process dynamics is determined by the conditional process density $p(x_t|x_{t-1})$.

Measurements: Let z_t be the measurement at time step t with a history $\mathcal{Z}_t = \{z_0, z_1, \dots, z_t\}$. Instead of the density $p(x_t)$, we distinguish between the density $p(x_t|\mathcal{Z}_{t-1})$ and the density $p(x_{t-1}|\mathcal{Z}_{t-1})$. Using the Bayes Rule, we can estimate the posterior density as follows:

$$\begin{aligned} p(x_t|\mathcal{Z}_t) &= \frac{p(z_t|x_t, \mathcal{Z}_{t-1})p(x_t|\mathcal{Z}_{t-1})}{p(z_t|\mathcal{Z}_{t-1})} \\ &\cong kp(z_t|x_t, \mathcal{Z}_{t-1})p(x_t|\mathcal{Z}_{t-1}) \\ &\cong kp(z_t|x_t)p(x_t|\mathcal{Z}_{t-1}) \end{aligned}$$

where k represents the normalization factor. The term $p(z_t|x_t)$ stands for the observation model. It represents the likelihood that a state x_t causes a specific measurement z_t . The term $p(x_t|\mathcal{Z}_{t-1})$ represents the prior density. It is the result of applying the dynamic model to the posterior density $p(x_{t-1}|\mathcal{Z}_{t-1})$ of the previous steps:

$$p(x_t|\mathcal{Z}_{t-1}) = \sum_{x_{t-1}} p(x_t|x_{t-1})p(x_{t-1}|\mathcal{Z}_{t-1})$$

A particle filter tracking scheme is made up of two major steps:

1. calculate the prior density function $p(x_t|\mathcal{Z}_{t-1})$ using the dynamic model
2. evaluate the posterior density function $p(x_t|\mathcal{Z}_t)$, given the measurements.

Weighted Sampling: In a particle filter we recursively approximate the posterior $p(x_{t-1}|\mathcal{Z}_{t-1})$ as a set of N weighted samples $\{x_{t-1}^{(i)}, \pi_{t-1}^{(i)}\}$, where $\pi_{t-1}^{(i)}$ is the

weight for particle $x_{t-1}^{(i)}$. Given this approximate representation, we obtain a Monte Carlo approximation of the Bayes filtering distribution:

$$p(x_t | \mathcal{Z}_t) \approx k p(z_t | x_t) \sum_i \pi_{t-1}^{(i)} p(x_t | x_{t-1}^{(i)}) \quad (4.12)$$

One way to view a particle filter is as an importance sampler. Specifically, N samples of $x_t^{(i)}$ are drawn from the following proposal distribution q

$$x_t^{(i)} \sim q((x_t) \triangleq \sum_i \pi_{t-1}^{(i)} p(x_t | x_{t-1}^{(i)}) \quad (4.13)$$

and then weighted by the likelihood, i.e.

$$\pi_j^{(j)} = p(\mathcal{Z}_t | x_t^{(j)}) \quad (4.14)$$

A particle filter for independent motion detection Our system employs an adaptive particle filter [45] to detect and track the independent motion region, which has been segmented by the algorithm described in Section 4.3.1.

The state density function for the object with the independent motion is represented by $p(x_t)$. The state vector at timestep t is x_t . The vector has three components $x_t = [r, \theta, V_R]^T$, where r, θ represent the radial and angular position of the mass point of the object and V_R the radial velocity.

The measurements are obtained from the residual image sequence, which will be represented with $I_R^0, I_R^1, \dots, I_R^t$. Using the above notation, we can write the posterior probability distribution of an independent motion region $p(x_t)$:

$$p(x_t) = p(x_t | I_R^0, I_R^1, \dots, I_R^t)$$

Using the Bayes Rule and defining α as the normalization constant, we can reformulate the filter as follows:

$$\begin{aligned} p(x_t) &= \alpha p(I_R^t | x_t, I_R^0, I_R^1, \dots, I_R^{t-1}) p(x_t | I_R^0, I_R^1, \dots, I_R^{t-1}) \\ &= \alpha p(I_R^t | x_t) p(x_t | I_R^0, I_R^1, \dots, I_R^{t-1}) \\ &= \alpha p(I_R^t | x_t) \sum_{x_{t-1}} p(x_t | I_R^0, I_R^1, \dots, I_R^{t-1}, x_{t-1}) p(x_{t-1} | I_R^0, I_R^1, \dots, I_R^{t-2}) \\ &= \alpha p(I_R^t | x_t) \sum_{x_{t-1}} p(x_t | x_{t-1}) p(x_{t-1} | I_R^0, I_R^1, \dots, I_R^{t-2}) \\ &= \alpha p(I_R^t | x_t) \sum_{x_{t-1}} p(x_t | x_{t-1}) p(x_{t-1}) \end{aligned}$$

The dynamic model of the state transition function considers not only the radial and angular displacement of the segmented image regions, but also their velocity. We assume that the acceleration of the independently moving objects are negligible compared to the frame rate.

The sample $\pi_{t+1}^{(i)}$ evolves as follows:

$$\pi_{t+1}^{(i)} = \pi_t^{(i)} + \delta t \times \dot{\pi}_t^{(i)} + N\left(\frac{\sigma}{w}\right) \quad (4.15)$$

where $N(\sigma/w)$ represents the Gaussian noise distribution of the measurement with a standard deviation σ and mean w . The mean value of this distribution is estimated directly from the measurement residual image. The time interval δt represents the time between two frames and in our calculation we set its value to 1.

Each particle is evaluated using the measurement density function. The image regions, which are segmented using the scheme described in Section 4.3, have higher grayvalues at the image coordinates with significant outliers. The outliers indicate independently moving scene objects or significant depth parallax. During the tracking scheme, the segmented image is normalized to represent a 2D probability field of having independent motion. The image regions with higher grayvalues represent higher probability of having independent motion. Then, the evaluation step of the particle filter estimates the mean value of this 2D probability field.

The resampling stage employs the adaptive particle filtering, as proposed in [45]. The adaptation scheme chooses a small number of samples if the posterior density is focused on a small part of the state space, and it chooses a large number of samples if the state uncertainty is high. It also allows to include newcoming scene objects into the tracking scheme.

The state and the measurement vectors in our system represent the 2D position and the velocity of the tracked objects. It is possible to expand this system, e.g. by adding 3D information obtained from a 3D sensor, which can facilitate the discrimination of depth parallax and independent motion regions.

The output of the particle filtering is a set of dense particle clouds, as can be seen in Fig.(4.10). The white dots represent the position of the particles. For the further processing, we employed a clustering technique and estimated the mass points of the clusters to determine the tracked object position. It is then possible to draw the circular or the rectangular region including the object to be tracked. The tracking system is tested on different situations. If the scene does not include objects with independent motion, the particles do not converge onto a region. This situation can be seen in Fig.(4.11), where the person originally seen in Fig.(4.10) moves out of the scene. The residual image does not contain significant

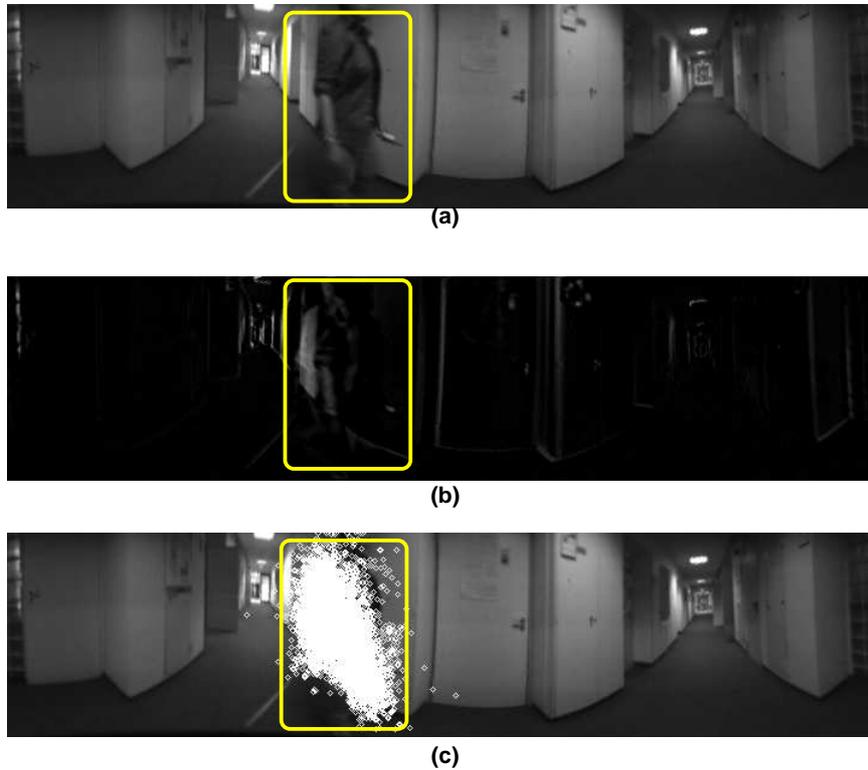


Figure 4.10: *Resulting particle image projected back on original image.*

regions, which detains the particle filter to converge. The system assumes that there is at least one region, which moves with significant different velocity than the background of the scene. Fig.(4.12) depicts the situation where this assumption does not hold. The person, seen in the figure moves in the same direction and with a very similar velocity as the camera. Also, the image region, where the person is located, is poorly illuminated. Therefore, the intensity gradient is not significant enough for a reliable optical flow estimation. In this example, the tracking system does not converge since the above mentioned errors cause the measurement image of the particle filter include only nonsignificant residual regions.

The system is tested on image sequences with one independent motion region at most. It is straight forward to extend it to track multiple objects by applying additional particle filters. The system should then suppress the region of the residual image progressively, as soon as one filter converges onto it. This allows the following particle filters to converge on other regions with independent motion. It is also possible to define regions of convergence for the parts of the scene where it is expected to find independent motion.



Scene without independent motion



Distribution of the particles on the image

Figure 4.11: *A scene without IM. The resulting particle image shows no convergence.*



Scene with a non-significant independent motion



Distribution of the particles on the image

Figure 4.12: *IM Object moves out of the scene and becomes non-significant. The resulting particle image shows no convergence.*

4.4 3D-Omnidirectional View

In addition to the temporal probabilistic analysis introduced in the previous section, one can estimate the 3D structure of a scene using additional sensory modalities to disambiguate the process of independent motion segmentation. In this section, we will briefly review the standard 3D detection techniques, as well as a recent technology to achieve 3D information in real time. A more detailed information on 3D image processing and reconstruction can be found in the respective literature, e.g. [44], [39] and [54].

The human eye can perceive the 3D-Structure of an environment by a sophisticated internal evaluation of the two 2D images, formed on two spatially separated retinas. Although the means of this internal biological evaluation is not exactly known, it is obvious that we use two slightly varying fields of view supplied by our left and right eye to determine the depth of a scene. This technique, called *stereo vision*, has also been conventionally exploited by most of the artificial 3D vision systems. The two images of the same scene taken from two distinct cameras are matched and using the known spatial distance between the two similar cameras, their vergence angle, and other internal parameters, the 3D structure of the scene can be estimated in principle.

The major problem in estimating the 3D information from 2D images, is caused by the projection operation itself. The projection of a 3D scene onto an image plane maps an infinite line onto a point in the image. For a fixed viewpoint and direction, infinitely many continuous and discontinuous 3D configurations of points could project on our retina to form an image of that environment. Therefore, the depth information of the scene is lost inevitably, when it is projected. But if two images from different viewpoints can be placed in correspondence, the intersection of the lines of sight from two matching image points determines a point for the triangulation (see Fig.(4.13)).

A simple 3D algorithm can be described as follows:

- Take two images separated by a baseline
- Identify the points to be matched between the two images
- Use the simple triangulation to derive the two lines on which the physical world point lies
- Intersect the lines, the resulting point is the 3D coordinates of the points

The challenging part of this simple algorithm is the second step: determining the correspondences between the feature points in two respective images. Even if we let the epipolar constraint limit the search space of all possible matches between

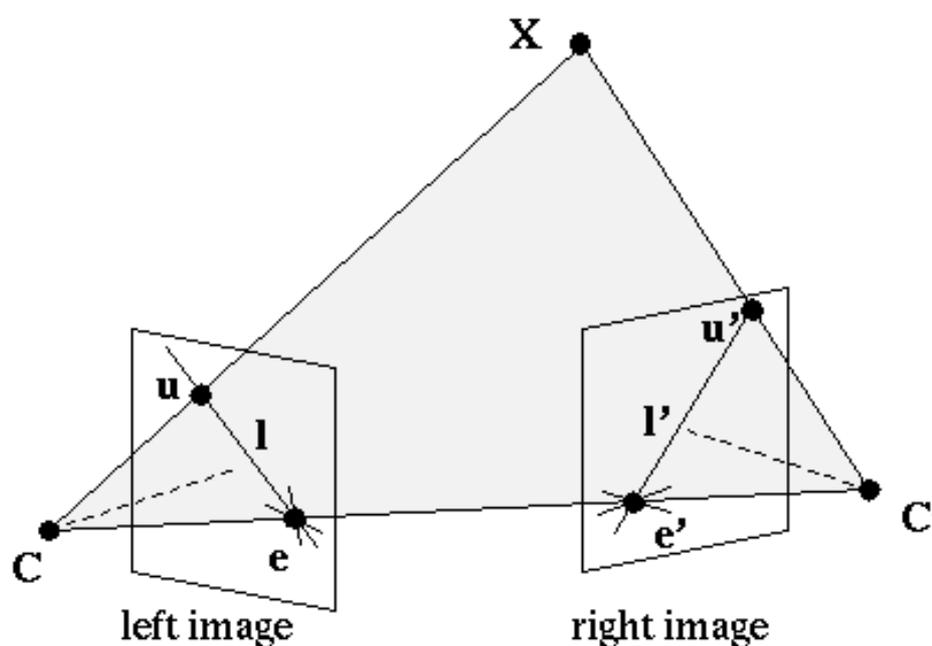


Figure 4.13: **Epipolar Geometry for Stereopsis:** *The line connecting optical centers C and C' is called the baseline. Any scene point X observed by the two cameras and the two corresponding rays from optical centers C and C' define an epipolar plane. This plane intersects the image planes in the epipolar lines l and l' . The epipoles are the intersections of the baseline with the respective image planes. Let u and u' be the projections of the scene point X in the left and the right image respectively. The ray CX represents all possible projections of the point X to the left image, and is also projected into the epipolar line l' in the right image. The point u' in the right image that corresponds to the projected point u in the left image must thus lie on the epipolar line l' in the right image. This fact constrains the search space for the correspondence of the feature points. (This figure is redrawn from [108])*

the images, there is a rest amount of uncertainty for a mismatch, that results from several factors. One of these factors is the camera calibration problem. If we use two different cameras, we have to calibrate them so that we get exactly the same camera distortion on the two images. In this way, we can assure that the features of the respective images that are to be matched, are distorted with the same effects and in the same amount. This calibration problem can become very complex and time consuming (for a more detailed information on the calibration problem please see the chapter on stereopsis in [44]). The other problem is that the change of the view point may cause the feature points to disappear or suddenly appear so that again a mismatching follows.

One can extract the 3D information not only using stereo vision from two calibrated cameras, but also using monocular and dynamical vision (optical flow). This method is called as *3D-From-Motion* or *Stereo-From-Motion*. While the general problem of correspondence remains, the calibration problem mentioned above is eliminated, since there is only one camera that takes the images at two different points in time. The optical flow field estimated from these sequential images inherits the depth information of the scene. The major drawback of the method is that the camera has to move with a single and continuous, non-accelerating, precisely controlled motion. Any other motion in the scene than the camera motion, disturbs the 3D detection and causes ambiguities. In the case of the independent motion detection task, these ambiguities are always present which motivates us to look for alternative 3D detection techniques.

4.4.1 Recent Technology for 3D Acquisition

There are several alternative 3D imaging techniques like employing structured light, ultrasonic ranging or laser range finders. The basic idea of structured light methods is to project patterns of light onto an object and reconstruct the surface geometry from the distortions of such patterns that are produced on object surfaces. Although this is a simple method to use, it has low spatial resolution. The patterns, which are to be detected become sparser with the distance. Laser range finders are active sensors, which work on the principle that the surface of objects reflects laser light back towards a receiver which then measures the time (or phase difference) between transmission and reception in order to calculate the depth. Most of the laser range finders work at long distances with low depth resolution in a specific geometrical dimension. Recent developments allow for wide angle depth measurements, e.g. the 3D laserscanner developed in Fraunhofer-AIS ([117]). Here, a new design allows for capturing wide fields of view, which is realized by means of mechanical rotations of a single 2D laser scanner in vertical dimensions. These laser-scanners are well suited for 3D reconstruction of

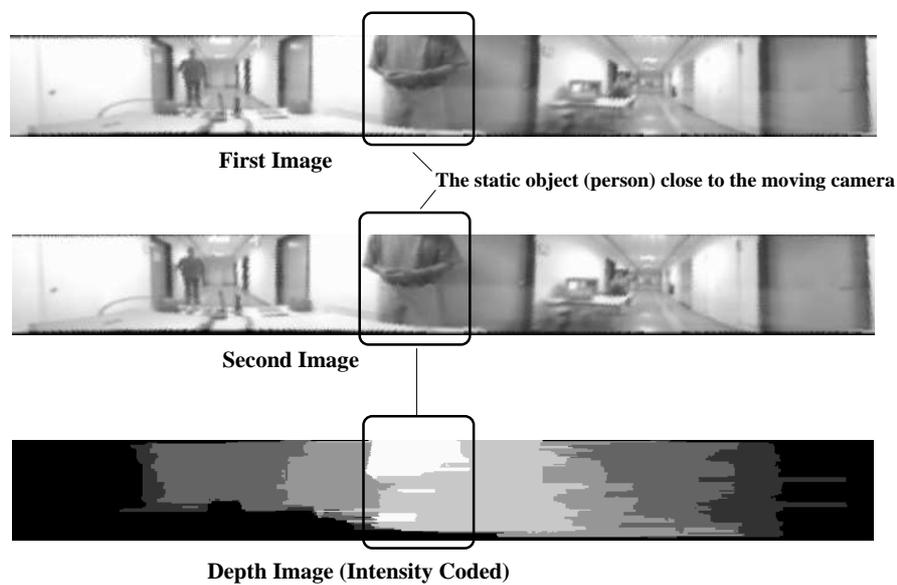


Figure 4.14: *The optical flow field inherits the 3D information about the static scene. The closer parts of the scene induce higher flow regions as the camera moves continuously in an indoor environment and captures sequential images. The resulting depth image pixels become brighter if the part of the scene gets closer to the camera.*

motionless scenery but not advantageous for time critical applications.

The recent developments on active range finding arrays and line sensors present encouraging results in real time active depth measurements, which are interesting for time critical scene analysis and motion detection applications. These sensors are capable of parallel and area-wide depth measurements. Their design is mostly based on the time-of-the-flight (TOF) principle. A modulated signal (mostly laser) is sent out and upon its return the echo time of the signal is measured. This echo time, which is measured at the receiver, is related to the distance of the structures in the environment.

The Photonic-Mixer-Device (PMD) consists of smart pixels which enable fast optical sensing and demodulation of incoherent light in one component. This active sensor illuminates the 3D scene by means of a modulated optical sender, and the returning signal is captured and mapped onto a PMD array for a parallel demodulation of the echo signal, TOF and distance evaluation. Due to the parallel operation of the array, the 3D measurement is in milliseconds-range, which allows for a real-time operation [102].

A single PMD sensor element has behind an optical input window, two photogates, which are isolated from each other by a thin oxide layer. Depending on the provided reference voltage, the photogates conduct in left or the right readout circuitry. These reference voltages control the output of the sensor element. If the incident light is constant (e.g. sunlight in the scene), the modulated signal generates charge carriers, which move to the left and the right at the same amount. The output of the sensor element, which consists of the difference of the reference modulation results in zero output. The output voltage depends directly on the phase delay between light and pixel modulation. This phase delay can then be evaluated to achieve the 3D information of the scene. The recent versions of the PMD sensor use Schottky-barrier photodiodes instead of photogates as indicated above. For more structural and detailed information on the PMD sensors, please see [102],[69].

Although the PMD sensor size is subject to ongoing research, the currently available array sizes map only a small portion of a real scene. This is a restriction for autonomous robot navigation. One possible solution is to expand the field of view using a reflective surface. The range preserving reflective surface family introduced by Chahl and Srinivasan([25]) seems to be a suitable candidate. This mirror family increases the field of view and preserves the range properties of the scene on the projected image. This property is achieved by the special mirror design, that keeps the linearity between the radial angle - the angle between the optical axis of the sensor and the incoming ray - and the angle of elevation - the angle between a ray of light impinging on the surface and the downward vertical-constant. This linearity is important for the 3D measurements using a PMD

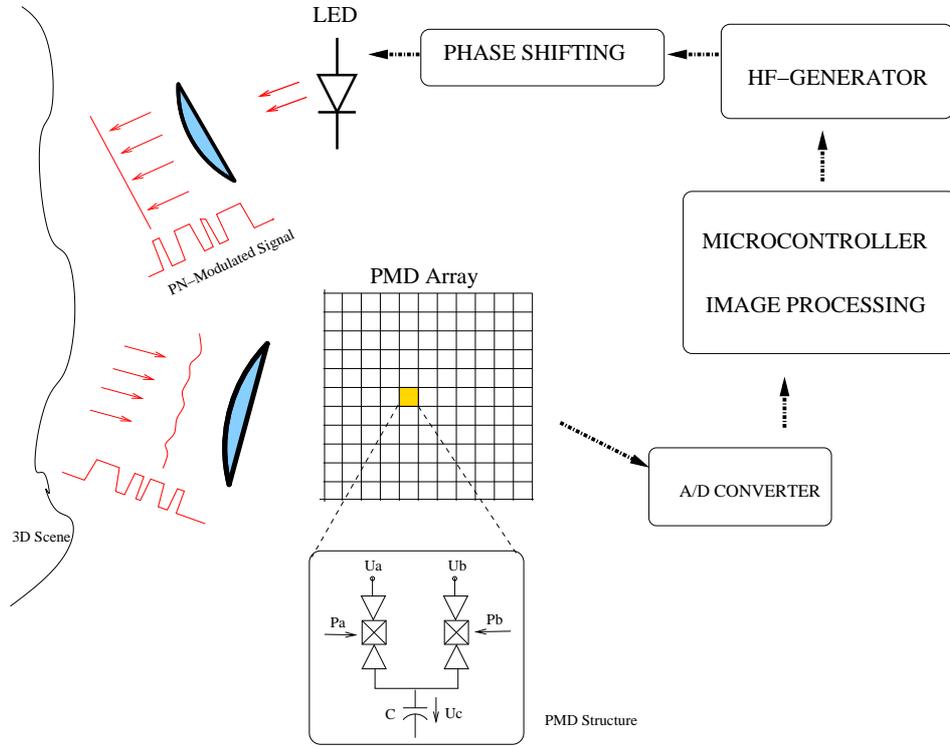


Figure 4.15: *PMD Principle (redrawn from [102])*

sensor, as the phase delay, which is essentially measured, is not distorted due to the refraction on the mirror surface.

4.4.2 Integrating 3D Measurement into the IME Scheme

The 3D information achieved from an active sensor can enhance the proposed visual analysis in several ways. Firstly, the egomotion detection can profit from a more accurate and direct depth estimation. In Section 4.2, we proposed to analyze the angular flow field distribution, and mentioned that the amplitude of the approximated sine distribution yields the translational motion component, which is scaled with the range of the 3D point. If the objects are equally distant, this scale factor (referred as r in Eqn. 4.4) does not effect the motion estimation. But, if 3D information about the objects of varying depth layers are available, one can determine the magnitude of the translational motion accurately and directly by solving for:

$$\|v_t\| = \frac{A}{r(\Theta)}, \quad (4.16)$$

where A represents the amplitude of the sine function and $r(\Theta)$ is the measured 3D distance of an object at an angular position Θ . The measurement of 3D position of the objects allows then for an accurate measurement of the translational motion of the camera-robot system.

The 3D information can be exploited for resolving the depth parallax problem, which is inherent in the independent motion estimation task. For the following discussion, consider the situation depicted in Fig.(4.16). A mobile robot equipped with an omnidirectional 3D sensor translates in the horizontal direction with a speed of v . The camera registers three points, P_1 , P_2 , P_3 , each of them located in different depth ranges from the center of projection. Let us consider that P_2 is on an object which moves independently in an arbitrary direction. Measuring the rate of change in three dimensions (elevation, longitudinal and depth) would yield velocity information on these three different points, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. Since point P_2 moves independently, it is expected that the velocity measurement \mathbf{v}_2 , yields significantly different magnitudes and directions in at least one of the three components of $\mathbf{v}_2 = [\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z]$. The rate of positional change of the 3D points P_1 and P_3 are similar even if they induce different 2D optical flow vectors. Segmenting the motion boundaries based on three components allows the unambiguous detection of independent motion regions for each frame.

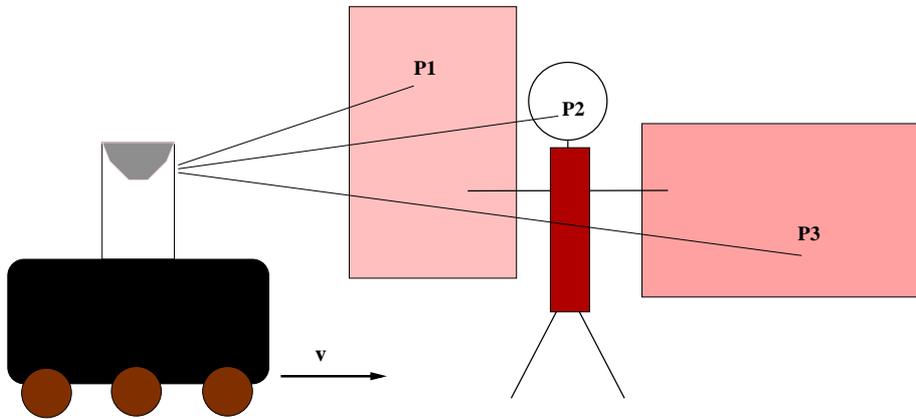


Figure 4.16: *Robot navigation in a sample 3D world*

Finally, the temporal tracking can be enhanced by integrating the 3D information into the particle filter. The state space \mathbf{x}_s of the filter can be given using the 2D position and the velocity of a tracked image object, as follows:

$$\mathbf{x}_s = [x, y, v_x, v_y] \quad (4.17)$$

The measurement model ($M(x^t|x^{t-1})$) of the system can then be represented with a multi-variate Gaussian, using the new measurement vector μ (including the 3D information coming from a 3D camera) as the mean value :

$$\mu = [x, y, z, v_x, v_y, v_z] \quad (4.18)$$

$$M(x^t|x^{t-1}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_g|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^t \Sigma_g^{-1}(\mathbf{x}-\mu)} \quad (4.19)$$

with the dimension d of the state vector, Σ_g representing the covariance matrix, which controls the measurement neighborhood in our system.

Chapter 5

Experimental Results

Catadioptric cameras provide extended views of an environment and capture larger parts of a background scene, which typically consist of static and non-static structures. Although the magnitudes of the optical flow vectors vary due to motion parallax or depth changes in a scene, the angular flow direction remains undisturbed during the egomotion on a horizontal plane, assuming that the significant part of the background consists of a static scene. This property of omnidirectional optical flow field facilitates the detection of egomotion direction and rotational velocity, and also gives clues about the regions where independent motion might have occurred. In the previous chapter, we exploited this fact theoretically and introduced methods for determining the egomotion of a mobile camera-robot system as well as the independent motion.

This chapter describes the quantitative analysis and the conducted experiments to substantiate the theoretical concepts. The first section of the chapter describes the experimental setup and the technical details of the mobile imaging system used for the experiments. The following sections are devoted to the quantitative results of the egomotion detection and the independent motion tracking experiments.

5.1 Experimental Setup

The mobile test platform is composed of the mobile autonomous robot, KURT2, and the omnidirectional catadioptric camera, VCAM 360, mounted vertically on the robot (see Fig.(5.1)). KURT2 is equipped with motor encoders and one electronic compass, which can deliver an estimation of the egomotion information. The catadioptric camera is designed so that the grayscale images include the 360° view of the horizontal plane. Its a specially curved mirror surface that preserves the linear angular relationship of incoming rays and their projections on the image



Figure 5.1: *KURT2 with the Omnidirectional Camera Setup*

plane (designed as proposed in [25]). The camera is placed 75cm above the upper plane of the robot platform. The center of the camera is placed at the distance of about 5cm from the front edge of the robot, where the rotation axis of the robot is at about 22cm, measured again from the front edge.

The camera sensor is capable of scanning progressively at 30 fps. This allows a reasonable navigation speed for the robot platform. The images have VGA-Resolution at 8 bit grayscale depth (640x480 pixels). For only some of the egomotion detection experiments, the direction measurements from the electronic compass and the odometry sensor (motor encoders) of KURT2 were available, and the visual measurements could be compared. The analysis of egomotion detection was designed for on-the-fly processing. Independent motion tracking was done in batch mode.

The structure of the implemented algorithm is shown in Fig.(5.2). The four implemented modules for image capturing, omnidirectional optical flow calculation, visual egomotion detection and independent motion detection interact with each other and yield regions-of-interest on each input image. The 2D maps of egomotion path and the independent motion regions can be obtained from the pixel positions of these regions-of-interest. The software, implemented in C++, runs on a computer board with the IntelPentiumM - 1.6GHz Processor. The egomotion detection and independent motion segmentation from one pair of images is completed in 0.71 sec. Note that the implementation aims at testing the theoretical concepts and is not optimized for real time applications, yet. One can

apply several strategies to reduce the above mentioned processing time:

- ⇒ The mentioned time measurement was achieved by measuring the system clock ticks of a standard Linux operating system. For relevant real time applications it is recommended to employ a real time operating system with a priority based task scheduling system.
- ⇒ The estimation time for optical flow depends directly on the image size and resolution since it is a pixelwise operation. Reducing the resolution and image size, or applying a content based analysis (e.g. directly assigning a constant background velocity for image regions without significant structures), should optimize the total processing time for time critical applications.
- ⇒ The processing time can also be reduced by using a specialized hardware design for the proposed algorithm.

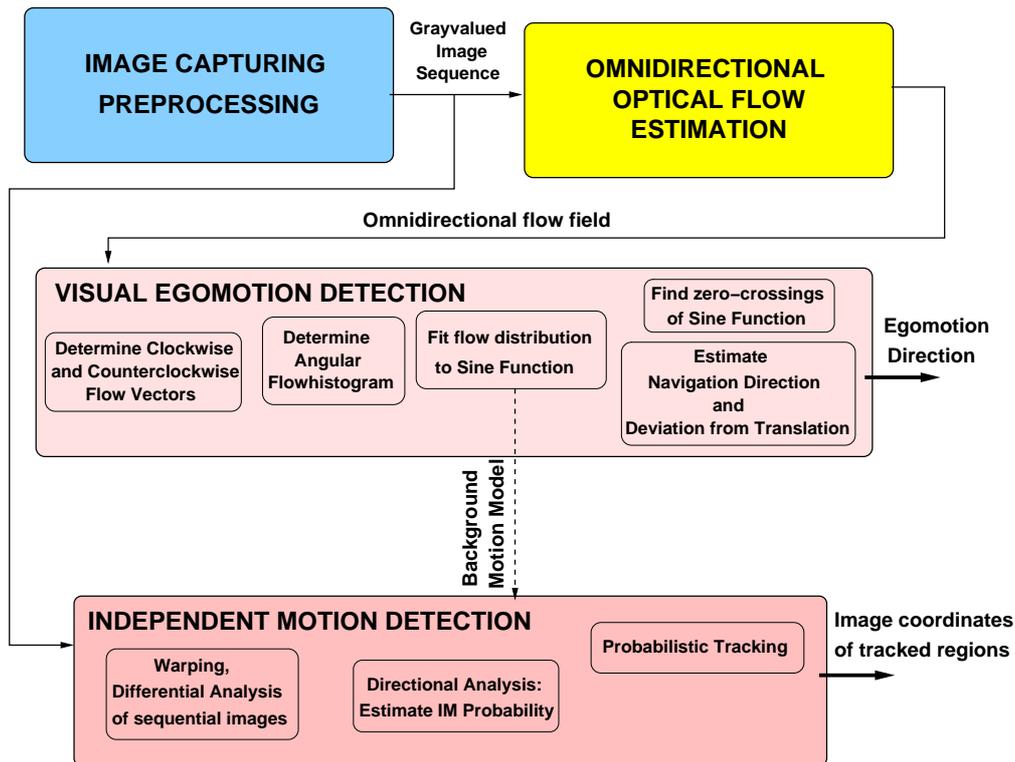


Figure 5.2: *The block structure of the implemented software*

5.2 Visual Egomotion Detection

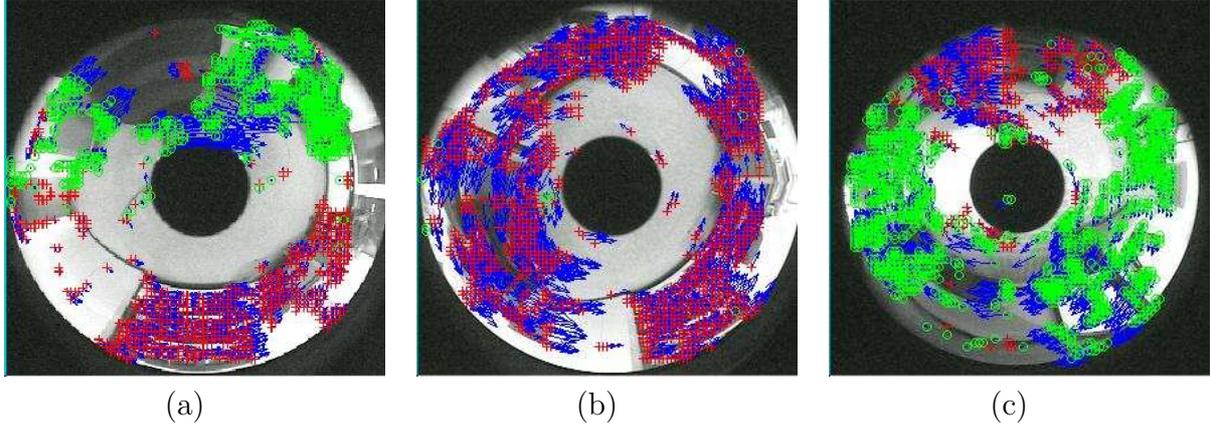


Figure 5.3: *Optical flow field in omnidirectional images. Red (+) indicate counterclockwise, green (o) indicate clockwise motion; (a) in the scene there is only translational motion to the left horizon with the door. (b) Here is only rotational motion which is counterclockwise. (c) Here the motion is "curvilinear". The robot-camera system rotates its vertical axis about 45° .*

The angular directions of the lines passing from the Focus-of-Expansion(FOE), Focus-of-Contraction(FOC) and the omnidirectional image center give clues about the egomotion of the mobile camera-robot system. Rotational components of the curvilinear egomotion change the relative position of the FOE and FOC lines in the global flow fields. In the case of purely translational motion, it is expected that the FOE and FOC lines have opposite directions, spanning an angle of 180° on the omnidirectional flow field. As the rotational component affects the motion field, it is expected that this angle decreases on one side of the field. Purely rotational motion causes the FOC and FOE points to vanish. The analysis of the clockwise (CW) and counterclockwise (CCW) flow vectors, marked in the images, render the regions with contra-directional flow fields. The proportions of these fields indicate the positions of the FOE and FOC. Ideally, this should lead to a unique detection of both of the lines. In Fig.(5.3), the resulting images of a sample run are shown. The proportionality of the clockwise and counterclockwise flow fields varies as the motion path changes from a purely translational (as shown in Fig.(5.3(a)), the proportional distribution of clockwise and counterclockwise flow field) to purely rotational motion (Fig.(5.3(b)), only counterclockwise flow field).

5.2.1 Estimation of translation direction in synthetical images

Synthetic images distorted with the given mirror properties simulating the omnidirectional camera at hand have been generated, to enable a quantitative and controlled analysis. In the simulation, the virtual camera-robot system navigates in an indoor corridor towards a given direction and with a constant speed. In the sequence from which the frame in Fig.(5.4) was taken, the camera moves to the angular direction of 270° , measured beginning from the right horizontal direction and incremented counterclockwise. In other words, the robot navigates in the direction of the 6:00 position of an analog clock and the beginning of the angular coordinate is at the 3:00 position. The angular distribution of the flow

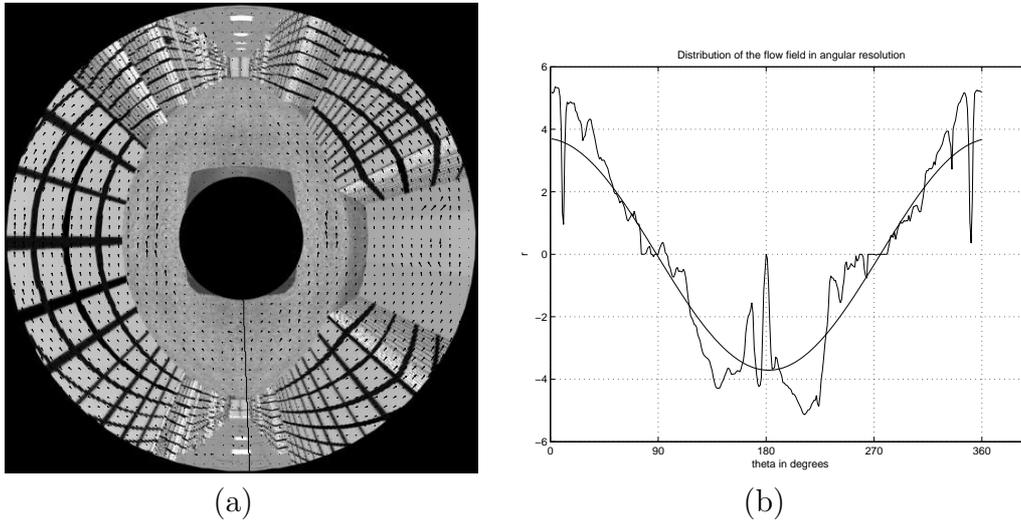


Figure 5.4: *Flow field calculated on synthetic image sequence with known FOE and FOC positions. The direction of the two singular points span exactly a 180° angle. The images also include the artificial scene distortion due to the mirror surface. (a) The omnidirectional optical flow field, (b) The angular flow distribution with the fitting sine function; the horizontal axis represents each angular direction in degrees, for which the flow vectors were averaged radially, the vertical axis represents the approximate speed in Pixel/Frame*

vectors on the omnidirectional field shows a sine-like pattern (Fig.(5.4(b))). The angular position of the minima of this distribution indicates the motion direction. In the pure translatory case, these minima are exactly 180° apart. Deviation from the pure translation causes the phase of the sine function to vary. Quantitative

analysis of the synthetical image sequences showed that it is possible to detect the navigation direction of the camera-robot system with up to a maximal error of 5° per image.

5.2.2 Egomotion Detection: Experiments with real images

The real omnidirectional images were captured as the mobile camera-robot system moves in indoor environments. As the robot moves on an even horizontal plane, the rolling motion in the vertical direction is negligible. In our experiments we let the camera-robot system follow several different paths. The three types of the motion paths, that the robot follows on the horizontal plane, are purely translational, purely rotational or complex, which means that the robot follows an arbitrary path with the combination of some translational and rotational motion.

Purely translational motion

The system captures images as the robot navigates straightforward in an indoor corridor. The scenery can be visualized as depicted in Fig.(5.5). During this purely translational motion, the omnidirectional flow field is distributed proportionally along the circumferential dimension of the captured images, which can be represented with the angular flow distribution. The rotational component is negligible. Fig.(5.5) shows one of the images of a sequence of 60 images, in polar and in cartesian coordinates. The angular flow distribution during the translatory motion remains similar to the one given in Fig.(5.5(c)), which can be fitted to a sine function with the zero crossings at around 0° and 180° . The zero crossings indicate the translation direction on the omnidirectional image plane.

The numerical fitting of the angular flow distribution, allows to estimate the translational and rotational component of the motion. This function can be written as previously described in Section 4.1:

$$\begin{aligned} F(\Theta) &= A \cdot \sin(\Theta - \Lambda) + \vec{\omega} \\ A &= \frac{-\|\vec{v}\|}{r} \end{aligned} \quad (5.1)$$

where r is the radial distance to the object at the angle Θ . The magnitude of the rotational component can be written as:

$$\|\vec{\omega}\| = R_m \cdot \tan(\omega_z)$$

where R_m represents the radius of the projected image, and ω_z the angular rotation of the system around its z-axis. In the experimental results, the rotational compo-

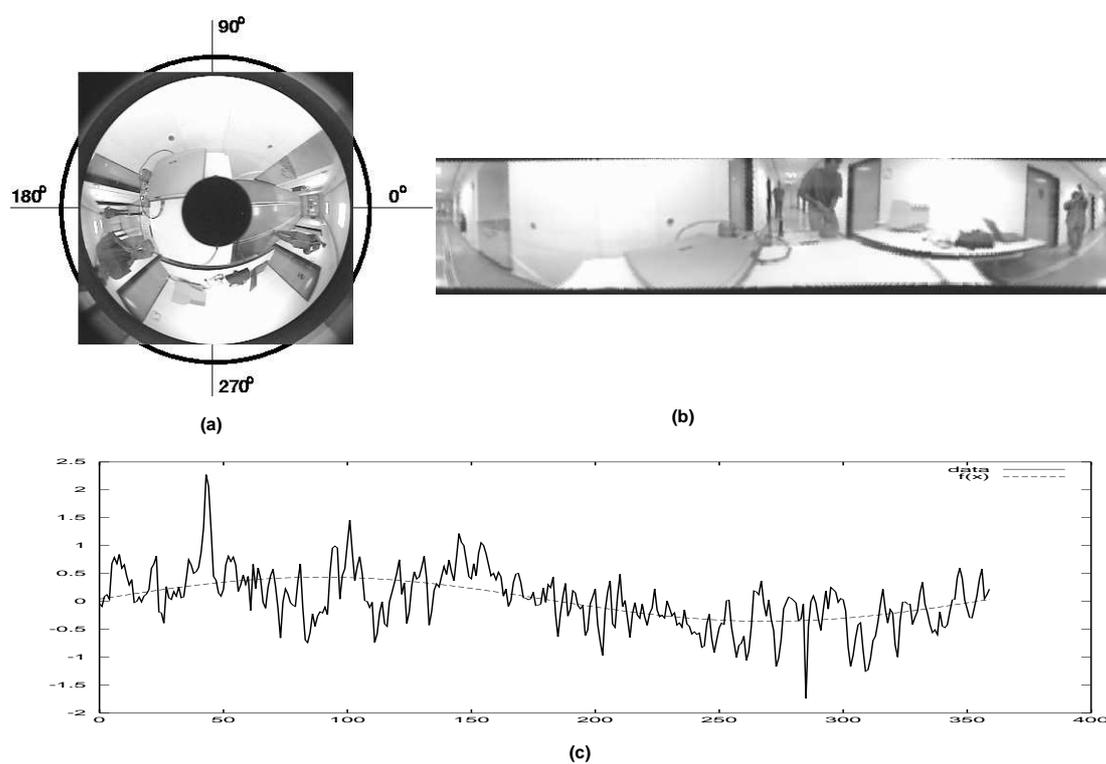


Figure 5.5: (a) One image from the sequence, which is captured as the robot translates in an indoor corridor. (b) The same image represented in cartesian coordinates. The unwrapping begins from the 0° of the image represented in polar coordinates as shown in (a). (c) Angular flow distribution and the fitting sine function.

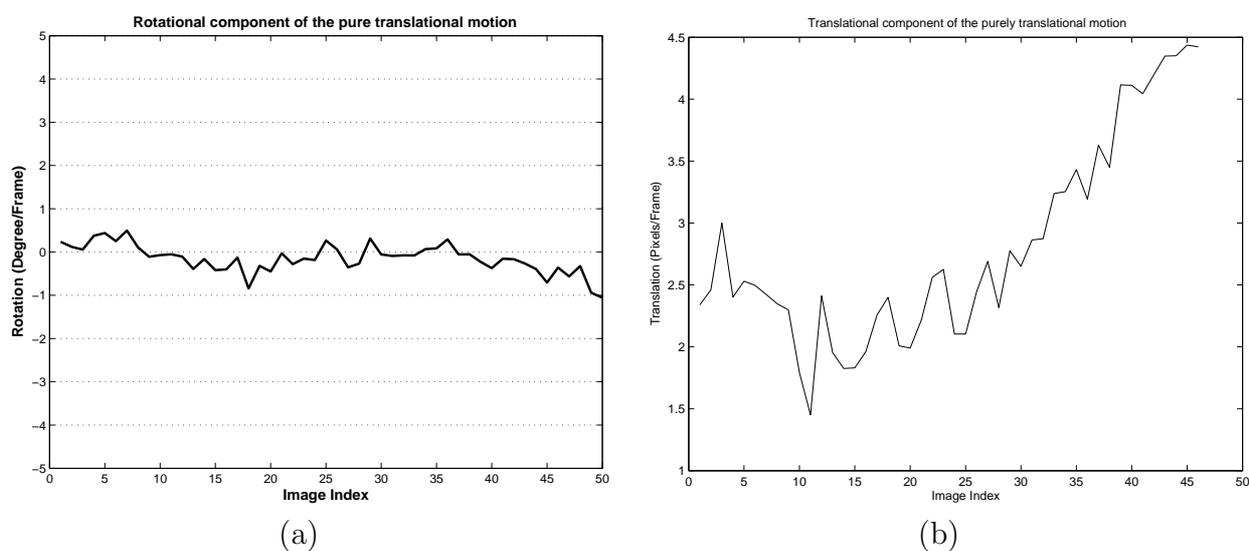


Figure 5.6: *The rotational and translational components of the pure translational motion. The first diagram depicts the rotational speed of the robot around its vertical axis. It varies between $\pm 1^\circ$ per Frame and therefore negligible, since the motion is purely translational. The second diagram (b) presents the translational speed during the navigation. The horizontal displacement between the frames is around 2.7 Pixels/Frame and the robot accelerates up to 4.5 Pixels/Frame during the last phase of the navigation.*

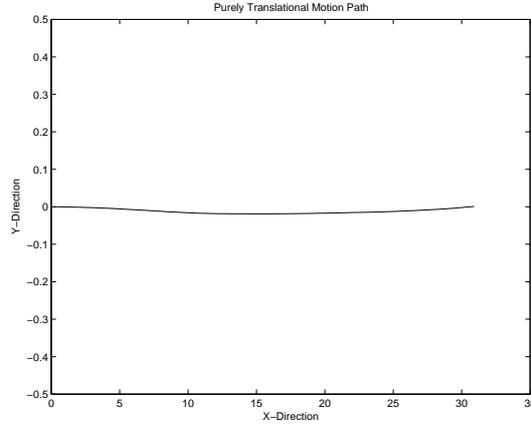


Figure 5.7: *The estimated robot path in 2D space. The robot translates on a purely horizontal path. The rotational component is negligible. The motion is assumed to be smooth.*

ment of the egomotion is determined from the $\vec{\omega}$ component of the fitting function. The translational speed is estimated from the amplitude, A of the sine component (given in Eqn.(5.1)). The estimated rotational component $\omega = [0, 0, \omega_z]$ and the translation speed \vec{v} for the sequential frames is depicted in Fig.(5.6). As expected, the rotational component ω_z is very small due to the pure translational motion and varies around $\pm 1^\circ/Frame$.

Using the estimated egomotion parameters, we reconstructed the motion path of the robot on a 2D map, given in Fig.(5.7). The path shows only a horizontal displacement in 2D space, as expected.

Purely rotational motion

The robot turning around its vertical axis has an angular flow distribution, which is constant around the rotational speed. In the fitting function (Eqn.(5.1)), the component representing the rotational velocity ($\vec{\omega}$) becomes significantly larger than the sine component ($A \cdot \sin(\Theta - \Lambda)$). Fig.(5.8) depicts one frame of the image sequence captured as the robot rotates around its vertical axis. The angular flow distribution (Fig.(5.8(c))) approximates to a constant function for all angular directions. The sign of this function indicates a counterclockwise egomotion.

In Fig.(5.9(a)-(b)), the estimated rotational and translational components are depicted. As opposed to the pure translational case, here the rotational motion component is more significant and varies around $8^\circ/Frame$. The translational component varies around $\pm 1 Pixels/Frame$. This is the negligible translational

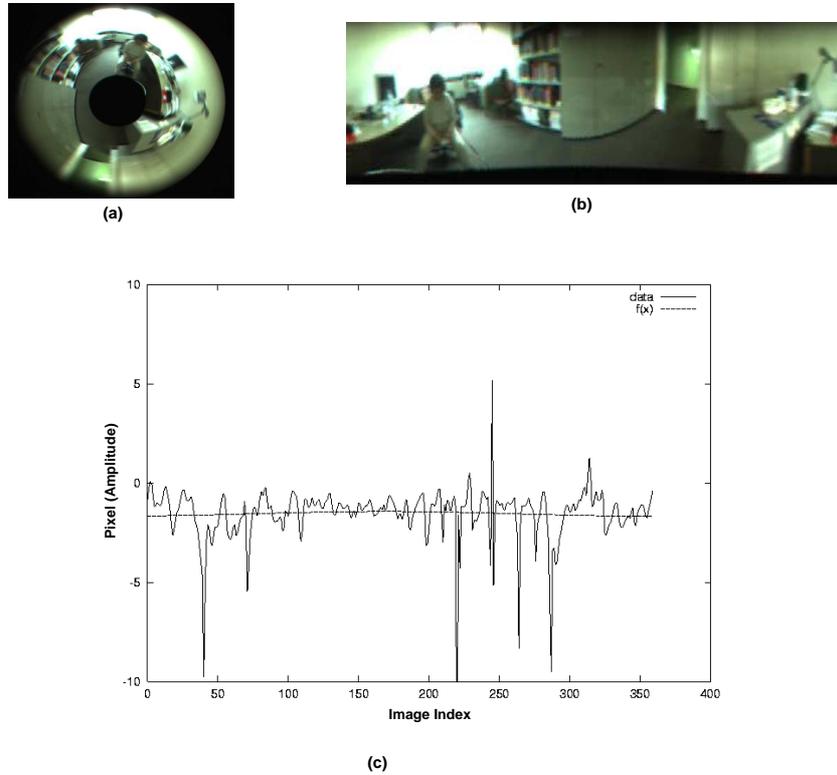


Figure 5.8: (a) One frame of the image sequence captured as the robot rotates counterclockwise around its vertical axis. (b) The image unwrapped onto the cartesian coordinates. (c) Angular flow distribution.

motion of the camera, which is due to the fact that it is not placed exactly at the center of the rotation, which is the center of the robot platform. Note that the ratio between the amplitude of the distribution and the estimated rotational motion ($|\frac{A}{\omega_z}|$) is lower than 1.0 for most of the images in the sequence, which indicates that the rotational component is more significant than the translational component.

Fig.(5.10) shows the path of the robot motion mapped on the 2D space. It can be seen that the translation motion in x and y directions have negligible components. The path shows only sharp turnings of the robot around itself with a small displacement in x and y directions, which is due to the wheel slippage and due to the fact that the camera is not placed exactly on the rotation axis of the robot.

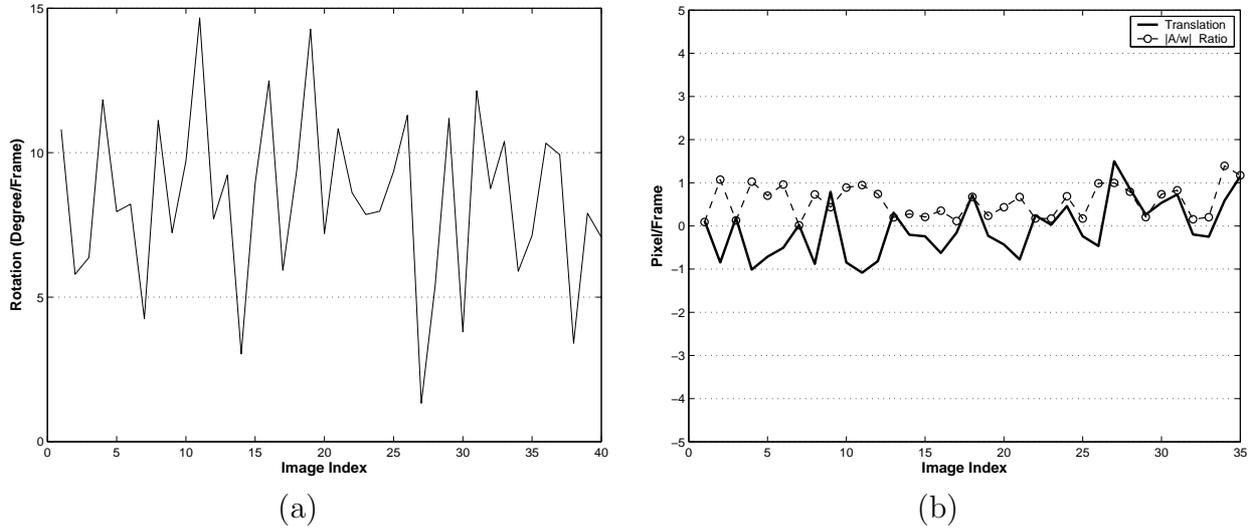


Figure 5.9: *The motion components of the purely rotational navigation. (a) The rotational component of the purely rotational motion. (b) The translational component of the purely rotational motion and the ratio of the amplitude and rotational component ($|\frac{A}{\omega_z}|$). The lower this ratio, the more significant is the rotational component of the motion.*

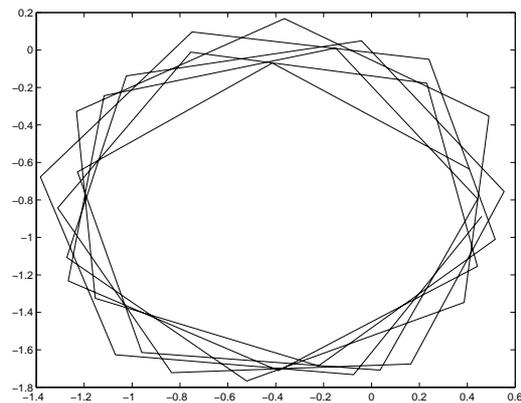


Figure 5.10: *Purely rotational motion path. The robot rotates around its Z-Axis. Translational component is negligible. The diagram shows the robot path on the 2D space.*

Curvilinear motion

The two examples of the pure translational and rotational motion are interesting only from a theoretical point of view. In practice, the real indoor navigation of a mobile robot system consists of a combination of these two types of motion. We call this as the curvilinear or complex motion. In the following experiments, the robot was programmed to follow a corridor; first in a direction towards a wall, moving around a static obstacle, correcting its position at the corridor wall and translating further. The visually estimated egomotion parameters are used to recover the 2D map of this motion path, as can be seen in Fig.(5.11(a)). Fig.(5.11(b)) depicts this angular variation during the navigation. The rotational component is significant when the robot takes a turn and corrects its position. During the horizontal displacement, the angular component is negligible.

Another complex motion path that the robot had followed, is shown in Fig.(5.12). The navigation starts from an office room. Then the robot moves out to a corridor, turns to the left, after a while turns back moving around its axis, translates to the reverse direction, takes two sharp turns around its axis and moves further horizontally. The thick lines indicate the real path of the robot, which were observed. The blue thin line represents the estimated path. The visual path approximates the observed ground-truth, especially during the smooth translational motion. The sharp turns of the robot where the rotational motion is significant can only be followed with instantaneous errors. Note that the image sequence includes several frames with pure rotational motion, where there are no significant directional information. Additionally, some frames are without any background motion, e.g. captured as the robot stopped for some seconds.

In the experiments conducted so far, the visual measurement results and the ground truth of the egomotion could only be compared with the observed 2D paths of the robot motion, since there was no gyroscopic data available. Fig.(5.13) shows the experimental results, which can be compared with the gyrodometric data, that the system reads off from the electronic gyroscope and the motor encoders. In [116] and [115], we give detailed descriptions of this sensory setup. The angular measurements coming from the gyrodometric measurements are comparable to the visual measurements. The largest measurement difference can be observed at the sharp turnings where there is no translational component and during the discontinuous motion (accelerating or stopping of the robot during navigation).

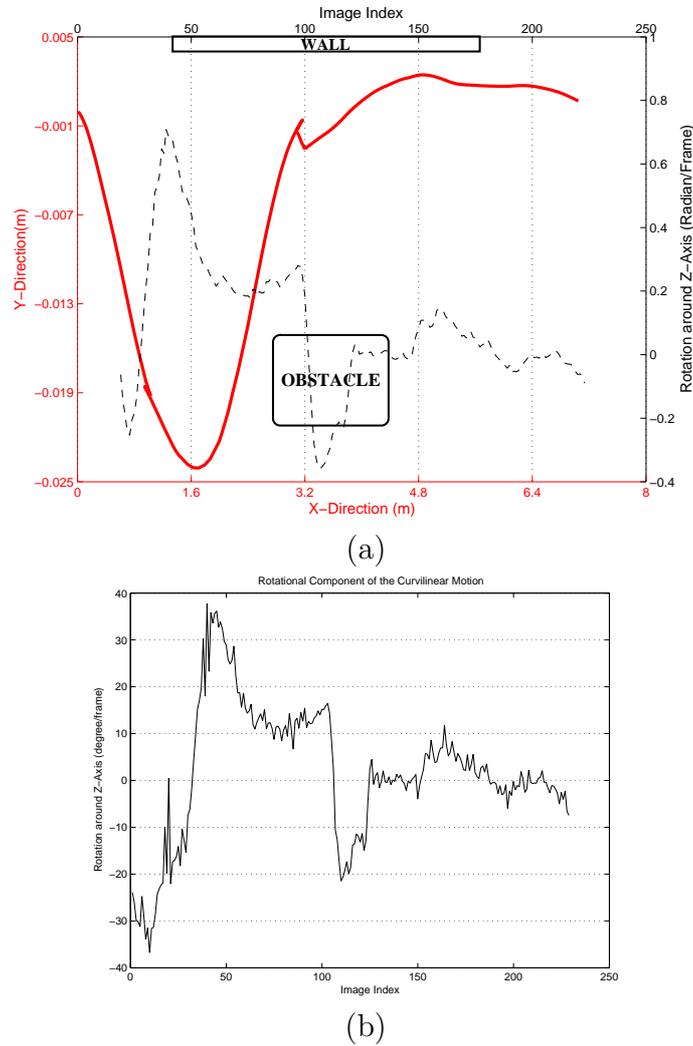


Figure 5.11: *The robot follows a curvilinear path in a corridor. (a) Visually estimated robot path on the 2D space in comparison with the rotational component of each image. The thick red curve indicates the path, the deviations in X and Y directions measured in meters from the starting point of the robot motion, located on the left upper part of the diagram. The dashed curve represents the rotational component of the motion for each image index represented with the upper axis of the figure and measured in Radians/Frame represented with the rightmost axis. The robot first follows a path towards a wall, making a sharp turn to avoid an obstacle and then corrects its position as it comes too close to the wall. (b) Rotational component of the curvilinear navigation. The rotation around the vertical axis becomes significant as the robot follows a curve and vanishes as the motion becomes more and more horizontal.*

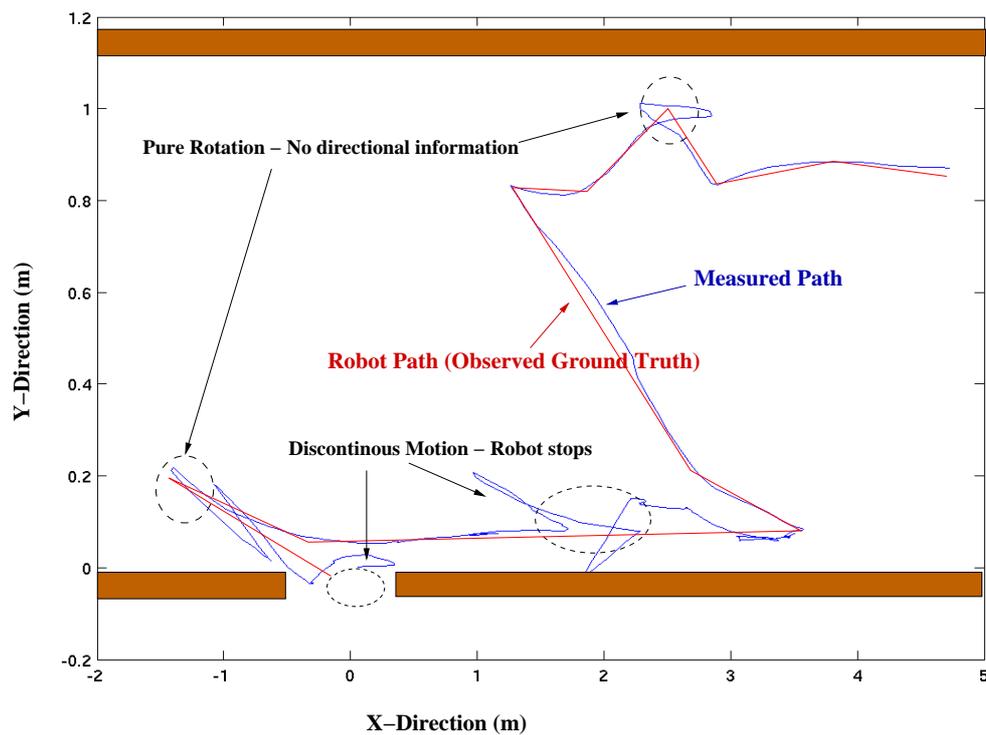


Figure 5.12: The robot follows a complex path in a corridor. The diagram shows the robot path on the 2D space. The red lines (thick) indicate the observed ground truth path, and the blue lines (thin) represent the visually estimated motion path. The deviations shown in circular regions are due to discontinuous motion or pure rotation of the robot where there were no directional information available.

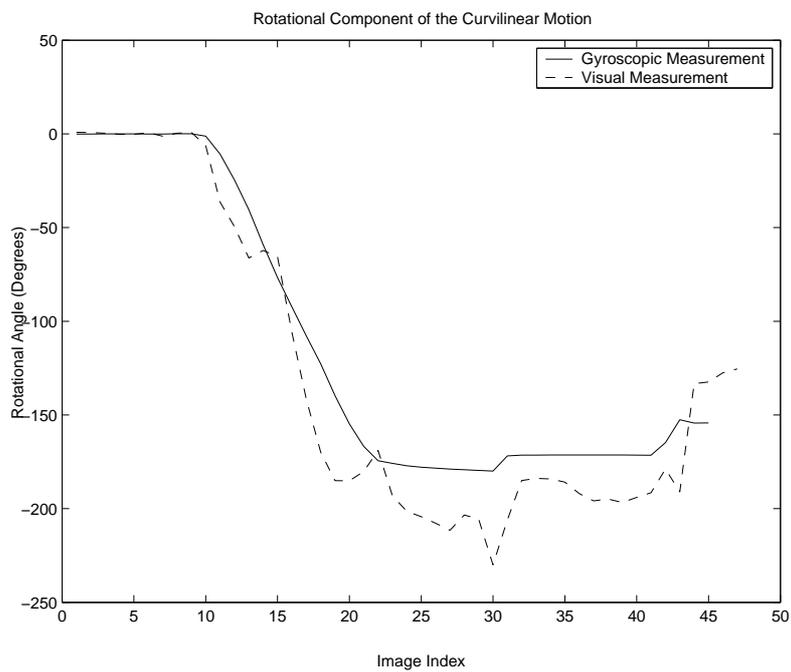


Figure 5.13: *Rotational component of the curvilinear motion: visual measurement (the dashed line) and gyroscopic measurement (solid line), are comparable during the smooth motion. Sudden turns and discontinuities in the motion of the robot caused measurement differences up to $20^\circ/\text{Frame}$.*

5.2.3 Summary of the egomotion detection experiments

The indoor experiments examined the feasibility of the visual egomotion detection algorithm, which was proposed in the previous chapter. The experiments showed that the intrinsic structural features of the dense flow fields of omnidirectional image sequences, which were captured during the navigation of the mobile robot moving through different paths, allow the detection of the rotational and translational components of the egomotion. Using these components the motion paths of the robot could be reconstructed.

In the majority of the test runs, we could estimate the pure translational and rotational egomotion of the robot with up to a $\pm 5^\circ/Frame$ accuracy. The curvilinear motion could be estimated with a sufficient accuracy that allowed a direct comparison with the observed 2D paths of the egomotion.

For the egomotion detection, we assumed a smooth motion of the camera robot system and equidistant range in the scene structure. During the experiments this assumption held only partly. Especially during the sharp turns of the robot around its axis, the smoothness assumption of the motion was not necessarily fulfilled. These situations were the major source of error in the results. During the last experiment, the gyrodometric data of the robot was available. The comparison of the two data set showed that the visual measurement errors and the accumulative errors of the gyrodometric data add up to a difference of up to $20^\circ/Frame$. But the visual measurements are more robust to measurement interruption since they are instantenous and framewise as opposed to the accumulative gyroscopic measurements. The fusion of the two sensory modi may improve the quality of the egomotion detection. The theoretical conception of a fusion task has been described in [116].

The visually estimated egomotion parameters are used to construct a background motion model, which will be exploited for the independent motion detection. The obtained experimental results are satisfactory to constitute such a model. More enhanced tasks, e.g. exact visual reconstruction of the scene, should consider more precise measurements of the scene structure and also the fusion of other sensory equipment.

In the following, the background model is used to stabilize the image and to detect the independently moving scene objects.

5.3 Independent Motion Detection

5.3.1 Segmentation of independently moving scene objects

Independent motion (IM) causes distinct changes on the global flow pattern of a moving camera-robot system. While the background motion occupies the larger part of the global flow field, IM arise regions of disturbances in this pattern. Static objects, which are located in the near region of the mobile robot, may also cause greater flow vectors, which differ in their values but not in the direction of the egomotion. The analysis of significant flow regions of the global field enables the segmentation of such static objects, that come closer to the center of projection, and also independently moving objects. Both tasks might be interesting from the navigational point of view. If the major goal is to determine the dynamically moving objects in the scene, then it might be necessary to look at the changes of the flow field in time. One efficient way to do so is to warp the given image by applying the background model to the previously estimated flow field. If the egomotion model of the robot can be determined, any object moving with an independent speed and direction other than the estimated model, will produce a residual region between the warped image and the real image. The regions where this warp error is high, may indicate IM. The procedure for the proposed detection method of the IM can be outlined as follows:

- calculate the flow field using the first two images and generate the model of the egomotion
- generate a pseudo image by warping the first image with the estimated egomotion model
- calculate the difference between the second image and the pseudo image and detect the residual regions

In Fig.(5.14), two real images of an indoor navigation sequence are shown. Fig.(5.14(d)) presents the results of the significant flow region detection. Note that in Fig.(5.14(d)) the algorithm detects regions with independent motion as well as the static nearby objects. One can distinguish these two object categories by applying the warping analysis, outlined above. The details of this algorithm was described in Section 4.3. In Fig.(5.14(e)), the detected silhouette of the person on the right, who is moving independently from the robot-camera system, is shown.

The IM segmentation can be effectively improved by the image warping, which considers background modeling and egomotion compensation. This can be visualized in Fig.(5.15). The first image, Fig.(5.15(a)), shows the difference of the sequential frames without the egomotion compensation. The residual regions,

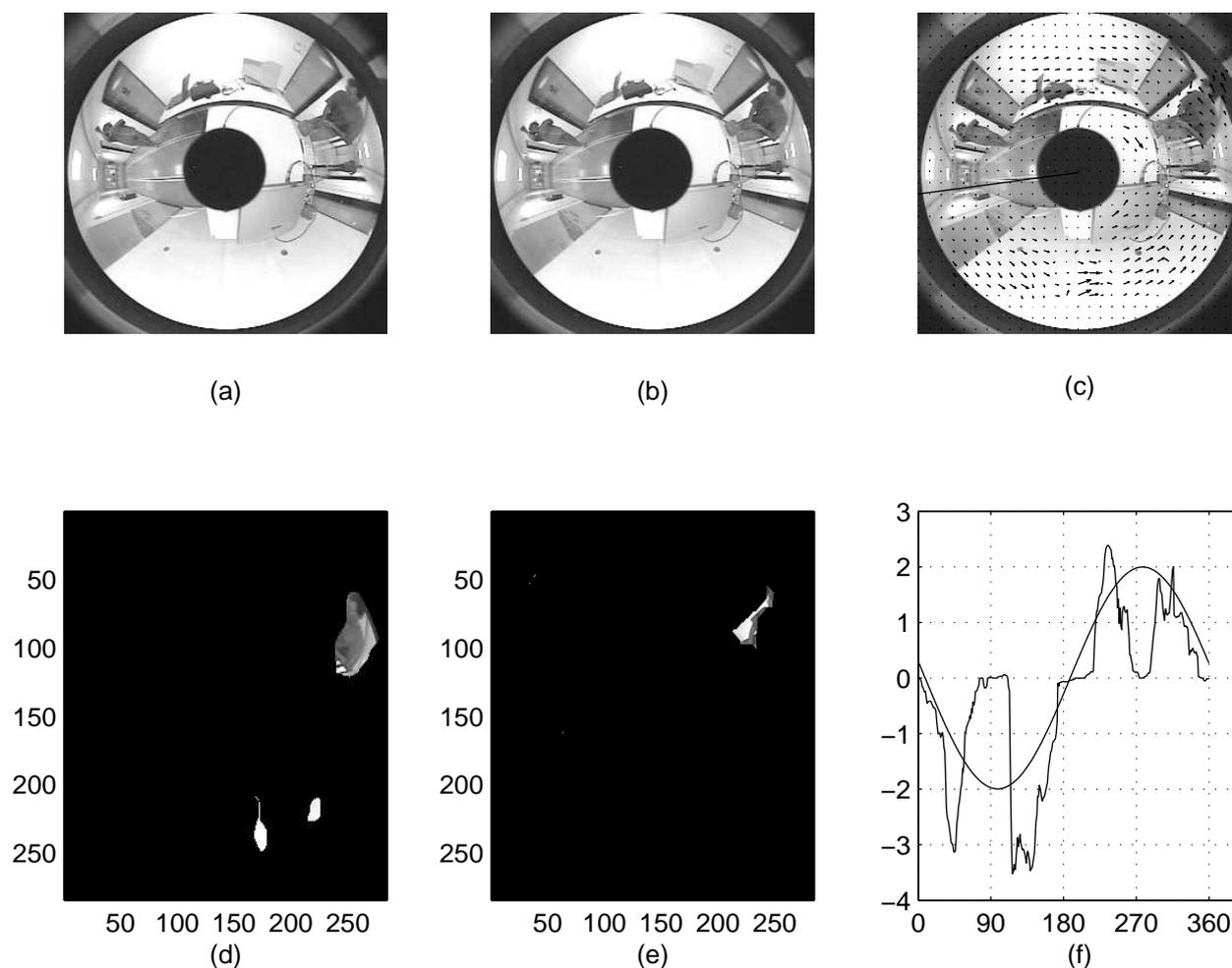


Figure 5.14: (a) First image of a sequence taken from a camera moving through a corridor. (b) Second image of the sequence, note that the person on the right moves to the left as the camera translates. (c) Optical flow field superposed onto the second image, the line on the left indicates the estimated translation direction. (d) Significant flow regions thresholded and segmented from their background. (e) Image region with the maximum warping error. (f) Angular flow distribution of the flow and the (sine) fitting function.

which appear brighter than the background, are not explicitly due to independent motion only. The depth parallax can be observed specially on the edges of the corridor and the doors and the resulting residual field is cluttered. The second image, Fig.(5.15(b)), visualizes the same sequence after the compensated differentiation as proposed in Section 4.3. The brightest regions are due to independent motion. The brightness of the image regions with the depth parallax are damped, which means that on such regions the probability of detecting IM is effectively low.

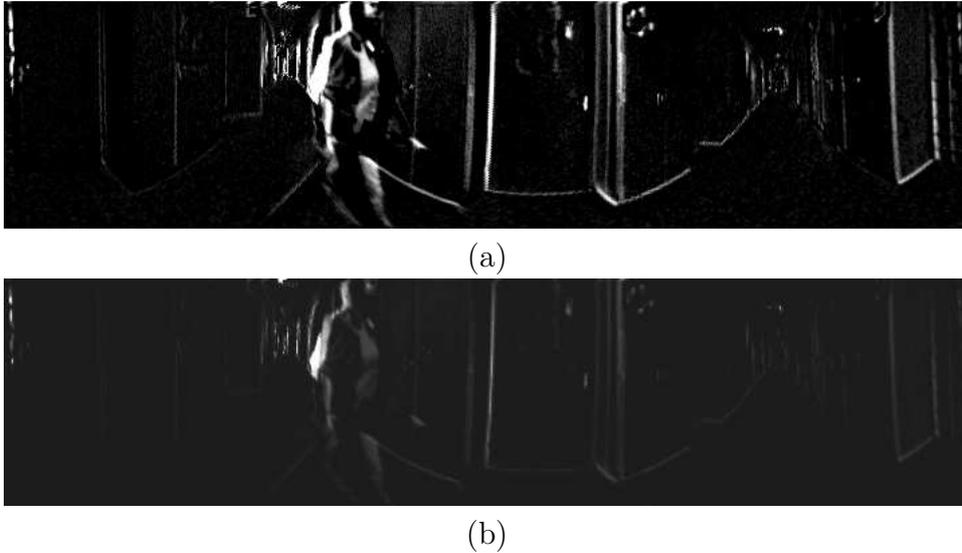


Figure 5.15: *Independent motion segmentation: (a) Flow outlier regions without background compensation; direct comparison of two sequential frames (b) Residual regions after background compensation. Here, it is evident that the clutter on the residual images are reduced significantly. The following tracking scheme profits from this clutter-reduced residual image.*

5.3.2 Tracking independent motion

The particle filter, presented in Section 4.3.2, utilizes the residual images to track the IM regions and allows a temporal analysis. The residual image, which is essentially a 2D probability distribution of independent motion, represents the measurement input of the filter. The state vector of the particle filter includes the pixel positions of a 2D tracked region (angular and radial coordinates of the region on the polar image), as well as the rates-of-change of these positions. Initially, the system generates a number of particles, that should converge on the 2D pixel positions of the scene objects with the highest probability of representing IM.

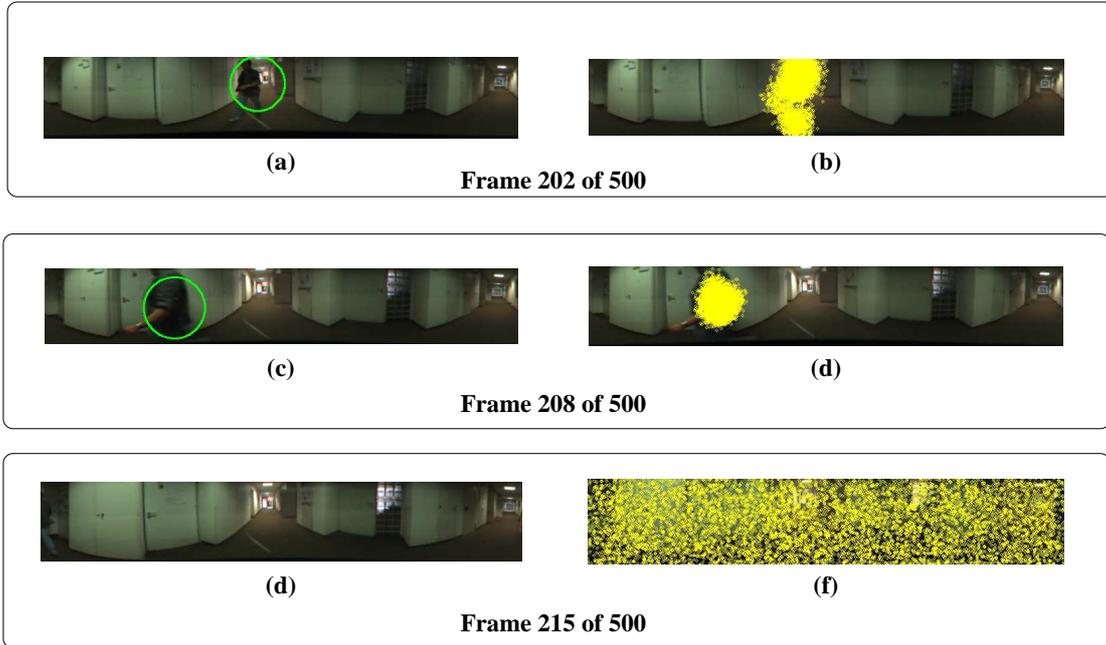


Figure 5.16: A correct detection and tracking of the person moving independently from the camera-robot system.

Depending on the residual measurements, these pixel positions are tracked through the image sequence. The tracking system assumes that the IM regions move with a constant speed on the horizontal plane, whereas the motion and the position of the depth parallax regions are not constrained.

In Fig.(5.16), the resulting three frames of a tracking experiment are shown. The camera-robot system translates in the opposite direction as the person passes by. The system keeps track of the person. The particles shown with yellow circles converge on the 2D positions of the independent motion. As soon as the person is out of the field of view, the convergence is disturbed and the particles are distributed homogeneously, as seen on the last frame.

In Section 4.3.1, it has been noted that the depth parallax yields similar residual regions on the motion field like the IM. In the case of significant depth parallax, the system converges onto those 2D regions with the static objects with varying depths. In Fig.(5.17), an example of this misinterpreted detection is shown. The first frame converges on a corner of the door and the system detects this region as IM falsely. In the following frame, the convergence is disturbed, and the particles are homogeneously distributed, indicating that the system could not keep the track of that object. The temporal continuity of the false detections is usually not

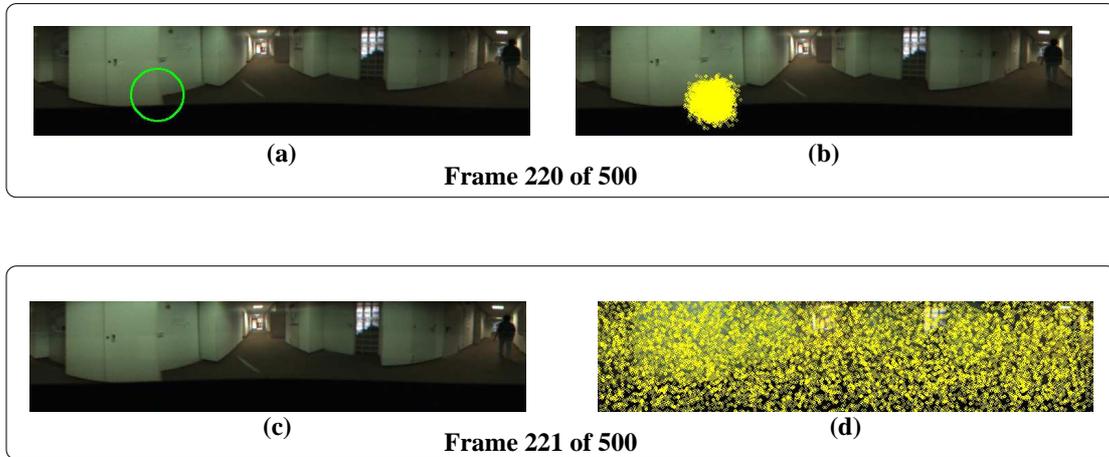


Figure 5.17: (a) *False detection of IM (depth parallax)* and (b) *the reaction of the particle filter in the following frame. The measurement model states in this case that the previous detection was a false detection and it was a depth parallax rather than IM. The new measurement corrects the particle distribution in the following frames, in this case causing it to distribute homogeneously on the whole image plane.*

sufficient for a decision.

Fig.(5.18) depicts the angular position of one independently moving object and the heading direction of the robot in sequential frames. The detected angular positions evolve differently from the heading directions, which were estimated by measuring the angular position of the FOE in each frame. There is also a crossing over of the two paths on frame 24, indicating that the IM region crosses the heading direction, gets out of the field of view from the left side of the image and appears from the right side of the image again. This is due to polar image mapping. The omnidirectional images are unwrapped counterclockwise, starting from right hand side horizontal direction, similar to the unwrapping utilized in Section 5.2.2.

In Fig.(5.19), the results of another IM tracking experiment are depicted. In the upper image (Fig.(5.19(a))), the results of a smooth IM region tracking and the angular positions of the FOE, are shown. Fig.(5.19(b)) shows the tracking of falsely detected regions, represented with red squares. It can be seen from the diagram that such regions are followed only for a short period of time, on the average for three frames.

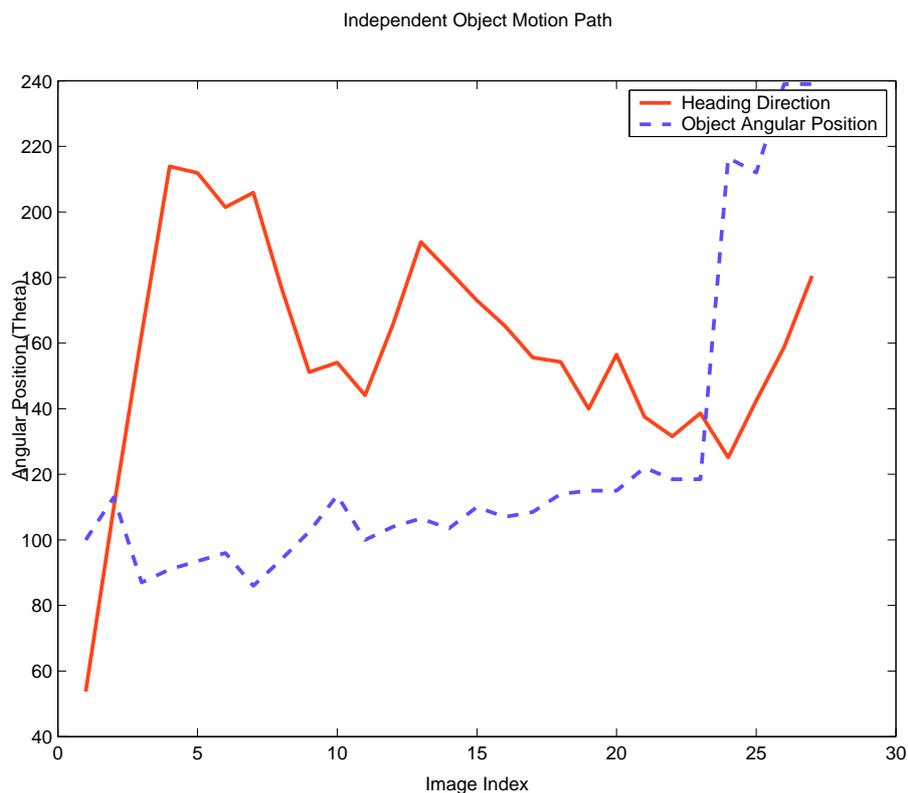
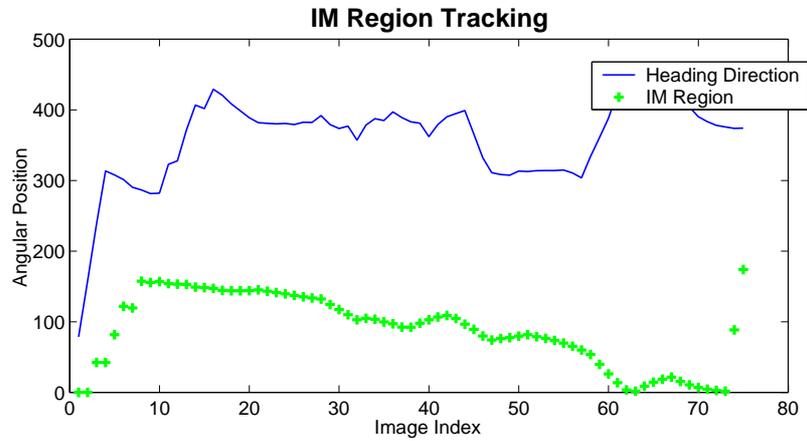
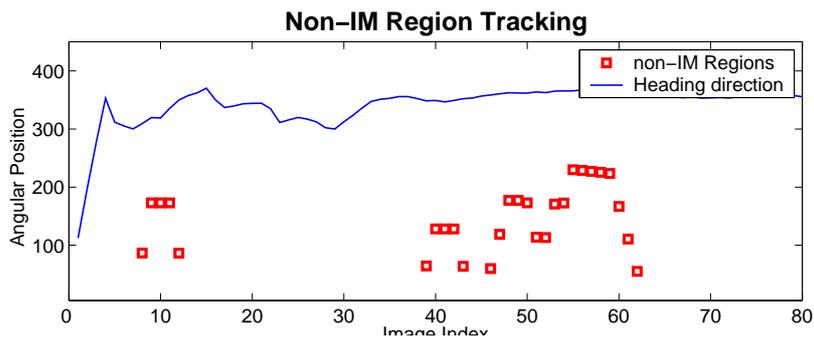


Figure 5.18: *The positions of the robot and the IM region are represented by the angular direction on the polar unwrapped image. The dashed line indicates the angular position of the IM region and the red line represents the position of the FOE, indicating the heading direction of a translating robot, in each frame. Both paths differ significantly so that the hypothesis of independent motion can be validated.*



(a)



(b)

Figure 5.19: (a) The results of an IM tracking experiment. The green (+) indicates the angular positions of the IM region on each polar image frame. The blue (solid) line represents the heading direction measured by absolute FOE pixel position. The slopes of the lines are different, which indicates that the egomotion and IM are in different directions. (b) The results of the experiment without IM regions. The vertical zero value means that there were no regions found for the respective frame. The red squares represent the angular position of falsely detected residual regions on the image. Note that the system tracks them for a short time and loses their track after three frames on the average.

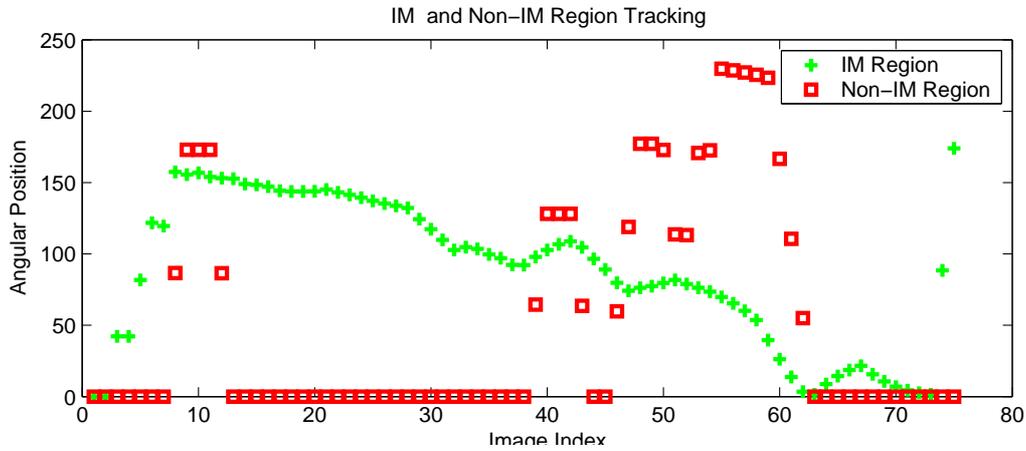


Figure 5.20: The diagram represents the superposed results of the two test runs, the IM and non-IM Region-Tracking experiments. The zero value of the vertical axis indicates that no regions have been detected, otherwise the y axis represents the angular position of the detected region on the polar image. The x-axis represents the frame number (image index). The pixel positions of the IM Regions are depicted with green + and the resulting angular positions show a smooth variation, indicating a reliable tracking. The non-IM regions (shown with red squares) are detected due to depth parallax but cannot be tracked smoothly. The system decreases the importance of such residual regions in every step, causing the tracker to loose the region positions every other frame.

5.3.3 Summary of the independent motion detection experiments

The experiments were conducted using a camera-robot system moving in an indoor corridor of an office building. Typically, in such an environment, there are people moving around, passing by the robot, and translating in arbitrary directions. The aim of the experiments was to detect those objects with the independent motion. The detection scheme is made up of two sequential steps; the segmentation of residual image regions with significant values, and the tracking of such regions using particle filter to be able to observe the motion paths during several frames.

The IM regions are segmented by analyzing the sequential frames and eliminating the background motion. The background motion model is constructed by analyzing the angular flow distributions and it is assumed that the all the static objects of the scene undergoes the same 2D positional changes imposed by the background motion. First, the effect of the background elimination for the resid-

ual region segmentation has been examined. There is a significant improvement on the residual images as depicted in Fig.(5.15). Here, the residual regions caused by the depth parallax are less significant than the IM regions. Since the residual images are utilized as a measure of the tracking system, a reliable segmentation in this stage, improves tracking, too.

The tracking scheme has been implemented using a particle filter, as discussed in Section 4.3.2 in detail. The particles converge on 2D regions, where the system expects to detect IM. The tracking system converges onto an IM region in 2-3 frames on the average and keeps track of the region until it moves out of the field of view. There are false detections caused by the significant depth parallax problem. Those false detections are tracked usually for three frames on the average, with varying angular positions. The discontinuity of the tracking is an indicator that the region does not represent IM.

The direct comparison of the two types of region tracking (IM and non-IM), is shown in Fig.(5.20). The tracking of IM region indicated by the green (+) is smooth, whereas the non-IM regions are tracked for a short period of time and lost again. The emphasis of the residual regions indicating IM and warping of the image with the background motion model affects these results.

Chapter 6

Conclusion and Outlook

6.1 Summary

The dynamical image processing on omnidirectional image sequences reveals information about the navigational prerequisites of an autonomous mobile robot. The combination of the two techniques, namely the *dynamical* and *omnidirectional* imaging, allows for a novel interpretation of the visual measurements. This combination is the major starting point of the thesis, which exploits the structural features of the omnidirectional optical flow field to determine the motion parameters of a mobile robot that navigates in real indoor environments with dynamical components.

Dynamical image processing is a common term for the methods of spatio-temporal image analysis, in which sequences of images captured in constant and small time intervals, are processed to extract motion parameters of a moving scenery. Thereby, the motion field, which is the 2D projection of the 3D motion in the scene, is represented with the optical flow field. There has been intensive research on the accurate and time efficient optical flow calculation. Chapter 2 reviews and compares the methods introduced so far, focusing on their efficiency for the robotic applications. The comparative evaluations suggest that the differential methods, like the method of Lucas and Kanade [74] represent an acceptable trade-off between the accuracy requirements and the computing complexity for most of the robotic applications.

The omnidirectional imaging enables capturing a wide field of view of the real scene, which is beneficial for the interpretation of the measurements and scene interpretation. The natural and the artificial methods of capturing a wide field of view have been reviewed in Chapter 2.6. In this thesis, we employed a catadioptrical sensor, that consists of a standard CCD sensory element and a convex, curved

mirror surface that reflects the 360° horizontal and 160° vertical view field. Even though the mirror keeps the angular ratio of the incoming and projected light rays, the geometrical distortion on the 2D images requires special considerations for calculating the optical flow. The differential method of Lucas and Kanade employs separable spatial filtering and linearizes the spatial and temporal gradients locally. The local linearity assumption, however, has to be modified for omnidirectional projection, since the mirror reflection causes non-affine distortions on the image plane. The projection from the mirror surface can be assumed to be spherical. Therefore, the local filters can be redefined on the spherical coordinates. The thesis proposed a redesign of the spatial gradient filters and local smoothness region on a curved grid on the sphere, keeping the advantage of using separable filters for the image data processing. The results showed that this modification yielded more accurate estimations of the omnidirectional optical flow field, without additional computing cost.

The omnidirectional optical flow field estimates the 3D motion field of a camera-robot system. The estimated motion can result from the self motion of the camera-robot system (referred throughout the thesis as the egomotion), or other moving objects like other robots, people, moving ball etc., or both. The omnidirectional optical flow field distribution in the angular direction, shows distinct structural features, depending on the egomotion of the camera-robot system. The pure horizontal translation, for example, induces an angular flow field distribution of proportionally distributed clockwise and counterclockwise optical flow vectors. The rotational motion on the other hand disturbs this proportionality and causes the flow vectors in the direction of the motion, either clockwise or counterclockwise, to overbalance the distribution. In the real navigation scenarios, the robot motion is a combination of the both motion types, which causes a sine like distribution of the angular flow histogram. The parameters of this sine function indicate the heading direction, the deviation from the purely translational motion and the approximate horizontal speed of the robot. Chapter 4, addresses the details of the proposed methods for egomotion and independent motion detection. In the experiments, which are summarized in Section 5.2.3, we applied these parameters to determine the motion path of the robot in an indoor environment.

Determining the egomotion of the robot facilitates the segmentation of independent motion, which arises as residual regions on the global omnidirectional optical flow distribution. The thesis has solved the independent motion detection problem by segmenting these residual regions. The algorithm, proposed in Section 4.3, compensates for the background motion using the egomotion parameters estimated in previous steps, and segments regions of outlier optical flow. However, these regions indicate only probabilistic information about the independent motion, because of depth parallax effects. Depth parallax comes into play when

the scene objects appear in different depth layers as the camera moves towards them. These depth variations cause residual regions on the optical flow distribution, too. We propose to redefine the problem of independent motion tracking as a probability based problem that can be solved in the temporal dimension by a probabilistic tracking.

The particle filter based tracking of outlier regions of the optical flow field validates or rejects the hypothesis of independent object motion. In Section 5.3, we give the experimental results, that support the summarized concepts.

6.1.1 Strengths and Limitations

Omnidirectional imaging facilitates capturing larger parts of a dynamic scene without mechanical rotations of the camera or additional exploratory motion of the robot. Therefore, it allows for a more efficient scene and motion analysis compared to the standard imaging sensors. The thesis took this advantage and developed robust techniques to detect egomotion and independent motion for omnidirectional image sequences based on the topology of the optical flow field. The modified method for the optical flow calculation allowed for more accurate estimation of flow vectors in magnitude and in angular direction.

The proposed visual egomotion detection is robust to local occlusions and noise since it is based on the global analysis of the optical flow field, rather than landmark tracking. Visual detection of egomotion direction suffers only in situations where the camera-robot motion is discontinuous and in the case of pure rotational motion. The proposed scheme is used to model the background motion of a scene. The resulting background motion models are satisfactory for most of the real robot navigation cases.

Using the proposed techniques, independently moving scene objects can be detected and followed for a longer time. The tracking is robust even in the case of cluttered background scene and nonreliable residual image. In particular scene constellations, non-independent motion with significant depth parallax may cause false detections. These false detections have different temporal evolution than the tracked independent motion regions.

The thesis finds out that dynamical image measurements and optical flow can be exploited to gain navigational parameters of an autonomous vision system. Gibsons theory stating that the perceptive structure of changing visual scene should be feasible to specify the structure of the environment, has been here verified and extended to detect independent motion. Specifically the new approach of interpreting topological structure of the dynamical image features rather than quantitative measurements of landmarks and scene structures, seem to be efficient and informative for navigational tasks.

6.2 Outlook

The following points remain to be revisited:

- ⇒ The thesis builds a segmentation stage for a more general image processing system, that might further consist of pattern recognition and classification stages. It is possible to enhance the system with a learning algorithm, e.g. to classify the type of the egomotion, translation or rotation, and decide on the types of independent motion that has been tracked. For example, a human motion detection module based on the periodicity analysis [90], can be used to identify the people that are moving independent of the camera-robot system.
- ⇒ The methods proposed in the thesis are not constrained on the use of gray-valued images. It is straightforward to add the color information as an additional feature in the tracking and segmentation system. Also, additional sensory input, like the odometry, can be integrated for a more robust robot navigation. Specifically, the independent motion detection task can benefit from such a fusion, since the background modeling and depth parallax may be solved or verified using additional sensory modi, which robustify the here proposed methods.
- ⇒ The implementation of the proposed algorithms did not aim to realize a real-time processing, but to test the feasibility of the theoretical concepts in general, it may be advantageous to optimize the software realization to achieve a more time efficient system.
- ⇒ **3D Omnidirectional Imaging** The redefinition of independent motion detection problem as a temporal tracking problem and the proposed solution—that employs the particle filter scheme for the probabilistic tracking, allows for the robust detection of motion in many real indoor navigation scenarios. Nevertheless, the ambiguity caused by the depth parallax remains to be an unsolved problem in generalizing the proposed methods. This problem can be solved by capturing the depth structure of the scene with an active sensor capable of measuring the 3D data in real time. The PMD-3D-Camera is an active sensor that fulfills these requirements. The sensor is capable of detecting 3D structure of a scene in real time, using spatially organized and in grid form arranged active sensing elements that measure the 3D data in parallel. The 3D data supplied by this sensor can be integrated into the particle filter scheme for an active tracking of objects. It is also possible to

integrate the 3D data in the earlier stages of the processing, namely at the stage of optical flow calculation, modifying the BCC(Eqn.(2.1)):

$$\begin{pmatrix} I_x(\vec{x}, t) \\ I_y(\vec{x}, t) \\ I_z(\vec{x}, t) \end{pmatrix} \cdot \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} + I_t(\vec{x}, t) = 0 \quad (6.1)$$

where the vector \vec{x} is now 3 dimensional. The equation yields not only the velocities in x, y but also in z direction. The rate of change in z direction can then be exploited to disambiguate the depth parallax problem. The existing active sensor is capable of capturing limited field of view. The possible modification of the sensor with a reflective surface would only deliver feasible measurement data, if the sensor dimensions and the magnification factor of the mirror surface are comparable.

Experiments with the Catadioptrical PMD-Camera:

We conducted several experiments to investigate the feasibility of 3D visual sensing with a wide angle catadioptric sensor. The 3D vision sensor (PMD-Camera), was enhanced by using the refractive surface to enlarge the field of view, as seen in Fig.(6.1).



Figure 6.1: *Pictures of the experimental setup and the environment*

The sensor is composed of a central receptive unit with 64 horizontal and 32 vertical pixels and two laterally positioned light sources. The pixels of the receptive unit are capable of capturing the (2D) grayvalue as well as the depth information. The depth information is measured by exploiting the Time-of-the-Flight Principle. Thus, the 3D scene is illuminated by means of a modulated optical sender (LED), and mapped onto the receptive array in

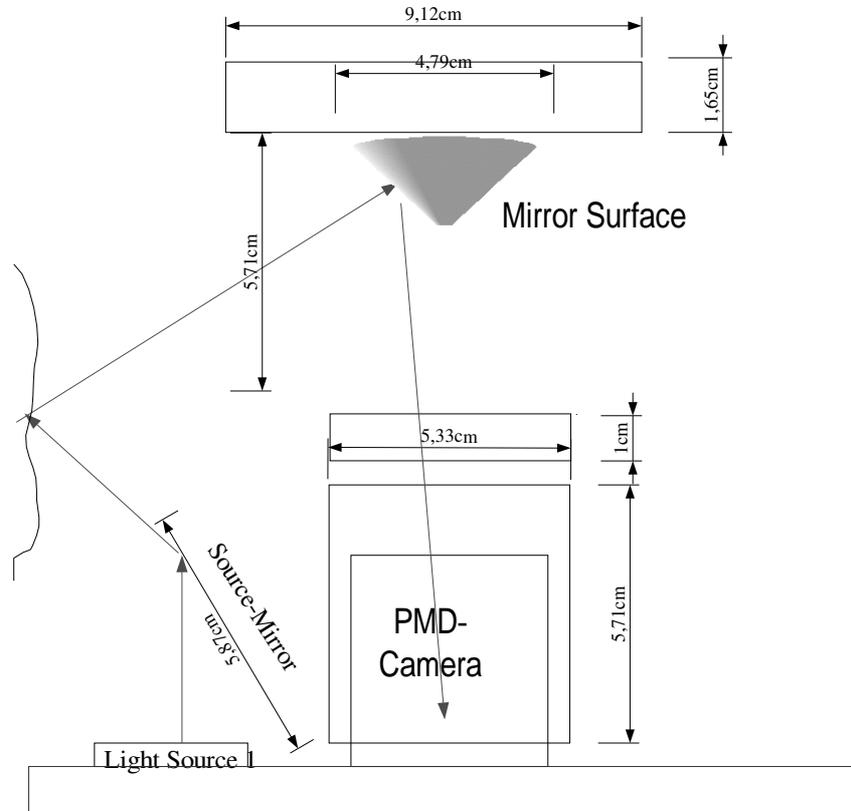


Figure 6.2: *Experimental setup of the 3D Omniscamera.*

the focal plane of the receiving optics for the echo time-of-flight and distance evaluation [101].

The camera optics included a lens with a 50° aperture. To widen this field of view, we modified this setup, adding a refractive surface to the top of the receptive unit. One of the lighting sources was also bent on one side using a plane mirror surface (see Fig.(6.2)). This setup enlarges the field of view with a constant factor of $\alpha = 3.6$. The magnification factor α is constant, due to the design of the hyperbolic mirror that was used.

Refracted Omnidirectional View The experimental environment was a homogeneously illuminated indoor scene with sufficiently structured objects. In the resulting images shown in Fig.(6.3), a pen was brought into the field of view. The white spot can be seen in Fig.(6.3(a)), which depicts the grayvalue of the scene. One can distinguish the pen as a curve, since it is

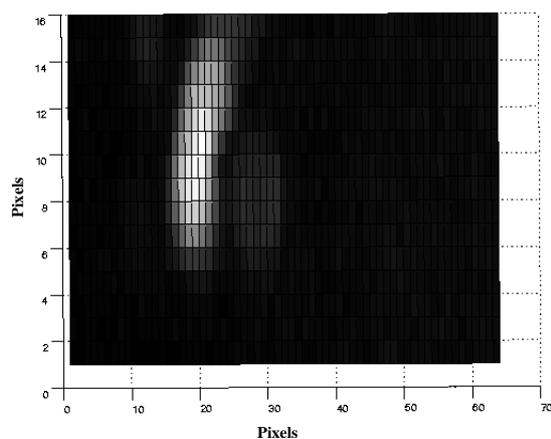
deformed by the refractive surface. In Fig.(6.3(b) - (c)), the distance maps are presented. One can spot the near region, which is where the pen was, and the background. Since we only used one of the light sources, only the half of the receptive unit receives structured light. This means that an expanded field of view ($50 * 3.6^\circ$) is projected onto the 32 x 16 pixel region of the receptive unit.

It should be noted that the depth measurements in this experiment are not distorted by the mirror surface. It is only the 2D shape of the pen which was bent due to the projection. Since the number of sensory elements are too small to capture the artificially enlarged environment, we only get a limited resolution of the 3D space. This problem could be solved by using the newer generations of the PMD-Camera with higher number of pixels (sensory elements).

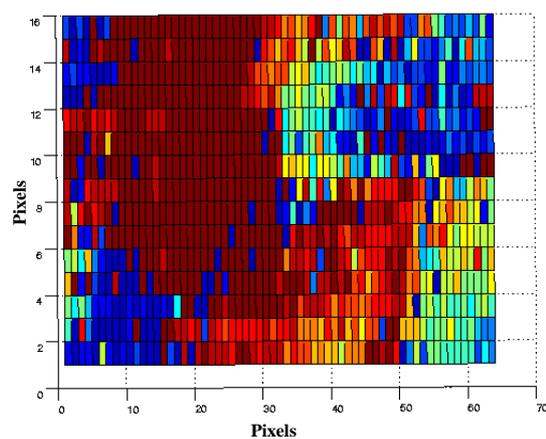
Omnidirectional View of a Rotated Camera In addition to the experiment with the refractive surface, we rotated the PMD-Camera around its vertical axis to achieve a panoramic view of the environment. The results are depicted in Fig.(6.4). Fig.(6.4(a)) shows the panoramic 2D scene constructed by stitching 36 grayvalue snapshots of the scene. The 3D image of the same scene is shown in Fig.(6.4(b)). The 3D image is color coded, red representing closer objects, and blue showing the farthest regions. Although postprocessing of the images to achieve an omnidirectional image is computationally more demanding, the 3D information can still be used for various segmentation tasks. However, such images cannot be employed to accomplish real time motion segmentation tasks, admittedly. The mechanical rotation of the camera around its axis and stitching the images to an omnidirectional image require additional computational processing, which is unfavorable for time critical applications. This experiment has been conducted only to investigate the advantages of enlarging the 3D field-of-view and to present the expected omnidirectional scene when the spatial resolution of the 3D sensor permits a reasonable projection.

Results and Future Work for 3D Omnidirectional Imaging:

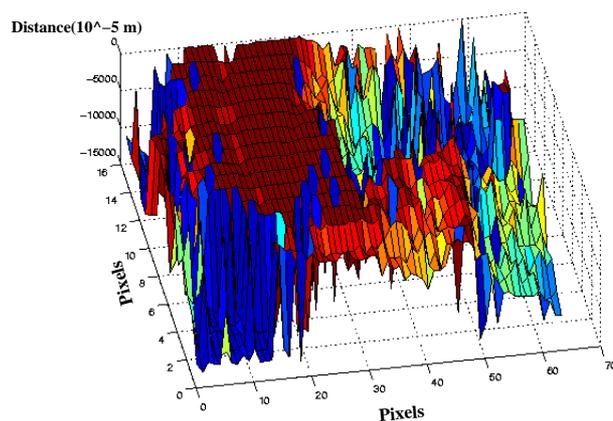
The first experiments show that widening the field of view using a refractive surface or mechanically, by rotating the camera on the vertical axis, allows a better interpretation of the scene. Nevertheless, using the refractive surface causes a mapping of the real scene onto a very small portion of the receptive unit of the camera. Additionally, the projective deformation of the scene due to the refraction on the mirror surface, makes a reliable interpretation of the scene difficult.



(a) Grayvalue image

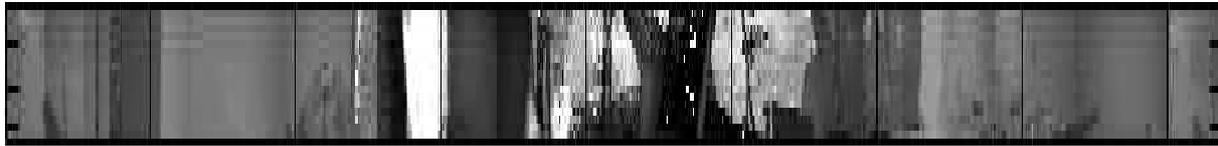


(b) Distance map

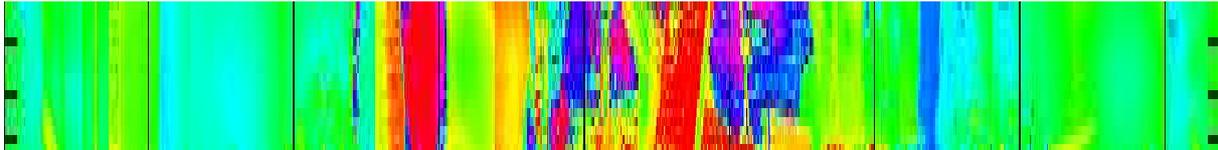


(c) Distance map from the perspective view

Figure 6.3: Images of a pen that has been brought nearer to the 3D omnidirectional camera. (a) The grayvalue image, where the brighter region indicates the pen (b) Distance map from the top view (c) Distance map from the perspective view; the red color indicate nearer regions. It can be seen that the region on the left, which is colored with dark red (the region with the pen), has smaller distance to the camera as the other parts of the map.



(a) Grayvalue Image



(b) Distance Image

Figure 6.4: *Results: 360° view of the indoor scene acquired by the PMD Camera, which has been rotated around its vertical axis. (a) The grayvalue image of the indoor lab scene. (b) 3D image of the 360° view. The red regions indicate nearer objects. One can spot some edges in the middle which represent a side view of a laptop-monitor placed in front of the camera.*

In our experiments, the panoramic images achieved by rotating the camera, provided more information about the real scene. In the future work, one can consider a camera setup with a larger receptive unit matrix. Also using both of the light sources provided by the PMD-Camera will be advantageous.

Appendix A

Egomotion Estimation

The egomotion problem can be formalized as follows:

$$\vec{V} = -\vec{T} - \vec{w} \times \vec{P} \quad (\text{A.1})$$

We can write the velocity vector V with the temporal derivatives of the three dimensional distance vectors (X, Y, Z) for a point P in space:

$$\vec{V} = [X', Y', Z']^T$$

The rotational (\vec{w}) and the translational (\vec{T}) components of the motion can also be defined as:

$$\begin{aligned} \vec{w} &= [\alpha, \beta, \mu] \\ \vec{T} &= [t_X, t_Y, t_Z] \end{aligned}$$

where α, β, μ represent the angular and t_X, t_Y, t_Z represent the translational displacement around each of the three cartesian axis respectively (see Fig.(A.1)). The egomotion equation (Eqn.(A.1)) can then be reformulated using the above definition:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = - \begin{bmatrix} t_X \\ t_Y \\ t_Z \end{bmatrix} - \begin{bmatrix} \alpha \\ \beta \\ \mu \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (\text{A.2})$$

or in open form:

$$\begin{aligned} X' &= -t_X - \beta \cdot Z + \mu \cdot Y \\ Y' &= -t_Y - \mu \cdot X + \alpha \cdot Z \\ Z' &= -t_Z - \alpha \cdot Y + \beta \cdot X \end{aligned} \quad (\text{A.3})$$

The image plane coordinates (x, y) can be written due to the perspective projection as follows:

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}$$

The estimated optical flow vectors u and v represent the spatiotemporal deviation on the image plane. Therefore:

$$u = \frac{dx}{dt}, \quad v = \frac{dy}{dt}$$

Substituting the 3D coordinates (X, Y, Z) into above formulation, we get:

$$\begin{aligned} u &= \frac{d}{dt} \left(\frac{X}{Z} \right) \Rightarrow \frac{X'}{Z} - \frac{X \cdot Z'}{Z^2} \\ v &= \frac{d}{dt} \left(\frac{Y}{Z} \right) \Rightarrow \frac{Y'}{Z} - \frac{Y \cdot Z'}{Z^2} \end{aligned}$$

Substituting A.3 for the components X', Y' and Z' , we get:

$$\begin{aligned} u &= \frac{1}{Z} (-t_X + xt_Z - (1 + x^2)\beta - \mu y - \alpha xy) \\ v &= \frac{1}{Z} (-t_Y + yt_Z - (1 + y^2)\alpha - \mu x - \beta xy) \end{aligned}$$

Writing this equation system in matrix form and defining the image point as $\vec{p} = [x, y, 1]^T$, we get the formula for the image velocity $\vec{v}(\vec{p})$, as referred in Eqn.(4.2) in Section 4.1:

$$\begin{aligned} \vec{v}(\vec{p}) &= \begin{bmatrix} u \\ v \end{bmatrix} \\ &= \left(-\frac{1}{Z} \right) \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix} (\vec{T} + \vec{w} \times \vec{p}) \end{aligned}$$

Appendix B

Nonlinear Least-Squares Fitting: Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm [77] provides a numerical solution to the problem of minimizing nonlinear functions. Formally, the problem of multidimensional nonlinear least-squares fitting requires the minimization of the squared residuals of n functions, f_i , in p parameters, x_i ,

$$\begin{aligned}\Phi(x) &= (1/2) \sum_{i=1}^n f_i(x_1, \dots, x_p)^2 \\ &= (1/2) \|F(x)\|^2\end{aligned}\tag{B.1}$$

The algorithm proceeds from an initial guess using the linearization,

$$\psi(q) = \|F(x + q)\| = \|F(x) + Jq\|\tag{B.2}$$

where x is the initial point, q is the proposed step and J is the Jacobian matrix $J_{ij} = df_i/dx_j$.

Generally, at a minimum of the sum of squares S , we should be able to write $\nabla_q S = 0$. Therefore, with the above linearization, this leads to the following equation:

$$(J^T J)q = -J^T F(x)\tag{B.3}$$

from which q can be obtained by inverting the term $(J^T J)$, which is done using the ‘damped version’ of the above equation:

$$(J^T J + \lambda)q = -J^T F(x).\tag{B.4}$$

The damping factor λ is adjusted each iteration. If the reduction of S is rapid, a smaller value can be used, whereas if an iteration gives insufficient reduction in the residual, λ can be increased giving a step closer to the gradient descent direction.

Bibliography

- [1] ADELSON, E. H., AND BERGEN, J. R. Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America* 2, 2 (February 1985), 284–299.
- [2] ADIV, G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7, 4 (July 1985), 384–401.
- [3] ALBRIGHT, T. D., DESIMONE, R., AND GROSS, C. G. Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology* 51 (1984), 16–31.
- [4] ALOIMONOS, Y., AND DURIC, Z. Estimating the heading direction using normal flow. *International Journal of Computer Vision* 13, 1 (September 1994), 33–56.
- [5] ANANDAN, P. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision* 2 (1988), 283–310.
- [6] ARGYROS, A., AND ORPHANOUDAKIS, S. Independent 3d motion detection based on depth elimination in normal flow fields. *Proc. CVPR'97* (1997), 672–677.
- [7] BAB-HADIASHAR, A., AND SUTER, D. Robust optical flow computation. *International Journal of Computer Vision* 29, 1 (1998), 59–77.
- [8] BARRON, J. L., BEAUCHEMIN, S. S., AND FLEET, D. J. On optical flow. In *6th Int. Conf. on Artificial Intelligence and Information-Control Systems of Robots (AIICSR)* (Bratislava, Slovakia, 1994), I. Plander, Ed., World Scientific, pp. 3–14. Sept. 12–16, 1994, Smolenice Castle, Slovakia.

-
- [9] BARRON, J. L., FLEET, D. J., AND BEAUCHEMIN, S. S. Performance of optical flow techniques. *International Journal of Computer Vision* 12, 1 (1994), 43–77.
- [10] BARRON, J. L., FLEET, D. J., BEAUCHEMIN, S. S., AND BURKITT, T. A. Performance of optical flow techniques. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '92)* (June 1992), IEEE Computer Society, pp. 236–242.
- [11] BEAUCHEMIN, S., AND BARRON, J. On discontinuous optical flow. *Computers and Artificial Intelligence* 20, 3 (2000).
- [12] BEAUCHEMIN, S. S., AND BARRON, J. L. The computation of optical flow. *ACM Computing Surveys* 27, 3 (1995), 433–467.
- [13] BERNARDINO, A., SANTOS-VICTOR, J., AND SANDINI, G. Tracking planar structures with log-polar images. Tech. Rep. VisLab-TR 06/2000, Instituto Superior Tecnico, Lisboa, Portugal, 2000.
- [14] BLACK, M. *Robust Incremental Optical Flow*. PhD thesis, Yale University, 1992.
- [15] BLACK, M. J., FLEET, D. J., AND YACOOB, Y. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding* 78, 1 (2000), 8–31.
- [16] BOLUDA, J. Motion detection independent of the camera movement with a logpolar sensor. In *Proc. of the 13th International Conference on Digital Signal Processing* (1997), pp. 809–812. Santorini, Greece, July 1997.
- [17] BRADLEY, D., MAXWELL, M., ANDERSEN, R. A., BANKS, M. S., AND SHENOY, K. V. Mechanisms of heading perception in primate visual cortex. *Science* 273, 5281 (1996), 1544–1547.
- [18] BRUSS, A., AND HORN, B. Passive navigation. *Computer Vision, Graphics, and Image Understanding* 21 (1983), 3–20.
- [19] BÜLTHOFF, H., LEE, S.-W., T. POGGIO, AND C. WALLRAVEN, Eds. *Modeling Insect Compound Eyes: Space-Variant Spherical Vision* (2002), Springer.
- [20] CAMUS, T. *Real-Time Optical Flow*. PhD thesis, Brown University, 1994.

-
- [21] CARTWRIGHT, B. A., AND COLLETT, T. S. Landmark learning in bees. *Journal of Comparative Physiology*, 151 (1983), 521–543.
- [22] CHAHL, J., AND SRINIVASAN, M. Navigation, path planning and homing for autonomous mobile robots using panoramic visual sensors. In *Proc. of AISB Workshop on Spatial Reasoning in Mobile Robots and Animals* (Manchester, May 1997), pp. 47–55.
- [23] CHAHL, J., AND SRINIVASAN, M. Range estimation with a panoramic visual sensor. *J. Opt. Soc. Am. A* 14, 9 (September 1997), 2144–2151.
- [24] CHAHL, J., AND SRINIVASAN, M. A complete panoramic vision system, incorporating imaging, ranging, and three dimensional navigation. In *OMNIVIS 2000* (2000), pp. 104–110.
- [25] CHAHL, J. S., AND SRINIVASAN, M. Reflective surfaces for panoramic imaging. *Applied Optics* 36, 31 (1997), 8275–8285.
- [26] CHANG, C., AND GAUDIANO, P. Editorial: Biomimetic robots. *Robotics and Autonomous Systems* 30 (2000), 1–2.
- [27] COX, I. J., ROY, S., AND HINGORANI, S. L. Dynamic histogram warping of image pairs for constant image brightness. In *IEEE International Conference on Image Processing* (1995), vol. II, IEEE Press, pp. 366–369. October 23–26, 1995, Washington. Extended version available as an NEC technical report.
- [28] CRONIN, T. W., AND MARSHALL, J. Parallel processing and image analysis in the eyes of mantis shrimps. *Biol. Bull.*, 200 (2001), 177–183.
- [29] DANIILIDIS, K. Attentive visual motion processing: Computations in the log-polar plane. In *Theoretical Foundations of Computer Vision* (1994), pp. 1–20.
- [30] DANIILIDIS, K. Computation of 3-d-motion parameters using the log-polar transform. In *Computer Analysis of Images and Patterns* (1995), pp. 82–89.
- [31] DANIILIDIS, K., AND GEYER, C. Omnidirectional vision: Theory and algorithms. In *Proc. Int. Conf. Pattern Recognition, ICPR'00* (2000), IEEE.
- [32] DANIILIDIS, K., AND KRUEGER, V. Optical flow computation in the log-polar-plane. In *Computer Analysis of Images and Patterns* (1995), pp. 65–72.

- [33] DANILIDIS, K., MAKADIA, A., AND BÜLOW, T. Image processing in catadioptric planes: Spatiotemporal derivatives and optical flow computation. In *Proc. IEEE Workshop on Omnidirectional Vision 2002* (June, 2002), pp. 3–10.
- [34] DUCHON, A., WARREN, W., AND KAEHLING, L. Ecological robotics. *Adaptive Behaviour* 6, Issue 3-4, Special Issue on biologically inspired models of navigation (1998), 473–507.
- [35] DUFFY, C. J., AND WURTZ, R. H. Sensitivity of MST neurons to optic flow stimuli. i. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology* 65 (1991), 1329–1345.
- [36] EGELHAAF, M., AND BORST, A. Motion computation and visual orientation in flies. *Comparative Biochemical Physiology* 104A, 4 (1993), 659–673.
- [37] EUCLID. The optics of euclid. *Journal of the Optical Society of America* 35, 5 (May 1945), 357–372. c. 300 B.C., English translation from the Greek by H. E. Burton.
- [38] FAJEN, B., AND WARREN, W. Go with the flow. *Trends in Cognitive Sciences* 4, 10 (2000), 369–370.
- [39] FAUGERAS, O., LUONG, Q.-T., AND PAPADOPOULOU, T. *The geometry of multiple images*. MIT Press, 2001.
- [40] FERMÜLLER, C., SHULMAN, D., AND ALOIMONOS, Y. The statistics of optical flow. *Computer Vision and Image Understanding* 82, 1 (2001), 1–32.
- [41] FLEET, D., AND LANGLEY, K. Recursive filters for optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 1 (1995), 61–67.
- [42] FLEET, D. J., AND JEPSON, A. D. Computation of component image velocity from local phase information. *International Journal of Computer Vision* 5, 1 (1990), 77–104.
- [43] FLEURY, M., CLARK, A., AND DOWNTON, A. Evaluating optical-flow algorithms on a parallel machine. *Image and Vision Computing* 19 (2001), 131–143.
- [44] FORSYTH, D., AND PONCE, J. *Computer Vision a Modern Approach*. Prentice Hall, 2003.

-
- [45] FOX, D. Adapting the sample size in particle filters through kld-sampling. *International Journal of Robotics Research* (2003).
- [46] FRANCESCHINI, N., PICHON, J., AND BLANES, C. From insect vision to robot vision. *Philosophical Transactions of the Royal Society London: Biological Sciences (SeriesB)*, 337 (1992), 283–294.
- [47] FRANZ, M. O., AND MALLOT, H. A. Biomimetic robot navigation. *Robotics and Autonomous Systems* 30 (2000), 133–153.
- [48] FRAZIER, J., AND NEVATIA, R. Detecting moving objects from a moving platform. In *DARPA90* (1990), pp. 348–355.
- [49] GALVIN, B., MCCANE, B., NOVINS, K., MASON, D., AND MILLS, S. Recovering motion fields: An evaluation of eight optical flow algorithms. In *Proc. of the 9th British Machine Vision Conference (BMVC '98)* (1998), P. H. Lewis and M. S. Nixon, Eds., vol. 1. Southampton, UK, September 14–17, 1998.
- [50] GIACHETTI, A. Matching techniques to compute image motion. *Image and Vision Computing* 18 (2000), 247–260.
- [51] GIACHETTI, A., CAMPANI, M., AND TORRE, V. The use of optical flow for the autonomous navigation. *Lecture Notes in Computer Science* 800 (1994), 146–151.
- [52] GIBSON, J. *The Perception of the Visual World*. Houghton Mifflin, Boston (MA), 1950.
- [53] GLUCKMAN, J., AND NAYAR, S. Egomotion with omni cameras. In *Proc. of IEEE International Conference on Computer Vision, ICCV 98* (Bombay, India, 1998), pp. 999–1005.
- [54] GONZALES, R., AND WOODS, R. *Digital Image Processing*. Prentice Hall, 2002.
- [55] GUPTA, N., AND KANAL, L. 3-D motion estimation from motion field. *Artificial Intelligence* 78, 1–2 (1995), 45–86.
- [56] HAUSSECKER, H., AND FLEET, D. Computing optical flow with physical models of brightness variation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)* (June 2000), vol. II, IEEE Computer Society, pp. 760–767.

-
- [57] HEEGER, D. J. Optical flow using spatiotemporal filters. *International Journal of Computer Vision* 1, 4 (1988), 279–301.
- [58] HONG, J. Image based homing. In *Proc. of IEEE Int. Conf. on Robotics and Automation* (1991).
- [59] HORN, B. K. P., AND SCHUNCK, B. G. Determining optical flow. *Artificial Intelligence* 17 (1981), 185–203.
- [60] HORRIDGE, G. The evolution of visual processing and the construction of seeing systems. In *Proc. Royal Soc. London B* (1987), vol. 230, pp. 279–292.
- [61] HUBEL, D. H., AND WIESEL, T. N. Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex. *Journal of Physiology (London)* 160 (1962), 106–154.
- [62] IRANI, M., AND ANANDAN, P. A unified approach to moving object detection in 2d and 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 6 (1998), 577–589.
- [63] JAIN, R. Segmentation of frame sequences obtained by a moving observer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 5 (1984), 624–629.
- [64] JEPSON, A., AND HEEGER, D. Linear subspace methods for recovering translational direction. Tech. rep., 1992.
- [65] KEARNEY, J. K., THOMPSON, W. B., AND BOLEY, D. L. Optical flow estimation: An error analysis of gradient-based methods with local optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 2 (1987), 229–244.
- [66] KOENDERINK, J. J. Optic flow. *Vision Research* 26 (1986), 161–180.
- [67] KOLLER, D., DANIILIDIS, K., AND NAGEL, H. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision* 10, 3 (June 1993), 257–281.
- [68] KONRAD, J., AND DUBOIS, E. Multigrid bayesian estimation of image motion fields using stochastic relaxation. In *Proc. Second International Conference on Computer Vision, ICCV’88* (December 1988), IEEE, pp. 354–362.

- [69] KRAFT, H., FREY, J., MOELLER, T., ALBRECHT, M., GROTHOF, M., SCHINK, B., HESS, H., AND BUXBAUM, B. 3d-camera of high 3d-frame rate, depth-resolution and background light elimination based on improved pmd (photonic mixer device)-technologies. In *Proc. OPTO-04* (2004).
- [70] KRAPP, H. G., AND HENGSTENBERG, R. Estimation of self-motion by optic flow processing in single visual interneurons. *Nature*, 384 (1996), 463–466.
- [71] LEE, D. N. The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society of London (Series B)* 290 (1980), 169–179.
- [72] LIU, H., HONG, T., HERMAN, M., CAMUS, T., AND CHELLAPPA, R. Accuracy vs efficiency trade-offs in optical flow algorithms. *Computer Vision and Image Understanding: CVIU* 72, 3 (1998), 271–286.
- [73] LOURAKIS, M., ARGYROS, A., AND ORPHANOUDAKIS, S. Independent 3d motion detection using residual parallax normal flow fields. In *Proc. Sixth Int. Conf. on Computer Vision, ICCV03* (2003), IEEE, pp. 1012–1017.
- [74] LUCAS, B. D., AND KANADE, T. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh Int. joint Conf. on AI, IJCAI.81* (Vancouver, B.C., Canada, August 1981), vol. 2, pp. 674–679.
- [75] MACLEAN, W., JEPSON, A., AND FRECKER, R. Recovery of egomotion and segmentation of independent object motion using the em algorithm. In *Proc. 5th British Machine Vision Conference* (York, 1994), E. Hancock, Ed., vol. 1, BMVA Press, pp. 175–184.
- [76] MARCHESE, F., AND SORRENTI, G. Mirror design of a prescribed accuracy omni-directional vision system. In *Proc. IEEE Workshop on Omnidirectional Vision, OMNIVIS2002* (2002), pp. 136–142.
- [77] MORÉ, J. The levenberg-marquardt algorithm: Implementation and theory. In *Numerical Analysis, Lecture Notes in Mathematics 630*, G. Watson, Ed. Springer Verlag, 1977, pp. 105–116.
- [78] M.V. SRINIVASAN, M. L., AND HERRIDGE, G. Visual figure- ground discrimination in the honeybee: the role of motion parallax at boundaries. *Proc. R. Soc. London B*, 238 (1990), 331–350.
- [79] M.V. SRINIVASAN, M. LEHRER, W. K., AND ZHANG, S. Range perception through apparent image speed in freely-flying honeybees. *Visual Neuroscience*, 6 (1991), 519–535.

-
- [80] NAGEL, H. H. Direct estimation of optical flow and of its derivatives. In *Artificial and biological vision systems*, G. Orban, Ed., ESPRIT basic research series. Berlin(u.a.):Springer, 1992, pp. 193–224.
- [81] NAKAYAMA, K. Biological image motion processing: A review. *Vision Research* 25, 5 (1985), 625–660.
- [82] NALWA, V. A true omnidirectional viewer. Tech. rep., Bell Laboratories, Holmdel, NJ 07733, USA, 1996, February.
- [83] NEGAHDARIPOUR, S. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 9 (1998), 961–979.
- [84] NELSON, R. Qualitative detection of motion by a moving observer. *International Journal of Computer Vision* 7, 1 (1991), 33–46.
- [85] NELSON, R., AND ALOIMONOS, J. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58 (1988), 261–273.
- [86] NESI, P., BIMBO, A. D., AND BEN-TZVI, D. A robust algorithm for optical flow estimation. *Computer Vision and Image Understanding* 62, 1 (July 1995), 59–68.
- [87] NEVEN, H., AND SCHONER, G. Dynamics parametrically controlled by image correlations organize robot navigation. *Biological Cybernetics* 75 (1996), 293–307.
- [88] OHYA, A., KOSAKA, A., AND KAK, A. Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing. *Transactions on Robotics and Automation* 14, 6 (December 1998), 969–978.
- [89] PLESS, A., BRODSKY, T., AND ALOIMONOS, Y. Independent motion: The importance of history. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99)* (1999), pp. 92–97.
- [90] POLANA, R., AND NELSON, R. Detection and recognition of periodic, nonrigid motion. *International Journal of Computer Vision* 23, 3 (June 1997), 261–282.

- [91] PRADZNY, K. Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer. *Computer Graphics and Image Processing 17* (1981), 238–248.
- [92] PRAZDNY, K. On the information in optical flow. *Computer Vision, Graphics and Image Processing 22* (1983), 239–259.
- [93] REDLICK, F. P., JENKIN, M., AND HARRIS, L. R. Humans can use optic flow to estimate distance of travel. *Vision Research 41* (2001), 213–219.
- [94] REICHARDT, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication*, W. A. Rosenblith, Ed. Wiley, New York, 1961, pp. 303–317.
- [95] ROUGEAUX, S. *Real-Time Active Vision for Versatile Interaction*. PhD thesis, Université d'Evry, France, 1999.
- [96] RÖWEKAMP, T., AND PETERS, L. Intelligent RealTime sensor system for optical flow estimation. In *Proc. 30th ISATA, Dedicated Conference on Robotics, Motion and Machine Vision* (1997). Florence, Italy, June 1997.
- [97] RUSSELL, S., AND NORVIG, P. *Artificial Intelligence: A Modern Approach*, 2nd edition ed. Pearson Education, Inc, publishing as Prentice Hall, 2003.
- [98] SANDINI, G., SANTOS-VICTOR, J., CUROTTO, F., AND GARIBALDI, S. Robotic bees. In *Proc. of IROS-93* (Yokohoma, Japan, July 1993).
- [99] SANTOS-VICTOR, J., AND SANDINI, G. Uncalibrated obstacle detection using normal flow. *Machine Vision and Applications 9*, 3 (1996), 130–137.
- [100] SCHRATER, P. R., KNILL, D. C., AND SIMONCELLI, E. P. Perceiving visual expansion without optic flow. *Nature 410*, 6830 (2001), 816–819.
- [101] SCHWARTE, R. Real time 3d-perception by tof-echoing, 3d-video cameras. Tech. rep., INV Inst. f \tilde{A} $\frac{1}{4}$ r Nachrichtenverarbeitung, ZESS, PMD Technologies, 2004.
- [102] SCHWARTE, R., ZHANG, Z., GROTHOF, M., FREY, J., KRAFT, H., MOELLER, T., AND HESS, H. Oep(opto electronic processor) for extremely fast multi-channel analog and digital oe-signal-processing. In *Proc. OPTO'04* (2004).

- [103] SEKULER, R., WATAMANIUK, S., AND BLAKE, R. Perception of visual motion. In *Stevens' Handbook of Experimental Psychology*, H. Pashler and S. Yantis, Eds., 3rd ed., vol. 1: Sensation and Perception. Wiley, New York, 2002.
- [104] SILVA, C., AND SANTOS-VICTOR, J. Egomotion estimation using log-polar images. Tech. Rep. VisLab-TR 01/98, Instituto Superior Tecnico, Lisboa, Portugal, 1998.
- [105] SIMONCELLI, E. *Distributed Representation and Analysis of Visual Motion*. Phd thesis, Massachusetts of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, January 1993.
- [106] SIMONCELLI, E., ADELSON, E., AND HEEGER, D. Probability distributions of optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '91)* (1991), IEEE Computer Society, pp. 310–315. Maui, HI, June 1991.
- [107] SINGH, A. Robust computation of image-motion and scene-depth. In *Proc. International Conference on Robotics and Automation, ICRA'91* (Sacramento, California, April 1991), IEEE, pp. 2730–2737.
- [108] SONKA, M., HLAVAC, V., AND BOYLE, R. *Image Processing, Analysis, and Machine Vision*. Brooks and Cole Publishing, 1998.
- [109] SRINIVASAN, M., ZHANG, S., LEHRER, M., AND COLLETT, T. Honeybee navigation en route to the goal: Visual flight control and odometry. *Journal of Experimental Biology*, 199 (1996), 237–244.
- [110] STEIN, G., AND SHASHUA, A. Model-based brightness constraints: On direct estimation of structure and motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 9 (2000), 992–1015.
- [111] STILLER, C., AND KONRAD, J. Estimating motion in image sequences: A tutorial on modeling and computation of 2d motion. *IEEE Signal Processing Magazine* 16 (July 1999), 70–91.
- [112] STOCKER, A., AND DOUGLAS, R. Computation of smooth optical flow in a feedback connected analog network. *Advances in Neural Information Processing Systems* 12 (1998).
- [113] STRATMANN, I. Omnidirectional imaging and optical flow. In *Proc. Workshop on Omnidirectional Vision 2002* (2002), pp. 104–111.

- [114] STRATMANN, I. Visual detection of motion using omnidirectional imaging and optical flow. In *Proc. of Mechatronics and Robotics 04* (September 2004), pp. 907–912.
- [115] STRATMANN, I., AND SOLDA, E. Omnidirectional vision and inertial clues for robot navigation. In *Proc. 11th Int. Conf. on Advanced Robotics, ICAR'03* (2003), pp. 1915–1921.
- [116] STRATMANN, I., AND SOLDA, E. Omnidirectional vision and inertial clues for robot navigation. *Journal of Robotic Systems* 21, 1 (January 2004), 33–39.
- [117] SURMANN, H., LINGEMANN, K., NÄCHTER, A., AND HERTZBERG, J. Fast acquiring and analysis of three dimensional laser range data. In *Proc. 6th Int. Fall Workshop Vision, Modelling and Visualization* (November 2001), pp. 59–66.
- [118] SVOBODA, T. *Central Panoramic Cameras, Design, Geometry, Egomotion*. PhD thesis, Czech Technical University, Center for Machine Perception, 1999.
- [119] TISTARELLI, M., AND SANDINI, G. Estimation of depth from motion using anthropomorphic visual sensor. *Image Vision Computation* 8 (1990), 271–278.
- [120] TISTARELLI, M., AND SANDINI, G. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 4 (April 1993), 401–410.
- [121] TSAI, R., AND HUANG, T. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis Machine Intelligence PAMI-6*, 1 (January 1984), 13–27.
- [122] ULLMAN, S. *The Interpretation of Visual Motion*. The MIT Press, 1979.
- [123] ULLMAN, S. Maximizing rigidity: The incremental recovery of 3-d structure from rigid and nonrigid motion. *Perception* 13, 3 (1984), 255–274.
- [124] VERRI, A., AND POGGIO, T. Motion field and optical flow: Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 5 (1989), 490–498.

- [125] VON HELMHOLTZ, H. *Handbuch der physiologischen Optik*, 3 ed., vol. 3. Verlag von Leopold Voss, Hamburg und Leipzig, 1910. Dritte Auflage ergänzt und herausgegeben in Gemeinschaft mit A. Gullstrand, J. von Kries und W. Nagel.
- [126] WALLACH, H., AND O'CONNELL, D. N. The kinetic depth effect. *Journal of Experimental Psychology* 45 (1953), 205–217.
- [127] WARREN, W. H., AND HANNON, D. J. Direction of self-motion is perceived from optical flow. *Nature* 336, 6195 (1988), 162–163.
- [128] WARREN, W. H., KAY, B., ZOSH, W. D., DUCHON, A. P., AND S.SAHUC. Optic flow is used to control human walking. *Nature Neuroscience* 4 (2001), 213–216.
- [129] WEBER, J., AND MALIK, J. Robust computation of optical flow in a multi-scale differential framework. *International Journal of Computer Vision* 14, 1 (1995), 67–81.
- [130] WEHNER, R., MICHEL, B., AND ANTONSEN, P. Visual navigation in insects: Coupling of egocentric and geocentric information. *Journal of Experimental Biology*, 199 (1996), 129–140.
- [131] WESTIN, C., BHALERAO, A., KNUTSSON, H., AND KIKINIS, R. Using local 3d structure for segmentation of bone from computer tomography images. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'97)* (June 1997), IEEE.
- [132] YAGI, Y., AND KAWATO, S. Panoramic scene analysis with conic projection. In *Proc. Int. Conf. on Robots and Systems, IROS* (1990).
- [133] YAGI, Y., NISHII, W., YAMAZAWA, K., AND YACHIDA, M. Rolling motion estimation for mobile robot by using omnidirectional image sensor hyperomnivision. In *Proc. 1996 International Conference on Pattern Recognition ICPR'96* (1996), pp. 946–948.
- [134] YAMAZAWA, K., YAGI, Y., AND YACHIDA, M. Obstacle detection with omnidirectional image sensor hyperomni vision. In *Proc. Int. Conf. Robotics and Automation* (1995), IEEE, pp. 1062–1067.