The International Journal of Robotics Research
http://ijr.sagepub.com/

Active vision in robotic systems: A survey of recent developments

Shengyong Chen, Youfu Li and Ngai Ming Kwok
The International Journal of Robotics Research 2011 30: 1343 originally published online 22 August 2011
DOI: 10.1177/0278364911410755

The International Journal of
Robotics Research
30(11) 1343–1377

DOI: 10.1177/0278364911410755
ijr.sagepub.com

Active vision in robotic systems: A survey of recent developments

Shengyong Chen[1], Youfu Li[2] and Ngai Ming Kwok[3]

Abstract

In this paper we provide a broad survey of developments in active vision in robotic applications over the last 15 years. With increasing demand for robotic automation, research in this area has received much attention. Among the many factors that can be attributed to a high-performance robotic system, the planned sensing or acquisition of perceptions on the operating environment is a crucial component. The aim of sensor planning is to determine the pose and settings of vision sensors for undertaking a vision-based task that usually requires obtaining multiple views of the object to be manipulated. Planning for robot vision is a complex problem for an active system due to its sensing uncertainty and environmental uncertainty. This paper describes such problems arising from many applications, e.g. object recognition and modeling, site reconstruction and inspection, surveillance, tracking and search, as well as robotic manipulation and assembly, localization and mapping, navigation and exploration. A bundle of solutions and methods have been proposed to solve these problems in the past. They are summarized in this review while enabling readers to easily refer solution methods for practical applications. Representative contributions, their evaluations, analyses, and future research trends are also addressed in an abstract level.

Keywords

Active vision, computer vision, purposive perception planning, robotics, sensor placement, uncertainty, viewpoint scheduling

1. Introduction

About 20 years ago, Bajcsy, Cowan, Kovesi, and others discussed the important concept of active perception. 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 initiated in active 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, while the latter considers it as the procedural process of the matter.

Therefore, active perception mostly encourages the idea of moving a sensor to constrain interpretation of its environment. Since multiple three-dimensional (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. Today, the roles of sensor planning can be widely found in most autonomous robotic systems (Chen et al. 2008a).

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, 3D data acquisition, and robot action planning are the essential steps in developing such active vision systems.

Research in active vision is concerned with determining the pose and configuration for the visual sensor, plays an important role in robot vision 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

[1]Department of Computer Science, Zhejiang University of Technology, Hangzhou, People's Republic of China
[2]Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong
[3]School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW, Australia

Corresponding author:
Shengyong Chen, Department of Computer Science, Zhejiang University of Technology, 288 Liuhe Road, Hangzhou 310023, People's Republic of China
Email: sy@ieee.org

have to be planned for many vision tasks to make the entire object strategically visible.

By taking active actions in robotic perception, the vision sensor is purposefully configured and placed at several positions to observe a target. The intentional actions in purposive perception planning introduce active behaviors or purposeful behaviors. The robots with semantic perception can take intentional actions according to its set goal such as going to a specific location or obtaining the full information of an object. The action to be taken depends on the environment and the robot's own current state. However, difficulties often arise due to sensor noise and the presence of unanticipated obstacles in the workplace. To this end, a strategic plan is needed to complete a vision task, such as navigating through an office environment or modeling an unknown object.

In this paper, we review the advances in active vision technology broadly. Overall, significant progress has been made in several areas, including new techniques for industrial inspection, object recognition, security and surveillance, site modeling and exploration, multi-sensor coordination, mapping, navigation, tracking, etc. Owing to space limitations, this review mainly focuses on the introduction of ideas and high-level strategies.

The scope of this paper is very broad across the field of robotics. The term active vision defined in this paper is equivalent to the situation if and only if the robots have to adopt strategies for decisions of sensor placement (replacement) or sensor configuration (reconfiguration). It can be used for either general purposes or specific tasks.

Actually, no review of this nature can cite every paper that has been published. We include what we believe to be a representative sample of important work and broad trends from the last 15 years. In many cases, we provide references in order to better summarize and draw distinctions among key ideas and approaches. For further information regarding the early contributions in this topic, we refer the interested reader to a previous review (Tarabanis et al. 1995).

The remainder of this paper is structured as follows. Section 2 briefly gives an overview of related contributions. Section 3 introduces the tasks, problems, and applications of active vision methods. Section 4 discusses the available methods and solutions to specific tasks. We conclude in Section 5 and offer our impressions of current and future trends on the topic.

2. Overview of contributions

In the literature, there are about 2000 research papers published during 1986–2010 that are closely related to the topic of active vision perception in robotics, including sensor modeling and optical constraints, definition of best next view, placement strategy, and illumination planning. The number of 2010 records is not complete since we searched the publications only in the first quarter and most articles have not come into the indexing databases yet. Figure 1 shows the yearly distribution of the published papers. We can find from the plot that: (1) the subject emerged around 1988 and developed rapidly in the first 10 years, thanks to the new concept of 'active vision'; (2) it reaches a peak in 1998; (3) the subject cools down a little, probably due to the reasons of many difficulties related to 'intelligence'; (4) it became very active again in the last 5 years because of its wide applications; (5) we are currently at a second peak.

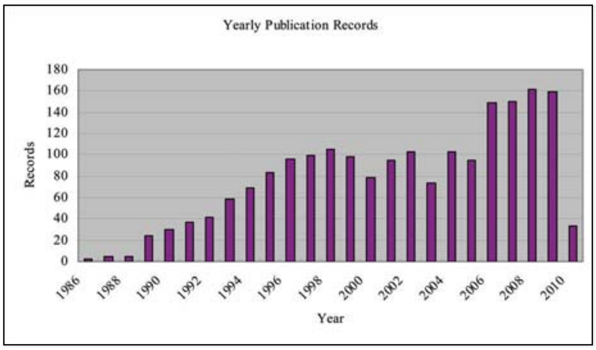

Fig. 1. Yearly published records from 1986 to 2010.

With regards to the research themes, there are several directions that researchers have adopted in the past. In Table 1, we list several classes that categorize the related work of active vision perception; according to target knowledge, sensor type, task or application, approach, evaluation objective, and planning dimensions.

2.1. Representative work

Active vision has very wide application in robotics. Here we summarize these applications in the following list where we can find the most significant roles: purposive sensing, object and site modeling, robot localization and mapping, navigation, path planning, exploration, surveillance, tracking, search, recognition, inspection, robotic manipulation, assembly and disassembly, and other purposes.

For the methods used in solving active vision problems, we can also find a tremendous diversity. The most commonly used methods are: generate and test, synthesis, sensor simulation, hypothesis and verification, graph theory, cooperative network, space tessellation, geometrical analysis, surface expectation, coverage and occlusion, tagged roadmap, visibility analysis, next best view, volumetric space, probability and entropy, classification and Bayesian reasoning, learning and knowledge-based, sensor structure, dynamic configuration, finite element, gaze and attention, lighting control, fusion, expert system, multi-agent, evolutionary computation, soft computing, fuzzy inference, neural network, basic constraints, and task-driven.

Among the huge variety of tasks and methods, we extract a few representative contributions in Table 2 to represent the state of the art.

Table 1. Categories of active vision applications.

Knowledge	model-based	partial	no *a priori*	
Sensor	intensity camera	range sensor	both	others
Task	inspection	navigation	modeling	recognition
Approach	generate and test	synthesis	graph	AI
Objective	visibility	accuracy	efficiency	cost
Dimensions	1	2	2.5	3

Table 2. Representative contributions

Purpose/Task	**Method**	**Representative**
Inspection	Constraint analysis	(Trucco et al. 1997; Tarabanis et al. 1995; Dunn, Olague and Lutton 2006)
Inspection	Genetic algorithm and graph	(Chen and Li 2004)
Surveillance	Linear programming, coverage	(Sivaram et al. 2009; Bottino et al. 2009)
Grasping	Kalman filter	(Motai and Kosaka 2008)
Search	Probability	(Shubina and Tsotsos 2010)
Tracking	Geometrical	(Barreto et al. 2010)
Exploration	Uncertainty driven	(Whaite and Ferrie 1997)
Reinforcement learning	Reinforcement learning	(Royer et al. 2007)
Site modeling	Prediction and verification	(Reed and Allen 2000; Chang and Park 2009; Blaer and Allen 2009; Marchand and Chaumette 1999b)
Object modeling	Next best view	(Banta et al. 2000; Chen and Li 2005; Pito 1999)
Object modeling	Information entropy,rule based	(Li and Liu 2005; Kutulakos and Dyer 1995)
Recognition	Optimal visibility	(de Ruiter et al. 2010)
Recognition	Probabilistic	(Farshidi et al. 2009; Roy et al. 2005)
Path planning	Roadmap	(Baumann et al. 2008; Zhang et al. 2009)
General	Random occlusion	(Mittal and Davis 2008)
Camera-lighting	Geometrical	(Marchand 2007)
Multirobot formation	Graph theory	(Kaminka et al. 2008)

2.2. *Further information*

For a quick understanding of the related work, it is recommended to read the representative contributions listed in Table 2. For further intensive tracking of the literature, the following reviews reflect different aspects of the topic:

1. review of active recognition (Arman and Aggarwal 1993);
2. review of industrial inspection (Newman and Jain 1995);
3. review of sensor planning in the early staged (Tarabanis et al. 1995);
4. review of 3D shape measurement with active sensing (Chen et al. 2000);
5. review of surface reconstruction from multiple range images (Zhang et al. 2000);
6. review and comparison of view planning techniques for 3D object reconstruction and inspection (Scott et al. 2003);
7. review of free-form surface inspection techniques (Li and Gu 2004);
8. review of active recognition through next view planning (Roy et al. 2004);
9. review of 3D measurement quality metrics by environmental factors (MacKinnon et al. 2008b);
10. review of multimodal sensor planning and integration for wide area surveillance (Abidi et al. 2008);
11. review of computer-vision-based fabric defect detection (Kumar 2008).

Of course, purposive perception planning remains an open problem in the community. The task of finding a suitably small set of sensor poses and configurations for specified reconstruction or inspection goals is extremely important for autonomous robots. The ultimate solution is unlikely to ever exist as the complicated problems always need better solutions along with the development of artificial intelligence.

3. Tasks and problems

Active vision endows the robot with the ability to actively place the sensor at several viewpoints through a planning strategy. It inevitably became a key issue in active systems because the robot had to decide ‘where to look’. 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 to the camera 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. On the other hand, a reconfigurable sensor can change its structural parameters to adapt itself to the scene to obtain maximum 3D information from the target. According to task conditions, the problem is roughly classified into two categories, i.e. model-based and non-model-based tasks.

3.1. Model and non-model-based approaches

For model-based tasks, especially for industrial inspections, the placements of the sensor need to be determined and optimized before carrying out operations. In general, in these tasks, the sensor planning problem is to find a set of admissible viewpoints within the permissible space, which satisfy all of the sensor placement constraints and can complete the required vision task. In most of the related works, the constraints in sensor placement are expressed as a cost function where the planning is aimed at achieving the minimum cost. However, the evaluation of a viewpoint has normally previously 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 (Boutarfa et al. 2008).

The most typical task of model-based application is for industrial inspection (Yang and Ciarallo 2001). Along with the CAD model of the target, a sensing plan is generated to completely and accurately acquire the geometry of the target (Sheng et al. 2001b; Olague 2002). The sensing plan comprises the set of viewpoints that defines the exact position and orientation of the camera relative to the target (Prieto et al. 2003). Sampling of the object surface and viewpoint space is characterized, including measurement and pose errors (Scott 2009).

For tasks of observing unknown objects or environments, the viewpoints have to be decided in runtime 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 the knowledge about the unseen portions of the viewing volume while satisfying all 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 10 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.

The most typical task of non-model-based application is for target modeling (Banta et al. 2000). Online planning is required to decide where to look (Lang and Jenkin 2000) for site modeling (Reed and Allen 2000) or real-time exploration and mapping (Kollar and Roy 2008). Of the published literature in active vision perception over the years, Cowan and Kovesi (1988) presented one of the earliest pieces of research on this problem although some primary works can be found in the period 1985–1987.

To date, there have been more than 2000 papers published. At the early stage, these works focused on sensor modeling and constraint analysis. In the first 10 years, most of these research works were model-based and usually for applications in automatic inspection or recognition. The generate-and-test method and the synthesis method are mostly used. In the recent 10 years, while optimization was still in development for model-based problems, the importance is being increasingly realized in planning viewpoints for unknown objects or no *a priori* environment because this is very useful for many active vision tasks such as site modeling, surveillance, and autonomous navigation. The tasks and problems are summarized in this section separately.

3.2. Purposive sensing

The aim of purposive sensing in robotic tasks is to obtain better images for robot understanding. Efficiency and accuracy are often the primary concerns in the acquisition of 3D images (Chen et al. 2008b; Fang et al. 2008; Li and Wee 2008). Taking the most common example of using stereo image sequences during robot movement, not all input images contribute equally to the quality of the resultant motion. Since several images may often contain similar and hence overly redundant visual information. This leads to unnecessarily increased processing times. On the other hand, a certain degree of redundancy can help to improve the reconstruction in more difficult regions of a model. Hornung and colleagues proposed an image selection scheme for multi-view stereo which results in improved reconstruction quality compared to uniformly distributed views (Hornung et al. 2008).

People have also sought methods for determining the probing points for efficient measurement and reconstruction of freeform surfaces (Li and Liu 2003). For an object that has a large surface or a local steep profile, a variable resolution optical profile measurement system that combined two CCD cameras with zoom lenses, one line laser and a three-axis motion stage was constructed (Tsai and Fan 2007). The measurement system can flexibly zoom the lens in or out to measure the object profile according to the slope distribution of the object. Model-based simulation system is helpful for planning numerically controlled surface scanning (Wu et al. 2005). The scanning-path determination is equivalent to the solution of next best view in this aspect (Sun et al. 2008).

In order to obtain a minimal error in 3D measurements (MacKinnon et al. 2008a), an optimization design of the camera network in photogrammetry is useful in 3D reconstruction from several views by triangulation (Olague and Mohr 2002). The combination of laser scanners and touch probes can potentially lead to more accurate, faster, and denser measurements. To overcome the conflict between efficiency and accuracy, Huang and Qian developed a dynamic sensing-and-modeling approach for integrating a tactile point sensor and an area laser scanner to improve the measurement speed and quality (Huang and Qian 2007).

Spatial uncertainty and resolution are the primary metrics of image quality; however, spatial uncertainty is affected by a variety of environmental factors. A review of how researchers attempted to quantify these environmental factors was given by MacKinnon et al. (2008b), along with spatial uncertainty and resolution, and an illustration of a wide range of quality metrics was provided.

For reconstruction in large scenes having large depth ranges with depth discontinuities, an idea is available to integrate coarse-to-fine image acquisition and estimation from multiple cues (Das and Ahuja 1996).

For the construction of realistic models, simultaneous capture of the geometry and texture (Treuillet et al. 2007) is inevitable. The quality of the 3D reconstruction depends not only on the complexity of the object but also on its environment. Good viewing and illumination conditions ensure image quality and thus minimize the measurement error. Belhaoua and colleagues investigated the placement problem of lighting sources moving within a virtual geodesic sphere containing the scene, with the aim of finding positions leading to minimum errors for the subsequent 3D reconstruction (Belhaoua et al. 2009; Liu 2009). It is also found that automatic light source placement plays an important role for maximum visual information recovery (Vazquez 2007).

3.3. Object modeling

In order to reconstruct an object completely and accurately (Shum et al. 1997; Banta et al. 2000; Lang and Jenkin 2000; Doi et al. 2005; Li and Liu 2005), and at the same time determine the scanning path (Larsson and Kjellander 2008; Wang et al. 2009), multiple images have to be acquired from different views (Pito 1999). An increasing number of views generally improve the accuracy of the final 3D model but it also increases the time needed to build the model. The number of the possible views can, in principle, be infinite. Therefore, it makes sense to try to reduce the number of required views to a minimum while preserving a certain accuracy of the model, especially in applications for which the efficiency is an important issue. Approaches to next view planning can not only can generate 3D shapes with minimal views (Sablatnig et al. 2003; Zhou et al. 2008), but also is especially useful for the acquisition of large-scale indoor and outdoor scenes (Blaer and Allen 2007) or interior and exterior models (Null and Sinzinger 2006), even with partial occlusions (Triebel and Burgard 2008).

To minimize the number of images for complete 3D reconstruction where no prior information about the objects is available, in the literature techniques are explored based on characterizing the shapes to be recovered in terms of visibility and number and nature of cavities (Pito 1999; Chen and Li 2005; Zetu and Akgunduz 2005; He and Li 2006a; Lin et al. 2007; Loniot et al. 2007).

Typically, Callieri and colleagues designed a system to reduce the three main bottlenecks in human-assisted 3D scanning: the selection of the range maps to be taken (view planning), the positioning of the scanner in the environment, and the range maps' alignment. The system is designed around a commercial laser-based 3D scanner moved by a robotic arm. The acquisition session is organized in two stages. First, an initial sampling of the surface is performed by automatic selection of a set of views. Then, some added views are automatically selected, acquired and merged with the initial set in order to fill the surface regions left unsampled (Callieri et al. 2004). Similar techniques of free-form surface scanning were presented by Huang and Qian (2008a, 2008b) and Fernandez et al. (2008).

The strategy of viewpoint selection for global 3D reconstruction of unknown objects presented by Jonnalagadda et al. (2003) has four steps: local surface feature extraction, shape classification, viewpoint selection and global reconstruction. An active vision system (Biclops) with two cameras constructed for independent pan/tilt axes, extracts 2D and 3D surface features from the scene. These local features are assembled into simple geometric primitives. The primitives are then classified into shapes, which are used to hypothesize the global shape of the object. The next viewpoint is chosen to verify the hypothesized shape. If the hypothesis is verified, some information about global reconstruction of a model can be stored. If not, the data leading up to this viewpoint are re-examined to create a more consistent hypothesis for the object shape.

Unless using 3D reconstruction from unordered viewpoints (Liang and Wong 2010), incremental modeling is the common choice for complete automation of scanning an unknown object. An incremental model, representing object surface and workspace occupancy, is combined together with an optimization strategy to select the best scanning viewpoints and generate adaptive collision-free scanning trajectories. The optimization strategy attempts to select the viewpoints that maximize the knowledge of the object taking into account the completeness of the current model and the constraints associated with the sensor (Martins et al. 2003).

Most methods for model acquisition require the combination of partial information from different viewpoints in order to obtain a single, coherent model. This, in turn, requires the registration of partial models into a common coordinate frame, a process that is usually done off-line.

As a consequence, holes due to undersampling and missing information often cannot be detected until after the registration. Liu and Heidrich introduced a fast, hardware-accelerated method for registering a new view to an existing partial geometric model in a volumetric representation. The method performs roughly one registration every second, and is therefore fast enough for on-the-fly evaluation by the user (Liu and Heidrich 2003). A procedure is also recently proposed to identify missing areas from the initial scanning data from default positions and to locate additional scanning orientations to fill in the missing areas (Chang and Park 2009). In contrast, He, Li and colleagues prefer to use a self-determination criterion to inform the robot when the model is complete (Li et al. 2005a, 2005b; He and Li 2006b).

3.4. Site modeling

It is very time-consuming to construct detailed models of large complex sites using a manual process. Therefore, in tasks of modeling unstructured environments (Craciun et al. 2008), especially in a wide outdoor area, perception planning is required to reduce unobserved portions (Asai et al. 2007). One of the main drawbacks is determining how to guide the robot and where to place the sensor to obtain complete coverage of a site (Reed and Allen 2000; Blaer and Allen 2009). To estimate the computational complexity, if the size of one dimension of the voxel space is n, then there could be $O(n^2)$ potential viewing locations. If there are m boundary unseen voxels, the cost of the algorithm could be as high as $O(n^2 \times m)$ (Blaer and Allen 2009).

For static scenes, the perception-action cycles can be handled at various levels: from the definition of perception strategies for scene exploration down to the automatic generation of camera motions using visual servoing. Marchand and Chaumette use a structure from controlled motion method which allows an optimal estimation of geometrical primitive parameters (Marchand and Chaumette 1999b). The whole reconstruction/exploration process has three main perception–action cycles (Figure 2). It contains the internal perception–action cycle which ensures the reconstruction of a single primitive, and a second cycle which ensures the detection, the successive selection, and finally the reconstruction of all the observed primitives. It partially solves the occlusion problem and obtains a high-level description of the scene.

In field environments, it is usually not possible to provide robotic systems with valid/complete geometric models of the task and environment. The robot or robot teams will need to create these models by performing appropriate sensor actions. In addition, the robot(s) will need to position its sensors in a task-directed optimum way. The Instant Scene Modeler (iSM) is a vision system for generating calibrated photo-realistic 3D models of unknown environments quickly using stereo image sequences (Se and Jasiobedzki 2007). Equipped with iSM, unmanned ground

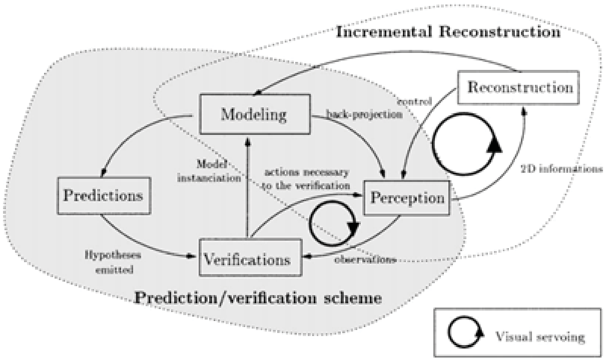

Fig. 2. The prediction/verification scheme for scene exploration (with kind permission from Springer Science + Business Media: Marchand E and Chaumette F (1999b) An autonomous active vision system for complete and accurate 3D scene reconstruction. *Int J Comput Vision* 32: 171–194).

vehicles (UGVs) can capture stereo images and create 3D models to be sent back to the base station, while they explore unknown environments. An algorithm based on iterative sensor planning and sensor redundancy is proposed by Sujan and Dubowsky to enable robots to efficiently position their cameras with respect to the task/target (Sujan and Dubowsky 2005a). Intelligent and efficient strategy is developed for unstructured environment (Sujan and Meggiolaro 2005).

A field robot for site modeling is usually equipped with range sensors, DGPS/compass, an inertial measurement unit (IMU), odometers, etc., such as the iSM (Se and Jasiobedzki 2007). More sensors are set up with the AVENUE, for localizing and navigating itself through various environments (Figure 3).

3.5. Surveillance

Since vision contains much higher information content than other sensors in describing the scene, cameras are frequently applied for surveillance purposes. In these tasks, cameras can be installed in fixed locations and directed, through pan–tilt manipulations, toward the target in an active manner. On the other hand, cameras can be installed on mobile platforms. Surveillance is also tightly connected with target search and tracking where the active vision principle is regarded as an important attribute.

3.5.1. Surveillance with a set of fixed cameras This problem was addressed by Sivaram et al. (2009) who were concerned with how to select the optimal combination of sensors and how to determine their optimal placement in a surveillance region in order to meet the given performance requirements at a minimal cost for a multimedia surveillance system. Therefore, the sensor configuration for surveillance applications calls for coverage optimization (Janoos et al. 2007; Yao et al. 2010). The goal in

Fig. 3. The ATRV-2 AVENUE-based mobile robot for site modeling (Blaer and Allen 2009 (with permission of Wiley Blackwell)).

such problems is to develop a strategy of network design (Saadatseresht and Varshosaz 2007).

Locating sensors in 2D can be modeled as an art gallery problem (Howarth 2005; Bottino and Laurentini 2006a; Bodor et al. 2007). Consider the external visibility coverage for polyhedra under the orthographic viewing model. The problem is to compute whether the whole boundary of a polyhedron is visible from a finite set of view directions, and if so, how to compute a minimal set of such view directions (Liu and Ramani 2009). Bottino and colleagues provide detailed formulation and solution in their research (Bottino and Laurentini 2008; Bottino et al. 2009).

3.5.2. Surveillance with mobile robots Mobile sensors can be used to provide complete coverage of a surveillance area for a given threat over time, thereby reducing the number of sensors required. The surveillance area may have a given threat profile as determined by the kind of threat, and accompanying meteorological, environmental, and human factors (Ma et al. 2009). UGVs equipped with surveillance cameras present a flexible complement to the numerous stationary sensors being used in security applications today (Ulvklo et al. 2004; Hernandez and Wang 2008). However, to take full advantage of the flexibility and speed offered by a group of UGV platforms, a fast way to compute desired camera locations to cover an area or a set of buildings, e.g., in response to an alarm, is needed (Nilsson et al. 2008, 2009).

Such surveillance systems aim to design an optimal deployment of vision sensors (Lim et al. 2006; Angella et al. 2007; Nayak et al. 2008). System reconfiguration is sometimes necessary for the autonomous surveillance of a target as it travels through a multi-object dynamic workspace with an *a priori* unknown trajectory (Bakhtari et al. 2006; Bakhtari and Benhabib 2007; Bakhtari et al. 2009).

Autonomous patrolling robots are to have significant contributions in security applications for surveillance purposes (Briggs and Donald 2000; Cassinis and Tampalini 2007). In the near future robots will also be used in home environments to provide assistance for the elderly and challenged people (Nikolaidis et al. 2009; Biegelbauer et al. 2010).

In monitoring applications (Sakane et al. 1995; Mackay and Benhabib 2008a), Schroeter et al. present a model-based system for a mobile robot to find an optimal pose for the observation of a person in indoor living environments. The observation pose is derived from a combination of the camera position and view direction as well as further parameters such as the aperture angle. The optimal placement of a camera is required because of the highly dynamic range of the scenes near windows or other bright light sources, which often results in poor image quality due to glare or hard shadows. The method tries to minimize these negative effects by determining an optimal camera pose based on two major models: A spatial free space model and a representation of the lighting (Schroeter et al. 2009). A recent review of multimodal sensor planning and integration for wide area surveillance can be found in (Abidi et al. 2008).

3.5.3. Search Object search is also a model-based vision task where the object 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 (Shimizu et al. 2005; Wang et al. 2008).

The task is often complicated by the fact that portions of the area are hidden from the camera view. Different viewpoints are necessary to observe the target. As a consequence, viewpoint selection for search tasks seems similar to viewpoint selection for data acquisition of an unknown scene. The problem of visual matching was shown to be NP complete (Ye and Tsotsos 1999). It has exponential time complexity relative to the size of the image. Suppose one wishes a robot to search for and locate a particular object in a 3D world. A direct search certainly suffices for the solution. Assuming that the target may lie with equal probability at any location, the viewpoint selection problem is resolved by moving a camera to take images of the previously not viewed portions of the full 3D space. This kind of exhaustive, brute-force approach can suffice for a solution; however, it is both computationally and mechanically prohibitive.

In practice, sensor planning is very important for object search since a robot needs to interact intelligently and effectively with the environment. Visual attention may be a mechanism that optimizes the search processes inherent in vision, but attention itself is a complex phenomenon (Shubina and Tsotsos 2010). The utility of a search operation **f** is given by

$$E(\mathbf{f}) = \frac{\sum \mathbf{p}(c_i, \tau)\,\mathbf{b}(c_i, \mathbf{f})}{t(\mathbf{f})} \tag{1}$$

where $t(\mathbf{f})$ is the time when action **f** takes place. The knowledge about the potential target locations is encoded as a target probability distribution $\mathbf{p}(c_i, \tau)$. The goal is to select an operation with the highest utility value. Since the cost of each action is approximately the same if the robot is stationary, the next action is selected in such a way that it maximizes the numerator of (1) (Shubina and Tsotsos 2010).

With the assumption of a realistic, high-dimensional and continuous state space for the representation of objects expressing their rotation, translation and class, Eidenberger et al. present an exclusively parametric approach for the state estimation and decision-making process to achieve very low computational complexity and short calculation times (Eidenberger et al. 2008).

3.5.4. Tracking Active tracking is part of the active vision paradigm (Riggs et al. 2010), where visual systems adapt themselves to the observed environment in order to obtain extra information or perform a task more efficiently. An example of active tracking is fixation, where camera control assures that the gaze direction is maintained on the same object over time. A general approach for the simultaneous tracking of multiple moving targets using a generic active stereo setup is studied by Barreto et al. (2010). The problem is formulated for objects on a plane, where cameras are modeled as line scan cameras, and targets are described as points with unconstrained motion.

Dynamically reconfigurable vision systems have been suggested in an online mode (Reddi and Loizou 1995; Wang et al. 2008), as effective solutions for achieving this objective, namely, relocating cameras to obtain optimal visibility for a given situation. To obtain optimal visibility of a 3D object of interest, its six-degree-of-freedom (six-DOF) position and orientation must be tracked in real time. An autonomous, real-time, six-DOF tracking system for *a priori* unknown objects should be able to (1) select the object, (2) build its approximate 3D model and use this model to (3) track it in real time (de Ruiter et al. 2010).

Zhu and Sakane developed an adaptive panoramic stereovision approach for localizing 3D moving objects (Zhou and Sakane 2003). The research focuses on cooperative robots involving cameras 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 drawn. (i) If the distance from the main camera to the target, $D1$, is significantly greater (e.g. a factor of 10 greater) than the size of the robot (R), the best geometric configuration is

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

where B is the best baseline distance for the minimum distance error and ϕ_1 is the main camera's inner angle of the triangle formed by the two robots and the target. (ii) The depth error of the adaptive stereovision is proportional to $D^{1.5}$ where D is the camera–target distance, 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).

Some problems such as camera fixation, object capturing and detecting, and road following involve tracking or fixating on 3D points and features (Biegelbauer et al. 2010). The solutions to these problems also require an analysis of depth and motion. Theoretical approaches based on optical flow are the most common solution to these problems (Raviv and Herman 1994; Han et al. 2008).

Vision tracking systems for surveillance and motion capture rely on a set of cameras to sense the environment (Chen and Davis 2008). There is a decision problem which corresponds to answering the question: can the target escape the observer's view? Murrieta-Cid et al. defined this problem and considered to maintain surveillance of a moving target by a non-holonomic mobile observer (Murrieta-Cid et al. 2005). The observer's goal is to maintain visibility of the target from a predefined, fixed distance. An expression derived for the target velocities is

$$\begin{pmatrix} \dot{x}_T(t) \\ \dot{y}_T(t) \end{pmatrix} = \begin{pmatrix} \cos\theta & -l\cos\phi \\ \sin\theta & l\cos\phi \end{pmatrix} \begin{pmatrix} u_1 \\ u_3 \end{pmatrix} \tag{3}$$

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

$$f(u_1, u_3) = u_1^2 + 2u_1u_3 l\sin(\theta - \phi) + l^2 u_3^2 = 1. \tag{4}$$

The above equation 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 assuming $\dot{x}_T^2 + \dot{y}_T^2 \leq 1$. They deal specifically with the situation in which the only constraint on the target's velocity is a bound on the speed, and the observer is a non-holonomic, differential drive system having bounded speed. The system model is developed to derive a lower bound for the required observer speed.

To dynamically manage the viewpoint of a vision system for optimal 3D tracking, Chen and Li adopt the effective sample size in the proposed particle filter as a criterion for evaluating tracking performance and employ it to guide the view-planning process to find the best viewpoint configuration. The vision system is designed and configured to maintain a largest number of effective particles, which minimizes tracking error by revealing the system to a better swarm of importance samples and interpreting posterior states in a better way (Chen and Li 2008, 2009).

3.6. *Mobile Robotics*

In applications involving the deployment of mobile robots, it is a fundamental requirement that the robot is able to take perception of its navigation environment. When cameras are equipped on mobile robots, it enables the robot to observe its workspace and active vision naturally becomes a very desirable ability to improve the autonomy of these machines.

3.6.1. Localization and mapping As a problem of determining the position of a robot or its vision sensor, localization has been recognized as one of the most fundamental problems in mobile robotics (Caglioti 2001; Flandin and Chaumette 2002). Mobile robots often determine their actions according to their positions. Thus, their observation strategies are mainly for self-localization (Mitsunaga and Asada 2006). The aim of localization is to estimate the position of a robot in its environment, given local sensorial data. Stereo vision-based 3D localization is used in a semi-automated excavation system for partially buried objects in unstructured environments by Maruyama et al. (2010). Autonomous navigation is also possible in outdoor situations with the use of a single camera and natural landmarks (Royer et al. 2007; Chang et al. 2010).

Zingaretti and Frontoni present an efficient metric for appearance-based robot localization (Zingaretti and Frontoni 2006). 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 (Gonzalez-Banos and Latombe 2002; Borrmann et al. 2008; Frintrop and Jensfelt 2008a; Nuchter and Hertzberg 2008).

In intelligent transportation systems, vehicle localization usually relies on global positioning system (GPS) technology; however the accuracy and reliability of GPS are degraded in urban environments due to satellite visibility and multipath effects. Fusion of data from a GPS receiver and a machine vision system can help to position the vehicle with respect to objects in its environment (Rae and Basir 2009).

In robotics, maps are metrical and sometimes topological. A map contains space-related information about the environment, i.e. not all that a robot may know or learn about its world need go into the map. Metric maps are supposed to represent the environment geometry quantitatively correctly, up to discretization errors (Nuchter and Hertzberg 2008).

Again for the SLAM problem (Ballesta et al. 2010; Kaess and Dellaert 2010), the goal is to integrate the information collected during navigation into the most accurate map possible. However, SLAM does not address the sensor-placement portion of the map-building task. That is, given the map built so far where should the robot go next? Gonzalez-Banos and Latombe (2002) proposed an algorithm to guide the robot through a series of 'good' positions, where 'good' refers to the expected amount and quality of the information that will be revealed at each new location. This is similar to the next-best-view (NBV) problem. However, in mobile robotics the problem is complicated by several issues, two of which are particularly crucial. One is to achieve safe navigation despite an incomplete knowledge of the environment and sensor limitations. The other is the need to ensure sufficient overlap between each new local model and the current map, in order to allow registration of successive views under positioning uncertainties inherent to mobile robots. They described an NBV algorithm that uses the safe-region concept to select the next robot position at each step. The new position is chosen within the safe region in order to maximize the expected gain of information under the constraint that the local model at this new position must maintain a minimal overlap with the current global map (Gonzalez-Banos and Latombe 2002).

In addition, because individual scans are registered into a coherent 3D geometry map by SLAM, semantic knowledge can help an autonomous robot act goal-directly, then, consequently, part of this knowledge has to be related to objects, functionalities, events, or relations in the robot's environment. A semantic map for a mobile robot is a map that contains, in addition to spatial information about the environment, assignments of mapped features to entities of known classes (Nuchter and Hertzberg 2008).

While considerable progress has been made in the area of mobile networks by SLAM or NBV, a framework that allows the vehicles to reconstruct a target based on a severely underdetermined data set is rarely addressed. Recently, Mostofi and Sen present a compressive cooperative mapping framework for mobile exploratory networks. The cooperative mapping of a spatial function is based on a considerably small observation set where a large percentage of the area of interest is not sensed directly (Mostofi and Sen 2009).

3.6.2. Navigation, path planning, and exploration For exploring unknown environments, many robotic systems use topological structures as a spatial representation. If localization is done by estimating the global pose from landmark information, robotic navigation is tightly coupled to metric knowledge. On the other hand, if localization is based on weaker constraints, e.g. the similarity between images capturing the appearance of places or landmarks, the navigation can be controlled by a homing algorithm. Similarity-based localization can be scaled to continuous metric localization by adding additional constraints (Whaite and Ferrie 1997; Hovland and McCarragher 1999; Sheng et al. 2001a; Kim and Cho 2003; Hubner and Mallot 2007; Baker and Kamgar-Parsi 2010).

If the environment is partially unknown, the robot needs to explore its work space autonomously. Its task is to incrementally build up a representation of its surroundings (Suppa and Hirzinger 2007; Wang and Gupta 2007). A local navigation strategy has to be implemented for unknown environment exploration (Amin et al. 2008; Radovnikovich et al. 2010; Thielemann et al. 2010).

For navigation in an active way, an UGV is usually equipped with a 'controllable' vision head, e.g. a stereo camera on a pan/tilt mount (Banish et al. 2010; Borenstein et al. 2010). Kristensen presented the problem of autonomous navigation in partly known environments (Kristensen 1997). 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 decision. In other aspects, indoor navigation using adaptive neuro-fuzzy controller is addressed by Budiharto et al. (2010) and path recognition for outdoor navigation is addressed by Shinzato et al. (2010).

The problem of path planning for a robotic sensor investigated by Zhang et al. (2009) is assumed with a platform geometry $A \subset R^2$, and a field-of-view geometry $S \subset R^2$, that navigates a workspace $W \subset R^2$ for the purpose of classifying multiple fixed targets based on posterior and prior sensor measurements, and environmental information (Figure 4). The robotic sensor path τ must simultaneously achieve multiple objectives including: (1) avoid all obstacles in W; (2) minimize the traveled distance; and (3) maximize the information value of path (τ), i.e. the measurement set along a path τ. The robotic sensor performance is defined by an additive reward function:

$$R(\tau) = w_V V(\tau) - w_D D(\tau) \tag{5}$$

where, $V(\tau)$ is the information value of path (τ), and $D(\tau)$ is the distance traveled along τ. The constants w_V and w_D weigh the trade-off between the values of the measurements and the traveled distance. Then, the *Geometric Sensor Path Planning Problem* is defined as follows.

Problem: Given a layout W and a joint probability mass function P, find a path τ^* for a robotic sensor with platform A and field-of-view S that connects the two ends, and maximizes the profit of information defined in (1) (Zhang et al. 2009).

In active perception for exploration, navigation, or path planning, there is a situation that the robot has to work in

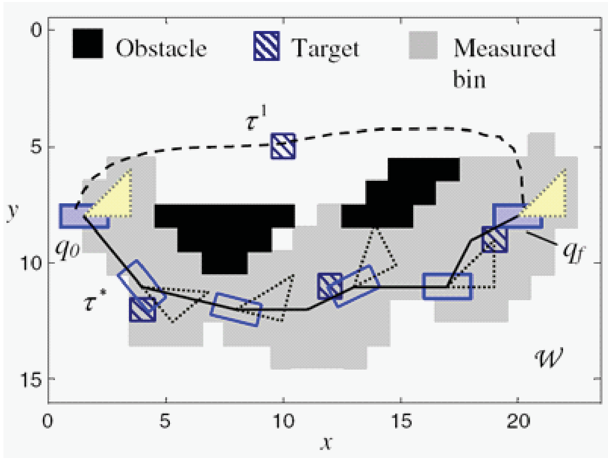

Fig. 4. An example of sensor path planning, where both the location and geometry of targets and obstacles must be accounted for in planning the sensor path (with kind permission from Springer Science + Business Media: Zhang G, Ferrari S and Qian M (2009) An information roadmap method for robotic sensor path planning. *J Intelligent Robotic Syst* 56: 69–98).

a dynamic environment and the sensing process may associate with many noises or uncertainties. Research in this issue has become the most active in recent years. A reinforcement learning scheme is proposed for exploration in Kollar and Roy (2008). Occlusion-free path planning was studied by Baumann et al. (2008, 2010), Nabbe and Hebert (2007), and Oniga and Nedevschi (2010).

3.7. Robotic manipulations

The use of robotic manipulators had shown a boost in manufacturing productivity. This increase depends critically on the simplicity that the robot manipulator can be re-configured or re-programmed to perform various tasks. To this end, actively placing the camera to guide the manipulator motion has become a key component of automatic robotic manipulator systems.

3.7.1. Robotic manipulation Vision-guided approaches are designed to robustly achieve high precision in manipulation (Miura and Ikeuchi 1998; Nickels et al. 2010) or to improve productivity (Park et al. 2006). For the assembly/disassembly tasks, 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 (Figure 5) (Thomas et al. 2007). Nelson et al. introduced a dynamic sensor planning method (Nelson and Papanikolopoulos 1996). 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 of the constraints into a system and resulted in a control law. Kececi et al. employed an independently mobile camera with a six-DOF robot to

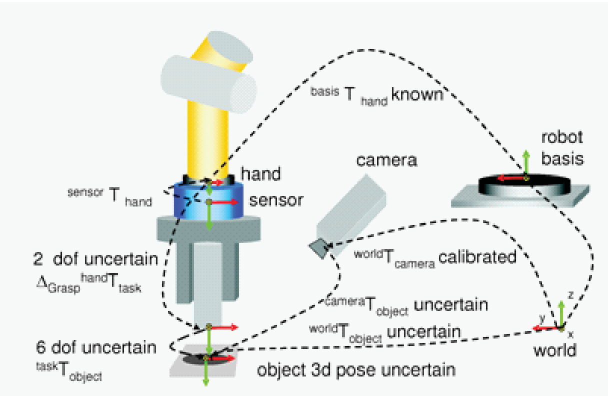

Fig. 5. Vision sensor for solving object poses and uncertainties in the assembly work cell (Thomas et al. 2007 (© 2007 IEEE)).

monitor a disassembly process so that it can be planned (Kececi et al. 1998). 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.

Stemmer et al. used 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 (Stemmer et al. 2006). 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 invokes the passivity properties of the robot and the environment.

Object verification (Sun et al. 2007), feature detectability (Zussman et al. 1994), and real-time accessibility analysis for robotics (Jang et al. 2007) are also major concerns in robotic manipulation. The access direction of the object to grasp can be determined through visibility query (Jang et al. 2008; Motai and Kosaka 2008).

3.7.2. Recognition In many cases, a single view may not contain sufficient features to recognize an object unambiguously (Byun and Nagata 1996). Therefore, another important application of sensor planning is active object recognition (AOR) which has recently attracted much attention within the computer vision community.

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 et al. 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 result, 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

Fig. 6. The objects for active recognition experiments (Farshidi et al. 2009 (with permission of Elsevier)).

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 by Roy et al. 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 (Roy et al. 2005).

In computer vision, object recognition problems are often based on single image data processing (Eggert et al. 1995; SyedaMahmood 1997). In various applications this processing can be extended to a complete sequence of images, usually received passively. Deinzer et al. (2009) selectively moved a camera around a target object. Reliable classification results are desirable with a clearly reduced amount of necessary views by optimizing the camera movement for the access of new viewpoints. The optimization criterion is the gain of class discriminative information when observing the appropriate next image (Gremban and Ikeuchi 1994; Roy et al. 2000).

While relevant research in active object recognition/pose estimation has mostly focused on single-camera systems, Farshidi et al. propose two multi-camera solutions that can enhance object recognition rate, particularly in the presence of occlusion. Multiple cameras simultaneously acquire images from different view angles of an unknown, randomly occluded object belonging to a set of *a priori* known objects (Farshidi et al. 2009). Eight objects, as illustrated in Figure 6, are considered in the experiments with four different pose angles, each 90° apart. Also, five different levels of occlusion have been designated for each camera's image.

In the early stage, Ikeuchi et al. developed a sensor modeler, called VANTAGE, to place the light sources and cameras for object recognition (Ikeuchi and Robert 1991; Wheeler and Ikeuchi 1995). It mostly solves the detectability (visibility) of both light sources and cameras. Borotschnig and Paletta summarized a framework for appearance-based AOR as in Figure 7 (Borotschnig and Paletta 2000).

Among the literature of recognition, many solutions are available (Kuno et al. 1991; Arman and Aggarwal 1993; Dickinson et al. 1997; Callari and Ferrie 2001). Typically,

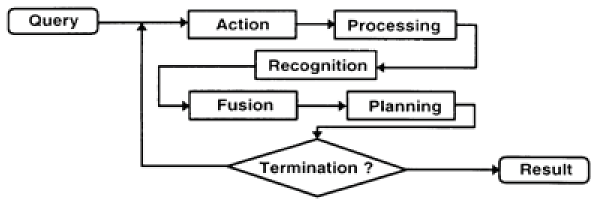

Fig. 7. The framework of appearance-based active object recognition (Borotschnig and Paletta 2000 (with permission of Elsevier)).

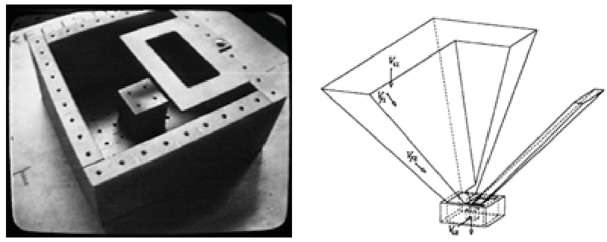

Fig. 8. The admissible domain of viewpoints (Tarabanis et al. 1995 (© IEEE 1995)).

we may refer to the fast recognition by learning (Grewe and Kak 1995) and function-based reasoning (Sutton and Stark 2008), as well as multi-view recognition of time-varying geometry objects (Mackay and Benhabib 2008b). A review of sensor planning for active recognition can be found in Roy et al. (2004).

3.7.3. Inspection Dimensional inspection using a contact-based coordinate measurement machine is time consuming because the part can only be measured on a point-by-point basis (Prieto et al. 2002). The automotive industry has been seeking a practical solution for rapid surface inspection using a 3D sensor. The challenge is the capability to meet all of the requirements including sensor accuracy, resolution, system efficiency, and system cost. A robot-aided sensing system can automatically allocate sensor viewing points, measure the freeform part surface, and generate an error map for quality control (Bardon et al. 2004; Shih and Gerhardt 2006; Shi et al. 2010). A geometric dimension and tolerance inspection process is also needed in industries to examine the conformity of manufactured parts with the part specification defined at the design stage (Gao et al. 2006; Sebastian et al. 2007).

In fact, in the literature, sensor planning for the model-based tasks is mostly related to industrial inspection, where 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 (Mason 1997; Trucco et al. 1997; Sheng et al. 2001b; Yang and Ciarallo 2001; Sheng et al. 2003; Wong and Kamel 2004). It was said that this problem in fact was a nonlinear multi-constraint optimization problem (Chen and Li 2004; Dunn and Olague 2004; Rivera-Rios et al. 2005; Taylor and Spletzer 2007). The problem comprises camera, robot, and environmental constraints. A viewpoint is optimized and evaluated by a cost function which uses a probability-based global search technique. It is difficult to compute robust viewpoints which satisfy all feature detectability constraints. Optimization methods such as tree annealing and genetic algorithms are commonly used to compute the viewpoints subjected to multi-constraints (Olague and Mohr 2002; Chen and Li 2004; Olague and Dunn 2007).

Tarabanis et al. developed a model-based sensor planning system, the machine vision planner (MVP), which works with 2D images obtained from a CCD camera (Tarabanis et al. 1995, 1996). The MVP system 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. A viewpoint is sought that is both globally admissible and central to the admissibility domain (Figure 8).

Based on the work on the MVP system, Abrams et al. made a further development for planning viewpoints for vision inspection tasks within a robot work cell (Abrams et al. 1999). 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). This 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. present a probabilistic analysis of the effect of the localization errors on the dimensional measurements of the line entities for a parallel stereo setup (Figure 9). 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 (Rivera-Rios et al. 2005).

In order to obtain a more complete and accurate 3D image of an object, Prieto et al. presented an automated acquisition planning strategy utilizing its CAD model. 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 (Prieto et al. 2001, 2003).

In addition the minimum number of viewpoints is desired in sensor planning, to further improve the efficiency of robot manipulation, we need to reduce the traveling cost of the robot placements (Wang et al. 2007; Martins et al. 2005; Chen and Li 2004). The whole procedure for generating a perception plan is described as: (1) generate a number

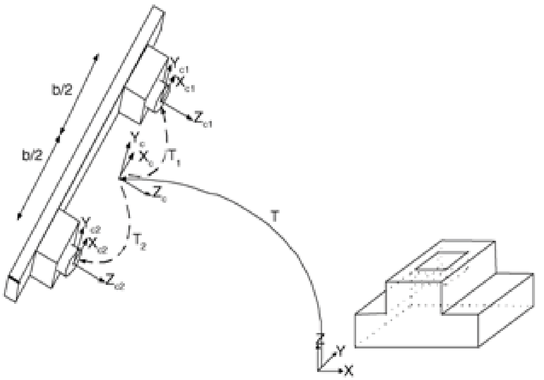

Fig. 9. Stereo pose determination for dimensional measurement (Rivera-Rios et al. 2005 (© IEEE 2005)).

of viewpoints; (2) reduce redundant viewpoints; (3) if the placement constraints are not satisfied, increase the number of viewpoints; (4) construct a graph corresponding to the space distribution of the viewpoints; and (5) find a shortest path to optimize robot operations.

Automated visual inspection systems are also developed for defect inspection, such as specular surface quality control (Garcia-Chamizo et al. 2007), car headlight lens inspection (Martinez et al. 2008), and others (Chen and Liao 2009; Martinez et al. 2009; Perng et al. 2010; Sun et al. 2010). Related techniques are useful to improve the productivity of assembly lines (Park et al. 2006). Self-reconfiguration (Garcia and Villalobos 2007a, 2007b) and self-calibration (Carrasco and Mery 2007; Treuillet et al. 2009) are also mentioned in some applications. Further information regarding early literatures can be found in the review by Newman and Jain (1995).

3.8. General-purpose tasks

The automatic selection of good viewing parameters is a very complex problem. In most cases the notion of good viewing strongly depends on the concrete application, but some general solutions still exist in a limited extent (Chu and Chung 2002; Zavidovique and Reynaud 2007). Commonly, two kinds of viewing parameters must be set for active vision perception: camera parameters and lighting parameters (number of light sources, its position, and eventually the orientation of the spot). The former determine how much of the geometry can be captured and the latter have an influence on how much of it is revealed (Vazquez 2007).

Some multiview strategies are proposed for different application prospects (Al-Hmouz and Challa 2005; Fiore et al. 2008). Mittal specifically addressed the state of the art in the analysis of scenarios where there are dynamically occurring objects capable of occluding each other. The visibility constraints for such scenarios are analyzed in a multi-camera setting. Also analyzed are other static constraints such as image resolution and field of view, and algorithmic requirements such as stereo reconstruction, face detection and background appearance. Theoretical analysis with the proper integration of such visibility and static constraints leads to a generic framework for sensor planning, which can then be customized for a particular task. The analysis may be applied to a variety of applications, especially those involving randomly occurring objects, and include surveillance and industrial automation (Mittal 2006).

In some robotic vision tasks, such as surveillance, inspection, image-based rendering, environment modeling, require multiple sensor locations, or the displacement of a sensor in multiple positions for fully exploring an environment or an object. Edge covering is sufficient for tasks such as inspection or image-based rendering. However, the problem is NP-hard, and no finite algorithm is known for its exact solution. A number of heuristics have been proposed, but their performances with respect to optimality are not guaranteed (Bottino et al. 2009). In 2D surveillance, the problem is modeled as an art gallery problem. A subclass of this general problem can be formulated in terms of planar regions that are typical of building floor plans. Given a floor plan to be observed, the problem is then to reliably compute a camera layout such that certain task-specific constraints are met. A solution to this problem is obtained via binary optimization over a discrete problem space (Erdem and Sclaroff 2006). It can also be applied in security systems for industrial automation, traffic monitoring, and surveillance in public places, such as museums, shopping malls, subway stations, and parking lots (Mittal and Davis 2004, 2008).

With visibility analysis and sensor planning in dynamic environments, in which the methods include computing occlusion-free viewpoints (Tarabanis et al. 1996) and feature detectability constraints (Tarabanis et al. 1994), applications are widely existing in product inspection, assembly, and design in reverse engineering (Tarabanis et al. 1995; Scott 2009; Yegnanarayanan et al. 2009).

In other aspects, an approach was proposed by Marchand (2007) to control camera position and/or lighting conditions in an environment using image gradient information. An auto-focusing technique is used by Quang et al. (2008) in a projector–camera system. Smart cameras are applied by Madhuri et al. (2009). A camera network is designed with dynamic programming by Lim et al. (2007).

4. Methods and solutions

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 a 2D space. 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 few decades.

Table 3. Sensor placement constraints (Chen and Li 2004).

Satisfaction	Constraint
G1	Visibility
G2	Viewing angle
G3	Field of view
G4	Resolution constraint
G5	In-focus or viewing distance
G6	Overlap
G7	Occlusion
G8	Image contrast (affect (*d, f, a*) settings)
G9	Kinematic reachability of sensor pose

Here the review scope is restricted to some common methods and solutions found in recently published contributions regarding view-pose determination and sensor parameter setting in robotics. It does not include: foveal sensing, hand–eye coordination, autonomous vehicle control, landmark identification, qualitative navigation, path following operation, etc., although these are also issues concerning the active perception problem. We give little consideration to contributions on experimental study (Treuillet et al. 2007), sensor simulation (Wu et al. 2005; Loniot et al. 2007), interactive modeling (Popescu et al. 2004), and semi-automatic modeling (Liu and Heidrich 2003) either.

For the methods and solutions listed in the following, they might be used independently, or as hybrids, in the above-mentioned applications and tasks.

4.1. Formulation of constraints

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., who highlighted the sensor placement problem, detailed descriptions of the acceptable viewpoints for satisfying many requirements (sensing constraints) have to be provided. Tarabanis et al. presented approaches to compute the viewpoints that satisfy many sensing constraints, i.e. resolution, focus, field of view, and detectability (Tarabanis et al. 1994, 1995, 1996). Abrams et al. also proposed to compute the viewpoints that satisfy the constraints of resolution, focus (depth of field), field of view, and detectability (Abrams et al. 1999).

A complete list of constraints is summarized and analyzed by Chen and Li (2004). An admissible viewpoint should satisfy as many as nine placement constraints, including the geometrical (G1, G2, G6), optical (G3, G5, G8), reconstructive (G4, G6), and environmental (G9) constraints. These are listed in Table 3. Figure 10 intuitively illustrates several constraints (G1, G2, G3, G5, G7). Considering the six points (A–F) on the object surface, it can be seen in the figure that only point A satisfies all five constraints, while all other points violated one or more of the constraints.

The formulation of perception constraints is mostly used in model-based vision tasks (Trucco et al. 1997), such as inspection, assembly/disassembly, recognition, and object

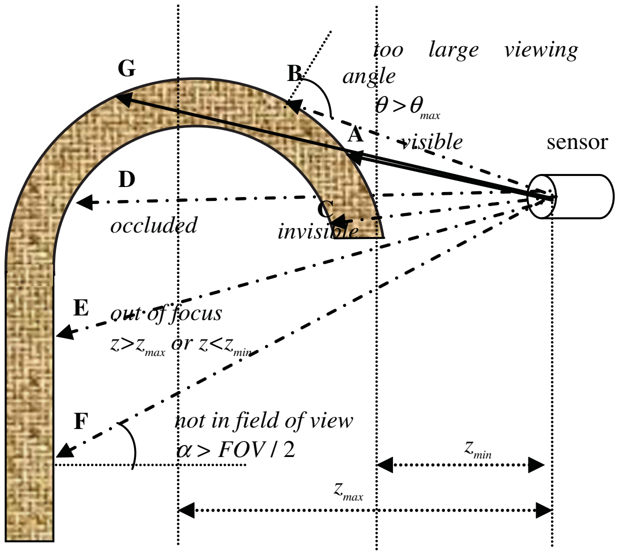

Fig. 10. Illustration of sensor placement constraints (Chen and Li 2004 (© IEEE 2004)).

search (Tarabanis et al. 1995), but a similar formulation is also valid in non-model-based tasks (Chen and Li 2005; Chen et al. 2008a). For the autonomous selection and modification of camera configurations during tasks, Chu and Chung consider both the camera's visibility and the manipulator's manipulability. The visibility constraint guarantees that the whole of a target object can be 'viewed' with no occlusions by the surroundings, and the manipulability constraint guarantees avoidance of the singular position of the manipulator and rapid modification of the camera position. The optimal camera position is determined and the camera configuration is modified such that visual information for the target object can be obtained continuously during the execution of assigned tasks (Chu and Chung 2002).

4.1.1. Cost functions Traditionally for sensor planning, a weighted function is often used for objective evaluation. It includes several components standing for placement constraints. For the object model, the NBV was defined as the next sensor pose which would enable the greatest amount of previously unseen three-dimensional information to be acquired (Banta et al. 2000; Li and Liu 2005). Tarabanis et al. chose to formulate the probing strategy as a function minimization problem (Tarabanis et al. 1995). The optimization function is given as a weighted sum of several component criteria, each of which characterizes the quality of the solution with respect to an associated requirement separately. The optimization function is written as

$$h = \max(\alpha_1 g_1 + \alpha_2 g_2 + \alpha_3 g_3 + \alpha_4 g_4) \qquad (6)$$

subject to $gi \geq 0$, to satisfy four constraints, i.e. the resolution, focus, field of view, and visibility.

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

$$V_{placement} = V_{g1} \bigcap V_{g2} \bigcap V_{g3} \bigcap V_{g4} \tag{7}$$

Marchand and Chaumette (1999a) took three factors into account in the strategy of viewpoint selection: (i) the new observed area volume $G(\phi_{t+1})$, (ii) the cost function F in order to reduce the total camera displacement $C(\phi_t, \phi_{t+1})$, and (iii) constraints to avoid unreachable viewpoints and to avoid positions near the robot joint limits $B(\phi)$. The cost function F_{next} to be minimized is defined as a weighted sum of the different measures:

$$F(\phi_{t+1}) = A(\phi) + a_1 g(\phi_{t+1}) + a_2 C(\phi_t, \phi_{t+1}) + a_3 B(\phi) \tag{8}$$

Ye and Tsotsos considered the total cost of object search via a function (Ye and Tsotsos 1999):

$$T[F] = \sum_{i=1}^{k} t_o(f_i) \tag{9}$$

where the cost $to(f)$ gives the total time needed to manipulate the hardware to the status specified by f, to take a picture, to update the environment and register the space, and to run the recognition algorithm. The effort allocation $F = (f_1, ..., f_k)$ gives the ordered set of operations applied in the search.

Chen and Li defined a criterion of lowest traveling cost according to the task execution time

$$T_{\cos t} = (T_1 + T_2)\,n + l_c \kappa \tag{10}$$

where T_1 and T_2 are constants reflecting the time for image digitalization, image preprocessing, 3D surface reconstruction, fusion and registration of partial models. n is the number of total viewpoints. k is the equivalent sensor moving speed. l_c is the total path length of robot operations, which is computed from the sensor placement graph (Chen and Li 2004).

4.1.2. Data driven In active perception, data-driven sensor planning makes sensing decisions according to local on-site data characteristics and to deal with environmental uncertainty (Whaite and Ferrie 1997; Miura and Ikeuchi 1998; Callari and Ferrie 2001; Bodor et al. 2007).

In model-based object recognition, SyedaMahmood presents an approach that uses color as a cue to perform selection either based solely on image data (data-driven), or based on the knowledge of the color description of the model (model-driven). The color regions extracted form the basis for performing data- and model-driven selection. Data-driven selection is achieved by selecting salient color regions as judged by a color-saliency measure that emphasizes attributes that are also important in human color perception. The approach to model-driven selection, on the other hand, exploits the color and other regional information in the 3D model object to locate instances of the object in a given image. The approach presented tolerates some of the problems of occlusion, pose, and illumination changes that make a model instance in an image appear different from its original description (SyedaMahmood 1997).

Mitsunaga and Asada investigated how a mobile robot selected landmarks to make a decision based on an information criterion. They argue that observation strategies should not only be for self-localization but also for decision making. An observation strategy is proposed to enable a robot equipped with a limited viewing angle camera to make decisions without self-localization. A robot can make a decision based on a decision tree and on prediction trees of observations constructed from its experiences (Mitsunaga and Asada 2006).

4.2. Expectation

Local surface features and expected model parameters are often used in active sensor planning for shape modeling (Flandin and Chaumette 2001). A strategy developed by Jonnalagadda et al. is to select viewpoints in four steps: local surface feature extraction, shape classification, viewpoint selection, and global reconstruction. When 2D and 3D surface features are extracted from the scene, they are assembled into simple geometric primitives. The primitives are then classified into shapes, which are used to hypothesize the global shape of the object and plan the next viewpoints (Jonnalagadda et al. 2003).

In purposive shape reconstruction, the method adopted by Kutulakos and Dyer is based on a relation between the geometries of a surface in a scene and its occluding contour: If the viewing direction of the observer is along a principal direction for a surface point whose projection is on the contour, surface shape (i.e. curvature) at the surface point can be recovered from the contour. They use an observer that purposefully changes viewpoint in order to achieve a well-defined geometric relationship with respect to a 3D shape prior to its recognition. The strategy depends on only curvature measurements on the occluding contour (Kutulakos and Dyer 1994).

Chen and Li developed a method by analyzing the target's trend surface, which is the regional feature of a surface for describing the global tendency of change. While previous approaches to trend analysis usually focused on generating polynomial equations for interpreting regression surfaces in three dimensions, they propose a new mathematical model for predicting the unknown area of the object surface. A uniform surface model is established by analyzing the surface curvatures. Furthermore, a criterion is defined to determine the exploration direction, and an algorithm is developed for determining the parameters of the next view (Chen and Li 2005).

On the other hand, object recognition does also obviously need to analyze local surface features. The appearance of

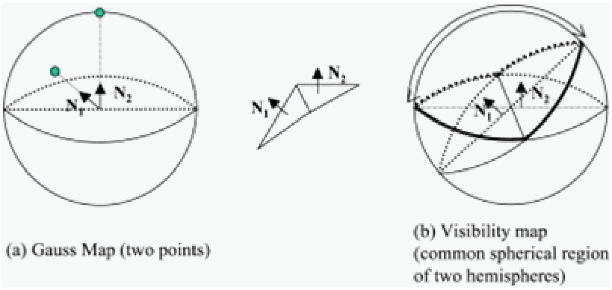

Fig. 11. The Gauss map and visibility map for scanning planning (Chang and Park 2009 (with permission of Elsevier)).

an object from various viewpoints is described in terms of visible 2D features, which are used for feature search and viewpoint decisions (Kuno et al. 1991).

4.2.1. Visibility A feature point target must be visible and not occluded in a robotic vision system (Briggs and Donald 2000; Chu and Chung 2002). It is essential for real-time robot manipulation in cluttered environments (Zussman et al. 1994; Jang et al. 2008), or adaptation to dynamic scenes (Fiore et al. 2008). Recognition of time-varying geometrical objects or subjects needs to maximize the visibility in a dynamic environment (Mackay and Benhabib 2008b).

As some missing areas may be found from the initial scans of an object (Fernandez et al. 2008), algorithms can be developed to compute additional scanning orientations (Chang and Park 2009). The algorithm by Chang and Park was designed by considering three major technological requirements of the problem: dual visibility, reliability, and efficiency. To satisfy the dual visibility requirement for the structured light vision sensor, the algorithm uses the concept of a visibility map as well as the diameter of a spherical polygon. Once dual visibility is satisfied, the algorithm attempts to locate the optimal scanning orientation to maximize the reliability. For a surface, the visibility map can be derived from a Gauss map, which is the intersection of the surface normal vectors and the unit sphere (Figure 11).

A model-based visibility measure for geometric primitives is called a visibility map. It is simple to calculate, memory efficient, accurate for viewpoints outside the convex hull of the object and versatile in terms of possible applications (Ellenrieder et al. 2005a). A global visibility map is a spherical image built to describe the complete set of global visible view directions for a surface. For the computation of global visibility maps, Liu et al. (2009) considered both the self-occlusions introduced by a region and the global occlusions introduced by the rest of the surfaces on the boundary of the polyhedron. The occluded view directions introduced between a pair of polyhedral surfaces can be computed from the spherical projection of the Minkowski sum of one surface and the reflection of the other. A suitable subset of the Minkowski sum, which shares the identical spherical projection with the complete Minkowski sum, is constructed to obtain the spherical images representing global occlusions (Liu et al. 2009).

4.2.2. Coverage, occlusion, and tessellation The sensor coverage problem for locating sensors in 2D can be modeled as an art gallery problem or museum problem (Bottino and Laurentini 2006a; Bottino et al. 2007; Bottino and Laurentini 2008). It originates from a real-world problem of guarding an art gallery with the minimum number of guards which together can observe the whole gallery. In the computational geometry, the layout of the art gallery is represented by a simple polygon and each guard is represented by a point in the polygon. A set S of points is said to guard a polygon if, for every point p in the polygon, there is some $q \in S$ such that the line segment between p and q does not leave the polygon.

The decision problem versions of the art gallery problem and all of its standard variations are NP complete. Regarding approximation algorithms, Eidenbenz et al. proved the problem to be APX hard, implying that it is unlikely that any approximation ratio better than some fixed constant can be achieved by a polynomial time approximation algorithm. Avis and Toussaint proved that a placement for these guards may be computed in $O(n \log n)$ time in the worst case, via a divide and conquer algorithm.

Recently, Nilsson et al. formulated the ‘minimum wall guard problem’ as follows.

Problem: Let $W = [(p_i, q_i) : p_i, q_i \in R_2]$ be a set of line segments corresponding to the walls that needs to be surveyed. Furthermore, let $O \subset R_2$ be the union of all obstacles. The problem is to find a minimum set $S \subset R_2$ of points on the ground plane such that every wall w_i in W is *guarded* by a point s_j in S. By *guarded* it means that s_j and w_i satisfy the constraints of visibility, resolution, and field of view.

They further proposed the following algorithm to find a solution.

Algorithm: (1) Find the candidate guard set S. (2) Calculate the walls guarded by each $s \in S$, using the three constraints. (3) Transcribe the problem of finding a subset of S that guards all walls W to a minimum set cover problem. (4) Solve the problem using a greedy approach.

In the algorithm, since the original problem is NP complete, they do not seek to find the true optimal set of guard positions. Instead, a near optimal subset of the candidate points is chosen with a known approximation ratio of $O(\log(n))$ (Nilsson et al. 2009).

As the art gallery problem is a well-studied visibility problem in computational geometry, many other solutions may be taken directly for visual sensor placement. Recently, a lower bound for the cardinality of the optimal covering solution, specific of a given polygon, has been proposed. It allows one to assess the performances of approximate

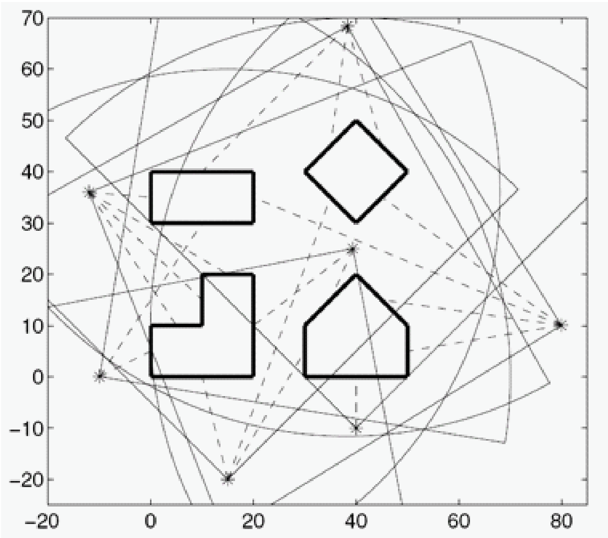

Fig. 12. A complex scenario with 19 walls to be guarded. The solution requires seven guards to guard all walls while satisfying occlusion, resolution and field of view constraints (with kind permission from Springer Science + Business Media: Nilsson U, Ogren P and Thunberg J (2009) Towards optimal positioning of surveillance UGVs. In *8th International Conference on Cooperative Control and Optimization*, pp. 221–233).

sensor location algorithms. It can be computed in reasonable time for environments with up to a few hundred edges (Bottino et al. 2009).

An example of a complex scenario is depicted in Figure 12. Seven UGVs are required to cover the 19 walls of the four buildings. Note that although no explicit obstacles are present, the buildings themselves serve as obstacles occluding the view of the UGVs.

However, if the problem is in three dimensions, then putting a guard at each vertex will not ensure that all of the museums are under observation. Although all of the surface of the polyhedron would be surveyed, for some polyhedra there are points in the interior which might not be under surveillance.

To determine minimal orthographic view covers for polyhedra, a global visibility map-based method was developed by Liu and Ramani (2009) to calculate an optimal or near-optimal solution using object space segmentation and viewpoint space sampling. The viewpoint space is sampled using a generate-as-required heuristic. The problem is then modeled as an instance of the classical set-cover problem and solved using a minimal visible set based branch-and-bound algorithm.

Coverage is also a cue in image selection for multi-view 3D sensing (Hornung et al. 2008), urban driving (Seo and Urmson 2008), multi-agent sensor planning (Bardon et al. 2004), and model acquisition session planning. Impoco et al. propose a solution to improve the coverage of automatically acquired objects. Rather than searching for the NBV in order to minimize the number of acquisitions, they propose a simple and easy-to-implement algorithm limiting our scope to closing gaps (i.e. filling unsampled regions) in roughly acquired models. The idea is to detect holes in the current model and cluster their estimated normals in order to determine new views (Impoco et al. 2004).

While most existing camera placement algorithms focus on coverage and/or visibility analysis, Yao et al. recently argued that visibility is insufficient for automated persistent surveillance. In some applications, a continuous and consistently labeled trajectory of the same object should be maintained across different camera views. Therefore, a sufficient uniform overlap between the cameras' fields of view should be secured so that camera handoff can successfully and automatically be executed before the object of interest becomes untraceable or unidentifiable. They propose sensor-planning methods that improve existing algorithms by adding handoff rate analysis and preserve necessary uniform overlapped fields of view between adjacent cameras for an optimal balance between coverage and handoff success rate (Lim et al. 2006; Yao et al. 2010).

There is a constraint in sensor planning that has not been thoroughly investigated in the literature, namely, visibility in the presence of random occluding objects (Mittal and Davis 2004, 2008). Such visibility analysis provides important performance characterization of multi-camera systems. Furthermore, maximization of visibility in a given region of interest yields the optimum number and placement of cameras in the scene. Mittal and Davis presented such primary contributions.

Although several factors contribute, occlusion due to moving objects within the scene itself is often the dominant source of tracking error. Chen and Davis introduced a configuration quality metric based on the likelihood of dynamic occlusion. Since the exact geometry of occluders cannot be known *a priori*, they use a probabilistic model of occlusion (Chen and Davis 2008).

There is another distinctive method used frequently for object modeling, i.e. spatial tessellation. Usually it tessellates a sphere or cylinder around the object to be modeled as a viewpoint space (MacKinnon et al. 2008a), look-up array, or grid maps (Se and Jasiobedzki 2007). 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.

4.2.3. Geometrical analysis Direct geometrical analysis is the most fundamental way in solving computer vision problems. For example, the robot configuration space (C-space) is adopted with pure geometric criteria (Wang and Gupta

2006). For simultaneous tracking of multiple moving targets using an active stereo, Barreto et al. propose to control the active system parameters in such a manner that the images of the targets in the two views are related by a homography. This homography is specified during the design stage and, thus, can be used to implicitly encode the desired tracking behavior. Such formulation leads to an elegant geometric framework that enables a systematic and thorough analysis of the problem. In the case of using two pan–tilt–zoom (PTZ) cameras with rotation and zoom control, it is proved that such a system can track up to three free-moving targets, while assuring that the image location of each target is the same for both views. If considering a robot head with neck pan motion and independent eye rotation, it is not possible to track more than two targets because of the lack of zoom (Barreto et al. 2010).

For optimal sensor placement in a surveillance region with a minimal cost, the problem is solved by Sivaram et al. by obtaining a performance vector, with its elements representing the performances of subtasks, for a given input combination of sensors and their placement. Then the optimal sensor selection problem can be converted into the form of Integer Linear Programming problem. The optimal performance vector corresponding to the sensor combination **n** (m-dimensional) is given by

$$\mathbf{P}^* = \mathbf{A} \times \mathbf{n} \tag{11}$$

where **A** is related to the performance matrix ($l \times m$) which is organized from the sensor types and surveillance subtasks. The performance constraints can be written as

$$\mathbf{A} \times \mathbf{n} \geq \mathbf{b} \tag{12}$$

where **b** is the required performance, an l-dimensional vector. To demonstrate the utility of our technique, a surveillance system is introduced which consists of PTZ cameras and active motion sensors for capturing faces (Sivaram et al. 2009).

In a robot motion-planning algorithm, Han et al. proposed to capture a moving object precisely using the single curvature trajectory. With the pre-determined initial states (i.e. position and orientation of the mobile robot and the final states), the mobile robot is made to capture a moving object (Han et al. 2008).

4.2.4. Volumetric space Out of the existing approaches, volumetric computation by region intersection is frequently used by researchers since the early stages (Cowan and Kovesi 1988). For example, it computes the region R_i of acceptable viewpoints for each constraint. If multiple surface features need to be inspected simultaneously, the region R_i is the intersection of the acceptable regions R_{ij} for each individual feature. Finally, the region of acceptable viewpoints is the intersection of all regions (Figure 13).

For scene reconstruction and exploration (Lang and Jenkin 2000), the quality of a new position ϕ_{i+1} is defined by the volume of the unknown regions that appear in the field of view of the camera (Marchand and Chaumette 1999a). The new observed region $G(\phi_{i+1})$ is given by

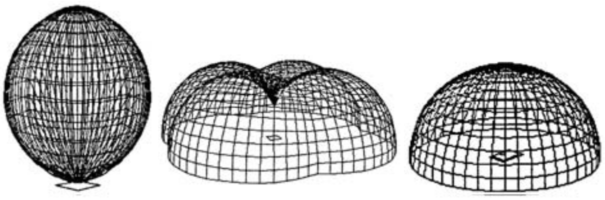

Fig. 13. The volumes of resolution, depth-of-field, and field-of-view constraints (Cowan and Kovesi 1988 (© IEEE 1988)).

$$G(\phi_{i+1}) = V(\phi_{i+1}) - V(\phi_{i+1}) \cap V(T_0^t) \tag{13}$$

where $V(\phi_{i+1})$ defines the part of the scene observed from the position ϕ_{i+1} and $V(\phi_{i+1}) \cap V(T_0^t)$ defines the subpart of $V(\phi_{i+1})$ that has already been observed.

Martins et al. presented a method to automate the process of surface scanning using optical range sensors and based on *a priori* known information from a CAD model. A volumetric model implemented through a 3D voxel map is generated from the object CAD model and used to define a sensing plan composed of a set of viewpoints and the respective scanning trajectories. Surface coverage with high data quality and scanning costs are the main issues in sensing plan definition (Martins et al. 2003, 2005).

Bottino and Laurentini presented a general approach to interactive, object-specific volumetric algorithms, based on a condition for the best possible reconstruction. The approach can be applied to any class of objects. As an example, an interactive algorithm is implemented for convex polyhedra (Bottino and Laurentini 2006b).

4.3. Multi-agent approach

4.3.1. Cooperative network Consider a mobile cooperative network that is given the task of building a map of the spatial variations of a parameter of interest, such as an obstacle map or an aerial map. Mostofi and Sen proposed a framework that allows the nodes to build a map with a small number of measurements. By compressive sensing, they studied how the nodes can exploit the sparse representation in the transform domain in order to build a map with minimal sensing (Mostofi and Sen 2009).

The surveillance of a maneuvering target with multiple sensors in a coordinated manner requires a method for selecting and positioning groups of sensors in real time (Naish et al. 2003). Heuristic rules are used to determine the composition of each sensor group by evaluating the potential contribution of each sensor. In the case of dynamic sensors, the position of each sensor with respect to the target is specified. The approach aims to improve the quality of the surveillance data in three ways: (i) the assigned sensors are maneuvered into 'optimal' sensing positions, (ii) the uncertainty of the measured data is mitigated through sensor fusion, and (iii) the poses of the unassigned sensors are adjusted to ensure that the surveillance system can react to future object maneuvers. If *a priori* target trajectory information is available, the system performance may be further improved by optimizing the initial pose of each sensor off-line.

As a single sensor system would not provide adequate information for a given sensor task, it is necessary to incorporate multiple sensors in order to obtain complete information. Hodge and Kamel presented an automated system for multiple sensor placement based on the coordinated decisions of independent, intelligent agents. The overall goal is to provide the surface coverage necessary to perform feature inspection on one or more target objects in a cluttered scene. This is accomplished by a group of cooperating intelligent sensors. In the system, the sensors are mobile, the target objects are stationary and each agent controls the position of a sensor and has the ability to communicate with other agents in the environment. By communicating desires and intentions, each agent develops a mental model of the other agents' preferences, which is used to avoid or resolve conflict situations (Hodge and Kamel 2003).

Bakhtari et al. developed another agent-based method for the dynamic coordinated selection and positioning of active-vision cameras for the simultaneous surveillance of multiple objects as they travel through a cluttered environment with unknown trajectories. The system dynamically adjusts the camera poses in order to maximize the system's performance by avoiding occlusions and acquiring images with preferred viewing angles (Bakhtari et al. 2006; Bakhtari and Benhabib 2007).

In other aspects, camera networking by dynamic programming is addressed by Lim et al. (2007). Issues of scalability and flexibility of multiple sensors are studied by Hodge et al. (2004). Cooperative localization using relative bearing constraints is used for error analysis by Taylor and Spletzer (2007).

4.3.2. Fusion When multiple optical sensors such as stereo vision, a laser-range scanner and a laser-stripe profiler are integrated into a multi-purpose vision system, fusion of range data into a consistent representation is necessary to allow for safe path planning and view planning. Suppa and Hirzinger dealt with such 3D sensor synchronization and model generation (Suppa and Hirzinger 2007). Cohen and Edan presented a sensor fusion framework for selecting online the most reliable logical sensors and the most suitable algorithm for fusing sensor data in a robot platform (Cohen and Edan 2008).

Visual sensors provide exclusively uncertain and partial knowledge of a scene. A suitable scene knowledge representation is useful to make integration and fusion of new, uncertain, and partial sensor measures possible. Flandin and Chaumette develop a method based on a mixture of stochastic and set membership models. Their approximated representation mainly results in ellipsoidal calculus by means of a normal assumption for stochastic laws and ellipsoidal over or inner bounding for uniform laws. These approximations allow us to build an efficient estimation process integrating visual data online. Based on this estimation scheme, optimal exploratory motions of the camera can be automatically determined (Flandin and Chaumette 2002).

While wide-area video surveillance is an important application, it is sometimes not practical to have video cameras that completely cover the entire region of interest. For obtaining good surveillance results in a sparse camera networks, it requires that they be complemented by additional sensors with different modalities, their intelligent assignment in a dynamic environment, and scene understanding using these multimodal inputs. Nayak et al. propose a probabilistic scheme for opportunistically deploying cameras to the most interesting parts of a scene dynamically given data from a set of video and audio sensors. Events are tracked continuously by combining the audio and video data. Correspondences between the audio and video sensor observations are obtained through a learned homography between the image plane and ground plane (Bakhtari et al. 2006; Nayak et al. 2008).

For 3D tracking, Chen and Li proposed a method to fuse sensing data of the most current observation into a 3D visual tracker with particle techniques. The importance density function in particle filter can be modified to represent posterior states by particle crowds in a better way. Thus, it makes the tracking system more robust to noise and outliers (Chen and Li 2008).

For vehicle localization, data fusion from GPS and machine vision is proposed by Rae and Basir (2009). Data association is needed to identity the detected objects, and to identity the road driven by the vehicle. For this purpose they employ multiple hypothesis tracking to consider multiple data association hypotheses simultaneously. Results show that using machine vision improves the localization accuracy and helps in the identification of the road being driven by the vehicle.

4.4. Statistical approaches

4.4.1. Probability and entropy Statistics, probability, Kalman filters, and associative Markov networks have been widely used in active object recognition (Wheeler and Ikeuchi 1995; Dickinson et al. 1997; Roy et al. 2000; Caglioti 2001), grasping (Motai and Kosaka 2008), and modeling (Triebel and Burgard 2008). In the research of multi-camera solutions, Farshidi et al. investigated the feasibilities of recognition algorithms to classify the object if its pose can be determined with a high confidence level, by processing the available information within a recursive Bayesian framework at each step. Otherwise, the algorithms compute the next most informative camera positions for capturing more images. The principle component analysis (PCA) is used to produce a measurement vector based on the acquired images. Occlusions in the images are handled by a probabilistic modeling approach that can increase the robustness of the recognition process with respect to structured noise. The camera positions at each recognition step are selected based on two statistical metrics regarding the quality of the observations, namely the mutual information (MI) and the Cramer–Rao lower bound (CRLB) (Farshidi et al. 2009). For the state *sn* being the variable of interest, the MI is a measure of the reduction in the uncertainty in *s*n due to the observation *g* and is defined as

$$I(s^n; g|a_n) = H(s^n|a_n) - H(s^n|g, a_n) \tag{14}$$

where a_n is the vector of camera positions, g is the observation vector, and $H(\ldots)$ is the entropy function defined in (16). CRLB is computed by

$$C_n = E\{(\hat{s}^n - s^n)(\hat{s}^n - s^n)^T\} \geq J_n^{-1} \tag{15}$$

where *Jn* is the Fisher information matrix.

Significant improvement is observed in the success rates of both MI-based and CRLB-based approaches. This enhancement was gained by incorporating a model of occlusion into the algorithms. The recognition rate in experiments without occlusion modeling is 48–50%, and improved to 98% if with occlusion modeling.

Borotschnig and Paletta also presented an active vision system for recognizing objects which are ambiguous from certain viewpoints (Borotschnig and Paletta 2000). 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 of attributing a score to each possible movement of the camera. The movement obtaining the highest score will be selected next. 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:

$$\Delta\psi_{n+1} := \arg\max s_n(\Delta\psi) \tag{16}$$

In sensor planning for object search, each robot action is defined by a viewpoint, a viewing direction, a field of view, and the application of a recognition algorithm. Ye and Tsotsos formulate it as an optimization problem: the goal is to maximize the probability of detecting the target with minimum cost. Since this problem is proved to be NP complete, in order to efficiently determine the sensing actions over time, the huge space of possible actions with fixed camera position is decomposed into a finite set of actions that must be considered. The next action is then selected from among these by comparing the likelihood of detection and the cost of each action. When detection is unlikely at the current position, the robot is moved to another position where the probability of target detection is the highest (Ye and Tsotsos 1999).

The Shannon entropy was also applied to the problem of automatic selection of light positions in order to automatically place light sources for maximum visual information recovery (Vazquez 2007).

The 3D site modeling of Wenhardt et al. (2007) is based on a probabilistic state estimation with sensor actions. The next best view is determined by a metric of the state estimation's uncertainty. Three metrics are addressed: D-optimality, which is based on the entropy and corresponds to the determinant of the covariance matrix of a Gaussian distribution, E-optimality, and T-optimality, which are based on eigenvalues or on the trace of matrices, respectively.

The entropy $H(\mathbf{q})$ of a probability distribution $p(\mathbf{q})$ is defined as (Li and Liu 2005; Farshidi et al. 2009)

$$H(\mathbf{q}) = \int p(\mathbf{q}) \log p(\mathbf{q})\, d\mathbf{q}. \tag{17}$$

For an n-dimensional Gaussian distribution, the entropy can be calculated in a closed form:

$$H(\mathbf{q}) = \frac{1}{2}[n + \log(2\pi^n|\mathbf{P}|)]. \tag{18}$$

The entropy depends only on the covariance $\mathbf{P}$ and the expected covariance is independent of the next observations. This allows us to use the entropy as an optimality criterion for sensor planning (Wenhardt et al. 2007).

For sensor-based robot motion planning, the robot plans the next sensing action to maximally reduce the expected C-space entropy, called the maximal expected entropy reduction (MER) criterion. From a C-space perspective, MER criterion consists of two important aspects: sensing actions are evaluated in C-space (geometric aspect); these effects are evaluated in an information theoretical sense (stochastic aspect). Wang and Gupta investigate how much of the performance is attributable to the paradigmatic shift to evaluating the sensor action in C-space and how much to the stochastic aspect, respectively (Wang and Gupta 2006).

In an intelligent and efficient strategy for unstructured environment sensing using mobile robot agents, a metric is derived from Shannon's information theory to determine optimal sensing poses (Sujan and Meggiolaro 2005). The map is distributed among the agents using an information-based relevant data reduction scheme. The method is particularly well suited to unstructured environments, where sensor uncertainty is significant. In their other contributions for site modeling and exploration, in addition to using Shannon's information theory to determine optimal sensing poses, the quality of the information in the model is used to determine the constraint-based optimum view for task execution. The algorithms are applicable for both an individual agent as well as multiple cooperating agents (Sujan and Dubowsky 2005b). The NBV is found by fusing a Kalman filter in the statistical uncertainty model with the measured environment map (Sujan and Dubowsky 2005a).

4.4.2. Bayesian reasoning Bayesian reasoning and classification methods are used in active perception for object recognition, search, surface reconstruction, and object modeling (Carrasco and Mery 2007; Mason 1997). Sutton and Stark also applied function-based reasoning for goal-oriented image segmentation (Sutton and Stark 2008) and Zhang et al. proposed a Bayesian network approach to sensor modeling and path planning (Zhang et al. 2009).

Bayesian inference is statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true. Bayes' theorem adjusts probabilities given new evidence in the following way:

$$H(H|E) = \frac{P(E|H)\,P(H)}{P(E)} \tag{19}$$

where H represents a specific hypothesis, which may or may not be some null hypothesis. Here $P(H)$ is called the prior probability of H that was inferred before new evidence, E, became available. We call $P(E|H)$ the conditional probability of seeing the evidence E if the hypothesis H happens to be true. It is also called a likelihood function when it is considered as a function of H for fixed E. We call $P(E)$ the marginal probability of E: the *a priori* probability of witnessing the new evidence E under all possible hypotheses.

For active recognition, the probability distribution of object appearance is described by multivariate mixtures of Gaussians which allows the representation of arbitrary object hypotheses (Eidenberger et al. 2008). In a statistical framework, Bayesian state estimation updates the current state probability distribution based on a scene observation which depends on the sensor parameters. These are selected in a decision process which aims at reducing the uncertainty in the state distribution (Eidenberger et al. 2008). For online recognition and pose estimation of a large isolated 3D object, Roy et al. propose a probabilistic reasoning framework for recognition and next-view planning (Roy et al. 2005).

Kristensen et al. proposed the sensor planning approach 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 (Kristensen 1997). Li and Liu adopted a B-spline for modeling the freeform surface. In the framework of Bayesian statistics for determining the probing points for efficient measurement and reconstruction, they developed a model selection strategy to obtain an optimal model structure for the freeform surface. In order to obtain reliable parameter estimation for the B-spline model, they analyzed the uncertainty of the model and used the statistical analysis of the Fisher information matrix to optimize the locations of the probing points needed in the measurements (Li and Liu 2003).

4.4.3. Hypothesis and verification The method of 'observation, modeling, hypothesis, and verification' is powerful for 3D model matching. In a semi-automated excavation system, the 3D object localization method used consists of three steps (Maruyama et al. 2010): (i) candidate regions are extracted from a range image obtained by an area-based stereo-matching method; (ii) for each region, multiple hypotheses for the position and orientation are generated for each object model; (iii) each hypothesis is verified and improved by an iterative method. The operator verifies the object localization results and then selects one of the objects as the best object that is suitable for grasping by the robot. The robot grasps objects based on the object localization result (Maruyama et al. 2010).

For path planning, a strategy by Nabbe and Hebert (2007) is based on a 'what-if' analysis of hypothetical future configurations of the environment. Candidate sensing positions are evaluated based on their ability to observe anticipated obstacles.

Hypothesis and verification is also used in viewpoint planning for 3D model reconstruction by Jonnalagadda et al. (2003) and Marchand and Chaumette (1999b). Jonnalagadda et al. presented a strategy to select viewpoints for global 3D reconstruction of unknown objects. The NBV is chosen to verify the hypothesized shape. If the hypothesis is verified, some information about global reconstruction of a model can be stored. If not, the data leading up to this viewpoint is re-examined to create a more consistent hypothesis for the object shape. The NBV algorithm uses only the local geometric features of an object and the visibility constraint is not used in the function to compute next viewpoint (Jonnalagadda et al. 2003).

To perform the complete and accurate reconstruction of 3D static scenes, Marchand and Chaumette used a structure from controlled motion method. As the method is based on particular camera motions, perceptual strategies able to appropriately perform a succession of such individual primitive reconstructions are proposed in order to recover the complete spatial structure of the scene. Two algorithms are suggested to ensure the exploration of the scene. The former is an incremental reconstruction algorithm based on the use of a prediction/verification scheme managed using decision theory and Bayes nets. It allows the visual system to get a high-level description of the observed part of the scene. The latter, based on the computation of new viewpoints, ensures the complete reconstruction of the scene (Marchand and Chaumette 1999b).

4.5. Soft and intelligent computation

4.5.1. Learning and expert system Interactive learning (Grewe and Kak 1995) or reinforcement learning (Kollar and Roy 2008) is frequently used for active recognition, localization, planning, and modeling (Wang et al. 2008). For example, inter-image statistics can be used for 3D environment modeling (Torres-Mendez and Dudek 2008). An expert knowledge-based sensor planning system was developed for car headlight lens inspection by Martinez et al. (2008).

For active viewpoint selection for object recognition, Deinzer et al. attempted an unsupervised reinforcement learning algorithm for modeling of continuous states, continuous actions, sequential fusion of gathered image information, and supporting rewards for an optimized recognition. The combined viewpoint selection and viewpoint fusion approach is to improve the recognition rates (Deinzer et al. 2009). Roy et al. attempted probabilistic reasoning for recognition of an isolated 3D object. Both the probability calculations and the next view planning have the advantage that the 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 et al. 2000).

For robotic real-time localization with a single camera and natural landmarks, Kwok employed an evolutionary computing approach (Kwok 2006) in the SLAM context to build a map simultaneously. Royer et al. gave a three-step approach. In a learning step, the robot is manually guided on a path and a video sequence is recorded with a front looking camera. Then a structure from motion algorithm is used to build a 3D map from this learning sequence. Finally in the navigation step, the robot uses this map to compute its localization in real time and it follows the learning path or a slightly different path if desired (Royer et al. 2007).

Consider the task of purposefully controlling the motion of an active, monocular observer in order to recover a global description of a smooth, arbitrarily-shaped object. Kutulakos and Dyer formulate global surface reconstruction as the task of controlling the motion of the observer so that the visible rim slides over the maximal, connected, reconstructible surface regions intersecting the visible rim at the initial viewpoint. They develop basic strategies that allow reconstruction of a surface region around any point in a reconstructible surface region. These strategies control viewpoint to achieve and maintain a well-defined geometric relationship with the object's surface, rely only on information extracted directly from images, and are simple enough to be performed in real time. Global surface reconstruction is then achieved by (i) appropriately integrating these strategies to iteratively grow the reconstructed regions, and (ii) obeying four simple rules (Kutulakos and Dyer 1995).

Robots often use topological structures as a spatial representation for exploring unknown environments. A method was developed by Hubner and Mallot (2007) to scale a similarity-based navigation system (the view-graph model) to continuous metric localization. Instead of changing the landmark model, they embed the graph into the 3D pose space. Therefore, recalibration of the path integrator is only possible at discrete locations in the environment. The navigation behavior of the robot is controlled by a homing algorithm which combines three local navigation capabilities, obstacle avoidance, path integration, and scene-based homing. This homing scheme allows automated adaptation to the environment. It is further used to compensate for path integration errors, and therefore allows a robot to derive globally consistent pose estimates based on weak metric knowledge. It is tested to explore a large, open, and cluttered environment.

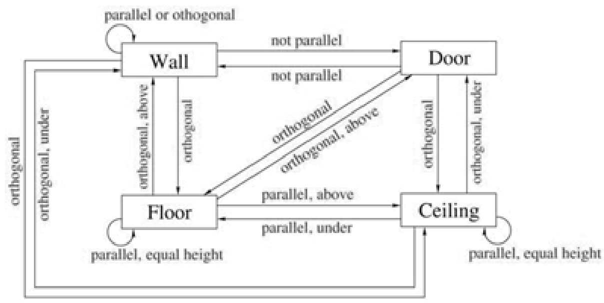

Fig. 14. Example of the constraint network with semantic mapping for scene interpretation (Nuchter and Hertzberg 2008 (with permission of Elsevier)).

Rule-based planning. Semantic maps and reasoning engines are useful in addition to geometry maps when the robot interacts with its environment in a goal-directed way. A semantic stance enables the robot to reason about objects; it helps disambiguate or round off sensor data; and the robot knowledge becomes reviewable and communicable. Nuchter and Hertzberg proposed an approach and an integrated robot system for semantic mapping. Coarse scene features are determined by semantic labeling. More delicate objects are then detected by a trained classifier and localized. Finally, the semantic maps can be visualized for inspection (Nuchter and Hertzberg 2008). Figure 14 shows an example of the object relationship for scene interpretation. The semantic mapping is performed through the following steps: (i) SLAM for acquiring 3D scans of the environment; (ii) scene interpretation by feature extracting and labeling; (iii) object detection for identification of known objects and their poses; and (iv) visualization of the semantic map.

4.5.2. Fuzzy and neural network Fuzzy inference and neural network are useful in sensor planning for prediction and recognition. Saadatseresht et al. solved automatic camera placement in vision metrology based on a fuzzy inference system (Saadatseresht et al. 2005). Martinez et al. recently proposed a methodology to include the inspection guideline in an automated headlamp lens inspection system. As the way in which the guideline includes the knowledge of an expert in the inspection of lenses is inherently qualitative and vague, a fuzzy rule-based system is developed to model this information (Martinez et al. 2009). Budiharto et al. used an adaptive neuro-fuzzy controller for servant robot indoor navigation (Budiharto et al. 2010).

Visibility uncertainty prediction was solved by an artificial neural network (ANN) by Saadatseresht and Varshosaz

(2007). For outdoor navigation, Shinzato et al. used ANN for path recognition (Shinzato et al. 2010).

4.5.3. Evolutionary computation The model-based sensor placement problem has been formulated as a nonlinear multi-constraint optimization problem as described in Section 4.2. It is difficult to compute robust viewpoints which satisfy all constraints. However, evolutionary computation is especially powerful in solving such problems. Chen and Li use a hierarchical genetic algorithm (GA) to determine the optimal topology in the sensor placements which will contain minimum number of viewpoints with the highest accuracy while satisfying all of the constraints. In the hierarchical chromosome, parametric genes represent the sensor poses and optical settings and control genes represent the topology of viewpoints. A plan of sensor placements is evaluated by a min–max criterion, which includes three objectives and a fitness evaluation formula (Chen and Li 2004). Similarly, a hybrid GA is used to solve the highly complicated optimization problem by Al-Hmouz and Challa (2005).

Although evolutionary computation is mostly used in model-based inspection (Olague 2002; Yang and Ciarallo 2001; Dunn and Olague 2003, 2004), the method has wider applications in active vision perception, e.g., path planning in assembly lines (Park et al. 2006) and monitoring (Sakane et al. 1995). Kang et al. applied the virus coevolutionary partheno-genetic algorithm (VEPGA), which combined a partheno-genetic algorithm (PGA) with virus evolutionary theory, for determining sensor placements (Kang et al. 2008).

4.6. Dynamic configuration

In an active vision system, since the robot needs to move from one place to another to perform a multi-view task, a traditional vision sensor with fixed structure is often inadequate for the robot to perceive the object features in an uncertain environment as the object distance and size are unknown before the robot sees it. A dynamically reconfigurable sensor can help the robot to control the configuration and gaze at the object surfaces. For example, with a structured light system, the camera needs to see the object surface illuminated by the projector, to perform the 3D measurement and reconstruction task. Active recalibration means that the vision sensor is reconfigurable during runtime to fit in the environment and can perform self-recalibration as required before visual perception (Chen et al. 2008a; Chu and Chung 2002).

In the literature, self-reconfiguration of automated visual inspection systems is addressed by Garcia and Villalobos (2007a, 2007b). Bakhtari et al. presented a reconfiguration method for the surveillance of an object as it travels through a multi-object dynamic workspace with unknown trajectory (Bakhtari et al. 2006, 2009).

In an environment of large scenes having large depth ranges with depth discontinuities, it is necessary to aim cameras in different directions and to fixate at different objects. An active approach is suggested by coarse-to-fine image acquisition by Das and Ahuja (1996), which involves the following steps. (i) A new fixation point is selected from among the non-fixated, low-resolution scene parts of current fixation. (ii) A reconfiguration of the cameras is initiated for re-fixation. As reconfiguration progresses, the images of the new fixation point are gradually deblurred and the accuracy of the position estimate of the point improves allowing the cameras to be aimed at it with increasing precision. (iii) The improved depth estimate is used to select focus settings of the cameras, thus completing fixation. Similarly, an active stereo head is implemented with visual behaviors by Krotkov and Bajcsy (1993), including functions of (i) aperture adjustment to vary depth of field and contrast, (ii) focus ranging followed by fixation, (iii) stereo ranging followed by focus ranging, and (iv) focus ranging followed by disparity prediction followed by focus ranging.

4.6.1. Gaze and attention Gaze and attention are important functions for a human to actively perceive in the environment, and as the same is true for robots. Visual perceptual capability starts with an early vision process that exhibits changes in visual sensitivity such as night vision and flash blindness under changing scene illumination. Visual attention directs the limited gaze resource to resolve visual competition with the cooperation of top-down attention and conspicuous bottom-up guidance. Grounded in psychological studies, it has four factors, i.e. conspicuity, mental workload, expectation and capacity, which determine successful attention allocation. For purposive perception, many devices and systems have been invented for robotics (Dickinson et al. 1997).

Active gaze control allows us to overcome some of the limitations of using a monocular system with a relatively small field of view. To implement active gaze control in SLAM, a system was addressed by Frintrop and Jensfelt (2008b), which specializes in creating and maintaining a sparse set of landmarks based on a biologically motivated feature-selection strategy. A visual attention system detects salient features that are highly discriminative and ideal candidates for visual landmarks that are easy to redetect. It supports (i) the tracking of landmarks that enable a better pose estimation, (ii) the exploration of regions without landmarks to obtain a better distribution of landmarks in the environment, and (iii) the active redetection of landmarks to enable close loop. It is concluded that active camera control outperforms the passive approach (Frintrop and Jensfelt 2008a).

Attention is often related to visual search. Consider the problem of visually finding an object in an unknown space. This is an optimization problem, i.e. optimizing the probability of finding the target given a fixed cost limit in terms of total number of robotic actions required to find the

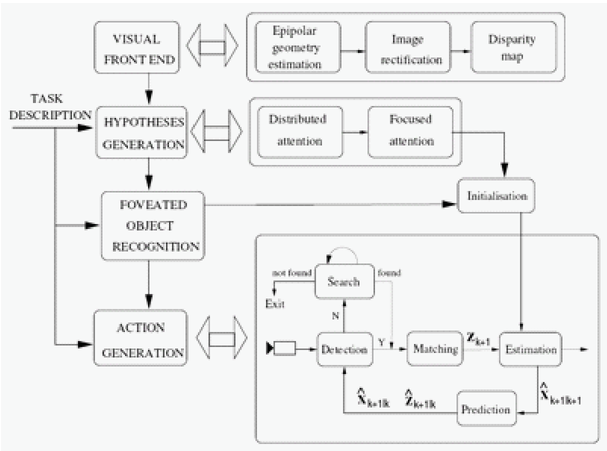

Fig. 15. The active vision system involving attention and gaze for action decision (Bjorkman and Kragic 2004 (© IEEE 2004)).

visual target. Shubina and Tsotsos present an approximate solution and investigate its performance and properties (Shubina and Tsotsos 2010).

With a pre-determined sensor lens, the system may be not able to deal with the scenes that have objects at different distances. Quang et al. presented a projector auto-focusing technique based on local blur information of the image that can overcome the above limitation. The algorithm is implemented on a projector–camera system, in order to focus the pattern which is projected by the projector on all objects in the scene sequentially. The proposed algorithm first obtains a blur map of the scene on the image by using a robust local blur estimator, and then the region of interest is decided by thresholding the obtained blur map. Since the main light source is provided by the projector, the proposed auto-focusing algorithm achieves a good performance with different light conditions (Quang et al. 2008).

With ego-motion (Shimizu et al. 2005), the robot is able to control the orientation of a single camera, while still allowing the robot to preview a wider area. In addition, controlling the orientation allows the robot to optimize its environment perception by only looking where the most useful information can be discovered (Radovnikovich et al. 2010).

Bjorkman and Kragic introduced a real-time vision system that consists of two sets of binocular cameras: a peripheral set for disparity-based attention and a foveal one for higher-level processes (Figure 15). Thus, the conflicting requirements of a wide field of view and high resolution can be overcome. The steps taken from task specification through object recognition to pose estimation are completely automatic, combining both appearance and geometric models. It was tested in a realistic indoor environment with occlusions, clutter, changing lighting and background conditions (Bjorkman and Kragic 2004).

4.6.2. Tagged roadmap Probabilistic roadmap methods are a class of randomized motion planning algorithms that have recently received considerable attention because they are capable of handling problems with many degrees of freedom, and large workspaces with many obstacles, for which other motion planning methods are computationally infeasible. Baumann et al. augments probabilistic roadmaps with vision-based constraints. The designed planner finds collision-free paths that simultaneously avoid occlusions of an image target and keep the target within the field of view of the camera (Baumann et al. 2008, 2010).

Another probabilistic roadmap method is presented for planning the path of a robotic sensor deployed in order to classify multiple fixed targets located in an obstacle-populated workspace (Zhang et al. 2009). Existing roadmap methods are not directly applicable to robots whose primary objective is to gather target information with an on-board sensor. In the proposed information roadmap, obstacles, targets, sensor's platform and field of view are represented as closed and bounded subsets of a Euclidean workspace. The information roadmap is sampled from a normalized information theoretic function that favors samples with a high expected value of information in the configuration space. The method is applied to a landmine classification problem to plan the path of a robotic ground-penetrating radar, based on prior remote measurements and other geospatial data. Results show that paths obtained from the information roadmap exhibit classification efficiency several times higher than that of other existing search strategies. Also, the information roadmap can be used to deploy non-overpass capable robots that must avoid targets as well as obstacles (Zhang et al. 2009; Oniga and Nedevschi 2010).

The research group of Allen et al. developed a system for automatic view planning called VuePlan. When combined with their mobile robot, AVENUE, the system is capable of modeling large-scale environments with minimal human intervention throughout both the planning and acquisition phases. The system proceeds in two distinct stages. In the initial phase, the system is given a 2D site footprint with which it plans a minimal set of sufficient and properly constrained covering views. It then uses a 3D laser scanner to take scans at each of these views. The planning system automatically computes and executes a tour of these viewing locations and acquires them with the robot's onboard laser scanner. These initial scans serve as an approximate 3D model of the site. The planning software then enters a second phase in which it updates this model by using a voxel-based occupancy procedure to plan the NBV (Blaer and Allen 2009). They have successfully used the two-phase system to construct precise 3D models of real-world sites located in New York City (Figure 16).

4.6.3. Solution for next best view problem A solution for the NBV problem is of particular importance for automated object modeling. Given a partial model of the target, we have to determine the sensor pose or scanning path to scan all of the visible surfaces of an unknown object. The solution to this problem would ideally allow the model to be

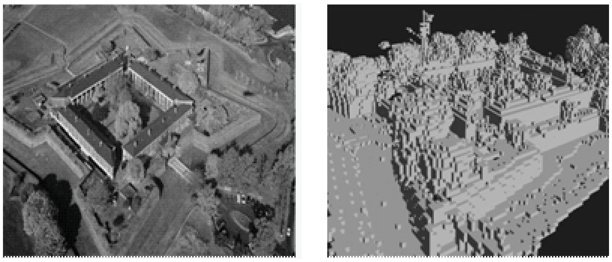

Fig. 16. 16 Complex site modeling by view planning with a footprint (Blaer and Allen 2009 (with permission of Wiley Blackwell)).

obtained from a minimum number of range images (Banta et al. 2000; Kim and Cho 2003; He and Li 2006a; Null and Sinzinger 2006; Blaer and Allen 2007; Huang and Qian 2008a, 2008b; Sun et al. 2008).

The NBV may be computed in two steps. First, the exploration direction for the next view is determined via a mass vector chain-based scheme. Then the accurate position of the next view is obtained by computing the boundary integral of the vectors fields. The position with the maximum integral value is selected as the NBV (Chen and Li 2005; Li et al. 2005a).

It is argued that solutions to the NBV problem are constrained by other steps in a surface acquisition system and by the range scanner's particular sampling physics. Another method for determining the unscanned areas of the viewing volume was presented by Pito (1999). The NBV is determined by maximizing the objective function $N(i)$

$$\max N(i) = o(o_v(i), o_s(i)), i \in [1, n] \tag{20}$$

where the parameters of $o(\dots)$ are understood to be the confidence-weighted area of the void patch and partial model visible by the scanner. The number of costly computations needed to determine whether an area of the viewing volume would be occluded from some scanning position is decoupled from the number of positions considered for the NBV, thus reducing the computational cost of choosing a viewpoint.

A self-termination criterion can be used for judging the completion condition in the measurement and reconstruction process. Li et al. derived such a condition based on changes in the volume computed from two successive viewpoints (Li et al. 2005a; He and Li 2006b).

4.6.4. Graph-based placement Graph theory played an important role in developing methods for automatic sensor placement (Sheng et al. 2001b; Kaminka et al. 2008; Yegnanarayanan et al. 2009). The general automatic sensor planning system (GASP) reported by Trucco et al. is to compute optimal positions for inspection tasks using feature-based object models (Trucco et al. 1997). This exploits a feature inspection representation which outputs an explicit solution off-line for the sensor position problem. The viewpoints are

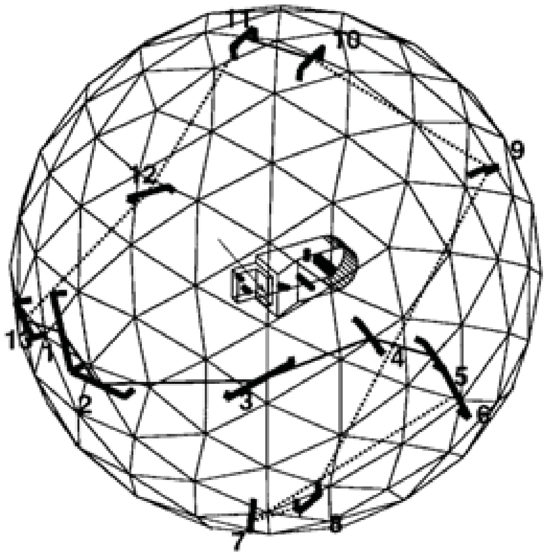

Fig. 17. The shortest path planned to take a stereo pair through the viewpoints for object inspection (Trucco et al. 1997 (© IEEE 1997)).

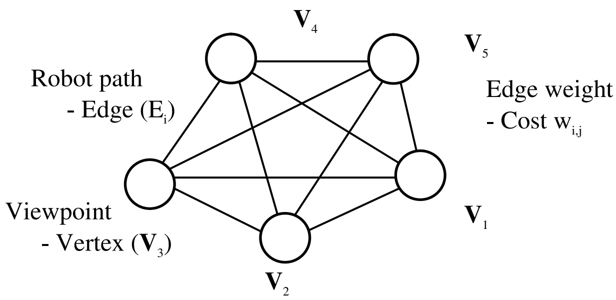

Fig. 18. Sensor placement graph (Chen and Li 2004 (© IEEE 2004)).

planned by computing the visibility and reliability. 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 in the constructed graph (Figure 17).

The method was further explicitly described by Chen and Li (2004), who gave detailed definition of the sensor placement graph and the traveling cost standard (Wang et al. 2007). A plan of viewpoints is mapped onto a graph $G = (V(G), E(G), \psi_G, w_E)$ with weight $\boldsymbol{w}$ on every edge $\boldsymbol{E}$, where the vertices $\boldsymbol{V}_i$ represent viewpoints. Edge $\boldsymbol{E}_{ij}$ represents a shortest collision-free path between viewpoint $\boldsymbol{V}_i$ and $\boldsymbol{V}_j$, and weight $\boldsymbol{w}_{ij}$ represents the corresponding distance. Figure 18 shows an example topology of a viewpoint plan. A practical solution to the sensor placement problem provides a number of viewpoints reachable by the robot and there must exist a collision-free path between every two acceptable viewpoints.

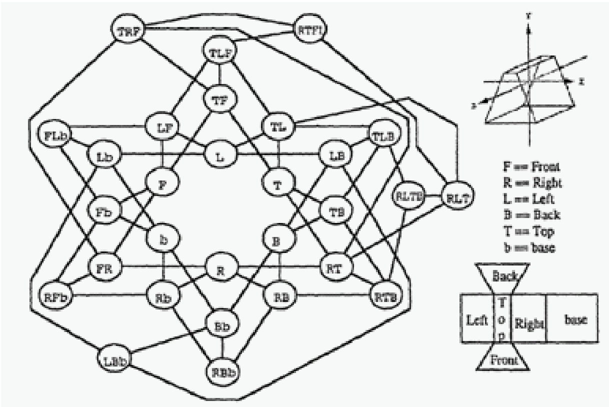

Fig. 19. The aspect graph of an object (with kind permission from Springer Science + Business Media: Eggert D, Stark L and Bowyer K (1995) Aspect graphs and their use in object recognition. *Ann Math Artificial Intelligence* 13: 347–375).

Eggert and colleagues attempted the use of the aspect graph for 3D object recognition (Eggert et al. 1995). The basic idea is that an iterative solution is generated for each of a set of candidate aspects and the best of these is chosen as the recognized view. Two assumptions are required: (i) the iterative search for the correct candidate aspect must converge to the correct answer, and (ii) the solution found for the correct aspect must be better than that found for any of the incorrect candidate aspects. In order to explore the validity of these assumptions, a simple aspect graph-based recognition system was implemented. The general definition of the aspect graph is that it is a graph structure in which: (i) there is a node for each general view of the object as seen from some maximal connected cell of viewpoint space, and (ii) there is an arc for each possible transition across the boundary between the cells of two neighboring general views, called an accidental view or a visual event (Figure 19).

In another method of object recognition by Kuno et al. (1991), features are ranked according to the number of viewpoints from which they are visible. The rank and feature extraction cost of each feature are used to generate a tree-like strategy graph. This graph gives an efficient feature search order when the viewpoint is unknown, starting with commonly occurring features and ending with features specific to a certain viewpoint. The system searches for features in the order indicated by the graph. After detection, the system compares a line representation generated from the 3D model with the image features to localize the object.

In 3D reconstruction and shape processing for reuse of the geometric models by Doi et al. (2005), a topology which defines the vertex (sampling point) connectivity and the shape of the mesh is assigned and conserved to meet the desired meshing. Stable meshing, and, hence, an accurate approximation free from the misconnection unavoidable in modeling, is then accomplished.

4.7. Active lighting

Basically, the light position should be determined to achieve adequate illumination, mathematically through the light path, i.e. surface absorption, diffused reflectance, specular reflectance, and image irradiance. Illumination now becomes the most challenging part of system design, and is a major factor when it comes to implementing color inspection (Garcia-Chamizo et al. 2007). Here, when illumination is also considered, the term ‘sensor’ has a border meaning (Quang et al. 2008; Scott 2009).

Eltoft and deFigueiredo found that illumination control could be used as a means of enhancing image features (Eltoft and deFigueiredo 1995). 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.

Measuring reflection properties of a 3D object is useful for active lighting control. Lensch et al. presented a method to select advantageous measurement directions based on analyzing the estimation of the bi-directional reflectance distribution function (BRDF) (Lensch et al. 2003). Ellenrieder et al. derived a phenomenological model of the BRDF of non-Lambertian metallic materials typically used in industrial inspection. They showed how the model can be fitted to measured reflectance values and how the fitted model can be used to determine a suitable illumination position. Together with a given sensor pose, this illumination position can be used to calculate the necessary shutter time, aperture, focus setting, and expected gray value to successfully perform a given inspection task (Ellenrieder et al. 2005b).

When the reflectance of the scene under analysis is uniform, the intensity profile of the image spot is a Gaussian and its centroid is correctly detected assuming an accurate peak position detector. However, when a change of reflectance occurs on the scene, the intensity profile of the image spot is no longer Gaussian. Khali et al. present two heuristic models to improve the sensor accuracy in the case of a variable surface reflectance (Khali et al. 2003).

To better describe the properties, Ikeuchi and Robert showed a sensor modeler, VANTAGE, to place the light sources and cameras for object recognition (Ikeuchi and Robert 1991). It was proposed to solve the detectability 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, e.g. a photometric stereo. It is represented as a sensor composition tree (SC tree). 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.

In order to automatically place light sources for maximum visual information recovery (Vazquez 2007) defined a metric to calculate the amount of information relative to an object that is effectively communicated to the user given a fixed camera position. This measure is based on an information-based concept, the Shannon entropy, and will be applied to the problem of automatic selection of light positions in order to adequately illuminate an object.

For the surveillance task of a mobile robot in indoor living space, in addition to the real conditions and poses, it was demonstrated that an illumination model is necessary for a planning behavior and good image quality results (Schroeter et al. 2009). The luminance of an object surface at position $(x; y)$ depends on the observer direction φ is modeled as $L = f(x; y; \varphi)$. The update of the illumination model can be done using a sequence of exposures with a standard camera.

To determine the optimal lighting position in view of 3D reconstruction error minimization, Belhaoua et al. proposed an evaluation criterion for each tentative position uses the contrast across object edges and the variance-based edge detection results. The best lighting position corresponds to the minimum variance and the maximum contrast values. Results show that the optimization of the lighting position leads indeed to minimization of the 3D measurement errors. The search procedure for optimal lighting source position is being fully automated using situation graph trees (SGTs) as a planning tool and is included in a complete dynamic re-planning tool for 3D automated vision-based reconstruction tasks (Belhaoua et al. 2009).

Marchand et al. 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 (Marchand 2007). For examples, illuminating the Venus de Milo is planned as in Figure 20.

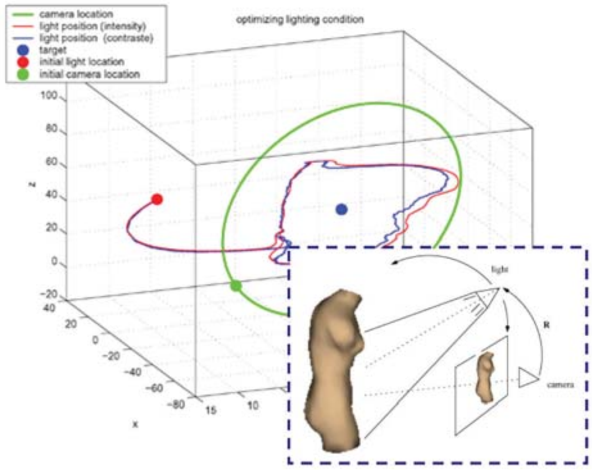

Fig. 20. An example of a camera and light source position control (Marchand 2007 (© IEEE 2007)).

5. Conclusions and future trends

In this paper we have summarized the recent development of active visual perception strategies in robotic applications. Typical contributions have been given for inspection, surveillance, recognition, search, exploration, localization, navigation, manipulation, tracking, mapping, modeling, assembly, and disassembly. Representative works have been listed for readers to have a general overview of the state of the art. A bundle of methods have been investigated with regards to solutions of visual perception acquisition problems, including visibility analysis, coverage and occlusion, spatial tessellation, data fusion, geometrical and graphic analysis, cost function evaluation, cooperative network, multi-agent, evolutionary computation, fuzzy inference, neural network, learning and expert system, information entropy, Bayesian reasoning, hypothesis and verification, etc. Issues of gaze, attention, dynamic configuration, and active lighting have also been addressed, while they are not emphasized in this survey. The largest volume of literature reviewed is related to inspection and object modeling, which correspond to model-based and non-model-based vision tasks. They contribute about 15% and 9% in the number of total publications, respectively.

Now let us look back from today at the survey by Tarabanis et al. (1995), where we can find that almost all of the 'future directions' pointed out 15 years ago have been studied with considerable advancements. While some typical problems still need to have better solutions, new challenges and requirements are emerging in the field. To make active perception even more effective in practical robotics, the challenges either are currently under investigation in research groups worldwide or need to be solved in the future. The following suggest some trends.

5.1. *Internet of things*

Internet of things refers to the networked interconnection of everyday objects whose purpose would be to make all things communicable. Every human being (as well as robot) is surrounded by 1,000 to 5,000 objects. The Internet of things would encode trillions of objects and follow the movement of those objects. If all objects of daily life can be identified and managed by computers in the same way humans can, the robots would have no difficulty in deciding their actions and would therefore be able to instantaneously identify any kind of object. Of course, it is impossible to encode all things. Robot vision can be a part of ambient intelligence between the environment and human beings. The visual knowledge obtained by active perception might be combined with other information from the Internet of things. Therefore, the robot itself should be included in the Internet of things and become the most intelligent object.

5.2. Data fusion and reliable decision

Today, multiple data sources are often obtained in a robotic system. When more than one kind of video camera, range sensor, sonar, infrared, ultrasound, GPS, compass, IMU, odometer, etc., are used together, vision perception can be made more reliable by data fusion. Consequently, a consistent representation should be developed so that fusion of positional data, range data, and appearance data can be realized to allow for safe path planning and effective view planning.

5.3. Cooperative networks

For a complex vision task in a large-scale environment, multiple robots can be adopted to accomplish the goal efficiently. This, however, requires a good scheme of system integration. Real-time data communication among all agents is required for systematic coordination. When exchanging detailed 2D/3D imagery data is impossible, extraction and representation of high-level abstract data should be implemented. Control and decision making in such systems will then become a critical issue.

5.4. On-site solution of uncertainty

In purposive perception planning for exploration, navigation, modeling, or other tasks, there is a situation that the robot has to work in a dynamic environment and the perception may associate with noise or uncertainties. Research in this issue has long been active in the field, but it seems that no complete solutions will be available in the near future.

5.5. Reconfigurable systems

As autonomous robots are expected to work in complex environments, fixed component structures are not capable of dealing with all situations. A flexible design makes the system reconfigurable during the task execution. Researchers are clearly aware of this issue, but it is a very slow progress to implement such a device due to high cost. In addition to the hardware mechanism, software for control and recalibration has to be developed concurrently.

5.6. Understanding and semantic representation

Relying solely on spatial data, active perception could not be very intelligent. Initially, the scene is seen in terms of a cloud of surface points, which would include millions of points. For scene interpretation, labeling can be processed to mark meaningful structures. Converting from source image data to geometrical shapes makes the scene understandable, and converting from geometrical shapes to semantic representation makes it much more understandable to the robot. By constructing a geometrical map and semantic map, knowledge of the spatial relationship about the environment can be used for reasoning to find objects and events. Such high-level representation and reasoning depend on, but also affect, the low-level vision perception.

5.7. Application in practical robots

In recent years, although researchers have continued working on the theoretical formulation of active sensor planning, many works tend to combine the existing methods with industrial applications such as inspection, recognition, search, modeling, tracking, exploration, assembly, and disassembly. Theoretical solutions are rarely perfect in practical engineering applications. Many sophisticated practical techniques have to be developed.

Funding

This work was supported in part by the National Natural Science Foundation of China and Microsoft Research Asia (NSFC No. 60870002), Zhejiang Provincial Natural Science Foundation (R1110679), and the Research Grants Council of Hong Kong (No. CityU118311).

References

Abidi BR, Aragam NR, Yao Y and Abidi MA (2008) Survey and analysis of multimodal sensor planning and integration for wide area surveillance. *ACM Comput Surveys* 41: 1–36.

Abrams S, Allen PK and Tarabanis K (1999) Computing camera viewpoints in an active robot work cell. *Int J Robotics Res* 18: 267–285.

Al-Hmouz R and Challa S (2005) Optimal placement for opportunistic cameras using genetic algorithm. *Proceedings of the 2005 Intelligent Sensors, Sensor Networks and Information Processing Conference*, pp. 337–341.

Amin S, Tanoto A, Witkowski U, Ruckert U and bdel-Wahab S (2008) Modified local navigation strategy for unknown environment exploration (*ICINCO 2008: Proceedings of the Fifth International Conference on Informatics in Control, Automation and Robotics*).*Robotics Automation* 1: 171–176.

Angella F, Reithler L and Gallesio F (2007) Optimal deployment of cameras for video surveillance systems. *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 388–392.

Arman F and Aggarwal JK (1993) Model-based object recognition in dense-range images - a review. *Comput Surveys* 25: 5–43.

Asai T, Kanbara M and Yokoya N (2007) Data acquiring support system using recommendation degree map for 3D outdoor modeling. *Proc SPIE* 4910: 64910H.

Baker P and Kamgar-Parsi B (2010) Using shorelines for autonomous air vehicle guidance. *Comput Vision Image Understand* 114: 723–729.

Bakhtari A and Benhabib B (2007) An active vision system for multitarget surveillance in dynamic environments. *IEEE Trans Syst Man Cybernet B* 37: 190–198.

Bakhtari A, Mackay M and Benhabib B (2009) Active-vision for the autonomous surveillance of dynamic, multi-object environments. *J Intelligent Robotic Syst* 54: 567–593.

Bakhtari A, Naish MD, Eskandari M, Croft EA and Benhabib B (2006) Active-vision-based multisensor surveillance - An implementation. *IEEE Trans Syst Man Cybernet C* 36: 668–680.

Ballesta M, Gil A, Reinoso O, Julia M and Jimenez L (2010) Multi-robot map alignment in visual SLAM. *WSEAS Trans Syst* 9: 213–222.

Banish M, Rodgers M, Hyatt B, Edmonson R, Chenault D, Heym J, Johnson J and Dobson K (2010) Exploiting uncalibrated stereo on a UAV platform. *Proc SPIE:* 76921T.

Banta JE, Wong LM, Dumont C and Abidi MA (2000) A next-best-view system for autonomous 3-D object reconstruction. *IEEE Trans Syst Man Cybernet A* 30: 589–598.

Bardon C, Hodge L and Kamel A (2004) A framework for optimal multi-agent sensor planning. *Int J Robotics Automat* 19: 152–166.

Barreto JP, Perdigoto L, Caseiro R and Araujo H (2010) Active stereo tracking of N <= 3 targets using line scan cameras. *IEEE Trans Robotics* 26: 442–457.

Baumann M, Leonard S, Croft EA and Little JJ (2010) Path planning for improved visibility using a probabilistic road map. *IEEE Trans Robotics* 26: 195–200.

Baumann MA, Dupuis DC, Leonard S, Croft EA and Little JJ (2008) Occlusion-free path planning with a probabilistic roadmap. In *2008 IEEE/RSJ International Conference on Robots and Intelligent Systems, Conference Proceedings*, pp. 2151–2156.

Belhaoua A, Kohler S and Hirsch E (2009) Determination of optimal lighting position in view of 3D reconstruction error minimization. In *The 10th European Congress of Stereology and Image Analysis*, pp. 408–414.

Biegelbauer G, Vincze M and Wohlkinger W (2010) Model-based 3D object detection. *Machine Vision Appl* 21: 497–516.

Bjorkman M and Kragic D (2004) Combination of foveal and peripheral vision for object recognition and pose estimation. In *IEEE International Conference Robotics and Automation*, pp. 5135–5140.

Blaer PS and Allen PK (2007) Data acquisition and view planning for 3-D modeling tasks. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 423–428.

Blaer PS and Allen PK (2009) View planning and automated data acquisition for three-dimensional modeling of complex sites. *J Field Robotics* 26: 865–891.

Bodor R, Drenner A, Schrater P and Papanikolopoulos N (2007) Optimal camera placement for automated surveillance tasks. *J Intelligent Robotic Syst* 50: 257–295.

Borenstein J, Borrell A, Miller R and Thomas D (2010) Heuristics-enhanced dead-reckoning (HEDR) for accurate position tracking of tele-operated UGVs. *Proc SPIE:* 76921R.

Borotschnig H and Paletta L (2000) Appearance-based active object recognition. *Image Vision Comput* 18: 715–727.

Borrmann D, Elseberg J, Lingemann K, Nuchter A and Hertzberg J (2008) Globally consistent 3D mapping with scan matching. *Robotics Auton Syst* 56: 130–142.

Bottino A and Laurentini A (2006a) Experimental results show near-optimality of a sensor location algorithm. In *2006 IEEE International Conference on Robotics and Biomimetics*, pp. 340–345.

Bottino A and Laurentini A (2006b) What's NEXT? An interactive next best view approach. *Pattern Recognition* 39: 126–132.

Bottino A and Laurentini A (2008) A nearly optimal sensor placement algorithm for boundary coverage. *Pattern Recognition* 41: 3343–3355.

Bottino A, Laurentini A and Rosano L (2007) A tight lower bound for art gallery sensor location algorithms. *ETFA 2007: 12th IEEE International Conference on Emerging Technologies and Factory Automation*, pp. 434–440.

Bottino A, Laurentini A and Rosano L (2009) A new lower bound for evaluating the performances of sensor location algorithms. *Pattern Recognition Lett* 30: 1175–1180.

Boutarfa A, Bouguechal NE and Emptoz H (2008) A new approach for an automated inspection system of the manufactured parts. *Int J Robotics Automat* 23: 220–226.

Briggs AJ and Donald BR (2000) Visibility-based planning of sensor control strategies. *Algorithmica* 26: 364–388.

Budiharto W, Jazidie A and Purwanto D (2010) Indoor navigation using adaptive neuro fuzzy controller for servant robot. In *Proceedings of the 2010 Second International Conference on Computer Engineering and Applications (ICCEA 2010)*, pp. 582–586.

Byun JE and Nagata T (1996) Active visual sensing of the 3-D pose of a flexible object. *Robotica* 14: 173–188.

Caglioti V (2001) An entropic criterion for minimum uncertainty sensing in recognition and localization - Part I: Theoretical and conceptual aspects. *IEEE Trans Syst Man Cybernet B* 31: 187–196.

Callari FG and Ferrie FP (2001) Active object recognition: Looking for differences. *Int J Comput Vision* 43: 189–204.

Callieri M, Fasano A, Impoco G, Cignoni P, Scopigno R, Parrini G and Biagini G (2004) RoboScan: an automatic system for accurate and unattended 3D scanning. In *2nd International Symposium on 3D Data Processing*, pp. 805–812.

Carrasco M and Mery D (2007) Automatic multiple visual inspection on non-calibrated image sequence with intermediate classifier block. In *Advances in Image and Video Technology*, pp. 371–384.

Cassinis R and Tampalini F (2007) AMIRoLoS an active marker internet-based robot localization system. *Robotics Auton Syst* 55: 306–315.

Chang MH and Park SC (2009) Automated scanning of dental impressions. *Computer-Aided Design* 41: 404–411.

Chang MS, Chou JH and Wu CM (2010) Design and implementation of a novel outdoor road-cleaning robot. *Advanced Robotics* 24: 85–101.

Chen F, Brown GM and Song MM (2000) Overview of three-dimensional shape measurement using optical methods. *Optical Eng* 39: 10–22.

Chen HY and Li YF (2008) Data fusion for three-dimensional tracking using particle techniques. *Opt Eng* 47: 016401.

Chen HY and Li YF (2009) Dynamic view planning by effective particles for three-dimensional tracking. *IEEE Trans Syst Man Cybernet B* 39: 242–253.

Chen SH and Liao TT (2009) An automated IC chip marking inspection system for surface mounted devices on taping machines. *J Sci Indust Res* 68: 361–366.

Chen SY and Li YF (2004) Automatic sensor placement for model-based robot vision. *IEEE Trans Syst Man Cybernet B* 34: 393–408.

Chen SY and Li YF (2005) Vision sensor planning for 3-D model acquisition. *IEEE Trans Syst Man Cybernet B* 35: 894–904.

Chen SY, Li YF and Zhang JW (2008a) *Active Sensor Planning for Multiview Vision Tasks* . New York: Springer, p. 283.

Chen SY, Li YF and Zhang JW (2008b) Vision processing for realtime 3-D data acquisition based on coded structured light. *IEEE Trans Image Process* 17: 167–176.

Chen X and Davis J (2008) An occlusion metric for selecting robust camera configurations. *Machine Vision Appl* 19: 217–222.

Chu GW and Chung MJ (2002) Autonomous selection and modification of camera configurations using visibility and manipulability measures. *J Robotic Syst* 19: 219–230.

Cohen O and Edan Y (2008) A sensor fusion framework for online sensor and algorithm selection. *Robotics Auton Syst* 56: 762–776.

Cowan CK and Kovesi PD (1988) Automatic sensor placement from vision task requirements. *IEEE Trans Pattern Anal Machine Intell* 10: 407–416.

Craciun D, Paparoditis N and Schmitt F (2008) Automatic pyramidal intensity-based laser scan matcher for 3D modeling of large scale unstructured environments. *Proceedings of the Fifth Canadian Conference on Computer and Robot Vision*, pp. 18–25.

Das S and Ahuja N (1996) Active surface estimation: Integrating coarse-to-fine image acquisition and estimation from multiple cues. *Artificial Intelligence* 83: 241–266.

de Ruiter H, Mackay M and Benhabib B (2010) Autonomous three-dimensional tracking for reconfigurable active-vision-based object recognition. *Proc IMechE B J Eng Manufacture* 224(B3): 343–360.

Deinzer F, Derichs C, Niemann H and Denzler J (2009) A framework for actively selecting viewpoints in object recognition. *Int J Pattern Recogn Artificial Intelligence* 23: 765–799.

Dickinson SJ, Christensen HI, Tsotsos JK and Olofsson G (1997) Active object recognition integrating attention and viewpoint control. *Comput Vision Image Understand* 67: 239–260.

Doi J, Sato W and Miyake T (2005) Topology conserved 3D reconstruction and shape processing for reuse of the geometric models. In *IEEE International Conference on Information Reuse and Integration*, pp. 410–414. 2005.

Dunn E and Olague G (2003) Evolutionary computation for sensor planning: The task distribution plan. *EURASIP J Appl Signal Process* 8: 748–756.

Dunn E and Olague G (2004) Multi-objective sensor planning for efficient and accurate object reconstruction. *Appl Evol Comput* 3005: 312–321.

Dunn E, Olague G and Lutton E (2006) Parisian camera placement for vision metrology. *Pattern Recognition Lett* 27: 1209–1219.

Eggert D, Stark L and Bowyer K (1995) Aspect graphs and their use in object recognition. *Ann Math Artificial Intelligence* 13: 347–375.

Eidenberger R, Grundmann T, Feiten W and Zoellner R (2008) Fast parametric viewpoint estimation for active object detection. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 464–469.

Ellenrieder MM, Kruger L, Stossel D and Hanheide M (2005a) A versatile model-based visibility measure for geometric primitives. In *Image Analysis (Lecture Notes in Computer Science* , vol. 3540). Berlin: Springer, pp. 669–678.

Ellenrieder MM, Wohler C and d'Angelo P (2005b) Reflectivity function based illumination and sensor planning for industrial inspection (*Optical Measurement Systems for Industrial Inspection IV*). *Proc SPIE* 5856: 89–98.

Eltoft T and deFigueiredo R (1995) Illumination control as a means of enhancing image features in active vision systems. *IEEE Trans Image Processing* 4: 1520–1530.

Erdem UM and Sclaroff S (2006) Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements. *Comput Vision Image Understand* 103: 156–169.

Fang SF, George B and Palakal M (2008) Automatic surface scanning of 3D artifacts. In *Proceedings of the 2008 International Conference on Cyberworlds*, pp. 335–341.

Farshidi F, Sirouspour S and Kirubarajan T (2009) Robust sequential view planning for object recognition using multiple cameras. *Image Vision Comput* 27: 1072–1082.

Fernandez P, Rico JC, Alvarez BJ, Valino G and Mateos S (2008) Laser scan planning based on visibility analysis and space partitioning techniques. *Int J Adv Manufacturing Technol* 39: 699–715.

Fiore L, Somasundaram G, Drenner A and Papanikolopoulos N (2008) Optimal camera placement with adaptation to dynamic scenes. In *2008 IEEE International Conference on Robotics and Automation*, pp. 956–961.

Flandin G and Chaumette F (2001) Vision-based control using probabilistic geometry for objects reconstruction. In *The 40th IEEE Conference on Decision and Control*, pp. 4152–4157.

Flandin G and Chaumette F (2002) Visual data fusion for objects localization by active vision. In *Proceedings of ECCV*, pp. 312–326.

Frintrop S and Jensfelt P (2008a) Active gaze control for attentional visual SLAM. In *2008 IEEE International Conference on Robotics and Automation*, pp. 3690–3697.

Frintrop S and Jensfelt P (2008b) Attentional landmarks and active gaze control for visual SLAM. *IEEE Trans Robotics* 24: 1054–1065.

Gao J, Gindy N and Chen X (2006) An automated GD&T inspection system based on non-contact 3D digitization. *Int J Production Res* 44(1): 117–134.

Garcia HC and Villalobos JR (2007a) Automated feature selection methodology for reconfigurable Automated Visual Inspection systems. In *2007 IEEE International Conference on Automation Science and Engineering*, pp. 703–708.

Garcia HC and Villalobos JR (2007b) Development of a methodological framework for the self reconfiguration of automated visual inspection systems. In *5th IEEE International Conference on Industrial Informatics*, pp. 207–212.

Garcia-Chamizo JM, Fuster-Guillo A and zorin-Lopez J (2007) Simulation of automated visual inspection systems for specular surfaces quality control. *Adv Image Video Technol* 4872: 749–762.

Gonzalez-Banos HH and Latombe JC (2002) Navigation strategies for exploring indoor environments. *Int J Robotics Res* 21: 829–848.

Gremban KD and Ikeuchi K (1994) Planning multiple observations for object recognition. *Int J Comput Vision* 12: 137–172.

Grewe L and Kak AC (1995) Interactive learning of a multiple-attribute hash table classifier for fast object recognition. *Comput Vision Image Understand* 61: 387–416.

Han S, Choi B and Lee J (2008) A precise curved motion planning for a differential driving mobile robot. *Mechatronics* 18: 486–494.

He BW and Li YF (2006a) A next-best-view method for automatic modeling of three dimensional objects. *Dynam Contin Discrete Impulsive Syst Ser B Appl Algorithms* 13E: 104–109.

He BW and Li YF (2006b) A next-best-view method with self-termination in active modeling of 3D objects. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5345–5350.

Hernandez OJ and Wang YF (2008) An autonomous off-road robot based on integrative technologies. In *2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 540–545.

Hodge L and Kamel M (2003) An agent-based approach to multisensor coordination. *IEEE Trans Syst Man Cybernet A* 33: 648–662.

Hodge L, Kamel M and Bardon C (2004) Scalability and optimality in a multi-agent sensor planning system. In *Soft Computing with Industrial Applications*, pp. 74–80.

Hornung A, Zeng BY and Kobbelt L (2008) Image selection for improved multi-view stereo. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2696–2703.

Hovland GE and McCarragher BJ (1999) Control of sensory perception in a mobile navigation problem. *Int J Robotics Res* 18: 201–212.

Howarth RJ (2005) Spatial models for wide-area visual surveillance: Computational approaches and spatial building-blocks. *Artificial Intelligence Rev* 23: 97–154.

Huang YB and Qian X (2008a) An efficient sensing localization algorithm for free-form surface digitization. *J Comput Inform Sci Eng* 8: 021008.

Huang YB and Qian XP (2007) A dynamic sensing-and modeling approach to three-dimensional point- and area-sensor integration. *J Manufacturing Sci Eng Trans ASME* 129: 623–635.

Huang YB and Qian XP (2008b) An efficient sensing localization algorithm for free-form surface digitization. In *27th Computers and Information in Engineering Conference, Vol 2, Pts A and B 2007 - Proceedings of the Asme International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 327–336.

Hubner W and Mallot HA (2007) Metric embedding of view-graphs - A vision and odometry-based approach to cognitive mapping. *Autonomous Robots* 23: 183–196.

Ikeuchi K and Robert JC (1991) Modeling sensor detectability with the Vantage Geometric Sensor Modeler. *IEEE Trans Robotics Automation* 7: 771–784.

Impoco G, Cignoni P and Scopigno R (2004) Closing gaps by clustering unseen directions. *Proceedings of the International Conference on Shape Modeling and Applications* , p. 307.

Jang HY, Moradi H, Le Minh P, Lee S and Han J (2008) Visibility-based spatial reasoning for object manipulation in cluttered environments. *Computer-Aided Design* 40: 422–438.

Jang HY, Moradi H, Lee S, Jang D, Kim E and Han J (2007) A graphics hardware-based accessibility analysis for real-time robotic manipulation. *Dynam Contin Discrete Impulsive Syst Ser B Appl Algorithms* 14: 97–106.

Janoos F, Machiraju R, Parent R, Davis JW and Murray A (2007) Sensor configuration for coverage optimization for surveillance applications. *Proc SPIE* 6491: 49105.

Jonnalagadda K, Lumia R, Starr G and Wood J (2003) Viewpoint selection for object reconstruction using only local geometric features. In *IEEE International Conference on Robotics and Automation*, pp. 2116–2122.

Kaess M and Dellaert F (2010) Probabilistic structure matching for visual SLAM with a multi-camera rig. *Comput Vision Image Understand* 114: 286–296.

Kaminka GA, Schechter-Glick R and Sadov V (2008) Using sensor morphology for multirobot formations. *IEEE Trans Robotics* 24: 271–282.

Kang F, Li JJ and Xu Q (2008) Virus coevolution partheno-genetic algorithms for optimal sensor placement. *Adv Eng Informatics* 22: 362–370.

Kececi F, Tonko M, Nagel H-S and Gengenbach V (1998) Improving visually servoed disassembly operations by automatic camera placement. In *IEEE International Conference on Robotics and Automation*, pp. 2947–2952.

Khali H, Savaria Y, Houle JL, Rioux M, Beraldin JA and Poussart D (2003) Improvement of sensor accuracy in the case of a variable surface reflectance gradient for active laser range finders. *IEEE Trans Instrum Meas* 52: 1799–1808.

Kim MY and Cho HS (2003) An active view planning method for mobile robots using a trinocular visual sensor (*Optomechatronic Systems IV*). *Proc SPIE* 5264: 74–83.

Kollar T and Roy N (2008) Trajectory optimization using reinforcement, learning for map exploration. *Int J Robotics Res* 27: 175–196.

Kristensen S (1997) Sensor planning with Bayesian decision theory. *Robotics Auton Syst* 19: 273–286.

Krotkov E and Bajcsy R (1993) Active vision for reliable ranging - cooperating focus, stereo, and vergence. *Int J Comput Vision* 11: 187–203.

Kumar A (2008) Computer-vision-based fabric defect detection: A survey. *IEEE Trans Indust Electron* 55: 348–363.

Kuno Y, Okamoto Y and Okada S (1991) Robot vision using a feature search strategy generated from a 3-D object model. *IEEE Trans Pattern Anal Machine Intell* 13: 1085–1097.

Kutulakos KN and Dyer CR (1994) Recovering shape by purposive viewpoint adjustment. *Int J Comput Vision* 12: 113–136.

Kutulakos KN and Dyer CR (1995) Global surface reconstruction by purposive control of observer motion. *Artificial Intelligence* 78: 147–177.

Lang J and Jenkin MRM (2000) Active object modeling with VIRTUE. *Autonomous Robots* 8: 141–159.

Larsson S and Kjellander JAP (2008) Path planning for laser scanning with an industrial robot. *Robotics Auton Syst* 56: 615–624.

Lensch HPA, Lang J, Sa AM and Seidel HP (2003) Planned sampling of spatially varying BRDFs. *Comput Graphics Forum* 22: 473–482.

Li XK and Wee WG (2008) Sensor error modeling and compensation for range images captured by a 3D ranging system. *Meas Sci Technol* 19: 125502.

Li YD and Gu PH (2004) Free-form surface inspection techniques state of the art review. *Computer-Aided Design* 36: 1395–1417.

Li YF, He B and Bao P (2005a) Automatic view planning with self-termination in 3D object reconstructions. *Sensors Actuators A Phys* 122: 335–344.

Li YF, He B, Chen S and Bao P (2005b) A view planning method incorporating self-termination for automated surface measurement. *Meas Sci Technol* 16: 1865–1877.

Li YF and Liu ZG (2003) Method for determining the probing points for efficient measurement and reconstruction of freeform surfaces. *Meas Sci Technol* 14: 1280–1288.

Li YF and Liu ZG (2005) Information entropy-based viewpoint planning for 3-D object reconstruction. *IEEE Trans Robotics* 21: 324–337.

Liang C and Wong KYK (2010) 3D reconstruction using silhouettes from unordered viewpoints. *Image Vision Comput* 28: 579–589.

Lim SN, Davis L and Mittal A (2007) Task scheduling in large camera networks. In *Computer Vision - ACCV*, pp. 397–407.

Lim SN, Davis LS and Mittal A (2006) Constructing task visibility intervals for video surveillance. *Multimedia Syst* 12: 211–226.

Lin HY, Liang SC and Wu JR (2007) 3D shape recovery with registration assisted stereo matching. In *Pattern Recognition and Image Analysis* (*Lecture Notes in Computer Science*, vol. 4478). Berlin: Springer, pp. 596–603.

Liu M, Liu YS and Ramani K (2009) Computing global visibility maps for regions on the boundaries of polyhedra using Minkowski sums. *Computer-Aided Design* 41: 668–680.

Liu M and Ramani K (2009) On minimal orthographic view covers for polyhedra. In *Proceedings SMI 2009: IEEE International Conference on Shape Modeling and Applications*, pp. 96–102.

Liu YH (2009) Replicator dynamics in the iterative process for accurate range image matching. *Int J Comput Vision* 83: 30–56.

Liu YS and Heidrich W (2003) Interactive 3D model acquisition and registration. In *Proceedings 11th Pacific Conference on Computer Graphics and Applications*, pp. 115–122.

Loniot B, Seulin R, Gorria P and Meriaudeau F (2007) Simulation for an automation of 3D acquisition and post-processing. In *Eighth International Conference on Quality Control by Artificial Vision* .

Ma CYT, Yau DKY, Chin JC, Rao NSV and Shankar M (2009) Matching and fairness in threat-based mobile sensor coverage. *IEEE Trans Mobile Comput* 8: 1649–1662.

Mackay M and Benhabib B (2008a) A multi-camera active-vision system for dynamic form recognition. In *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering* , pp. 26–31.

Mackay M and Benhabib B (2008b) Active-vision system reconfiguration for form recognition in the presence of dynamic obstacles. In *Articulated Motion and Deformable Objects*, pp. 188–207.

MacKinnon D, Aitken V and Blais F (2008a) Adaptive laser range scanning. In *2008 American Control Conference*, pp. 3857–3862.

MacKinnon D, Aitken V and Blais F (2008b) Review of measurement quality metrics for range imaging. *J Electron Imag* 17: 033003.

Madhuri P, Nagesh AS, Thirumalaikumar M, Varghese Z and Varun AV (2009) Performance analysis of smart camera based distributed control flow logic for machine vision applications. In *2009 IEEE International Conference on Industrial Technology*, pp. 90–95.

Marchand E (2007) Control camera and light source positions using image gradient information. In *IEEE International Conference on Robotics and Automation*, pp. 417–422.

Marchand E and Chaumette F (1999a) Active vision for complete scene reconstruction and exploration. *IEEE Trans Pattern Anal Machine Intell* 21: 65–72.

Marchand E and Chaumette F (1999b) An autonomous active vision system for complete and accurate 3D scene reconstruction. *Int J Comput Vision* 32: 171–194.

Martinez SS, Ortega JG, Garcia JG and Garcia AS (2008) An expert knowledge based sensor planning system for car headlight lens inspection. In *Computational Intelligence in Decision and Control*, pp. 1123–1128.

Martinez SS, Ortega JG, Garcia JG and Garcia AS (2009) A sensor planning system for automated headlamp lens inspection. *Expert Syst Applications* 36: 8768–8777.

Martins FAR, Garcia-Bermejo JG, Casanova EZ and Gonzalez JRP (2005) Automated 3D surface scanning based on CAD model. *Mechatronics* 15: 837–857.

Martins FAR, Garcia-Bermejo JG, Zalama E and Peran JR (2003) An optimized strategy for automatic optical scanning of objects in reverse engineering. *Proc IMechE B J Eng Manufacture* 217: 1167–1171.

Maruyama K, Takase R, Kawai Y, Yoshimi T, Takahashi H and Tomita F (2010) Semi-automated excavation system for partially buried objects using stereo vision-based three-dimensional localization. *Advanced Robotics* 24: 651–670.

Mason S (1997) Heuristic reasoning strategy for automated sensor placement. *Photogram Eng Remote Sens* 63: 1093–1102.

Mitsunaga N and Asada M (2006) How a mobile robot selects landmarks to make a decision based on an information criterion. *Autonomous Robots* 21: 3–14.

Mittal A (2006) Generalized multi-sensor planning. In*Proceedings of ECCV*, pp. 522–535.

Mittal A and Davis LS (2004) Visibility analysis and sensor planning in dynamic environments. In *Computer Vision - ECCV*, pp. 175–189.

Mittal A and Davis LS (2008) A general method for sensor planning in multi-sensor systems: Extension to random occlusion. *Int J Comput Vision* 76: 31–52.

Miura J and Ikeuchi K (1998) Task-oriented generation of visual sensing strategies in assembly tasks. *IEEE Trans Pattern Anal Machine Intell* 20: 126–138.

Mostofi Y and Sen P (2009) Compressive cooperative sensing and mapping in mobile networks. In *2009 American Control Conference*, pp. 3397–3404.

Motai Y and Kosaka A (2008) Hand–eye calibration applied to viewpoint selection for robotic vision. *IEEE Trans Indust Electron* 55: 3731–3741.

Murrieta-Cid R, Muoz L and Alencastre M (2005) Maintaining visibility of a moving holonomic target at a fixed distance with a non-holonomic robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Nabbe B and Hebert M (2007) Extending the path-planning horizon. *Int J Robotics Res* 26: 997–1024.

Naish MD, Croft EA and Benhabib B (2003) Coordinated dispatching of proximity sensors for the surveillance of manoeuvring targets. *Robotics Comput Integr Manufacturing* 19: 283–299.

Nayak J, Gonzalez-Argueta L, Song B, Roy-Chowdhury A and Tuncel E (2008) Multi-target tracking through opportunistic camera control in a resource constrained multimodal sensor network. In *2008 Second ACM/IEEE International Conference on Distributed Smart Cameras*, pp. 77–86.

Nelson BJ and Papanikolopoulos NP (1996) Robotic visual servoing and robotic assembly tasks. *IEEE Robotics Automat Mag* 58: 23–31.

Newman TS and Jain AK (1995) A survey of automated visual inspection. *Comput Vision Image Understand* 61: 231–262.

Nickels K, DiCicco M, Bajracharya M and Backes P (2010) Vision guided manipulation for planetary robotics - position control. *Robotics Auton Syst* 58: 121–129.

Nikolaidis S, Ueda R, Hayashi A and Arai T (2009) Optimal camera placement considering mobile robot trajectory. In *2008 IEEE International Conference on Robotics and Biomimetics*, pp. 1393–1396.

Nilsson U, Ogren P and Thunberg J (2008) Optimal positioning of surveillance UGVs. In *2008 IEEE/RSJ International Conference on Robots and Intelligent Systems*, pp. 2539–2544.

Nilsson U, Ogren P and Thunberg J (2009) Towards optimal positioning of surveillance UGVs. In *8th International Conference on Cooperative Control and Optimization* , pp. 221–233.

Nuchter A and Hertzberg J (2008) Towards semantic maps for mobile robots. *Robotics Auton Syst* 56: 915–926.

Null BD and Sinzinger ED (2006) Next best view algorithms for interior and exterior model acquisition. In *Advances in Visual Computing*, pp. 668–677.

Olague G (2002) Automated photogrammetric network design using genetic algorithms. *Photogram Eng Remote Sens* 68: 423–431.

Olague G and Dunn E (2007) Development of a practical photogrammetric network design using evolutionary computing. *Photogram Record* 22: 22–38.

Olague G and Mohr R (2002) Optimal camera placement for accurate reconstruction. *Pattern Recognition* 35: 927–944.

Oniga F and Nedevschi S (2010) Processing dense stereo data using elevation maps: road surface, traffic isle, and obstacle detection. *IEEE Trans Vehicular Technol* 59: 1172–1182.

Park TH, Kim HJ and Kim N (2006) Path planning of automated optical inspection machines for PCB assembly systems. *Int J Control Automat Syst* 4: 96–104.

Perng DB, Chen SH and Chang YS (2010) A novel internal thread defect auto-inspection system. *Int J Adv Manufacturing Technol* 47: 731–743.

Pito R (1999) A solution to the next best view problem for automated surface acquisition. *IEEE Trans Pattern Anal Machine Intell* 21: 1016–1030.

Popescu V, Sacks E and Bahmutov G (2004) Interactive modeling from dense colour and sparse depth. In *2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 430–437.

Prieto F, Lepage R, Boulanger P and Redarce T (2003) A CAD-based 3D data acquisition strategy for inspection. *Machine Vision Appl* 15: 76–91.

Prieto F, Redarce T, Boulanger P and Lepage R (2001) Tolerance control with high resolution 3D measurements. In *Third International Conference on 3-D Digital Imaging and Modeling*, 339–346.

Prieto F, Redarce T, Lepage R and Boulanger P (2002) An automated inspection system. *Int J Adv Manufacturing Technol* 19: 917–925.

Quang LB, Kim D and Lee S (2008) Auto-focusing technique in a projector–camera system. In *2008 10th International Conference on Control Automation Robotics and Vision (ICARV 2008)*, pp. 1914–1919.

Radovnikovich M, Vempaty P and Cheok K (2010) Auto-preview camera orientation for environment perception on a mobile robot. *Proc SPIE:* 75390Q.

Rae A and Basir O (2009) Reducing multipath effects in vehicle localization by fusing GPS with machine vision. In *12th International Conference on Information Fusion* , pp. 2099–2106.

Raviv D and Herman M (1994) A unified approach to camera fixation and vision-based road following. *IEEE Trans Syst Man Cybernet* 24: 1125–1141.

Reddi S and Loizou G (1995) Analysis of camera behavior during tracking. *IEEE Trans Pattern Anal Machine Intell* 17: 765–778.

Reed MK and Allen PK (2000) Constraint-based sensor planning for scene modeling. *IEEE Trans Pattern Anal Machine Intell* 22: 1460–1467.

Riggs T, Inanc T and Weizhong Z (2010) An autonomous mobile robotics testbed: construction, validation, and experiments. *IEEE Trans Control Syst Technol* 18: 757–766.

Rivera-Rios AH, Shih FL and Marefat M (2005) Stereo camera pose determination with error reduction and tolerance satisfaction for dimensional measurements. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 423–428.

Roy SD, Chaudhury S and Banerjee S (2000) Isolated 3-d object recognition through next view planning. *IEEE Trans Syst Man Cybernet A* 30: 67–76.

Roy SD, Chaudhury S and Banerjee S (2004) Active recognition through next view planning: a survey. *Pattern Recognition* 37: 429–446.

Roy SD, Chaudhury S and Banerjee S (2005) Recognizing large isolated 3-D objects through next view planning using inner camera invariants. *IEEE Trans Syst Man Cybernet B* 35: 282–292.

Royer E, Lhuillier M, Dhome M and Lavest JM (2007) Monocular vision for mobile robot localization and autonomous navigation. *Int J Comput Vision* 74: 237–260.

Saadatseresht M, Samadzadegan F and Azizi A (2005) Automatic camera placement in vision metrology based on a fuzzy inference system. *Photogram Eng Remote Sens* 71: 1375–1385.

Saadatseresht M and Varshosaz M (2007) Visibility prediction based on artificial neural networks used in automatic network design. *Photogrammetric Record* 22: 336–355.

Sablatnig R, Tosovic S and Kampel M (2003) Next view planning for a combination of passive and active acquisition techniques. In *Fourth International Conference on 3-D Digital Imaging and Modeling*, pp. 62–69.

Sakane S, Kuruma T, Omata T and Sato T (1995) Planning focus of attention for multifingered hand with consideration of time-varying aspects. *Comput Vision Image Understand* 61: 445–453.

Schroeter C, Hoechemer M, Mueller S and Gross HM (2009) Autonomous robot cameraman - observation pose optimization for a mobile service robot in indoor living space. In *IEEE International Conference on Robotics and Automation*, pp. 2199–2204.

Scott WR (2009) Model-based view planning. *Machine Vision Appl* 20: 47–69.

Scott WR, Roth G and Rivest JF (2003) View planning for automated three-dimensional object reconstruction and inspection. *ACM Comput Surveys* 35: 64–96.

Se S and Jasiobedzki P (2007) Stereo-vision based 3D modeling for unmanned ground vehicles (*Unmanned Systems Technology IX*). *Proc SPIE* 6561: X5610.

Sebastian JM, Garcia D, Traslosheros A, Sanchez FM and Dominguez S (2007) A new automatic planning of inspection of 3D industrial parts by means of visual system. In *Image*

Analysis and Recognition (*Lecture Notes in Computer Science*, vol. 4633). Berlin: Springer, pp. 1148–1159.
Seo YW and Urmson C (2008) A perception mechanism for supporting autonomous intersection handling in urban driving. In *2008 IEEE/RSJ International Conference on Robots and Intelligent Systems*, pp. 1830–1835.
Sheng WH, Xi N, Song M and Chen YF (2001a) CAD-guided robot motion planning. *Indust Robot* 28: 143–151.
Sheng WH, Xi N, Song MM and Chen YF (2001b) Graph-based surface merging in CAD-guided dimensional inspection of automotive parts. In *2001 IEEE International Conference on Robotics and Automation*, pp. 3127–3132.
Sheng WH, Xi N, Song MM and Chen YF (2003) CAD-guided sensor planning for dimensional inspection in automotive manufacturing. *IEEE-ASME Trans Mechatron* 8: 372–380.
Shi Q, Xi N and Zhang C (2010) Develop a robot-aided area sensing system for 3D shape inspection. *J Manufacturing Sci Eng Trans ASME* 132: 014502.
Shih CS and Gerhardt LA (2006) Integration of view planning with nonuniform surface sampling techniques for three-dimensional object inspection. *Opt Eng* 45: 113601.
Shimizu S, Yamamoto K, Wang CH, Satoh Y, Tanahashi H and Niwa Y (2005) Detection of moving object by mobile Stereo Omnidirectional System (SOS). *Elect Eng Japan* 152: 29–38.
Shinzato P, Fernandes L, Osorio F and Wolf D (2010) Path recognition for outdoor navigation using artificial neural networks: case study. In *2010 IEEE International Conference on Industrial Technology (ICIT 2010)*, pp. 1457–1462.
Shubina K and Tsotsos JK (2010) Visual search for an object in a 3D environment using a mobile robot. *Comput Vision Image Understand* 114: 535–547.
Shum HY, Hebert M, Ikeuchi K and Reddy R (1997) An integral approach to free-form object modeling. *IEEE Trans Pattern Anal Machine Intell* 19: 1366–1370.
Sivaram GSVS, Kankanhalli MS and Ramakrishnan KR (2009) Design of multimedia surveillance systems. *ACM Trans Multimedia Comput Commun Appl* 5: 23.
Stemmer A, Schreiber G, Arbter K and Albu-Schaffer A (2006) Robust assembly of complex shaped planar parts using vision and force. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 493–500.
Sujan VA and Dubowsky S (2005a) Efficient information-based visual robotic mapping in unstructured environments. *Int J Robotics Res* 24: 275–293.
Sujan VA and Dubowsky S (2005b) Visually guided cooperative robot actions based on information quality. *Autonomous Robots* 19: 89–110.
Sujan VA and Meggiolaro MA (2005) Intelligent and efficient strategy for unstructured environment sensing using mobile robot agents. *J Intelligent Robotic Syst* 43: 217–253.
Sun C, Wang P, Tao L and Chen S (2008) Method of scanning-path determination for color three-dimensional laser measurement. *Opt Eng* 47: 01360.
Sun J, Sun Q and Surgenor BW (2007) Adaptive visual inspection for assembly line parts verification. In *WCECS 2007: World Congress on Engineering and Computer Science*, pp. 575–580.
Sun TH, Tseng CC and Chen MS (2010) Electric contacts inspection using machine vision. *Image Vision Comput* 28: 890–901.
Suppa M and Hirzinger G (2007) Multisensory exploration of robot workspaces. *Tm-Technisches Messen* 74: 139–146.
Sutton MA and Stark L (2008) Function-based reasoning for goal-oriented image segmentation. *Towards Affordance-Based Robot Control* 4760: 159–172.
SyedaMahmood TF (1997) Data and model-driven selection using color regions. *Int J Comput Vision* 21: 9–36.
Tarabanis K, Tsai RY and Allen PK (1994) Analytical characterization of the feature detectability constraints of resolution, focus, and field-of-view for vision sensor planning. *CVGIP-Image Understanding* 59: 340–358.
Tarabanis K, Tsai RY and Kaul A (1996) Computing occlusion-free viewpoints. *IEEE Trans Pattern Anal Machine Intell* 18: 279–292.
Tarabanis KA, Allen PK and Tsai RY (1995) A survey of sensor planning in computer vision. *IEEE Trans Robotics Automat* 11: 86–104.
Tarabanis KA, Tsai RY and Allen PK (1995) The MVP sensor planning system for robotic vision tasks. *IEEE Trans Robotics Automat* 11: 72–85.
Taylor CJ and Spletzer J (2007) A bounded uncertainty approach to cooperative localization using relative bearing constraints. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2506–2512.
Thielemann J, Breivik G and Berge A (2010) Robot navigation and obstacle detection in pipelines using time-of-flight imagery. *Proc SPIE:* 75260O.
Thomas U, Molkenstruck S, Iser R and Wahl FM (2007) Multi sensor fusion in robot assembly using particle filters. In *IEEE International Conference on Robotics and Automation*, pp. 3837–3843.
Torres-Mendez LA and Dudek G (2008) Inter-image statistics for 3D environment modeling. *Int J Comput Vision* 79: 137–158.
Treuillet S, Albouy B and Lucas Y (2007) Finding two optimal positions of a hand-held camera for the best reconstruction. In *2007 3Dtv Conference*, pp. 173–176.
Treuillet S, Albouy B and Lucas Y (2009) Three-dimensional assessment of skin wounds using a standard digital camera. *IEEE Trans Med Imaging* 28: 752–762.
Triebel R and Burgard W (2008) Recovering the shape of objects in 3D point clouds with partial occlusions. In *Field and Service Robotics: Results of the 6th International Conference*, pp. 13–22.
Trucco E, Umasuthan M, Wallace AM and Roberto V (1997) Model-based planning of optimal sensor placements for inspection. *IEEE Trans Robotics Automat* 13: 182–194.
Tsai TH and Fan KC (2007) An image matching algorithm for variable mesh surfaces. *Measurement* 40: 329–337.
Ulvklo M, Nygards J, Karlholm J and Skoglar P (2004) Image processing and sensor management for autonomous UAV surveillance (*Airborne Intelligence*). *Proc SPIE* 5409: 50–65.
Vazquez PP (2007) Automatic light source placement for maximum visual information recovery. *Comput Graphics Forum* 26: 143–156.
Wang P, Zhang ZM and Sun CK (2009) Framework for adaptive three-dimensional acquisition using structured light vision system. *J Vacuum Sci Technol B* 27: 1418–1421.
Wang PP and Gupta K (2006) A configuration space view of view planning. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1291–1297.
Wang PP and Gupta K (2007) View planning for exploration via maximal C-space entropy reduction for robot mounted range sensors. *Advanced Robotics* 21: 771–792.

Wang PP, Krishnamurti R and Gupta K (2007) View planning problem with combined view and traveling cost. In *IEEE International Conference on Robotics and Automation*, pp. 711–716.

Wang Y, Hussein II and Erwin RS (2008) Awareness-based decision making for search and tracking. In *2008 AMERICAN CONTROL CONFERENCE* , pp. 3169–3175.

Wenhardt S, Deutsch B, Angelopoulou E and Niemann H (2007) Active visual object reconstruction using D-, E-, and T-Optimal next best views. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2810–2816.

Whaite P and Ferrie FP (1997) Autonomous exploration: Driven by uncertainty. *IEEE Trans Pattern Anal Machine Intell* 19: 193–205.

Wheeler MD and Ikeuchi K (1995) Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Trans Pattern Anal Machine Intell* 17: 252–265.

Wong C and Kamel M (2004) Comparing viewpoint evaluation functions for model-based inspectional coverage. In *1st Canadian Conference on Computer and Robot Vision* , pp. 287–294.

Wu P, Suzuki H and Kase K (2005) Model-based simulation system for planning numerical controlled multi-axis 3D surface scanning machine. *JSME Int J Ser C Mech Syst Machine Elements Manufacturing* 48: 748–756.

Yang CC and Ciarallo FW (2001) Optimized sensor placement for active visual inspection. *J Robotic Syst* 18: 1–15.

Yao Y, Chen CH, Abidi B, Page D, Koschan A and Abidi M (2010) Can you see me now? Sensor positioning for automated and persistent surveillance. *IEEE Trans Syst Man Cybernet B* 40: 101–115.

Ye YM and Tsotsos JK (1999) Sensor planning for 3D object search. *Comput Vision Image Understand* 73: 145–168.

Yegnanarayanan V, Umamaheswari GK and Lakshmi RJ (2009) On a graph theory based approach for improved computer vision. In *Proceedings of the 2009 International Conference on Signal Processing Systems*, pp. 660–664.

Zavidovique B and Reynaud R (2007) The situated vision: a concept to facilitate the autonomy of the systems. *Traitement Signal* 24: 309–322.

Zetu D and Akgunduz A (2005) Shape recovery and viewpoint planning for reverse engineering. *Int J Adv Manufacturing Technol* 26: 1370–1378.

Zhang G, Ferrari S and Qian M (2009) An information roadmap method for robotic sensor path planning. *J Intelligent Robotic Syst* 56: 69–98.

Zhang ZG, Peng X, Shi WQ and Hu XT (2000) A survey of surface reconstruction from multiple range images. In *Proceedings of Systems Integrity and Maintenance*, pp. 519–523.

Zhou H and Sakane S (2003) Learning Bayesian network structure from environment and sensor planning for mobile robot localization. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems* , pp. 76–81.

Zhou XL, He BW and Li YF (2008) A new view planning method for automatic modeling of three dimensional objects. *Intell Robotics Appl I* 5314: 161–170.

Zingaretti P and Frontoni E (2006) Appearance based robotics. *IEEE Robotics Automat Mag* 13: 59–68.

Zussman E, Schuler H and Seliger G (1994) Analysis of the geometrical features detectability constraints for laser-scanner sensor planning. *Int J Adv Manufacturing Technol* 9: 56–64.