Active Sensor Planning for Multiview Vision Tasks

Shengyong Chen · Youfu Li ·
Jianwei Zhang · Wanliang Wang

Active Sensor Planning for Multiview Vision Tasks

Springer

Shengyong Chen
College of Information Engineering
Zhejiang University of Technology
Hangzhou 310014, PR China
sy@ieee.org

Jianwei Zhang
AB TAMS, Department of Informatics
University of Hamburg
Vogt-Koelln-Strasse 30, D-22527
Hamburg, Germany
zhang@informatik.uni-hamburg.de

Y. F. Li
Dept of Manufacturing Engineering
and Engineering Management
City University of Hong Kong
83 Tat Chee Avenue, Kowloon
Hong Kong
meyfli@cityu.edu.hk

Wanliang Wang
College of Software Engineering
Zhejiang University of Technology
Hangzhou 310014
PR China

ISBN: 978-3-540-77071-8 e-ISBN: 978-3-540-77072-5

Library of Congress Control Number: 2007940478

Cover design: WMXDesign GmbH

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The problem of active sensor planning was firstly addressed about 20 years ago and attracted many people after then. Recently, active sensing becomes even more important than ever since a number of advanced robots are available now and many tasks require to act actively for obtaining 3D visual information from different aspects. Just like human beings, it's unimaginable if without active vision even only in one minute. Being active, the active sensor planner is able to manipulate sensing parameters in a controlled manner and performs active behaviors, such as active sensing, active placement, active calibration, active model construction, active illumination, etc. Active vision perception is an essential means of fulfilling such vision tasks that need take intentional actions, e.g. entire reconstruction of an unknown object or dimensional inspection of an industrial workpiece.

The intentional actions introduce active or purposive behaviors. The vision system (the observer) takes intentional actions according to its mind, the mind such as going to a specific location and obtaining the useful information of the target, in an uncertain environment and conditions. It has a strategic plan to finish a certain vision task, such as navigating through an unfamiliar environment or modeling of an unknown object. It is capable of executing the plan despite the presence of unanticipated objects and noisy sensors.

A multi-view strategy is often required for seeing object features from optimal placements since vision sensors have limited field of view and can only "see" a portion of a scene from a single viewpoint. This means that the performance of a vision sensor depends heavily both on the type and number of sensors and on the configuration of each sensor. What important is that the sensor is active. Compared with the typical passive vision where it is limited to what is offered by the preset visual parameters and environmental conditions, the active planner can instead determine how to view by utilizing its capability to change its visual parameters according to the scene for a specific task at any time.

From this idea, many problems have to be considered in constructing an active perception system and these important problems lead our motivation of the research on active sensor and sensing techniques. For many practical vision tasks, because, it is very necessary to develop a multiview plan of control strategy, and these viewpoints can be decided either offline or in run-time.

The purpose of this book is to introduce the challenging problems and propose some possible solutions. The main topics addressed are from both theoretical and technical aspects, including sensing activity, configuration, calibration, sensor modeling, sensing constraints, sensing evaluation, viewpoint decision, sensor placement graph, model based planning, path planning, planning for unknown environment, incremental 3D model construction, measurement, and surface analysis.

With our acknowledgements, the research work in this book was supported by the Research Grants Council of Hong Kong, the Natural Science Foundation of China, and the Alexander von Humboldt Foundation of Germany. Several colleagues, Z. Liu, B. He, B. Zhang, H. Zhang, and T. Tetsis, have contributed in part with the research work, experiments, and writing.

Shengyong Chen
Zhejiang University of Technology (China)

Contents

Chapter 1
Introduction

Active sensor planning is an important means for fulfilling vision tasks that require intentional actions, e.g. complete reconstruction of an unknown object or dimensional inspection of a workpiece. Constraint analysis, active sensor placement, active sensor configuration, and three-dimensional (3D) data acquisition are the essential steps in developing such active vision systems. This chapter presents the general motivations, ideas for solutions, and potential applications of active sensor planning for multiview vision tasks.

1.1 Motivations

The *intentional actions* in active sensor planning for visual perception introduce *active behaviors* or *purposeful behaviors*. The vision agent (the observer) takes intentional actions according to its set goal such as going to a specific location or obtaining the full information on an object, in the current environment and its own state. A strategic plan is needed to finish a vision task, such as navigating through an office environment or modeling an unknown object. In this way, the plan can be executed successfully despite the presence of unanticipated objects and noisy sensors.

Therefore there are four aspects that need to be studied in developing an active observer, i.e. the *sensor itself* (its type and measurement principle), the *sensor state* (its configuration and parameters), the *observer state* (its pose), and *the planner* for scene interpretation and action decision. Although many other things have to be considered in constructing an active perception system, these important issues lead to the research on active visual perception and investigations on visual sensing, system reconfiguration, automatic sensor planning, and interpretation and decision.

1.1.1 The Tasks

A critical problem in modern robotics is to endow the observer with a strategic plan to finish certain tasks. The multi-view strategy is an important means of taking active actions in visual perception, by which the vision sensor is purposefully placed at several positions to observe a target.

Sensor planning which determines the pose and configuration for the visual sensor thus plays an important role in active vision perception, not only because a 3D sensor has a limited field of view and can only see a portion of a scene from a single viewpoint, but also because a global description of objects often cannot be reconstructed from only one viewpoint due to occlusion. Multiple viewpoints have to be planned for many vision tasks to make the entire object (or all the features of interest) visible strategically.

For tasks of observing unknown objects or environments, the viewpoints have to be decided in run-time because there is no prior information about the targets. Furthermore, in an inaccessible environment, the vision agent has to be able to take intentional actions automatically. The fundamental objective of sensor placement in such tasks is to increase knowledge about the unseen portions of the viewing volume while satisfying all the placement constraints such as in-focus, field-of-view, occlusion, collision, etc. An optimal viewpoint planning strategy determines each subsequent vantage point and offers the obvious benefit of reducing and eliminating the labor required to acquire an object's surface geometry. A system without planning may need as many as seventy range images for recovering a 3D model with normal complexity, with significant overlap between them. It is possible to reduce the number of sensing operations to less than ten times with a proper sensor planning strategy. Furthermore, it also makes it possible to create a more accurate and complete model by utilizing a physics-based model of the vision sensor and its placement strategy.

For model-based tasks, especially for industrial inspections, the placements of the sensor need to be determined and optimized before robot operations. Generally in these tasks, the sensor planning is to find a set of admissible viewpoints in the acceptable space, which satisfy all of the sensor placement constraints and can finish the vision task well. In most of the related works, the constraints in sensor placement are expressed as a cost function with the aim to achieve the minimum cost. However, previously the evaluation of a viewpoint has normally been achieved by direct computation. Such an approach is usually formulated for a particular application and is therefore difficult to be applied to general tasks. For a multi-view sensing strategy, global optimization is desired but was rarely considered in the past.

In an active vision system, the visual sensor has to be moved frequently for purposeful visual perception. Since the targets may vary in size and distance and the task requirements may also change in observing different objects or features, a structure-fixed vision sensor is usually insufficient. For a structured light vision sensor, the camera needs to be able to "see" just the scene illuminated by the projector. Therefore the configuration of a vision setup often needs to be changed to reflect the constraints in different views and achieve optimal acquisition performance. A reconfigurable sensor can change its structural parameters to adapt itself to the scene to obtain maximum 3D information from the target.

In practical applications, in order to reconstruct an object with high accuracy, it is essential that the vision sensor be carefully calibrated. Traditional calibration methods are mainly for static uses in which a calibration target with specially designed features needs to be placed at precisely known locations. However, when the sensor is reconfigured, it must be re-calibrated again. To avoid the tedious and laborious procedures in such traditional calibrations, a self-recalibration method is needed to perform the task automatically so that 3D reconstruction can follow immediately without a calibration apparatus and any manual interference.

Finally, 3D reconstruction is either an ultimate goal or a means to the goal in 3D computer vision. For some tasks, such as reverse engineering and constructing environments for virtual reality, the 3D reconstruction of the target is the goal of the vision perception. A calibrated vision sensor which applies the 3D sensing techniques is used to measure the object surfaces in the scene. Then the obtained local models are globally integrated into a complete model for describing the target shape. For some other vision tasks, such as object recognition and industrial inspection, the 3D reconstruction is an important means to achieve the goal. In such a case, the 3D measurement is performed at every step for making a decision or drawing a conclusion about the target.

1.1.2 From a Biological View

I move, therefore I see. (Hamada 1992)

Active sensor planning now plays a most important role in practical vision systems because generally a global description of objects cannot be reconstructed from only one viewpoint due to occlusion or limited Field Of View (FOV). For example, in the case of object modeling tasks, because there is no prior information about the objects or environments, it is obviously very necessary to develop a multi-view plan of a controlling strategy, and these views can be decided either in run-time or off-line.

To illustrate the strong relationship between active perception and multi-view sensor planning, we may begin the explanation with a look of human behaviors. In humans, the operations and informational contents of the global state variable, which are sensations, images, feelings, thoughts and beliefs, constitute the experience of causation. In the field of neuro-dynamics and causality, Freeman (1999a) used circular causality to explain neural pattern formation by self-organizing dynamics. It explained how stimuli cause consciousness by referring to causality. An aspect of intentional action is causality, which we extrapolate to material objects in the world. Thus causality is a property of mind, not matter.

In the biological view (Fig. 1.1), in a passive information processing system a stimulus input gives information (Freeman 1999b), which is transduced by receptors into trains of impulses that signify the features of an object. Symbols are processed according to rules for learning and association and are then bound into a representation, which is stored, retrieved and matched with new incoming representations. In active systems perception begins with the emergence of a goal that is implemented by the search for information. The only input accepted is that which is consistent with the goal and anticipated as a consequence of the searching actions. The key component to be modeled in brains provides the dynamics that constructs goals and the adaptive actions by which they are achieved.

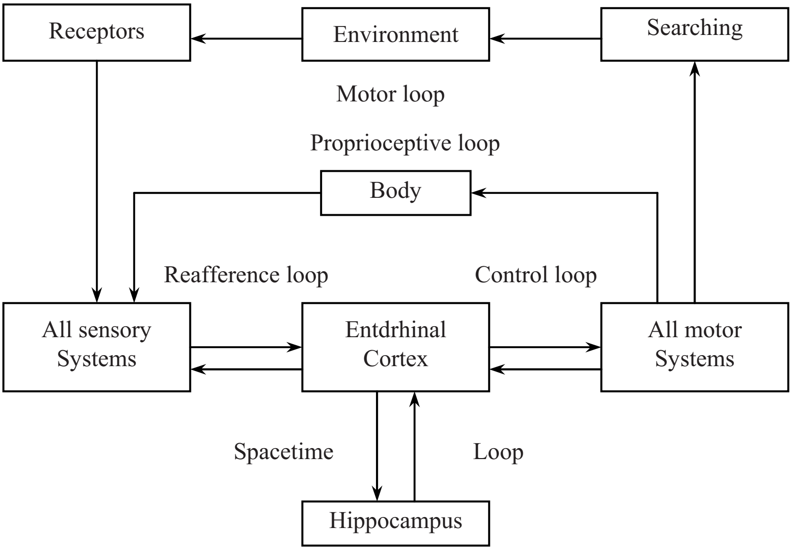

Fig. 1.1. The proprioceptive loop of human beings (Freeman 1999b)

1.1.3 The Problems and Goals

Many applications in robotics involve a good knowledge of the robot environment. 3D machine vision is the technology which allows computers to measure the three-dimensional shape of objects or environments, without resorting to physically probing their surfaces. The object/environment model is constructed in three stages. First, apply a computer vision technique called “shape from X” (e.g. shape from stereo) to determine the shapes of the objects visible in each image. The second stage is to integrate these image-based shape models into a single, complete shape model of the entire scene. Third, finally the shape model is rendered with the same color of the real object.

In developing such a technique, sensor planning is a critical issue since a typical 3D sensor can only sample a portion of an object at a single viewpoint. Using a vision sensor to sample all of the visible surfaces of any but the most trivial of objects, however, requires that multiple 3D images be taken from different vantage points and integrated, i.e. merged to form a complete model (Fig. 1.2). An optimal viewpoint planning strategy (or next best view – NBV algorithm) determines each subsequent vantage point and offers the obvious benefit of reducing and eliminating the labor required to acquire an object’s surface geometry.

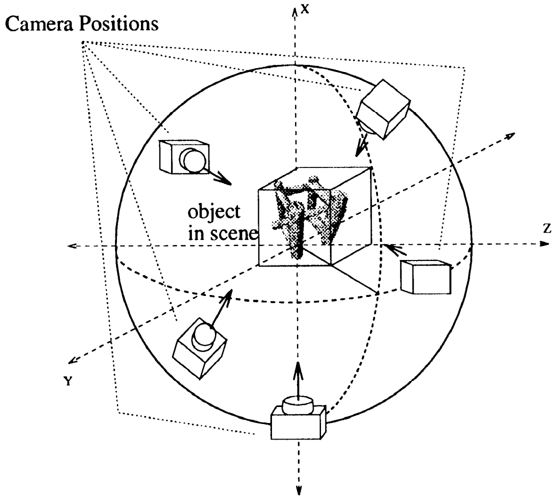

Fig. 1.2. A sequence of sensor pose placements for object modeling (Banta and Abidi 1996), p1→p2→...→pn

On the other hand, the performance of the vision perception of a robot, and thus the quality of knowledge learnt, can be significantly affected by the properties of illumination such as intensity and color. "Where to place the illuminants and how to set the illuminant parameters" for improving the process of vision tasks becomes an increasingly important problem that needs be solved. Drawing on the wide body of knowledge in radiometry and photometry will prove useful. For this purpose, this book also focuses attention on the topic of illumination planning in the robot vision system. Illumination planning can be considered as a part of the sensor planning problem for vision tasks. This book presents a study on obtaining the optimal illumination condition (or most comfortable condition) for vision perception. The "comfortable" condition for a robot eye is defined as: the image has a high signal-to-noise ratio and high contrast, is within the linear dynamic range of the vision sensor, and reflects the natural properties of the concerned object. "Discomfort" may occur if any of these criteria are not met because some scene information may not be recorded. This book also proposes appropriate methods to optimize the optical parameters of the luminaire and the sensor (including source radiant intensity, sensor aperture and focus length) and the pose parameters of the luminaire, with emphasis on controlling the intensity and avoiding glare.

The proposed strategy to implement placements of vision sensors and light sources requires an eye-and-hand setup allowing the sensors (a pair of stereo cameras or a structure of projector+camera) to be moving around and looking/shining at an object from different viewpoints. The sensor or luminaire is mounted on the end-effector of a robot to achieve arbitrary spatial position and orientation. The purpose of moving the camera is to arrive at viewing poses, such that required details of the unknown object can be acquired by the vision system for reconstruction of a 3D model. This book ignores the problems of the relative orientation between cameras and manipulator coordinate systems, and how to control the sensor movement.

Another goal of this book is to demonstrate the interdependence between a solution to the sensor planning problem and the other stages of vision image process, as well as the necessity and benefits of utilizing a model of the sensor when determining sensor setting and viewpoint.

1.1.4 Significance and Applications

The techniques described in this book may have outstanding significance in the many applications of computer vision. Using artificial vision for 3D object reconstruction and modeling is the technology which allows computers to obtain the three-dimensional shape of objects, without resorting to physically probing their surfaces. This is best suited for tasks where non-contact nature, a fast measurement rate and cost are of primary concern, especially for such applications as:

- medical applications,
- archeology,
- quality ensurance,
- reverse engineering,
- rapid product design,
- robotics, etc.

The technology of active sensor planning has its significance in both model based and non-model based applications of computer vision. Typical non-model based applications include:

- 3D object reconstruction and modeling,
- target searching,
- scene exploration,
- autonomous navigation, etc.

Model-based applications, where the object's geometry and a rough estimate of its pose are known, are widely used in:

- product assembly/disassembly,
- feature detection,
- inspection,
- object recognition,
- searching,
- dimensional measurement,
- surveillance,
- target tracking,
- monitoring, etc.

1.2 Objectives and Solutions

The general aim of this book is to introduce the ideas for possible solutions for the above-motivated problems in an active visual system. The objectives of the research include the following:

- To introduce the guideline for setting up typical active vision systems and applying it to 3D visual perception;
- To develop methods of active reconfiguration for purposive visual sensing;
- To investigate methodologies of automatic sensor planning for industrial applications;
- To propose strategies for sensor planning incorporation with illumination planning;
- To find solutions for exploring the 3D structure of an unknown target.

Among all of the above, this book places its emphasis on the last three issues. In the study of sensor planning, previous approaches mainly focused on the modeling of sensor constraints and calculating a “good” viewpoint for observing one or several features on the object. Little consideration was given to the overall efficiency of a generated plan with a sequence of viewpoints. In model-based vision tasks, researchers have made efforts to find an admissible domain of viewpoints to place the sensor to look at one or several object features. However, this method is difficult to apply in a multi-feature-multi-viewpoint problem as it cannot determine the minimum number of viewpoints and their relative distribution.

In non-model-based vision tasks, previous research efforts often concentrated on finding the best next views by volumetric analysis or occlusion as a guide. However, since no information about the unknown target exists, it is actually impossible to give the true best next view. It exists but can only be determined after the complete model has been obtained. Therefore a critical problem is still not well solved: the global optimization of sensor planning. When multiple features need to be observed and multiple viewpoints need to be planned, the minimum number of viewpoints needs to be determined. To achieve high efficiency and quality, the optimal spatial distribution of the viewpoints should be determined too. These are also related to the sensor configuration and environmental constraints. Furthermore, to make it flexible in practical applications, we need to deal with arbitrary object models without assumptions on the object features.

In this book, ideas are presented to solve the relevant issues in active sensing problems. Novel methodologies are developed in sensor configuration, 3D reconstruction, sensor placement, and viewpoint planning for multiview vision tasks.

In setting up the active vision system for the study, both traditional stereo cameras and coded structured light systems are investigated. The coded light approach can be adopted for the digital projector to generate binary patterns with light and dark stripes which are switchable during the operation. This method features high accuracy and reliability.

The sensor planning presented in this book is an effective strategy to generate a sequence of viewing poses and corresponding sensor configurations for optimal completion of a multiview vision task. Methods are proposed to solve the problems for both model-based and non-model-based vision tasks. For model-based applications, the method involves the determination of the optimal sensor placements and the shortest path through the viewpoints for automatic generation of a perception plan. A topology of the viewpoints is achieved by a genetic algorithm in which a min-max criterion is used for evaluation. The shortest path is also determined by graph theory. The sensing plan generated by the proposed methods leads to global optimization. For non-model-based applications, the method involves the decision of the exploration direction and the determination of the best next view and the corresponding sensor settings. Some cues are proposed to predict the unknown portion of an object or environment and the next best viewpoint is determined by the expected surface. The viewpoint determined in such a way is predictably best. Information Entropy Based Planning, uncertainty-driven planning, and self-termination conditions are also discussed.

Numerical simulations and practical experiments are conducted to implement the proposed methods for the active sensing in the multi-view vision task. The implementation results obtained in these initial experiments are only intended for showing the validity of proposed methods and the feasibility for practical applications. Using the active visual perception strategy, 3D reconstruction can be achieved without the constraints on the system configuration parameters. This allows optimal system configuration to be employed to adaptively sense an environment. The proposed methods will give the active vision system the adaptability needed in many practical applications.

1.3 Book Structure

This book is organized as follows.

- *Chapter 2* presents the sensing fundamentals, measurement principles, and 3D reconstruction methods for active visual sensing. These will be used in the next chapters in formulating the methods of sensor planning. It also describes the methods for dynamic reconfiguration and recalibration of a stripe light vision system to overcome practical scene challenge.
- *Chapter 3* summarizes the relevant works on 3D sensing techniques and introduces the active 3D visual sensing systems developed in the community. Both stereo sensors and structured light systems are mainly considered in this book, although extensions of other types of visual sensors such as laser scanners are straightforward.
- *Chapter 4* presents the sensor models, summarizes the previous approaches to the sensor planning problem, formulates sensor placement constraints, and proposes the criteria for plan evaluation. The method for the model-based sensor placement should meet both the optimal sensor placements and the shortest path

through these viewpoints. The plan for such sensor placements is evaluated based on the fulfillment of three conditions: low order, high precision, and satisfying all constraints.

- *Chapter 5* presents a method for automatic sensor placement for model-based robot vision. The task involves determination of the optimal sensor placements and a shortest path through these viewpoints. During the sensor planning, object features are resampled as individual points attached to surface normals. The optimal sensor placement graph is achieved by a genetic algorithm in which a min-max criterion is used for the evaluation. A shortest path is determined by graph theories. A viewpoint planner is developed to generate the sensor placement plan.
- *Chapter 6* presents a sensing strategy for determining the probing points for achieving efficient measurement and reconstruction of freeform surfaces. The B-spline model is adopted for modeling the freeform surface. In order to obtain reliable parameter estimation for the B-spline model, we analyze the uncertainty of the model and use the statistical analysis of the Fisher information matrix to optimize the locations of the probing points needed in the measurements.
- *Chapter 7* presents the issues regarding sensor planning for incrementally building a complete model of an unknown object or environment by an active visual system. It firstly lists some typical approaches to sensor planning for model construction, including the multi-view strategy and existing contributions. The standard procedure for modeling of unknown targets is provided. A self-termination judgment method is suggested based on Gauss' Theorem by checking the variations of the surface integrals between two successive viewpoints so that the system knows when the target model is complete and it is necessary to stop the modeling procedure.
- *Chapter 8* presents an information entropy-based sensor planning approach for the reconstruction of freeform surfaces of 3D objects. In the framework of Bayesian statistics, it proposes an improved Bayesian information criterion (BIC) for determining the B-spline model complexity. Then, the uncertainty of the model is analyzed using entropy as the measurement. Based on this analysis, the information gain for each cross section curve is predicted for the next measurement. After predicting the information gain of each curve, we can obtain the information change for all the B-spline models. This information gain is them mapped into the view space. The viewpoint that contains maximal information gain about the object is selected as the Next Best View.
- *Chapter 9* also deals with the sensor placement problem, but for the tasks of non-model based object modeling. The method involves the decision of the exploration direction and the determination of the best next view and the corresponding sensor settings. The trend surface is proposed as the cue to predict the unknown portion of an object.
- *Chapter 10* presents some strategies of adaptive illumination control for robot vision to achieve the best scene interpretation. It investigates how to obtain the most comfortable illumination conditions for a vision sensor. Strategies are proposed to optimize the pose and optical parameters of the luminaire and the sensor, with emphasis on controlling the image brightness.

Chapter 2
Active Vision Sensors

This chapter presents the sensing fundamentals, measurement principles, and 3D reconstruction methods for active visual sensing. An idea of sensor reconfiguration and recalibration is also described which endows a robot with the ability of actively changing its sensing parameters according to practical scenes, targets, and purposes. These will be used in the next chapters in formulating the methods of sensor reconfiguration and sensor planning.

2.1 3D Visual Sensing by Machine Vision

Similar to human perception, machine vision perception is one of the most important ways for acquiring knowledge of the environment. The recovery of the 3D geometric information of the real world is a challenging problem in computer vision research. Active research in the field in the last 30 years has produced a huge variety of techniques for 3D sensing. In robotic applications, the 3D vision technology allows computers to measure the three-dimensional shape of objects or environments, without resorting to physically probing their surfaces.

2.1.1 Passive Visual Sensing

One class of visual sensing methods is called passive visual sensing where no other device besides cameras is required. These methods were usually developed at the early stage of computer vision research. By passive, no energy is emitted for the sensing purpose and the images are the only input data. The sensing techniques were often supposed to reflect the way that human eyes work. The limited equipment cost constitutes a competitive advantage of passive techniques compared with active techniques that require extra devices.

Such passive techniques include stereo vision, trinocular vision (Lehel et al. 1999, Kim 2004, Farag 2004), and many monocular shape-from-X techniques, e.g. 3D shape from texture, motion parallax, focus, defocus, shadows, shading, specularities, occluding contours, and other surface discontinuities. The problem is that recovering 3D information from a single 2D image is an ill-posed problem (Papadopoulos 2001). Stereo vision is still the single passive cue that gives reasonable accuracy. Human has two eyes, and precisely because of the way the world is projected differently onto the eyes, human is able to obtain the relative distances of objects. The setup of a stereo machine vision system also has two cameras, separated by a baseline distance *b*. The 3D world point may be measured by the two projection equations, in a way that is analogous to the way the human eyes work. To interpret disparity between images, the matching problem must be solved, which has been formulated as an ill-posed problem in a general context and which anyway is a task difficult to automate. This correspondence problem results in an inaccurate and slow process and reduces its usefulness in many practical applications (Blais 2004). The other major drawback of this passive approach is that it requires two cameras and it cannot be used on un-textured surfaces which are common for industrially manufactured objects. The requirement of ambient light conditions is also critical in passive visual sensing. The advantage of stereo vision is that it is very convenient to implement and especially suitable for natural environments. A few applications are illustrated in Figs. 2.1 to 2.3.

The structure-from-motion algorithms solve the following problem: given a set of tracked 2D image features captured by a moving camera, find the 3D positions and orientations of the corresponding 3D features (structure) as well as the camera motion. Pose estimation, on the other hand, solves the problem of finding the position and orientation of a camera given correspondences between 3D and 2D features. In both problems two-dimensional line features are advantageous because they can be reliably extracted and are prominent in man-made scenes. Taylor and Kriegman (1995) minimized a nonlinear objective function with respect to camera rotation, camera translation and 3D lines parameters. The objective function measures the deviation of the projection of the 3D lines on the image planes from the extracted image lines. This method provides a robust solution to the high-dimensional non-linear estimation problem. Fitzgibbon and Zisserman (1998) also worked towards the automatic construction of graphical models of scenes when the input was a sequence of closely spaced images. The point features were matched in triples of consecutive images and the fundamental matrices were estimated from pairs of images. The projective reconstruction and camera pose estimation was upgraded to a Euclidean one by means of auto-calibration techniques (Pollefeys et al. 1998). Finally, the registration of image coordinate frames was based on the algorithm of iterative closest points (Besl and Mckay 1992).

Fig. 2.1. Stereo vision for industrial robots

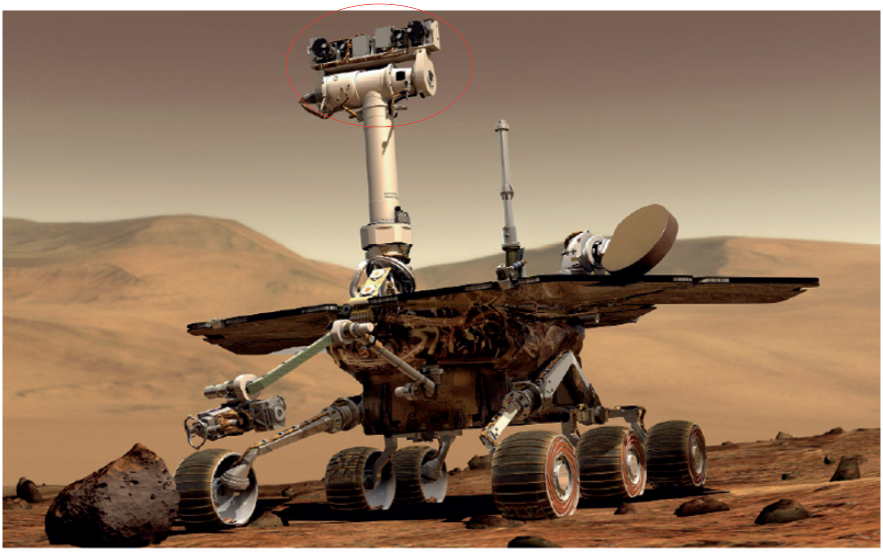

Fig. 2.2. Mars Rover in 3D (NASA mission in 2003–2004) (Pedersen 2003, Miller 2003, Madison 2006, Deen and Lore 2005)

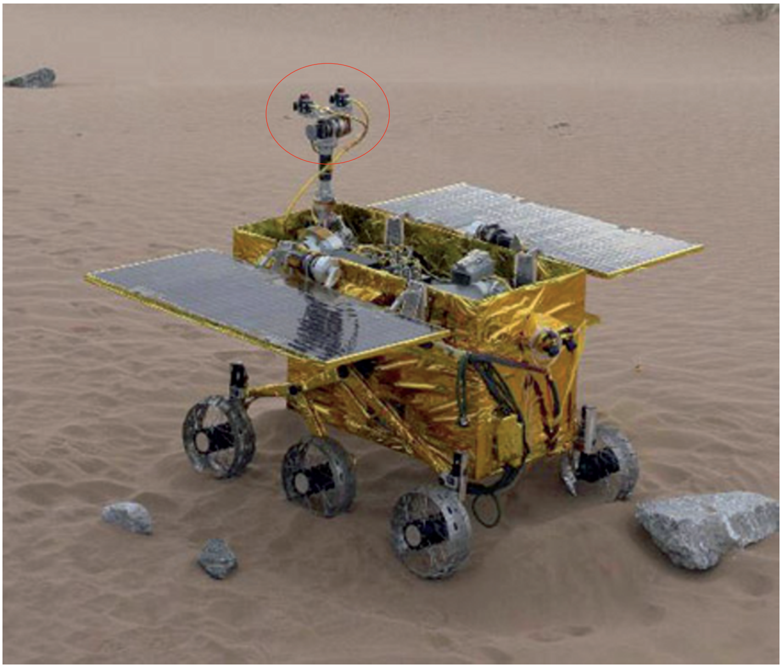

Fig. 2.3. MR-2 (Prototype of Chinese Moon Explorer in 2007–2008)

2.1.2 Active Visual Sensing

In contrast to passive visual sensing, the other class of visual sensing techniques is called active visual sensing. For the above cases of passive techniques (that use ambient light), only visible features with discernable texture gradients like on intensity edges are measured. For the example of the stereo setup, there is a corresponding problem. Matching corresponding points is easy if the difference in position and orientation of the stereo views is small, whereas it is difficult if the difference is large. However, the accuracy of the 3D reconstruction tends to be poor when the difference in position and orientation of the stereo views is small. To overcome the shortcomings of passive sensing, active sensing techniques have been developed in the recent years. These active systems usually do not have the correspondence problem and can measure with a very high precision.

By active sensing, an external projecting device (e.g. laser or LCD/DLP projector) is used to actively emit light patterns that are reflected by the scene and detected by a camera. That is to say they rely on probing the scene in some way rather than relying on natural lighting. Compared with the passive approach, active visual sensing techniques are in general more accurate and reliable.

Generally active 3D vision sensors can resolve most of the ambiguities and directly provide the geometry of an object or an environment. They require minimal operator assistance to generate the 3D coordinates. However, with laser-based approaches, the 3D information becomes relatively insensitive to background illumination and surface texture. Therefore, active visual sensing is ideal for scenes that do not contain sufficient features. Since it requires lighting control, it is usually suitable for indoor environments and both camera and projector need to be pre-calibrated.

Typically, properly formatted light, or another form of energy, is emitted in the direction of an object, reflected on its surface and received by the sensor; the distance to the surface is calculated using triangulation or time-of-flight (Papadopoulos 2001). Typical triangulation-based methods include single/multi-point projection, line projection, fringe and coded pattern projection, and moire effect (Figs. 2.4–2.6). Typical time-of-flight based methods are interferometers and laser range finders.

Moire devices work on the principle that: effectively projecting a set of fringe patterns on a surface using an interference technique, tracking the contours of the fringes allows the range to be deduced. Systems that use point projection, line scanning, and moiré effect are highly accurate, but can be slow. Moire devices are best suited to digitizing surfaces with few discontinuities.

Interferometers work on the principle that: if a light beam is divided into two parts (reference and measuring) that travel different paths, when the beams are combined together interference fringes are produced. With such devices, very small displacements can be detected. Longer distances can also be measured with low measurement uncertainty (by counting wavelengths).

For laser range finders, the distance is measured as a direct consequence of the propagation delay of an electromagnetic wave. This method usually provides good distance precision with the possibility of increasing accuracy by means of longer measurement integration times. The integration time is related to the number of samples in each measurement. The final measurement is normally an average of the sample measures, decreasing therefore the noise associated to each single measure. Spatial resolution is guaranteed by the small aperture and low divergence of the laser beam (Sequeira et al. 1995, 1996, 1999). Basically laser range finders work in two different techniques: pulsed wave and continuous wave. Pulsed wave techniques are based on the emission and detection of a pulsed laser beam. A short laser pulse is emitted at a given frequency and the time elapsed between the emission and the received echo is measured. This time is proportional to the distance from the sensor to the nearest object. In a continuous wave laser ranging system, rather than using a short pulse, a continuous laser beam modulated with a reference waveform is emitted and the range is determined as a result of the comparison of the emitted and received laser beams. This type of system can use either amplitude modulation (e.g. sinusoidal signal) or frequency modulation.

Among various 3D range data acquisition techniques in computer vision, the structured light system with coded patterns is based on active triangulation. A very simple technique to achieve depth information with the help of structured light is to scan a scene with a laser plane and to detect the location of the reflected stripe. The depth information can be computed out of the distortion along the detected profile. More complex techniques of structured light project multiple stripes (Fig. 2.7) or a pattern of grids at once onto the scene. In order to distinguish between stripes or grids they are coded either with different brightness or different colors (Fig. 2.8) (e.g. Coded Light Approach (Inokuchi et al. 1984, Stahs and Wahl 1992) and unique color encoding method). The structured light systems, as well as laser range finders, map directly the acquired data into a 3D volumetric model having thus the ability to avoid the correspondence problem associated with passive sensing techniques. Indeed, scenes with no textural details can be easily modeled. A drawback with the technique of coded stripes is that because each projection direction is associated with a code word, the measurement resolution is low. Fortunately, when this approach is combined with a phase-shift approach, a theoretically infinite height resolution can be obtained. For available products, Fig. 2.9 illustrates some examples of 3D laser scanners and Fig. 2.10 illustrates some examples of 3D Structured Light System.

Therefore, although there are many types of vision sensors available to measure object models by either passive or active methods, structured-light is one of the most important methods due to its many advantages compared with other methods, and thus it is successfully used in many areas for recovering 3D information of an industrial object. This chapter considers typical setups of the structured light system for active visual sensing, using stripe light vision or color-encoded vision. Their system configurations and measurement principles are presented in the following sections.

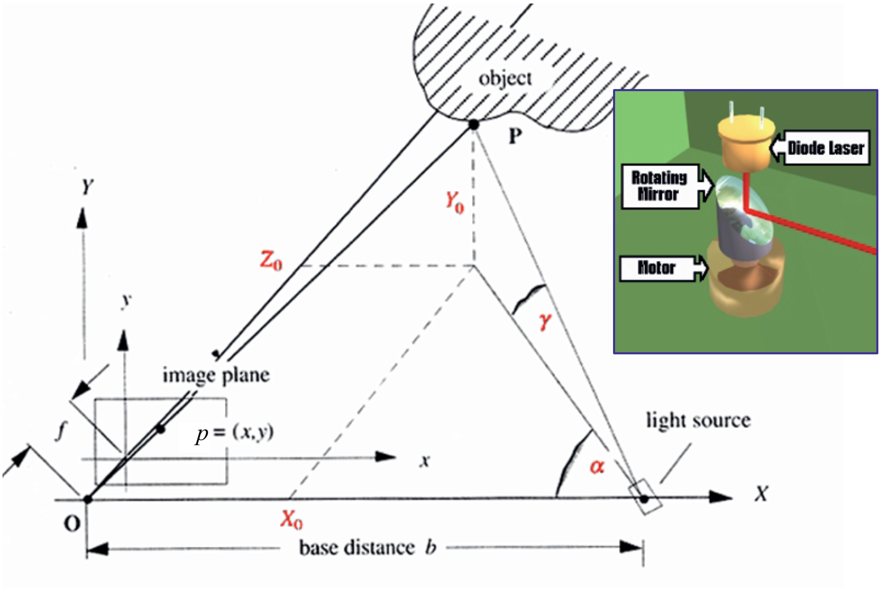

Fig. 2.4. Light spot projection

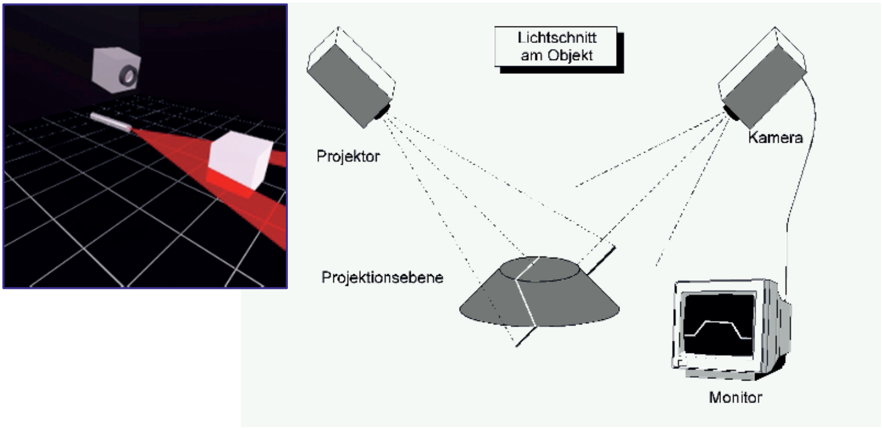

Fig. 2.5. A stripe light scanning system (Intersecting the projection ray with an additional ray or plane will lead to a unique reconstruction of the object point.)

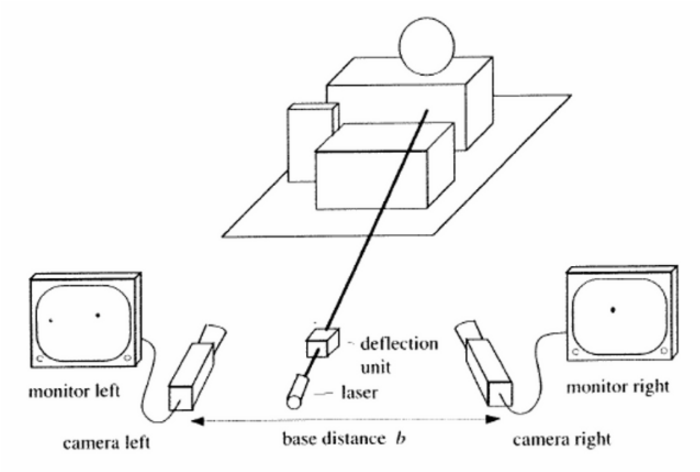

Fig. 2.6. Single spot stereo analysis

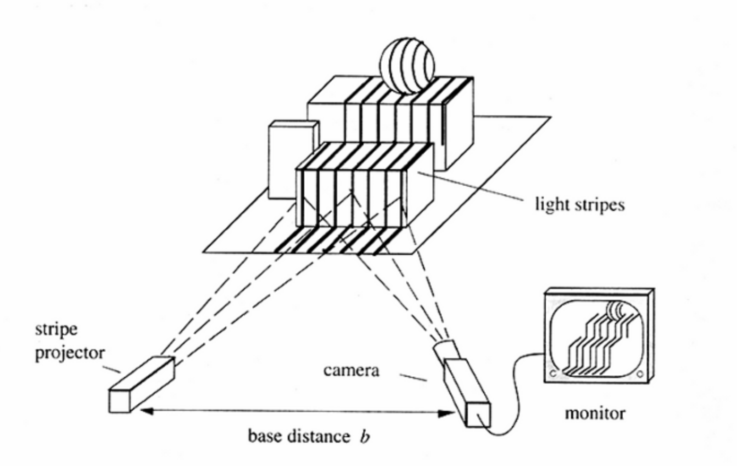

Fig. 2.7. Stripe light vision system

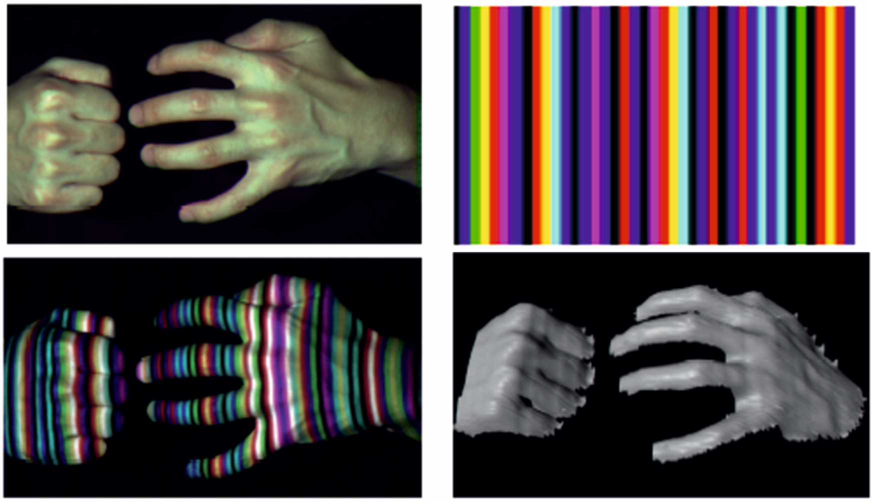

Fig. 2.8. Coded structured light vision: project a light pattern into a scene and analyze the modulated image from the camera