World Scientific Series in Robotics and Automated Systems – Vol. 7

# VISUAL SERVOING

## Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback

Edited by

**Koichi Hashimoto**
*Osaka University*



**World Scientific**
*Singapore • New Jersey • London • Hong Kong*

*Published by*

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 9128
*USA office:* Suite 1B, 1060 Main Street, River Edge, NJ 07661
*UK office:* 73 Lynton Mead, Totteridge, London N20 8DH

**VISUAL SERVOING**
**Real-Time Control of Robot Manipulators Based on Visual Sensory Feedback**



ISBN 981-02-1364-6

Printed in Singapore by Utopia Press.

# VISUAL SERVOING

**WORLD SCIENTIFIC SERIES IN ROBOTICS AND AUTOMATED SYSTEMS**

**Editor-in-Charge:** Prof T M Husband
(*Vice Chancellor, University of Salford*)

---

Vol. 1: Genetic Algorithms and Robotics — A Heuristic Strategy for Optimization (*Y Davidor*)

Vol. 2: Parallel Computation Systems for Robotics: Algorithms and Architectures (*Eds. A Bejczy and A Fijany*)

Vol. 3: Intelligent Robotic Planning Systems (*P C-Y Sheu*)

Vol. 4: Computer Vision, Models and Inspection (*A D Marshall and R R Martin*)

Vol. 5: Advanced Tactile Sensing for Robotics (*Ed. H R Nicholls*)

Vol. 6: Intelligent Control: Aspects of Fuzzy Logic and Neural Nets (*C J Harris, C G Moore and M Brown*)

Vol. 7: Visual Servoing (*Ed. K Hashimoto*)

Vol. 8: Modelling and Simulation of Robot Manipulators: A Parallel Processing Approach (*A Y Zomaya*)

***Forthcoming volumes:***

Vol. 9: Advanced Guided Vehicles – Aspects of the Oxford AGV Project (*S Cameron and P Probert*)

Vol. 10: Recent Trends in Mobile Robots (*Ed. Y F Zheng*)

Vol. 11: Cellular Robotics and Micro Robotics Systems (*T Fukuda and T Ueyama*)

# Preface

This book on visual servoing treats various aspects of automatic control of mechanical systems with visual sensory feedback. Image processing techniques and robot control schemes as well as the latest investigation on design of visual servo mechanism based on modern control schemes, e.g., *linear*, *nonlinear*, *adaptive*, *fuzzy logic* and *neural networks* are extensively described. New concepts for utilizing visual sensory information on real-time robotic control are discussed and the performance is evaluated through various simulations and real-time experiments.

This book consists of 11 articles by active leading researchers in the area of visual servoing. The authors describe the problems in visual servoing, possible approaches to solve the problems and future research subjects.

A review of this area and its extensive bibliography are presented by Peter Corke. He gives a clear and comprehensive summary of the research done in vision-related robotics. Two approaches have been proposed to realize visual servoing. One is *position-based* and the other is *feature-based*.

Position-based schemes predict the relative position and orientation between the object and the robotic hand. The prediction algorithm and the robot control strategies are discussed in the articles by Peter Allen *et al.* and William Wilson.

Feature-based schemes use the object features in the visual sensor output to generate robot control commands. There are four articles focusing on the feature-based approach. The article by John Feddema *et al.* gives a feature-based trajectory generator and criteria for image feature selection for robust visual feedback. The article by Brad Nelson *et al.* presents various aspects of the integration of visual servoing into robotic assembly systems. A control theoretic formulation of the visual servoing system is discussed and the linear quadratic as well as nonlinear controller design methods are developed in the article by Koichi Hashimoto and Hidenori Kimura. Classification of vision-based tasks and the task-function approach to visual servoing is described in the article by Francois Chaumette *et al.*

Problems on dynamic sensing with applications to robot juggling are discussed by Alfred Rizzi and Dan Koditschek. Peter Corke describes the hardware and software aspects which are necessary to realize the video-rate visual servoing. Fuzzy and neural network approaches are described by Il Hong Suh and Tae Won Kim. Finally, Zeungnam Bien and his group extensively discuss "what are features" and propose a fuzzy self-organizing scheme to visual servoing.

Firstly, I would like to thank the authors, who gladly agreed to contribute to this book, for their eagerness in producing good articles in quite a short period. Among them, special thanks are due to Dan Koditschek, Pradeep Khosla and Nick Papanikolopoulos for the valuable discussions during my visit to New Haven and Pittsburgh and also to Peter Corke for the helpful suggestions during the preparation of this book. I wish to express my gratitude for the helpful comments from Hidenori Kimura, Takumi Ebine, Takahiro Inoue and Saori Ogura of Osaka University.

I would also like to thank Professor T. M. Husband of Imperial College of Science, Technology and Medicine, who is also the Editor-in-Charge of the *Series on Robotics and Automated Systems*, for offering me the opportunity to edit this review volume for the series. And finally, I thank Dr. K.K. Phua and Ms. R. A. Hassan of World Scientific Publishing Company who encouraged me in preparing the manuscripts.

Koichi Hashimoto
Osaka, Japan
February 1993

# Contents

# VISUAL CONTROL OF ROBOT MANIPULATORS – A REVIEW

Peter I. Corke
CSIRO Division of Manufacturing Technology,
Preston, Victoria, 3072. Australia.

**Abstract**

This paper attempts to present a comprehensive summary of research results in the use of visual information to control robot manipulators and related mechanisms. An extensive bibliography is provided which also includes important papers from the elemental disciplines upon which visual servoing is based. The research results are discussed in terms of historical context, commonality of function, algorithmic approach and method of implementation.

# 1 Introduction

This paper presents the history, and reviews current research into the use of visual information for the control of robot manipulators and mechanisms. Visual control of manipulators promises substantial advantages when working with targets whose position is unknown, or with manipulators which may be flexible or inaccurate. The reported use of visual information to guide robots, or more generally mechanisms, is quite extensive and encompasses manufacturing applications, teleoperation, missile tracking cameras, fruit picking as well as robotic ping-pong, juggling, and balancing. Section 2 will introduce the topic of visual servoing and describe its relationship to other significant research areas.

Categorization of techniques is important in providing a structure for discussion. However in this field there are potentially many ways of classifying the reported results; fixed or end-effector-mounted cameras, monocular or binocular vision, planar or complete 3D motion control, algorithms for image processing, feature extraction and interpretation. The approach that has been adopted is to cover all reports of visual servoing, albeit briefly, in section 3, which is a comprehensive summary of literature on the topic. Sections 4 and 5 then discuss in greater detail the issues involved in position-based and image-based visual servoing respectively. The work of some researchers will thus be referred to several times in the paper.

Section 6 summarizes the variety of approaches used in implementing visual servoing. Finally, section 7 presents some conclusions. For further details of any particular technique the reader is always referred to the references. The bibliography is large, and attempts to encompass all research results in visual servoing, as well as important papers from the elemental disciplines upon which visual servoing is based.

# 2 Concepts of visual control

This section will introduce the important concepts of visual servoing and describe its relationship to other research areas such as active vision, and structure from motion. Terminology used in the paper will then be introduced, followed by a brief introduction to image-based and position-based visual servoing.

The use of vision with robots has a long history[110] and today vision systems are available from major robot vendors that are highly integrated with the robot's programming system. Capabilities range from simple binary image processing to more complex edge and feature based systems capable of handling overlapped parts.[16] However the feature in common with all these systems is that they are static, and typically image processing time is of the order of 0.1 to 1 second.

Traditionally visual sensing and manipulation are combined in an open-loop fashion, 'looking' then 'moving'. The accuracy of the operation depends directly on the accuracy of the visual sensor and the manipulator and its controller. An alternative to increasing the accuracy of these subsystems is to use a visual-feedback control loop, which will increase the overall accuracy of the system: a principle concern in any application. The term *visual servoing* appears to have been introduced by Hill and Park[46] in 1979 to distinguish their approach from earlier 'blocks world' experiments where the system alternated between picture taking and moving. Prior to the introduction of this term, the less specific term *visual feedback* was generally used.

Visual servoing is the fusion of results from many elemental areas including high-speed image processing, kinematics, dynamics, control theory, and real-time computing. It has much in common with research into *active vision* and *structure from motion*, but is quite different to the often described use of vision in hierarchical task-level robot control systems.

Some robot systems[60,65] which incorporate vision are designed for task level programming. Such systems are generally hierarchical, with higher levels corresponding to more abstract data representation and lower bandwidth. The highest level is capable of reasoning about the task, given a model of the environment. In general a look-then-move approach is used. Firstly, the target location and grasp sites are determined from calibrated stereo vision or laser rangefinder images, and then a sequence of moves are planned and executed. Vision sensors have tended to be used in this fashion because of the richness of the data they can produce about the world, in contrast to an encoder or limit switch which would be dealt with at the lowest level. Visual servoing is no more than the use of vision at the lowest level, with simple

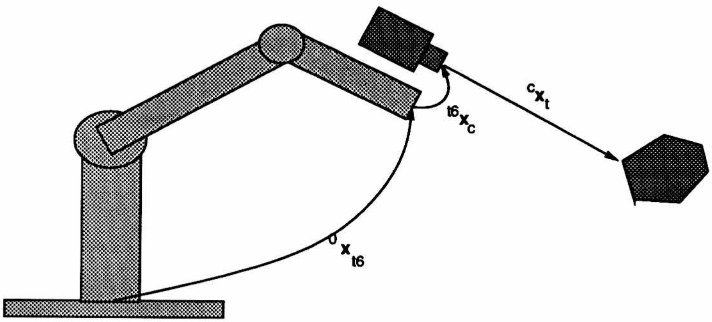


Figure 1: Relevant coordinate frames; world$^0$, end-effector$^{t6}$, camera$^c$ and target$^t$

image processing to provide reactive or reflexive behaviour.

Visual servoing has much in common with *active computer vision*,[13,8] which proposes that a set of simple visual behaviours can accomplish tasks through action, such as controlling attention or gaze.[21] The fundamental tenet of active vision is not to interpret the scene and then model it, but to direct attention to that part of the scene relevant to the task at hand. If the system wishes to learn something of the world, rather than consult a model, it should consult the world by directing the sensor. The benefits of an *active* robot-mounted camera include the ability to avoid occlusion, resolve ambiguity and increase accuracy.

Literature related to *structure from motion* is also relevant to visual servoing. Structure from motion attempts to infer the 3D structure and the relative motion between object and camera, from a sequence of images. In robotics however, we generally have considerable a priori knowledge of the target and the spatial relationship between feature points is known. Aggarwal[2] provides a comprehensive review of this active field.

## 2.1 Definitions

The task in visual servoing is to control the *pose* of the robot's end-effector, $\underline{x}_{t6}$, using visual information, *features*, extracted from the image. Pose, $\underline{x}$, is represented by a six element vector encoding position and orientation in 3D space. The camera may be fixed, or mounted on the robot's end-effector in which case there exists a constant relationship, ${}^{t6}\underline{x}_c$, between the pose of the camera and the pose of the end-effector. The image of the target is a function of the relative pose between the camera and the target, ${}^{c}\underline{x}_t$. Some relevant poses, are shown in Figure 1. The distance between the camera and target is frequently referred to as *depth* or *range*.

The camera contains a lens which forms a 2D projection of the scene onto the image

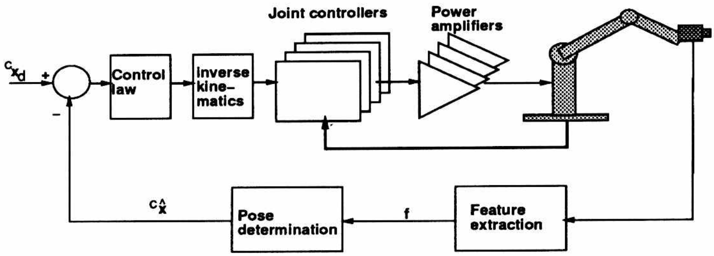


Figure 2: Dynamic position-based look-and-move structure

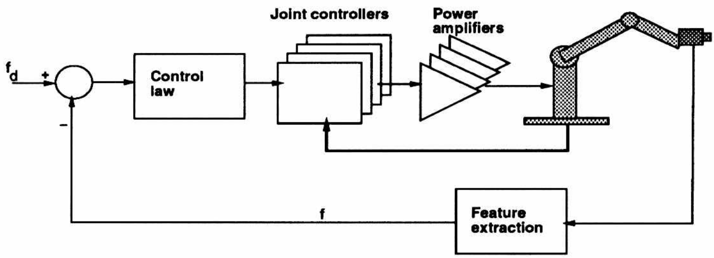


Figure 3: Dynamic image-based look-and-move structure

plane where the sensor is located. This projection causes direct depth information to be lost, and each point on the image plane corresponds to a ray in 3D space. Some additional information is needed to determine the 3D coordinate corresponding to an image plane point. This information may come from multiple views, or knowledge of the geometric relationship between several feature points on the target.

Robots typically have 6 degrees of freedom (DOF), allowing the end-effector to achieve any pose in 3D space. Visual servoing systems may control up to 6DOF. Planar positioning involves only 2DOF control, and may be sufficient for some applications.

A feature is defined generally as any measurable relationship in an image and examples include, moments, relationships between regions or vertices, polygon face areas, or local intensity patterns. Jang[53] provides a formal definition of features as image functionals. Most commonly the coordinates of a feature point or a region centroid are used. A good feature point is one that can be located unambiguously in different views of the scene, such as a hole in a gasket[34,33] or a contrived pattern.[77,27] Three or more features can be used to determine the pose (not necessarily uniquely, see Section 4.1) of the target relative to the camera, given knowledge of the geometric relationship between the feature points. A feature vector, $\underline{f}$, is a one dimensional

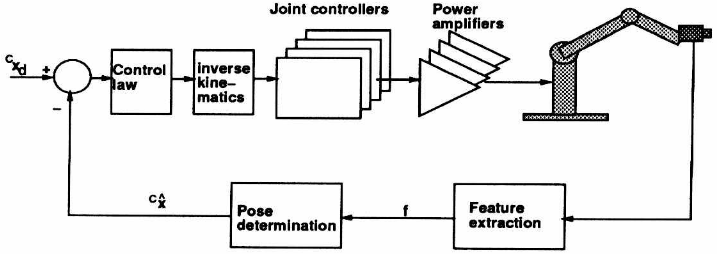


Figure 4: Position-based visual servo (PBVS) structure as per Weiss

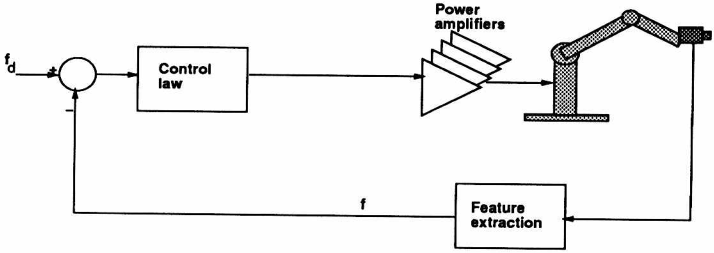


Figure 5: Image-based visual servo (IBVS) structure as per Weiss

vector containing feature information as described above.

## 2.2 Position versus image based servoing

Sanderson and Weiss[82] introduced an important classification of visual servo structures, and these are shown schematically in Figures 2 through 5. In position-based control, features are extracted from the image, and used in conjunction with a geometric model of the target to determine the pose of the target with respect to the camera. In image-based servoing the last step is omitted, and servoing is done on the basis of image features directly. The structures referred to as 'dynamic look and move' make use of joint feedback, whereas the PBVS and IBVS structures use no joint feedback information at all.

The image-based approach may reduce computational delay, eliminate the necessity for image interpretation and eliminate errors in sensor modeling and camera calibration. However it does present a significant challenge to controller design since the process is non-linear and highly coupled.

# 3 Summary

This section summarizes research and applications of visual servoing, from the pioneering work of the early 1970s to the present day. The discussion is generally chronological, but related applications or approaches are grouped together. Due to technological limitations of the time, some of the significant early work fails to meet the strict definition of visual servoing given above, and would now be classed as *look-then-move* robot control. Progress has however been rapid, and by the end of the 1970s systems had been demonstrated which were capable of 10Hz servoing and 3D position control for tracking, seam welding and grasping moving targets.

One of the earliest references is by Shirai and Inoue[84] in 1973 who describe how a visual feedback loop can be used to correct the position of a robot to increase task accuracy. A system is described which allows a robot to grasp a square prism and place it in a box using visual servoing. Edge extraction and line fitting is used to determine the position and orientation of the box. The camera is fixed, and a servo cycle time of 10s is reported.

Considerable work on the use of visual servoing was conducted at SRI International during the late 1970s. Early work[79,80] describes the use of visual feedback for bolt-insertion and picking moving parts from a conveyor. Hill and Park[46] describe visual servoing of a Unimate robot in 1979. Binary image processing is used for speed and reliability, and can provide planar position, and simple depth estimation from the apparent distance between known features. Experiments were also conducted using a projected light stripe to provide more robust depth determination as well as surface orientation. Their experiments demonstrated planar and 3D visually-guided motion, as well as tracking and grasping of moving parts. They also investigated some of the dynamic issues involved in closed-loop visual control. Prajoux[76] demonstrated visual servoing of a 2DOF mechanism for following a swinging hook. The system used a predictor to estimate the future position of the hook, and achieved settling times of the order of 1s.

A bolt-insertion task is also described by Geschke,[39] using stereo vision and a Stanford arm. The system features automated threshold setting, software image feature searching at 10Hz, and setting of position loop gains according to the visual sample rate. Stereo vision is achieved with a single camera and a novel mirror arrangement.

Simple hand-held light stripers of the type proposed by Agin[3] have been used in planar applications such as connector acquisition,[69] weld seam tracking,[20] and sealant application.[83] The latter lays a bead at 400mm/s with respect to a moving car-body, and shows a closed-loop bandwidth of 4.5Hz. More recently Venkatesan and Archibald[96] describes the use of two hand-held laser scanners for real-time 5DOF robot control.

Gilbert[41] describes an automatic rocket tracking camera which keeps the target centered in the camera's image plane by means of pan/tilt controls. The system uses video-rate image processing hardware to identify the target, and update the camera orientation at 60Hz. Dzialo and Schalkoff[31] discuss the effects of perspective on the control of a pan-tilt camera head for tracking.

Weiss[101] proposed the use of adaptive control for the non-linear time varying relationship between robot pose and image features in image-based servoing. Detailed simulations of image-based visual servoing are described, for a variety of manipulator structures of up to 3DOF.

Stability, accuracy and tracking speed for a Unimate-based visual-servo system are discussed by Makhlin.[63] Coulon and Nougaret[26] address similar issues and also provide a detailed imaging model for the vidicon sensor's memory effect. They describe a digital video processing system for determining the location of one target within a processing window, and use this information for closed-loop position control of an XY mechanism to achieve a settling time of around 0.2s to a step demand.

Weber and Hollis[100] developed a high-bandwidth planar-position controlled micro-manipulator. It is required to counter room and robot motor vibration effects with respect to the workpiece in a precision manufacturing task. Correlation is used to track workpiece texture. To achieve the high sample rate, yet maintain resolution, two orthogonal linear CCDs are used to observe projections of the image. Since the sample rate is high, 300Hz, the image shift between samples is small and reduces the size of the correlation window needed.

Image projections are also used by Kabuka.[55] Fourier phase differences in the vertical and horizontal binary image projections are used for centering a target in the image plane, and determining its rotation. This is applied to control of a two-axis camera platform using an IBM-PC/XT,[55] and takes 30s to settle on a target. An extension to this approach[56] uses adaptive control techniques to minimize performance indices on grey-scale projections. The approach is presented generally, but with simulations for planar positioning only.

Road vehicle guidance is described by Dickmanns.[28] An experimental system using real-time feature tracking and gaze controlled cameras, has guided a 5 ton experimental road vehicle at speeds of up to 96km/h along 20km of test track. Control of underwater robots using visual reference points[70] has also been proposed.

Visually guided machines have been built to emulate human skills at ping-pong,[11] juggling,[78] inverted pendulum balancing[28,9] and controlling a labyrinth game.[9] The latter is a wooden board mounted on gimbals on which a ball bearing rolls, the aim being to move the ball through a maze and not fall into a hole. The ball's position is observed at 40ms intervals, and a Kalman filter is used to reconstruct the ball's state. State feedback control gives a closed loop bandwidth of 1.3Hz. The ping-pong playing robot[11] does not use visual servoing, rather a model of the ball's trajectory is built and input to a dynamic path planning algorithm which attempts to strike the ball.

Visual servoing has also been proposed for catching flying objects on Earth or in space. Bukowski *et al.*[18] report the use of a Puma 560 to catch a ball with an end-effector mounted net. The robot is guided by a fixed-camera stereo-vision system and a 386 PC. Skofteland *et al.*[86] discuss capture of a free-flying polyhedron in space with a vision guided robot. Skaar *et al.*[85] uses as an example a 1DOF robot to catch a ball. Lin *et al.*[62] proposes a two-stage algorithm for catching moving targets; coarse positioning to approach the target in near-minimum time and 'fine tuning' to match robot acceleration and velocity with the target.

There have been several reports of the use of visual servoing for grasping moving targets. The earliest work appears to have been at SRI in 1978.[80] Recently Zhang *et al.*[109] present a tracking controller for visually servoing a robot to pick items from a fast moving conveyor belt (300mm/s). The camera is hand-held and the visual update interval used is 140ms. Allen *et al.*[7] use a 60Hz fixed-camera stereo vision system, to track a target moving at 250mm/s. Later work[6] extends this to grasping a toy train moving on a circular track. Houshangi[48] uses a fixed overhead camera, and a visual sample interval of 196ms, to enable a Puma 600 robot to grasp a moving target.

Fruit picking is a non-manufacturing application of visually guided grasping, where the target may be moving. Harrell[42] describes a hydraulic fruit-picking robot which uses visual servoing to control 2DOF as the robot reaches toward the fruit, prior to picking. The visual information is augmented by ultrasonic sensors to determine distance during the final phase of fruit grasping. The visual servo gains are continuously adjusted to account for changing camera target distance. This last point is significant but mentioned by few authors.[22,31]

Part mating has also been investigated using visual servoing. Ahluwalia and Fogwell[4] describe a system for mating two parts, each held by a robot and observed by a fixed camera. Only 2DOF for the mating are controlled, and a Jacobian approximation is used to relate image-plane corrections to robot joint-space actions. On a larger scale, a visually servoed robot has used for mating an umbilical connector from the service gantry to the US Space Shuttle.[27]

A number of dynamic effects become important at high sample rate with an eye-in-hand configuration, but effects such as oscillation and lag tend to be mentioned only in passing.[33,53] Corke[24,25] describes a system capable of 60Hz planar positioning. An image-based control scheme is used to close the robot's position loop, and independent control is used for each Cartesian DOF – no trajectory generator is used. The dynamics of this configuration are investigated and modeled. Later work[22] investigates the effect of varying camera object-distance on the closed-loop performance, and the direct use of image moments for orientation control.[23]

The image-based servo approach has been investigated experimentally by a number of researchers, but unlike Weiss they use closed-loop joint control, see Figure 3. Feddema[34,33,32] uses an explicit feature-space trajectory generator and closed-loop joint control, to overcome problems due to low visual sampling rate. Experimental work demonstrates image-based visual servoing for 4DOF. Rives *et al.*[77,19] describe a similar approach using the task function method,[81] and show experimental results for robot positioning using a target with four circle features. Hashimoto *et al.*[44] present simulations to compare position-based and image-based approaches, and experiments demonstrate image-based servoing of a Puma 560 tracking a target moving in a circle at 30mm/s – the visual servo interval is 250ms. Jang *et al.*[52] describe a generalized approach to servoing on image features, with trajectories specified in feature space – leading to trajectories (tasks) that are independent of target geometry. Experiments demonstrate tracking with a Puma 560, but some lag is evident.

Westmore and Wilson[102] demonstrate 3DOF planar tracking and achieve a settling time of around 0.5s to a step input. This is extended[99] to determine the 3D pose of the target using extended Kalman filtering. Experiments verify the pose determination, but not closed-loop control. Papanikolopoulos *et al.*[73] demonstrates tracking of a target undergoing planar motion with the CMU DDArm II robot system. Later work[74] demonstrates 3D tracking of static and moving targets, and adaptive control is used to estimate the target distance.

The use of visual servoing in a telerobotic environment has been discussed by Yuan *et al.*,[108] Papanikolopoulos *et al.*[74] and Tendick*et al.*.[93] Visual servoing can allow the task to be specified by the human operator in terms of visual features selected as a reference for the task.

Approaches based on neural networks,[66,43,59] and general learning algorithms,[67] have been used to achieve robot hand-eye coordination. A fixed camera observes objects and the robot within the workspace, and can learn relationship between robot joint angles and 3D position of the end-effector. Such systems require training, but the need for complex analytic relationships between image features and joint angles is eliminated.

# 4 Position-based visual servoing

A broad definition of position-based servoing will be adopted that includes all methods, whether based on analysis of features or 3D sensors, that determine the relative pose of the target in order to guide the robot. The simplest form of visual servoing involves robot motion in a plane orthogonal to the optical axis of the camera and can be used for tracking planar motion such as a conveyor belt. However tasks such as grasping and part mating require control over the relative distance and orientation to the target.

Humans use a variety of vision-based depth cues including texture, perspective, stereo disparity, parallax, occlusion and shading. For a moving observer, apparent motion of features is an important depth cue. The use of multiple cues, selected according to visual circumstance, helps to resolve ambiguity. Approaches suitable for computer vision are reviewed by Jarvis.[54]

Active range sensors project a controlled energy beam, generally ultrasonic or optical, and detect the reflected energy. Commonly a pattern of light is projected on the scene, which a vision system interprets to determine depth and orientation of the surface. Such sensors usually determine depth along a single stripe of light,