Chapter 3
Active Sensor Planning – the State-of-the-Art

The aim of sensor planning is to determine the pose and settings of a vision sensor for undertaking a vision task that usually requires multiple views. Planning for robot vision is a complex problem for an active system due to its sensing uncertainty and environmental uncertainty. This chapter describes the problem of active sensor planning formulated from practical applications and the state-of-the-art in this field.

3.1 The Problem

An active visual system is a system which is able to manipulate its visual parameters in a controlled manner in order to extract useful data about the scene in time and space. (Pahlavan et al. 1993)

Active sensor planning endows the observer capable of actively placing the sensor at several viewpoints through a planning strategy. In the computer vision community, when active perception became an important attention to researchers, sensor planning inevitably became a key issue because the vision agent had to decide "where to look". According to task conditions, the problem is classified into two categories, i.e. model-based and non-model-based vision tasks.

About 20 years ago, Bajcsy discussed the important concept of active perception (Bajcsy 1988). Together with other researchers" initial contributions at that time, the new concept (compared with the Marr paradigm in 1982) on active perception, and consequently the sensor planning problem, was thus issued in vision research. The difference between the concepts of active perception and the Marr paradigm is that the former considers vision perception as the intentional action of the mind but the latter considers it as the procedural process of matter.

Therefore, research of sensor planning falls into the area of active perception (Bajcsy 1988). It introduces the idea of moving a sensor to constrain interpretation of its environment. Since multiple 3D images need to be taken and integrated from different vantage points to enable all features of interest to be measured, sensor placement which determines the viewpoints with a viewing strategy thus becomes critically important for achieving full automation and high efficiency.

The problem of sensor placement in computer vision was addressed by Tarabanis et al. (1995) as: "for a given information concerning the environment (object under observation, sensor available) and concerning the task that the system

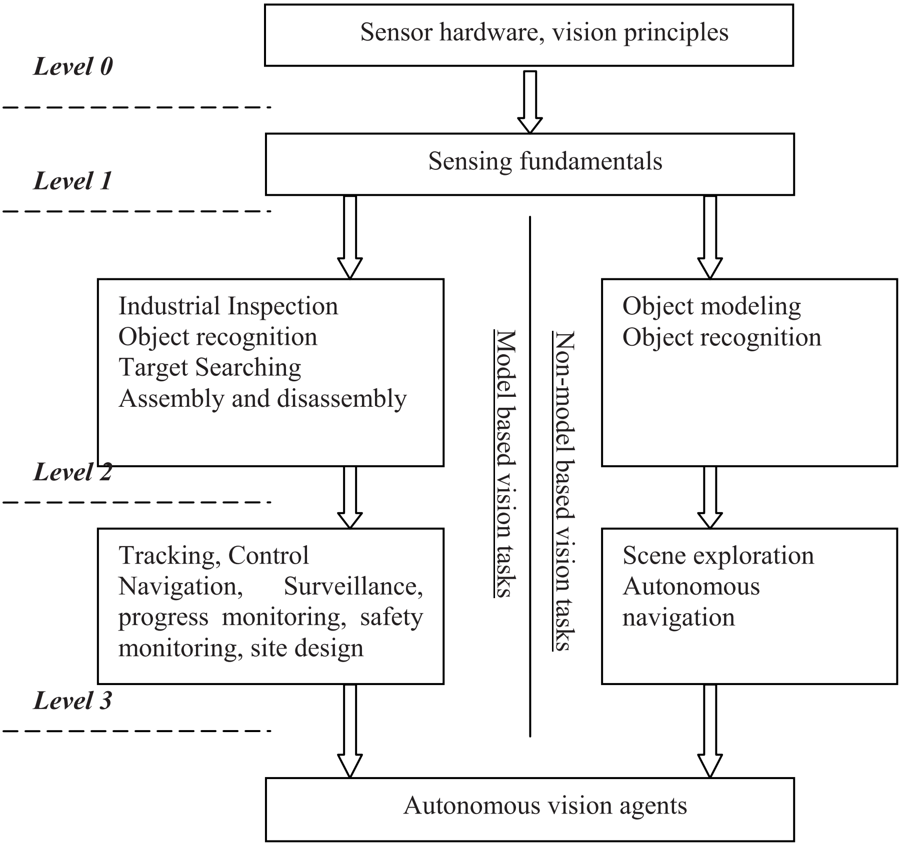

Fig. 3.1. The roles of active sensor planning in autonomous robots

must achieve (detection of characteristics, object recognition, scene reconstruction), to develop some automatic strategy to determine the sensor parameters (the position, the orientation and the optical parameters of the sensor) to carry out the task satisfying some criteria."

Today, the roles of sensor planning can be widely found in most autonomous robotic systems. According to the task conditions, the planning scheme can be applied on different levels of vision perception as illustrated in Fig. 3.1.

3.2 Overview of the Recent Development

The early work on sensor planning was mainly focused on the analysis of placement constraints, such as resolution, focus, field of view, visibility, and conditions for light source placement in 2D space (Lin et al. 1996). A viewpoint has to be placed in an acceptable space and a number of constraints should be satisfied. The fundamentals in solving such a problem were established in the last decades. Tarabanis et al. (1995) presented an intensive survey on sensing strategies developed in the early stage, concentrated upon the period between 1987 and 1991. Among them, Cowan (1988) gave detailed descriptions on computing the acceptable viewpoints for satisfying many requirements (sensor placement constraints). In Cowan (1988), lens aperture setting was also considered by computing the diffraction limit. The light position region was determined to achieve adequate illumination, mathematically through the light path, i.e. surface absorption, diffused reflectance, specular reflectance, and image irradiance. Abrams et al. (1999) also proposed to compute the viewpoints that satisfy the optical constraints, i.e. resolution, focus (depth of field), field-of-view, and detectability. Rosenfeld discussed some techniques and the relationship between object recognition and known or unknown viewpoints (Rosenfeld 1988). More extensive surveys of the early works can be found in Banta (1996), Marchand (1997, 1999), and Kelly et al. (2000).

Here the scope is restricted to recently published approaches to view-pose determination and sensor optical settings in the robotics community. It does not include: attention, gaze control, foveal sensing, hand-eye coordination, autonomous vehicle control, localization, landmarks, qualitative navigation, path following operation, etc., although these are also issues concerning the active perception problem.

Of the published literature in the recent years, Cowan (1988) is one of the earliest research on this problem in 1988 although some primary works can be found in the period 1985–1987. To date, there are more than two hundred research papers which mainly focus on sensor placement or viewpoint planning. At the early stage, these works were focused on sensor modeling, analysis of sensors" optical and geometrical parameters, and sensor placement constraints. From 1990 to 1995, most of these research works were CAD model-based and usually for applications in computer inspection or recognition. The generate-and-test method and the synthesis method are major contributions at that stage. From 1996 to 2000, while optimization was still necessary for model-based sensor placement, it is increasingly important to plan viewpoints for unknown objects or no a priori environment because this is very useful for many active vision tasks such as model reconstruction and autonomous navigation. In recent years, although researchers have to continue working on the theoretical formulation of active sensor planning, many works tend to combine the existing methods with specific application such as inspection, recognition, search, object modeling, tracking, exploration, navigation, localization, assembly and disassembly, etc.

Two outstanding methods have been widely used previously. They are the weighted function and tessellated space. The former uses a function that includes several components standing for placement constraints, e.g.

$$h = \max(\alpha_1 g_1 + \alpha_2 g_2 + \alpha_3 g_3 + \alpha_4 g_4) \tag{3.1}$$

Equivalently with constraint-based space analysis, for each constraint (such as visibility, resolution, field-of-view, and depth-of-field), the sensor pose is limited to a possible region. Then the viewpoint space is the intersection of these regions and the optimization solution is determined by the above function in the viewpoint space, i.e.,

$$V_{\text{placement}} = V_{g1} \cap V_{g2} \cap V_{g3} \cap V_{g4} \tag{3.2}$$

This method is usually used in model-based planning (Trucco 1997) tasks, such as inspection, assembly/disassembly, recognition, and object search.

The latter method tessellates a sphere or cylinder around the object to be modeled as a viewpoint space (or look-up array (Morooka et al. 1999)). Each grid point is a possible sensor pose for viewing the object. The object surface is partitioned as void surface, seen surface, unknown surface, and uncertain surface. The working space is also partitioned into void volume and viewing volume. Finally an algorithm is employed for planning a sequence of viewpoints so that the whole object can be sampled. This method is effective in dealing with some small and simple objects, but it is difficult to model a large and complex object with many concave areas because it cannot solve occlusion constraint.

More precisely, a number of approaches have been applied in deciding the placement of the vision sensor, including:

- geometrical/ volumetric computation
- tessellated sphere/space -TS
- generate-and-test approach (Kececi 1998, Trucco 1997)
- synthesis approach
- sensor simulation
- expert system
- rules (Liu and Lin 1994)
- iterative optimization method (Lehel et al. 1999)
- Bayesian decision (Zhou and Sakane 2001, Kristensen 1997)
- probabilistic reasoning (Roy 2000)
- tree annealing (Yao 1995)
- genetic algorithm (Chen et al. 2004).

Out of these approaches, volumetric computation by region intersection is most frequently used by researchers, e.g. (Cowan 1988). For each constraint, it computes the region Ri of acceptable viewpoints. If multiple surface features need to be inspected simultaneously, the region Ri is the intersection of the acceptable regions Rij for each individual feature. Finally, the region of acceptable viewpoints is the intersection of all regions.

3.3 Fundamentals of Sensor Modeling and Planning

Table 3.1 lists some fundamental works on sensor modeling and vision planning for robotic tasks. It provides an overview of typically used sensors, controllable parameters, proposed methods, and applied tasks.

Table 3.1. Summary of typical works on fundamental sensor planning

Reference	Sensors	Parameters	Method	Task
Cowan 1988	Camera; Extension to laser scanner	Resolution, focus (depth of field), field of view, visibility, view angle; 6 extrinsic parameters of the sensor	Geometrical computation	General model based vision task
Tarabanis 1991	Camera	Optical constraints (resolution, focus/ depth-of-field, field-of-view, and detectability)	Volume intersection method VIM	General purpose
Remagnino1 995	Camera	Position, look direction (pan/tilt), focal length	Geometrical computation	General task in partially known environment
Giraud 1995	General sensors	Perception number, sensor location	Geometrical approach, Bayesian statistics	Equipment design, general task
Triggs 1995	Camera	Task, camera, robot and environment	Probabilistic heuristic search, combined evaluation function	General model based vision task
Yao 1995	Camera	Generalized viewpoint, depth of field, field of view, resolution	Tree annealing	General model based vision task
Tarabanis 1995	Camera	Camera pose, optical settings, task constraints	VIM	Model based vision task
Stamos1998	Camera	Field-of-view, visibility	Interactive	General model based vision task
Lehel et al. 1999	Trinocular sensor (CardEye)	Relative intrinsic translation, pan, tilt, field of view angle	Iterative optimization	General vision tasks
Li and Liu 2003	Structured light	Reconfigured pose	Geometrical	Recalibration for active vision
Zanne et al. 2004	Eye-in-hand camera	Path	Constraint-based control	Visibility

(Continued)

Table 3.1. (Continued)

Reference	Sensors	Parameters	Method	Task
Farag 2004	Trinocular	Center, zoom, focus and vergence	SIFT algorithm	Mobile vision system
Mariottini 2005	Pinhole and panoramic cameras	Camera intrinsic and relative parameters	Geometrical modeling	Camera models and epipolar geometry
LaValle 2006	general	NA	Algorithms using information space, differential constraints, etc.	Motion planning
Hua et al. 2007	Panoramic Camera	Wide FOV, high-resolution	Mirror pyramid	Maximize the panoramic FOV

For active sensor planning, an intended view must first satisfy some constraints, either due to the sensor itself, the robot, or its environment. From the work by Cowan et al. (1988) who made a highlight on the sensor placement problem, detailed descriptions of the acceptable viewpoints for satisfying many requirements (sensor placement constraints) have to be provided. Cowan and Kovesi (1988) presented an approach to automatically generating camera locations (viewpoints), which satisfied many requirements (we term it sensor placement constraints) including resolution, in-focus, field-of-view, occlusion, etc. Shortly after that, they (Cowan and Bergman 1989) further described an integrated method to position both a camera and a light source. Besides determining the camera placement region to satisfy the resolution, field of view, focus, and visibility, lens aperture setting was also considered by computing the diffraction limit. The light position region was determined to achieve adequate illumination, mathematically through the light path, i.e. surface absorption, diffused reflectance, specular reflectance, and image irradiance. Similar concepts were also presented by Tarabanis et al. (1991) to compute the viewpoints that satisfy the sensing constraints, i.e. resolution, focus, field-of-view, and detectability. A complete list of constraints will be summarized and analyzed in Chap. 4.

To better describe the sensor properties, Ikeuchi et al. (1991) presented a sensor modeler, called VANTAGE, to place the light sources and cameras for object recognition. It mostly proposed to solve the detectability (visibility) (Zanne et al. 2004) of both light sources and cameras. It determined the illumination/observation directions using a tree-structured representation and AND/OR operations. The sensor is defined as consisting of not only the camera, but multiple components (G-sources), e.g. a photometric stereo. It is represented as a sensor composition tree (SC tree), as in Fig. 3.2. Finally, the appearance of object surfaces is predicted by applying the SC tree to the object and is followed by the action of sensor planning.

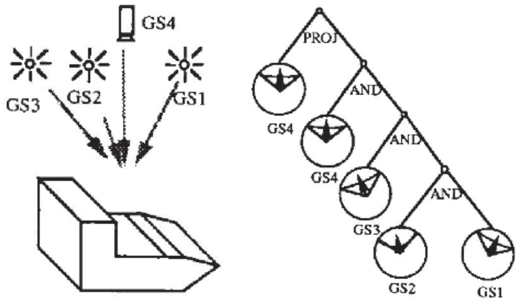

Fig. 3.2. The photometric stereo sensor and its SC tree (Ikeuchi and Robert 1991)

In some other typical works on constraint formulation, Remagnino et al. (1995) proposed to set the viewpoint, look direction, and focal length of a camera. With a partially known environment, it dealt with two problems: how to determine the sensor"s pose (in the bootstrap phase) and how to determine the next-look direction (in the run phase). It took into account errors in the object position stored in the memory and errors due to image segmentation. Rosenfeld et al. (1988) discussed some techniques and relationship between object recognition and known or unknown viewpoints. In fact, an intensive survey on sensing strategies developed in the first stage, i.e. the period from 1987 to 1992, was summarized by Tarabanis et al. (1995).

To a relatively higher level, Giraud and Jouvencel (1995) addressed the sensor selection at an abstract level for equipment design and perception planning. It is formulated with (1) the number of uses of a sensor; (2) the selection of multi-sensors; (3) discarding useless sensors; and (4) the location of the sensors. It used an approach based on geometrical interaction between a sensor and an environment and Bayes reasoning to estimate the achieved information. Later, Kristensen et al. (1997) proposes the sensor planning approach also using the Bayesian decision theory. The sensor modalities, tasks, and modules were described separately and the Bayes decision rule was used to guide the behavior.

The model-based sensor placement problem in fact is formulated as a nonlinear multi-constraint optimization problem. It is difficult to compute robust viewpoints which satisfy all feature detectability constraints. Yao and Allen (1995) presented a tree annealing (TA) method to compute the viewpoints with multi-constraints. They also investigated the stability and robustness while considering the constraints with the different scale factors and noises. Another way is done by Triggs and Laugier (1995) who described a planner to produce heuristically good static viewing positions. It combined many task, camera, robot and environmental constraints. A viewpoint is optimized and evaluated by a function which uses a probability-based global search technique.

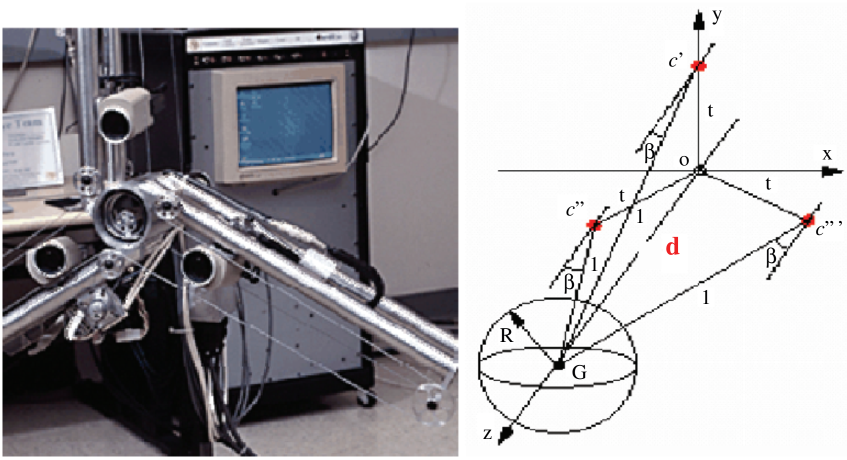

Fig. 3.3. The CardEye trinocular vision sensor and its model (with the Computer Vision and Image Processing Lab (CVIP) at the University of Louisville (Farag 2004))

In a recent book by Steve LaValle (2006), many different kinds of planning algorithms can be found related to visibility and sensor-based planning, e.g. information space, differential constraints, decision-theoretic planning, sampling-based planning, combinatorial planning, etc.

For active sensing purpose, many devices and systems have recently been invented for robotics, e.g. (Colin 2007, Hou et al. 2006). An ATRV-2 based AVENUE mobile robot is used by Blaer and Allen (2006) for automated site modeling (Fig. 3.4), at the Columbia University. Sheng et al. (2006) develop an automated, intelligent inspection system for these engineered structures, which

Fig. 3.4. The ATRV-2 based AVENUE mobile robot for automated site modeling (Blaer and Allen 2006)

employs a team of intelligent climbing robots and a command robot to collaboratively carry out the inspection task. To support autonomous navigation, a miniature active camera (MoCam) module is designed, which can be used in the pose calibration of the robot. Farag (2004) solves the planning problem for a mobile active system with a trinocular vision sensor (Fig. 3.3). An algorithm is proposed to combine a closed-form solution for the translation between the three cameras, the vergence angle of the cameras as well as zoom and focus setting with the results of the correspondences between the acquired images and a predefined target obtained using the Scale Invariant Feature Transform (SIFT) algorithm. There are two goals. The first is to detect the target objects in the navigation field. The second goal is setting the cameras in the best possible position with respect to the target by maximizing the number of correspondences between the target object and the acquired images. The ultimate goal for the algorithm is to maximize the effectiveness of the 3D reconstruction from one frame.

For fast development of sensor modeling, Mariottini and Prattichizzo (2005) develop an Epipolar Geometry Toolbox (EGT) on MATLAB which is a software package targeted to research and education in computer vision and robotic visual servoing (Fig. 3.5). It provides the user with a wide set of functions for designing multicamera systems for both pinhole and panoramic cameras. Several epipolar geometry estimation algorithms have been implemented. They introduce the toolbox in tutorial form, and examples are provided to demonstrate its capabilities. The complete toolbox, detailed manual, and demo examples are freely available on the EGT Web site (http://egt.dii.unisi.it/).

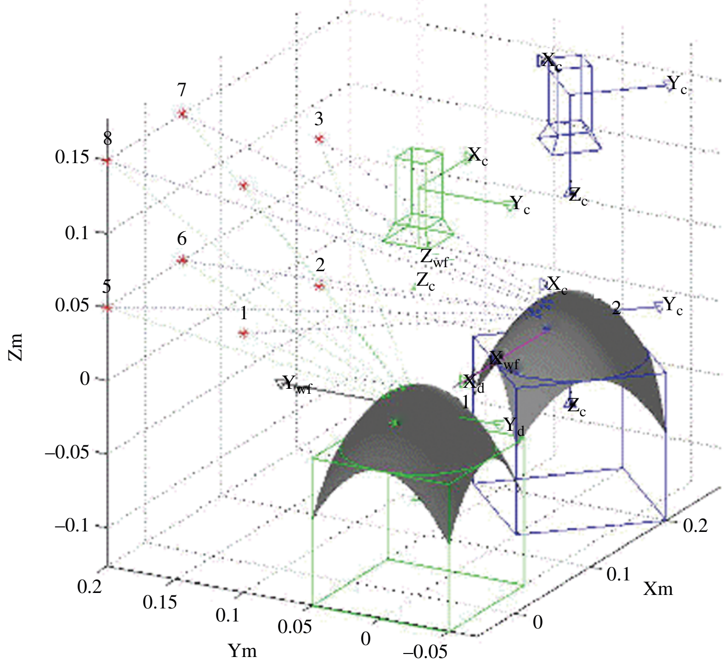

Fig. 3.5. The camera model for visual servoing (Mariottini and Prattichizzo 2005)

3.4 Planning for Dimensional Inspection

In many vision tasks, there exists an object model in the system. For example, in assembly (Nelson 1996), model-based recognition (Okamoto 1998), object searching, dimensional measurement, inspection, and semi-automated scene reconstruction, the object"s geometry and a rough estimate of its pose are known. Especially for the inspection tasks, using either range sensors (Prieto 1999) or intensity cameras (Gu et al. 1999, Abrams 1999), a nearly perfect estimate of the object"s geometry and possibly its pose are known and the task is to determine how accurately the object has been manufactured. Table 3.2 lists some typical works on sensor planning for automated inspection.

On object inspection, Yao and Allen argued that this problem in fact was a nonlinear multi-constraint optimization problem (Yao 1995). Triggs and Laugier (1995) described a planner to produce heuristically good static viewing positions. It combined many task, camera, robot and environmental constraints. A viewpoint is optimized and evaluated by a function which uses a probability-based global search technique. It is difficult to compute robust viewpoints which satisfy all feature detectability constraints. Yao and Allen (1995) presented a Tree Annealing (TA) method to compute the viewpoints with multi-constraints. They also investigated the stability and robustness while considering the constraints with the different

Table 3.2. Summary of typical works on sensor planning for dimensional inspection

Reference	Sensors	Parameters	Method	Task
Tarabanis 1995	Camera	Camera pose, optical settings, task constraints	VIM	Model based vision task
Abrams 1996	Camera	Detectability, in focus, field-of-view, visibility, and resolution	VIM	Inspection
Trucco 1997	Generalized sensor	Visibility, reliability, shortest path	Generate-and-test, VIM, FIR, CCAO	Inspection
Prieto 1999	Range sensor	Viewing distance, incident angle	Direct computation	Inspection
Sheng et al. 2003				
Hodge 2003	Multiple cameras	Positions	Agent-based coordination	Inspection
Chen et al. 2004	Camera, structured light	Camera pose, settings, task constraints	Genetic algorithm, graph theory	Model-based inspection, robot path
Rivera-Rios 2005	Stereo	Camera pose	Probabilistic analysis	Dimensional measurements
Bodor 2005	Cameras	Internal and external camera parameters	Analytical formulation	Observability

scale factors and noises. Elsewise, Olague and Mohr (1998) chose to use genetic algorithms to determine the optimal sensor placements.

In order to obtain a quality control close to the accuracy obtained with a coordinate measuring machine in metrology for automatic inspection, F. Prieto et al. suggest improving the accuracy of the depth measurements by positioning the sensor"s head according to a strategy for optimum 3D data acquisition (Prieto 1999). This strategy guarantees that the viewpoints found meet the best accuracy conditions in the scanning process. The proposed system requires the part"s CAD model to be in IGES format.

Several sensor planning systems have been developed by researchers. For example, Trucco et al. (1997) developed a general automatic sensor planning (GASP) system. Tarbox and Gottschlich (1999) had an Integrated Volumetric Inspection System (IVIS) and proposed three algorithms for inspection planning. Tarabanis et al. (1995) developed a model-based sensor planning system, the machine vision planner (MVP), which works with 2D images obtained from a CCD camera.

Compared with other vision sensor planning systems, the MVP system is notable in that it takes a synthesis rather than a generate-and-test approach, thus giving rise to a powerful characterization of the problem. In addition, the MVP system provides an optimization framework in which constraints can easily be incorporated and combined. The MVP system attempts to detect several features of interest in the environment that are simultaneously visible, inside the field of view, in focus, and magnified, by determining the domain of admissible camera locations, orientations, and optical settings (Fig. 3.6). A viewpoint is sought that is both globally admissible and central to the admissibility domain.

Based on the work on the MVP system (Tarabanis 1995), Abrams et al. (1996) made a further development for planning viewpoints for vision tasks within a robot work-cell. The computed viewpoints met several constraints such as detectability, in-focus, field-of-view, visibility, and resolution. The proposed viewpoint computation algorithm also fell into the "volume intersection method" (VIM). The planning procedure was summarized as: (1) Compute the visibility volumes Vivis; (2) compute the volumes ViFR combined with field-of-view and resolution constraints; (3) compute the overall candidate volume Vc as the intersection of all ViFR and Vivis; (4) find a position within Vc; (5) find the orientation; (6) compute the focus and maximum aperture; (7) verify that the parameters are all valid.

These is generally a straightforward but very useful idea. Many of the latest implemented planning systems can be traced back to this contribution. For example, Rivera-Rios et al. (2005) presents a probabilistic analysis of the effect of the localization errors on the dimensional measurements of the line entities for a parallel stereo setup (Fig. 3.7). The probability that the measurement error is within an acceptable tolerance was formulated as the selection criterion for camera poses. The camera poses were obtained via a nonlinear program that minimizes the total mean square error of the length measurements while satisfying the sensor constraints.

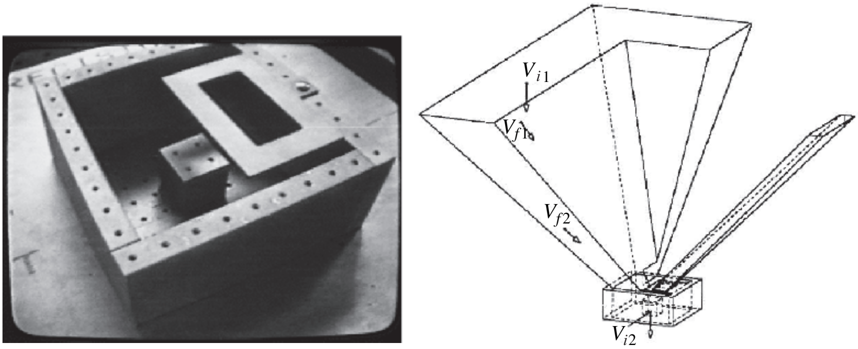

Fig. 3.6. The admissible domain of viewpoints (Tarabanis 1995)

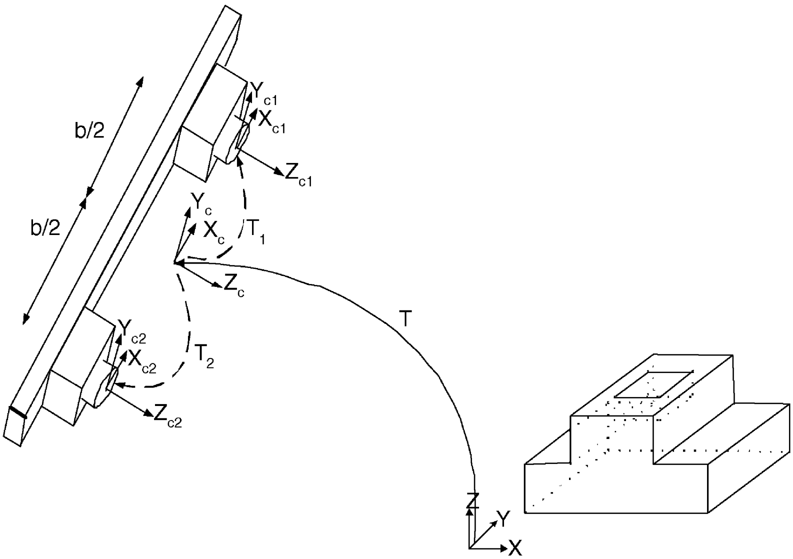

Fig. 3.7. Stereo pose determination for dimensional measurement (Rivera-Rios 2005)

The general automatic sensor planning system (GASP) reported by Trucco et al. (1997) is to compute optimal positions for inspection tasks using a known imaging sensor and feature-based object models. This exploits a feature inspection representation (FIR) which outputs an explicit solution off-line for the sensor position problem. A generalized sensor (GS) model was defined with both the physical sensor and the particular software module. The viewpoints are planned by computing the visibility and reliability. The reliability of the inspection depends on the physical sensors used and the processing software. In order to find a shortest path through the viewpoints in space, they used the Convex hull, Cheapest insertion, angle selection, Or-optimization (CCAO) as the algorithm to solve the traveling salesman problem (Fig. 3.8).

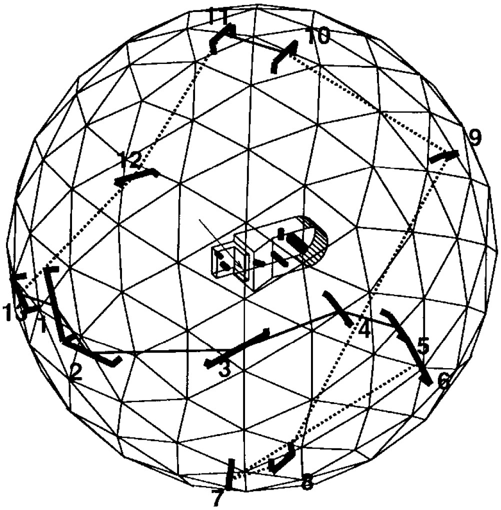

Fig. 3.8. The shortest path planned to take a stereo pair through the viewpoints (Trucco 1997)

In order to obtain a more complete and accurate 3D image of an object, Prieto et al. (1999) presented an automated acquisition planning strategy utilizing its CAD model in IGES format. The work was focused on improving the accuracy of the 3D measured points which is a function of the distance to the object surface and of the laser beam incident angle.

3.5 Planning for Recognition and Search

In many cases, a single view of an object may not contain sufficient features to recognize it unambiguously. Therefore another important application of sensor planning is active object recognition (AOR) which recently attracts much attention within the computer vision community. Object search is also considered a model-based vision task concerned with finding a given object in a known or unknown environment. The object search task not only needs to perform the object recognition and localization, but also involves sensing control, environment modeling, and path planning. Sensor planning is very important for 3D object search since a robot needs to interact intelligently and effectively with the 3D environment. Table 3.3 lists the typical works on sensor planning for vision-based recognition and search.

Table 3.3. Summary of typical works on sensor planning for recognition and search

Reference	Sensors	Parameters	Method	Task
Ikeuchi and Robert 1991	Light source, camera	Illumination/ observation directions	Tree-structured, logical operation	Object recognition
Ye 1995	Camera + range finder	Sensing pose, search space	Probability, tessellated sphere (TS)	Object search
Liu and Lin 1994 Lin et al. 1996	Structured light	View pose	Rules, feature prediction, MERHR	Recognition
Madsen and Christensen 1997	Camera	Viewing direction	Direct computation	Active object recognition (AOR)
Borotschnig 2000	Camera, illuminant	View pose	Probabilistic object classifications, score ranking	AOR
Deinzer 2000	Camera	Classification and localization	Reinforcement learning	AOR
Roy 2000	Camera	View pose, object features	Probabilistic reasoning, Bayes rule	AOR
Sarmiento et al. 2005	General sensor	Sensing locations	Convex cover algorithm	Object search
Xiao et al. 2006	Sonar and omni-directional camera	Path	Fuzzy logic algorithm	Search

In fact, two objects may have all views in common with respect to a given feature set, and may be distinguished only through a sequence of views (Roy 2000). Further, in recognizing 3D objects from a single view, recognition systems often use complex feature sets. Sometimes, it may be possible to achieve the same, incurring less error and smaller processing cost by using a simpler feature set and suitably planning multiple observations. A simple feature set is applicable for a larger class of objects than a model base with a specific complex feature set. Model base-specific complex features such as 3D invariants have been proposed only for special cases. The purpose of AOR is to investigate the use of suitably planned multiple views for 3D object recognition. Hence the AOR system should also take a decision on “where to look”. The system developed for this task is an iterative active perception system that executes the acquisition of several views of the object, builds a stochastic 3D model of the object and decides the best next view to be acquired. Okamoto et al. (1998) proposed such a method based on an entropy measure. Liu and Lin (1994), Lin et al. (1996), and Madsen and Christensen (1997) proposed their sensor planning strategies for recognition using rules to automatically predict and detect object features and calculate the next sensor pose, and they applied the maximum expected rate of hypothesis reduction (MERHR) to minimize the sensing actions. Madsen and Christensen's strategy (Madsen and Christensen 1997) was to determine the true angle on the object surface. It automatically guided a movable camera to a position where the optical axis is perpendicular to a plane spanned by any two intersecting edges on a polyhedral object, so that it could determine the true angle of a junction and align the camera. Ye and Tsotsos (1999) used a strategy for object search by planning the sensing actions on the sensed sphere or layered sensed sphere. It was based on a mobile platform, an ARK robot, equipped with a Laser Eye with pan and tilt capabilities. They combined the object recognition algorithm and the target distribution probability for the vision task.

Ikeuchi et al. (1991) developed a sensor modeler, called VANTAGE, to place the light sources and cameras for object recognition. It mostly solves the detectability (visibility) of both light sources and cameras. Borotschnig et al. (2000) also presented an active vision system for recognizing objects which are ambiguous from certain viewpoints. The system repositions the camera to capture additional views and uses probabilistic object classifications to perform view planning. Multiple observations lead to a significant increase in recognition rate. The view planning consists in attributing a score to each possible movement of the camera. The movement obtaining the highest score will be selected next (Fig. 3.9). It was based on the expected reduction in Shannon entropy over object hypotheses given a new viewpoint, which should consist in attributing a score $s_n(\Delta\psi)$ to each possible movement $\Delta\psi$ of the camera. The movement obtaining the highest score will be selected next:

$$\Delta\psi_{n+1} := \arg\max s_n(\Delta\psi) \tag{3.3}$$

Reinforcement learning has been attempted by Deinzer et al. (2000) for viewpoint selection for active object recognition and for choosing optimal next views for improving the classification and localization results. Roy et al. (2000) attempted probabilistic reasoning for recognition of an isolated 3D object. Both the probability calculations and the next view planning have the advantage that the

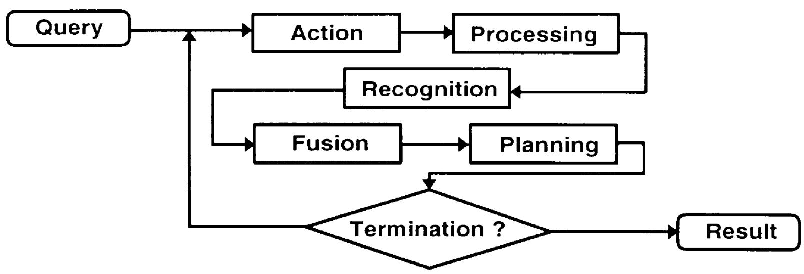

Fig. 3.9. The framework of appearance-based active object recognition (Borotschnig 2000)

knowledge representation scheme encodes feature-based information about objects as well as the uncertainty in the recognition process. The probability of a class (a set of aspects, equivalent with respect to a feature set) was obtained from the Bayes rule (Roy 2000):

$$P(B_i|E^k) = \frac{P(B_i)P(E^k|B_i)}{\sum_j P(B_j)P(E^k|B_j)} \tag{3.4}$$

where $P(B_i \mid E^k)$ is the post-probability of the given subtask done by the action agent.

In the next view planning, two possible moves may be followed from one view to another, i.e. primary move and auxiliary move. A primary move represents a move from an aspect, the minimum angle needed to move out of it. An auxiliary move represents a move from an aspect by an angle corresponding to the primary move of another competing aspect.

3.6 Planning for Exploration, Navigation, and Tracking

On sensor planning for exploration, navigation, and tracking, there is a similar situation that the robot has to work in a dynamic environment and the sensing process may associate with many noises or uncertainties. This issue has become the most active for many applications in recent years. For example, Bhattacharya et al. (2007), Gutmann et al. (2005), Kim (2004), Parker et al. (2004), Steinhaus et al. (2004), Giesler (2004), Yamaguchi et al. (2004), Wong and Jarvis (2004), and Bekris et al. (2004) are related to sensor planning for navigation; Yang et al. (2007), Deng et al. (2005), Chivilo et al. (2004), Harville and Dalong (2004), Thompson (2003), Nishiwaki (2003), and Saeedi et al. (2006) are related to sensor planning for tracking; Huwedi (2006), Leung and Al-Jumaily (2004), and Isler (2003) are related to sensor planning for exploration; Reitinger et al. (2007), Blaer (2006), Ikeda (2006), Park (2003), and Kagami (2003) are related to sensor planning for modeling; and Lim (2003) is for surveillance. Table 3.4 lists the typical works on sensor planning for these topics.

Table 3.4. Some typical works on sensor planning for navigation and modeling

Reference	Sensors	Parameters	Method	Task
Remagnino 1995	Camera	Position, look direction (pan/tilt), focal length	Direct computation	General vision task in partially known environment
Kristensen 1997	General sensor/ actuator	Sensor actions	Bayesian decision	Autonomous navigation in partly known environments
Gracias 2003			Mosaic-based	Underwater navigation

Table 3.4. (Continued)

Reference	Sensors	Parameters	Method	Task
Zhu 2004	Panoramic stereo	Position, orientation	Adaptive	Tracking and localization
Chen et al. 2005	General	Sensor pose	Trend surface	Object modeling
Skrzypc-zynski 2005	Cameras	Position	Landmarks	Positioning, navigation
Murrieta-Cid 2005	Range sensor	Visibility, distance, speed	Differential, system model	Surveillance; maintaining visibility
Hughes and Lewis 2005	Cameras	Camera placement, field of view	Simulation	Exploration
Belkhouche and Belkhouche 2005	General	Robot position and orientation	Guidance laws	Tracking, navigation
Kitamura 2006	Camera, other sensor	Human intervention	Biologically inspired, learning	Navigation
Ludington et al. 2006	Aerial camera	Position	Vision-aided inertial, probability	Navigation, search, tracking
Bhattacharya et al. 2007	Camera	Path, field of view, camera pan	Region based	Landmark-based navigation

For navigation in an active way, a robot is usually equipped with a “controllable” vision head, e.g. a stereo camera on pan/tilt mount (Fig. 3.10). Kristensen (1997) presented the problem of autonomous navigation in partly known environments (Fig. 3.11). Bayesian decision theory was adopted in the sensor planning approach. The sensor modalities, tasks, and modules were described separately and Bayes decision rule was used to guide the behavior. The decision problem for one sensor was constructed with a standard tree for myopic

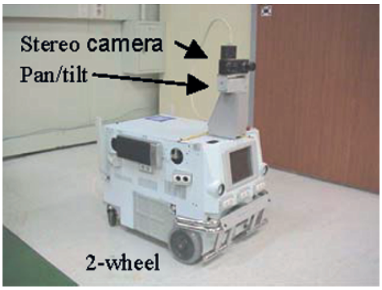

Fig. 3.10. The robot with an active stereo head (e.g. rotation, pan/tilt mount) (Parker et al. 2004)

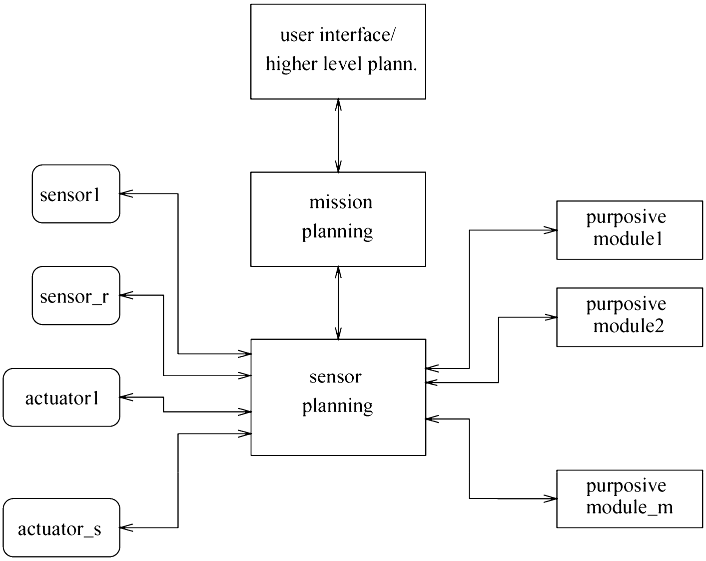

Fig. 3.11. The planning architecture with three levels of abstraction, illustrating that the planner mediates the sensors to the purposive modules (Kristensen 1997)

decision. Object search is also a model-based vision task which is to find a given object in a known or unknown environment. The object search task not only needs to perform object recognition and localization, but also involves sensing control, environment modeling, and path planning.

Zhuang et al. (2004) developed an adaptive panoramic stereovision approach for localizing 3D moving objects at the department of computer science at the University of Massachusetts at Amherst. The research focuses on cooperative robots involving cameras (residing on different mobile platforms) that can be dynamically composed into a virtual stereovision system with a flexible baseline in order to detect, track, and localize moving human subjects in an unknown indoor environment. It promises an effective way to solve the problems of limited resources, view planning, occlusion, and motion detection of movable robotic platforms. Theoretically, two interesting conclusions are given:

1. If the distance from the main camera to the target, D_1, is significantly greater (e.g., ten times greater) than the size of the robot (R), the best geometric configuration is

$$B \approx 2\sqrt{D_1 R},\ \cos\phi_1 = \frac{3BD_1}{2D_1^2 + B^2} \tag{3.5}$$

where B is the best baseline distance for minimum distance error and ϕ_1 is the main camera"s inner angle of the triangle formed by the two robots and the target.

2. The depth error of the adaptive stereovision is proportional to 1.5 the power of the camera-target distance ($D^{1.5}$), which is better than the case of the best possible fixed baseline stereo in which depth error is proportional to the square of the distance (D^2).

On the visual tracking problem, Belkhouche and Belkhouche (2005) pointed out that the traditional control algorithms based on artificial vision suffered from two problems:

1. The control algorithm has to process in real time a huge flow of data coming from the camera. This task may be difficult, especially for fast tracking problems. Thus, the maximum computational power for image processing is an important issue.
2. The target (or the lead car) is detected only when it appears in the camera"s field of view. Thus, the target must stay in the camera scope of the pursuer. This requirement is necessary to implement a vision-based algorithm.

Therefore, they make a mathematical formulation for modeling and controlling a convoy of wheeled mobile robots. The approach is based on guidance laws strategies, where the robotic convoy is modeled in terms of the relative velocities of each lead robot with respect to its following robot. This approach results in important simplifications to the sensory system as compared to artificial vision algorithms.

Concerning the surveillance problem, there is a decision problem which corresponds to answering the question: can the target escape the observer"s view? Murrieta-Cid et al. (2005) defined this problem and considered to maintain surveillance of a moving target by a nonholonomic mobile observer. The observer"s goal is to maintain visibility of the target from a predefined, fixed distance. The target escapes if

(a) it moves behind an obstacle to occlude the observer"s view,
(b) it causes the observer to collide with an obstacle, or
(c) it exploits the nonholonomic constraints on the observer"s motion to increase its distance from the observer beyond the surveillance distance.

An expression derived for the target velocities is:

$$\begin{pmatrix} \dot{x}_T(t) \\ \dot{y}_T(t) \end{pmatrix} = \begin{pmatrix} \cos\theta & -l\cos\phi \\ \sin\theta & l\cos\phi \end{pmatrix} \begin{pmatrix} u_1 \\ u_3 \end{pmatrix} \tag{3.6}$$

where θ and ϕ are the observer"s orientation, u_1 and u_3 are moving speeds, and l is the predefined surveillance distance.

To maintain the fixed required distance between the target and the observer, the relationship between the velocity of the target and the linear velocity of the observer is

$$f(u_1, u_3) = u_1^{\ 2} + 2u_1 u_3 l \sin(\theta - \phi) + l^2 u_3^2 = 1 \tag{3.7}$$

Equation (3.7) defines an ellipse in the u_1–u_3 plane and the constraint on u_1 and u_3 is that they should be inside the ellipse while supposing $\dot{x}_T^2 + \dot{y}_T^2 \le 1$. They deal specifically with the situation in which the only constraint on the target"s velocity is a bound on speed (i.e., there are no nonholonomic constraints on the target"s motion), and the observer is a nonholonomic, differential drive system having bounded speed. The system model is developed to derive a lower bound for the required observer speed. It"s also considered the effect of obstacles on the observer"s ability to successfully track the target.

Biologically inspired, Kitamura and Nishino (2006) use a consciousness-based architecture (CBA) for the remote control of an autonomous robot as a substitute for a rat. CBA is a developmental hierarchy model of the relationship between consciousness and behavior, including a training algorithm (Fig. 3.12). This training algorithm computes a shortcut path to a goal using a cognitive map created on the basis of behavior obstructions during a single successful trial. However, failures in reaching the goal due to errors of the vision and dead reckoning sensors require human intervention to improve autonomous navigation. A human operator remotely intervenes in autonomous behaviors in two ways: low-level intervention in reflexive actions and high-level ones in the cognitive map.

A survey has recently been carried out by Jia et al. (2006). It summarizes the developments of the last 10 years in the area of vision-based target tracking for autonomous vehicle navigation. It concludes that it is very necessary to develop robust visual target tracking based navigation algorithms for the broad applications of autonomous vehicles. Including the recent techniques in vision-based tracking and navigation, some trends of using data fusion for visual target tracking are also discussed. It is especially pointed out that through data fusion the tracking performance is improved and becomes more robust.

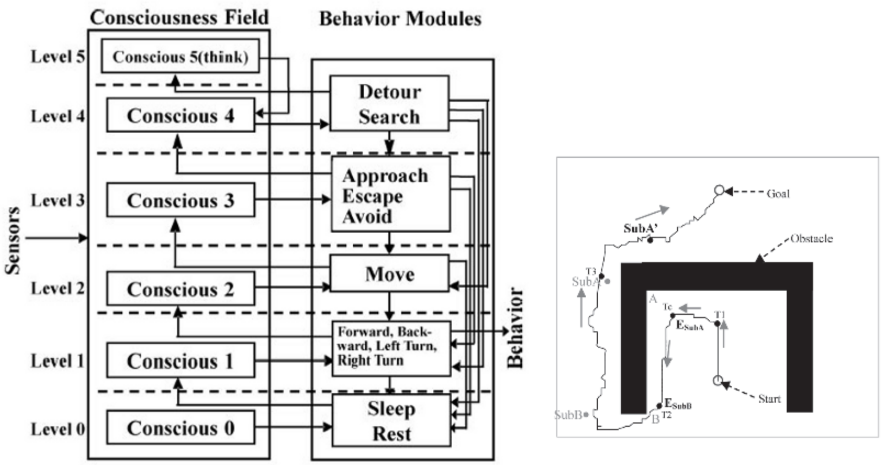

Fig. 3.12. The structure of six-layered consciousness-based architecture and an example of behavior track with intervention (right side)

3.7 Planning for Assembly and Disassembly

For the assembly/disassembly tasks (Table 3.5), a long-term aim in robot programming is the automation of the complete process chain, i.e. from planning to execution. One challenge is to provide solutions which are able to deal with position uncertainties (Thomas et al. 2007, Fig. 3.13). Nelson et al. (1996) introduced a dynamic sensor planning method. They used an eye-in-hand system and considered the resolution, field-of-view, depth-of-view, occlusions, and kinematic singularities. A controller was proposed to combine all the constraints into a system and resulted in a control law. Kececi et al. (1998) employed an independently mobile camera with a 6-DOF robot to monitor a disassembly process so that it can be planned. A number of candidate view-poses are being generated and subsequently evaluated to determine an optimal view pose. A good view-pose is defined with the criterion which prevents possible collisions, minimizes mutual occlusions, keeps all pursued objects within the field-of-view, and reduces uncertainties.

Takamatsu et al. (2002) developed an "assembly-plan-from-observation" (APO) system. The goal of the APO system is to enable people to design and develop a robot that can perform assembly tasks by observing how humans perform those tasks. Methods of contact relations configuration space (C-space) are used to clean up observation errors. Stemmer et al. (2006) use a vision sensor, with color segmentation and affine invariant feature classification, to provide the position estimation within the region of attraction (ROA) of a compliance-based assembly strategy. An assembly planning toolbox is based on a theoretical analysis and the maximization of the ROA. This guarantees the local convergence of the assembly process under consideration of the geometry in part. The convergence analysis uses the passivity properties of the robot and the environment.

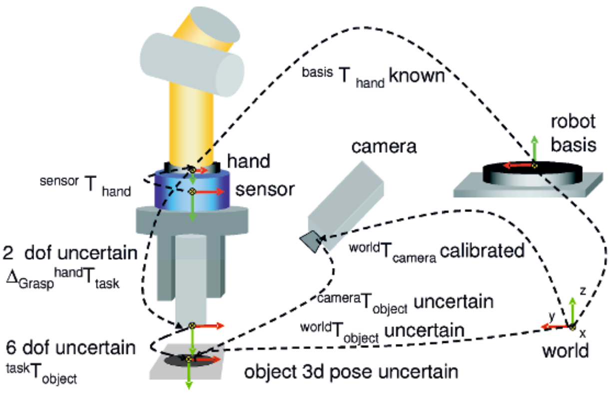

Fig. 3.13. Vision sensor for solving object poses and uncertainties in the assembly work cell (Thomas et al. 2007)

Table 3.5. Some typical works on sensor planning for assembly and disassembly

Reference	Sensors	Parameters	Method	Task
Nelson 1996	Camera	Resolution, FOV, depth-of-view, occlusions, kinematics	Controller (dynamic control law)	Assembly
Kececi 1998	Camera	FOV, view pose, occlusion, uncertainties	Generate-and-test, view-pose assessment/ evaluation	Disassembly
Molineros 2001	Camera	Position	Appearance-based	Assembly planning
Takamatsu 2002	General	Spatial relation	C-space	Assembly, recognition
Hamdi and Ferreira 2004	Virtual	Position	Physical-based	Microassembly
Kelsey et al. 2006	Stereo	Pose	Model-based, synthetic	Pose estimation and tracking
Thomas et al. 2007	Cameras	Relative poses	Multi sensor fusion	Assembly

3.8 Planning with Illumination

Table 3.6. Summary of typical works on sensor planning with illumination

Reference	Sensors	Parameters	Method	Task
Cowan 1989	Camera, light source	Camera, light position region	Illumination computation via reflectance	General model based tasks
Ikeuchi and Robert 1991	Light source, camera	Illumination/ observation directions	Tree-structured, logical operation	Object recognition
Eltoft 1995				Enhancing image features
Solomon 1995	Light source, camera	Positions	Model-based	Lambertian polyhedral objects
Racky and Pandit 1999	Light source	Position	Physics	Segmentation
Xu and Zhang 2001	Light source	Pose, intensity, and distribution of light sources	Neural-network	Surgical applications; general vision tasks

Table 3.6. (Continued)

Reference	Sensors	Parameters	Method	Task
Qu 2003				
Spence 2006	Photometric stereo	Position	Sensitivity analysis	Surface measurement
Yang and Welch 2006	Light source	Illumination variance	Illumination estimation	Tracking
Chen et al. 2007	Light source	Intensity, glares	PID-controller	General tasks
Marchand 2007	Light, camera	Positions	Brightness, contrast	Visual servoing

The light source for a natural scene is its illumination. For many machine-vision applications, illumination now becomes the most challenging part of system design, and is a major factor when it comes to implementing color inspection. Table 3.6 lists the typical works on sensor planning with illumination, recently carried out in the robot vision community. Here, when illumination is also considered, the term "sensor" has a border meaning "sensor/actuator/illuminant".

Eltoft and deFigueiredo (1995) found that illumination control could be used as a means of enhancing image features. Such features are points, edges, and shading patterns, which provide important cues for the interpretation of an image of a scene and the recognition of objects present in it. Based on approximate expressions for the reflectance map of Lambertian and general surfaces, a rigorous discussion on how intensity gradients are dependent on the direction of the light is presented. Subsequently, three criteria for the illumination of convex-shaped cylindrical surfaces are given. Two of these, the contrast equalization criterion and the max-min equalization criterion, are developed for optimal illumination of convex polyhedrons. The third, denoted shading enhancement, is applicable for the illumination of convex curved objects. Examples illustrate the merit of the criteria presented

Xu and Zhang (2001) and Zhang (1998) apply a method of modeling human strategy in controlling a light source in a dynamic environment to avoid a shadow and maintain appropriate illumination conditions. Ikeuchi et al. (1991) investigate the illumination conditions with logical operations of illuminated regions. Their developed sensor modeler, VANTAGE, determines the illumination directions using a tree-structured representation and AND/OR operations (Fig. 3.14).

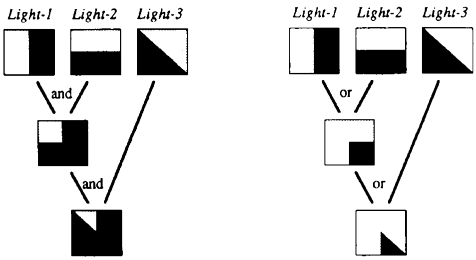

Fig. 3.14. Set operations ("AND" and "OR") among illuminated regions (Ikeuchi and Robert 1991)

Qu et al. (2003) discussed that irradiance distribution and intensity of the test object play a key role in accuracy and stability of the vision measuring system. They proposed a luminance transfer function to design the illumination so that it could adjust light radiation automatically by ways of Neural Networks and Pulse-Width Modulation switch power. They concluded that the illumination could greatly improve the accuracy and robustness of the vision measuring system.

Marchand et al. (2007) recently proposed an approach to control camera position and/or lighting conditions in an environment using image gradient information. The goal is to ensure a good viewing condition and good illumination of an object to perform vision-based tasks such as recognition and tracking. Within the visual servoing framework, the solution is to maximize the brightness of the scene and maximize the contrast in the image. They consider arbitrary combinations of either static or moving lights and cameras. The method is independent of the structure, color and aspect of the objects. For examples, illuminating the Venus of Milo is planned as in Fig. 3.15.

With regard to the placement of the illumination vectors for photometric stereo, Drbohlav and Chantler (2005) discussed the problem of optimal light configurations in the presence of camera noise. Solomon and Ikeuchi proposed an illumination planner for Lambertian polyhedral objects. Spence and Chantler (2006) also found the optimal difference between tilt angles of successive illumination vectors to be 120°. Such a configuration is therefore to be recommended for use with 3-image photometric stereo. Ignoring shadowing, the optimal slant angle was found to be 90° for smooth surfaces and 55° for rough surfaces. The slant angle selection therefore depends on the surface type.

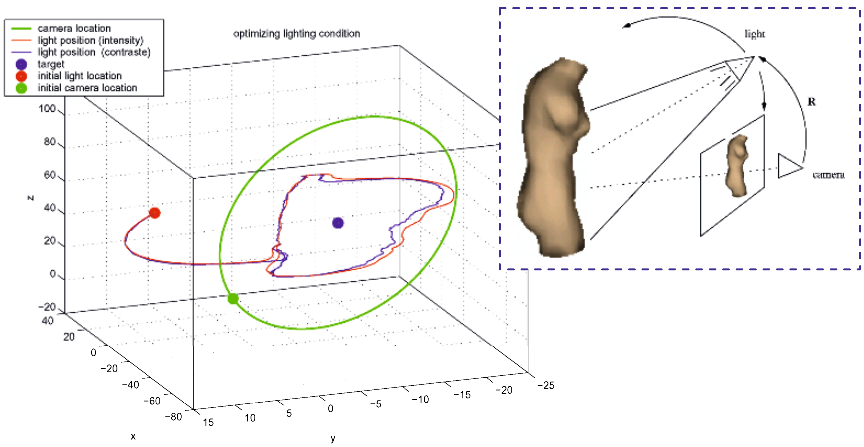

Fig. 3.15. En example of camera and light source position control

3.9 Other Planning Tasks

Besides the tasks already presented in this chapter, there are some other interesting works related to active sensor planning (Table 3.7). For example, Navarro-Serment et al. (2004) describe a method for observing maneuvering targets using a group of mobile robots equipped with video cameras. The cameras seek to observe the target while facing it as much as possible from their respective viewpoints. The work considers the problem of scheduling and maneuvering the cameras based on the evaluation of their current positions in terms of how well can they maintain a frontal view of the target. Some contributions such as interactive planning, virtual placement, robot localization, attention and gaze are briefly introduced below.

3.9.1 Interactive Sensor Planning

In cluttered and complex environments such as urban scenes, it can be very difficult to determine where a sensor should be placed to view multiple objects and regions of interest. Based on their earlier sensor planning results (Tarabanis 1995, Abrams 1999), Stamos and Allen (1998) and Blaer and Allen (2006) extended to build an interactive sensor planning system that can be used to select viewpoints subject to camera visibility, field of view and task constraints. Given a description of the sensor"s characteristics, the objects in the 3D scene, and the targets to be viewed, the algorithms compute the set of admissible view points that satisfy the constraints. The system first builds topologically correct solid models of the scene from a variety of data sources. Viewing targets are then selected, and visibility volumes and field of view cones are computed and intersected to create viewing volumes where cameras can be placed. The user can interactively manipulate the scene and select multiple target features to be viewed by a camera. VRML graphic models and then solid CAD models are assumed as the site models of the scenes (Fig. 3.16).

Table 3.7. Some other interesting works related to active sensor planning

Reference	Sensors	Parameters	Method	Task
Stamos 1998	Camera	Visibility, FOV, task constraints	Interactive	General purpose
Navarro-Serment 2004	Cameras	Positions	Evaluation function	Observation
Zingaretti 2006	Cameras	Relative intrinsic translation, pan, tilt, field of view angle	Partially observable Markov decision	Self-localization
State 2006	Cameras	Visibility, overlap, resolution	Simulation	3D reconstruction in VR
Lidoris et al. 2006	Camera	Gaze direction	Information gain	SLAM

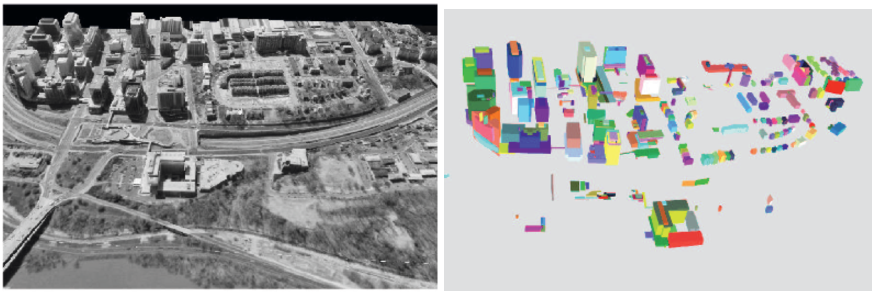

Fig. 3.16. The scene model in which the user can interactively select the target for sensor planning (Stamos and Allen 1998)

With similar tasks, a city model was generated from an incomplete graphics model of Rosslyn VA and was translated by the system to a valid solid model which the planner can use. Overlaid on the city model are the viewing volumes generated for different viewpoints on a selected target face in the scene. The object models and targets can be interactively manipulated while camera positions and parameters are selected to generate synthesized images of the targets that encode the viewing constraints. They extended this system to include resolution constraints (Tarabanis 1995, Allen and Leggett 1995, Reed et al. 1997, Stamos 1998, Abrams 1999).

3.9.2 Placement for Virtual Reality

Interactive camera planning is sometimes also used for virtual reality or simulation. Typical examples can be found from Williams and Lee (2006) and State et al. (2006). For example, the work by State et al. is to simulate in real time multi-camera imaging configurations in complex geometric environments. The interactive visibility simulator helps to assess in advance conditions such as visibility, overlap between cameras, absence of coverage and imaging resolution everywhere on the surfaces of a pre-modeled, approximate geometric dataset of the actual real-world environment the cameras are to be deployed in. A simulation technique is applied to a task involving real-time 3D reconstruction of a medical procedure. It has proved useful in designing and building the multi-camera acquisition system as well as a remote viewing station for the reconstructed data. The visibility simulator is a planning aid requiring a skilled human system designer to interactively steer a simulated multi-camera configuration towards an improved solution.

3.9.3 Robot Localization

As a problem of determining the position of a robot, localization has been recognized as one of the most fundamental problems in mobile robotics. The aim of localization is to estimate the position of a robot in its environment, given local sensorial data. Zingaretti and Frontoni (2006) present an efficient metric for appearance-based robot localization. This metric is integrated in a framework that uses a partially observable Markov decision process as position evaluator, thus allowing good results even in partially explored environments and in highly perceptually aliased indoor scenarios. More details of this topic are related to the research on simultaneous localization and mapping (SLAM) which is also a challenging problem and has been widely investigated (Eustice et al. 2006, Ohno et al. 2006, Lidoris et al. 2006, Herath et al. 2006, Zhenhe and Samarabandu 2005, Jose and Adams 2004, Takezawa et al. 2004, Prasser 2003).

3.9.4 Attention and Gaze

The general concept of active sensor planning should include attention and gaze. This book, however, does not place much emphasis on this issue. Some related works can be found from Bjorkman and Kragic (2004) and (Lidoris et al. 2006). Especially, Bjorkman et al. introduce a real-time vision system that integrates a number of algorithms using monocular and binocular cues to achieve robustness in realistic settings, for tasks such as object recognition, tracking and pose estimation (Fig. 3.17). The system consists of two sets of binocular cameras; a peripheral set for disparity-based attention and a foveal one for higher-level processes. Thus the conflicting requirements of a wide field of view and high resolution can be overcome. One important property of the system is that the step from task specification through object recognition to pose estimation is completely automatic, combining both appearance and geometric models. Experimental evaluation is performed in a realistic indoor environment with occlusions, clutter, changing lighting and background conditions.

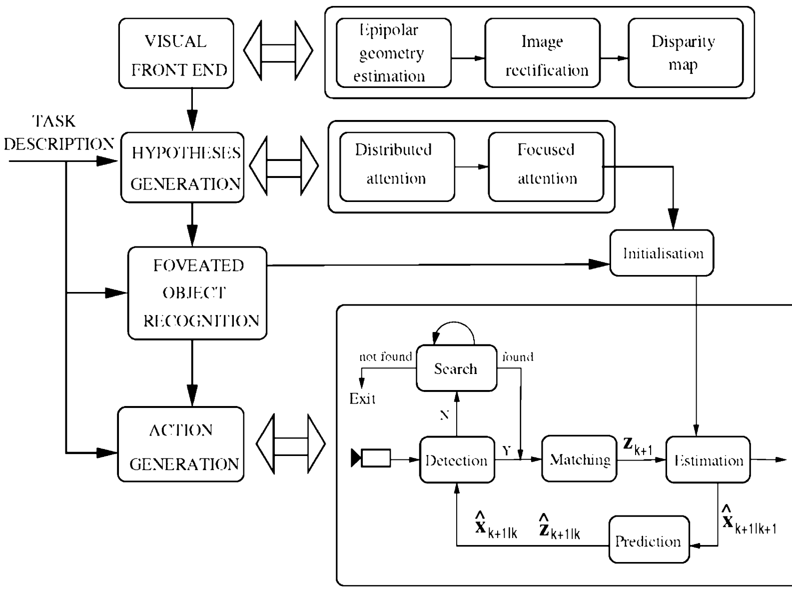

Fig. 3.17. The active vision system involving attention and gaze for action decision (Bjorkman and Kragic 2004)

3.10 Summary

This chapter summarizes the recent development related to the active sensor planning problem. Typical works for inspection, recognition, search, exploration, navigation, tracking, assembly, and disassembly are listed for readers to have a general overview of the state-of-the art.

In model-based tasks, the viewpoint planning is to find a set of admissible viewpoints in the acceptable space, which satisfy a set of the sensor placement constraints and can well finish the vision task. However, the previous approaches are normally formulated for a particular application and are therefore difficult to apply to general tasks. They mainly focus on modeling of sensor constraints and calculating a "good" viewpoint to observe one or several features on the object. Little consideration is given to the overall efficiency of a generated plan with a sequence of viewpoints. 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.

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. These problems will be discussed in the following chapters.