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

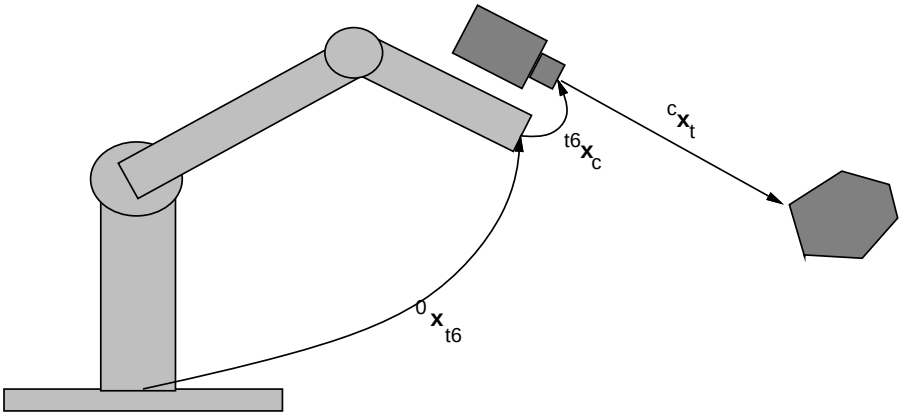

Figure 1: Relevant coordinate frames; world^0, end-effector^{t6}, camera^c and target^t

.

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

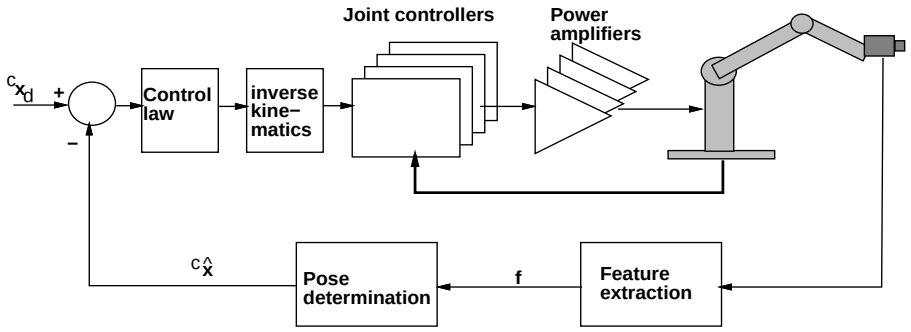

Figure 2: Dynamic position-based look-and-move structure

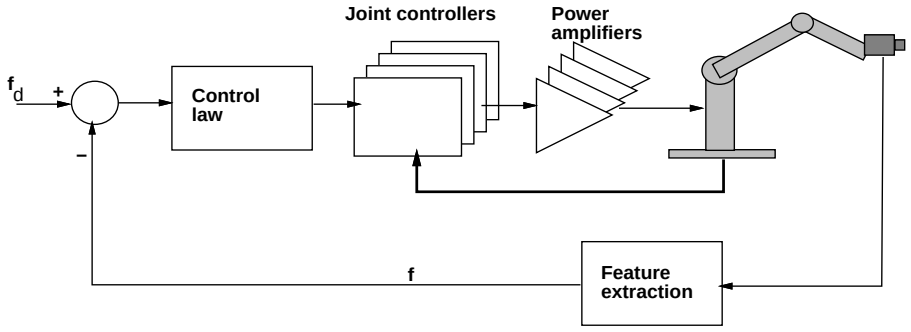

Figure 3: Dynamic image-based look-and-move structure

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

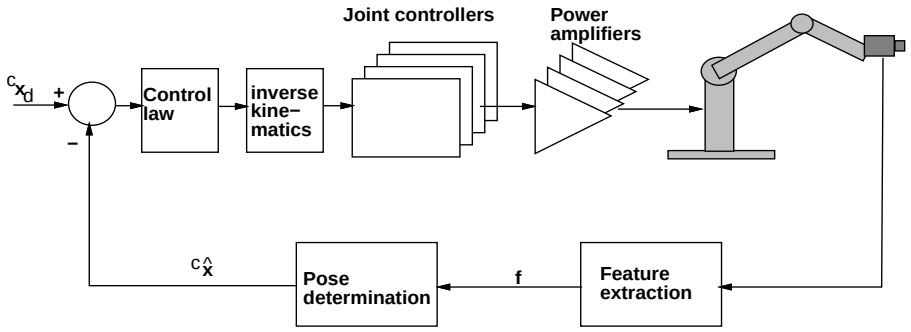

Figure 4: Position-based visual servo (PBVS) structure as per Weiss

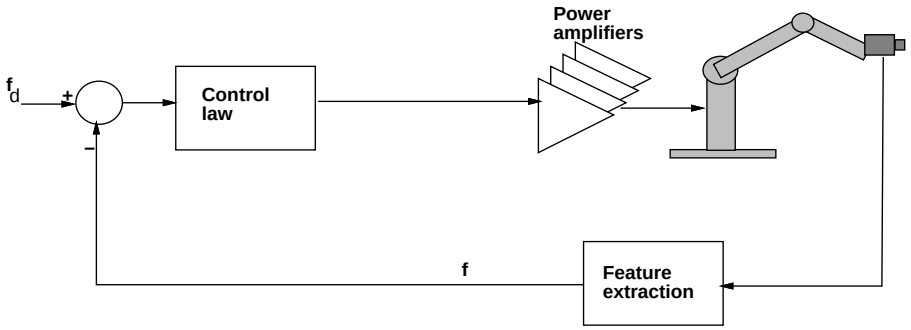

Figure 5: Image-based visual servo (IBVS) structure as per Weiss

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 micromanipulator. 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,

multiple stripes or a dense grid of points. If the sensor is small and mounted on the robot[3, 96, 30] the depth and orientation information can be used for servoing. Besl[14] provides a comprehensive survey which includes the operation and capability of many commercial active range sensors.

In contrast, passive techniques rely only on observation under ambient illumination to determine the depth or pose of the object. Some approaches relevant to visual servoing will be discussed in the following sections.

4.1 Photogrammetric techniques

Photogrammetry is the science of obtaining information about physical objects via photographic images.[105] Close-range, or terrestrial, photogrammetry is concerned with target distances less than 100m from the camera, and is thus highly relevant to the task of determining the relative pose of a target.

In order to determine the 3D pose of an object from 2D image plane coordinates some additional information is needed. This data includes knowledge of the relationship between the observed feature points (perhaps from a CAD model), and also the camera's calibration parameters.

The perspective imaging model of the lens and sensor is characterized by two sets of parameters, referred to as *intrinsic* and *extrinsic* parameters.[105] The intrinsic parameters include focal length, pixel scaling factors and the coordinate of the optical axis on the image plane. The extrinsic parameters specify the pose of the camera in the world coordinate space. Camera calibration is the process of determining these parameters,[61, 94, 90, 64] which are generally expressed in the form of a 3×4 homogeneous transformation matrix, known as the *calibration matrix*. The inverse problem, *camera location determination*, is to find the camera pose given an image, known relationship between feature points, and the intrinsic calibration parameters.

In the photogrammetric approach, the image plane coordinates of a number of image features, $\underline{f}$, plus knowledge of the camera's intrinsic calibration parameters, are used to solve for the camera's pose relative to the target, ${}^{c}\underline{x}$. It can be shown that at least three feature points are needed to solve for pose. Intuitively, the coordinate of each feature point yields two equations, and six equations are needed to solve for the six elements of the pose vector. In fact three feature points yield multiple solutions, but four coplanar points yields a unique solution.

Analytic solutions for three and four points are given by Fischler and Bolles[35] and Ganapathy.[36] Unique solutions exist for four coplanar, but not collinear, points (even for partial intrinsic parameters).[36] Six or more points always yield unique solutions, as well as the intrinsic camera calibration parameters. A technique that directly solves for the pose of a quadrangle (four coplanar feature points) is given by Hung *et al.*,[49] but is not robust in the presence of noise. Alternatively the camera calibration matrix can be computed from features on the target, then decomposed[89, 37] to yield the camera's pose.

Yuan[107] describes a general iterative solution independent of the number or distribution of feature points. For tracking moving targets the previous solution can be used as the initial estimate for iteration. Wang and Wilson[99] uses an extended Kalman filter to update the pose estimate given measured image plane feature locations. The filter convergence is analogous to the iterative solution.

Once the target pose relative to the camera is determined, it is then necessary to determine the target pose relative to the robot's end-effector. A technique to determine the transformation between robot end-effector and the camera frame is given by Tsai and Lenz.[95]

The cited drawbacks of the photogrammetric approach are the complex computation, and the necessity for camera calibration and a model of the target. None of these objections are overwhelming, and systems based on this principle have been demonstrated using iteration,[108] Kalman filtering,[102] and analytic solution.[38]

4.2 Stereo vision

Stereo vision[106] is the interpretation of two views of the scene taken from known different viewpoints, to resolve depth ambiguity. The location of feature points in one view must be matched with the location of the same feature points in the other view. This matching, or correspondence, problem is not trivial, and is subject to error. Another difficulty is the missing parts problem, where a feature point is visible in only one of the views, and therefore its depth cannot be determined.

Matching may be done on a few feature points, such as region centroids, corner features, or on fine feature detail such as surface texture. In the absence of significant surface texture, a random texture pattern can be projected onto the scene.

Implementations of 60Hz stereo-vision systems have been described by Andersson,[11] Rizzi *et al.*,[78] Allen *et al.*[7] and Bukowski *et al.*.[18] The first two operate in a simple contrived environment, with a single white target against a black background. The last two use optical flow or image differencing to eliminate static background detail. All use fixed, rather than end-effector-mounted cameras.

4.3 Depth from motion

Closely related to stereo vision is *monocular* or *motion stereo*,[71] also known as *depth from motion*. Sequential monocular views, taken from different viewpoints, are interpreted to derive depth information. Such a sequence may be obtained from a robot hand-mounted camera during robot motion. It must be assumed that targets in the scene do not move significantly between the views. Vernon and Tistarelli[97] use two contrived trajectories, one along the optical axis, and one rotation about a fixation point, to construct a depth map of a bin of components. The AIS visual-servoing scheme of Jang *et al.*[53] reportedly uses motion stereo to determine depth of feature points.

Self motion, or egomotion, produces rich depth cues from the apparent motion of features, and is important in biological vision.[98] Dickmanns[29] proposes an integrated spatio-temporal approach to analyzing scenes with relative motion so as to determine depth and structure. Based on tracking features between sequential frames, he terms it *4D vision*.

Research into insect vision[87] indicates that insects use self-motion to infer distances to targets for navigation and obstacle avoidance. Compared to mammals, insects have relatively simple, but effective visual systems, which may be a practical alternate model upon which to base future robot-vision systems.[88]

Fixation occurs when a sequence of images is taken with the camera moving so as to keep one point in the scene, the interest point, at the same location in the image plane. Knowledge of camera motion during fixation can be used to determine the 3D position of the target. Such a mechanism operates in animal vision. For example, when we follow a moving target we move our head and eyes in order to keep the image of a point of interest stationary on the retina. Stabilizing one point in a scene that is moving relative to the observer induces the target to stand out from the non-stabilized and blurred parts of the scene, and is thus a basis for scene segmentation. Coombs and Brown[21] describes a binocular system capable of fixating on a target, and some of the issues involved in control of the camera motion.

4.4 Depth from dynamics

Many reported experiments utilize fixed camera-target distances – perhaps due to focus or depth of field problems. The closed-loop transfer function of an image-based eye-in-hand visual servo system includes a gain term due to perspective.[22, 31, 42] Loop gain, and thus the closed-loop response is a function of the distance between the end-effector-mounted camera, and the target feature. Harrell[42] uses depth measured by an ultrasonic sensor to set an appropriate gain for the visual-servo control loop. Conversely, the identified closed-loop dynamics can be used to derive an estimate of depth[22] for a single point. More usefully, adaptive control or a self-tuning regulator would maintain the desired dynamic response as target distance changed, and the parameter values would be a function of target distance.

5 Image based servoing

Image-based visual servo control uses the location of features on the image plane directly for feedback. For example, consider Figure 6, where it is desired to move the robot so that the camera's view changes from initial to final view, and the feature vector from $\underline{f}_0$ to $\underline{f}_d$. $\underline{f}$ may comprise coordinates of vertices, or areas of the faces. Implicit in $\underline{f}_d$ is that the robot is normal to, and centered over the face of the cube, at the desired distance. Elements of the task are thus specified in image space, not

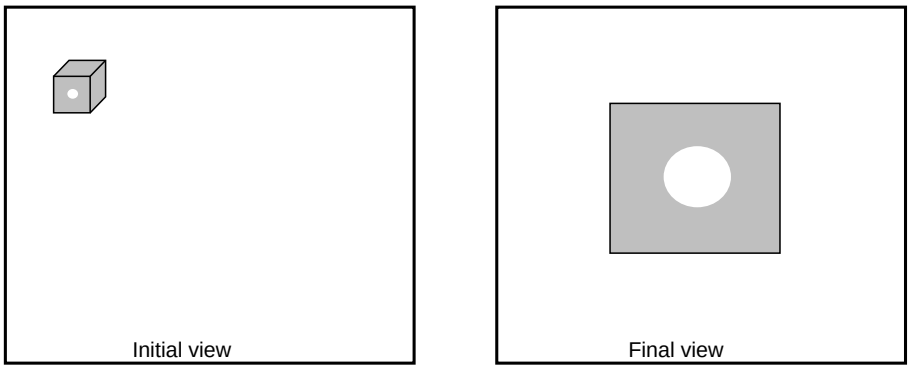

Figure 6: Example of initial and desired view of a cube

world space. Skaar *et al.*[85] propose that many real-world tasks may be described by one or more camera-space tasks, for instance by aligning visual cues in the scene.

For a robot with an end-effector-mounted camera, the viewpoint, and hence the features will be a function of the relative pose of the camera to the target, ${}^{c}\underline{x}_t$. In general this function is non-linear and cross-coupled such that motion of one end-effector DOF will result in the complex motion of many features. For example, camera rotation can cause features to translate horizontally and vertically on the image plane. This relationship

$$\underline{f} = f({}^{c}\underline{x}_t) \tag{1}$$

may be linearized about the operating point

$$\delta\underline{f} = {}^{f}\mathbf{J}_c({}^{c}\underline{x}_t)\ \delta{}^{c}\underline{x}_t \tag{2}$$

where

$${}^{f}\mathbf{J}_c({}^{c}\underline{x}_t) = \frac{\partial \underline{f}}{\partial {}^{c}\underline{x}_t} \tag{3}$$

is a Jacobian matrix, relating rate of change in pose to rate of change in feature space. This Jacobian is referred to variously as the *feature Jacobian*, *image Jacobian*, *feature sensitivity matrix*, or *interaction matrix*. Assume for the moment that the Jacobian is square and non-singular, then

$${}^{c}\dot{\underline{x}}_t = {}^{f}\mathbf{J}_c^{-1}({}^{c}\underline{x}_t)\ \dot{\underline{f}} \tag{4}$$

and a simple proportional control law

$${}^{c}\dot{\underline{x}}_t(t) = \mathbf{k}\ {}^{f}\mathbf{J}_c^{-1}({}^{c}\underline{x}_t)\ \left(\underline{f}_d - \underline{f}(t)\right) \tag{5}$$

will tend to move the robot towards the desired feature vector. $\mathbf{k}$ is a diagonal gain matrix, and (t) indicates a time varying quantity.

Pose rates ${}^{c}\dot{\underline{x}}_t$ may be converted to robot end-effector rates via a constant Jacobian, ${}^{t6}\mathbf{J}_c$ derived from the relative pose[75] between the end-effector and camera, ${}^{t6}\underline{x}_c$. A

technique to determine the transformation between the robot's end-effector and the camera frame is given by Tsai and Lenz.[95]

In turn, the end-effector rates may be converted to manipulator joint rates using the manipulator's Jacobian[104]

$$\dot{\underline{\theta}} = {}^{t6}\mathbf{J}_{\theta}^{-1}(\underline{\theta})\ {}^{t6}\dot{\underline{x}}_t \tag{6}$$

where $\underline{\theta}$ represent the joint angles of the robot. The complete equation is

$$\dot{\underline{\theta}}(t) = \mathbf{k}\ {}^{t6}\mathbf{J}_{\theta}^{-1}(\underline{\theta})\ {}^{t6}\mathbf{J}_{c}\ {}^{f}\mathbf{J}_{c}^{-1}({}^{c}\underline{x})\ \left(\underline{f}_d - \underline{f}(t)\right) \tag{7}$$

Such a closed-loop system is relatively robust in the presence of image distortions[26] and kinematic parameter variations in the manipulator Jacobian.[68]

A number of researchers have demonstrated results using this image-based approach to visual servoing. The significant problem is computing or estimating the feature Jacobian, and a variety of approaches will be described next.

5.1 Approaches to image-based visual servoing

The proposed IBVS structure of Weiss, Figure 5, controls robot joint angles directly using measured image features. The non-linearities include the manipulator kinematics and dynamics as well as the perspective imaging model. Adaptive control is proposed, since the gain, ${}^{f}\mathbf{J}_{c}^{-1}({}^{c}\underline{\theta})$, is pose dependent. The changing relationship between robot pose, and image feature change is learned during the motion. Weiss uses independent single-input single-output (SISO) model-reference adaptive control (MRAC) loops for each DOF, citing the advantages of modularity and reduced complexity compared to multi-input multi-output (MIMO) controllers. The proposed SISO MRAC requires one feature to control each joint and no coupling between features, and a scheme is introduced to select features so as to minimize coupling. This last constraint is difficult to meet in practice; for 6DOF servoing it is not possible to avoid coupling between camera rotation and the translation of features.

Weiss[101] presents detailed simulations of various forms of image-based visual servoing with a variety of manipulator structures of up to 3DOF. Sample intervals of 33ms and 3ms are investigated, as is control with measurement delay. With non-linear kinematics (revolute robot structure) the SISO MRAC scheme has difficulties. Solutions proposed, but not investigated, include MIMO control and a higher sample rate, or the dynamic-look-and-move structure, Figure 3.

Weiss found that even for a 2DOF revolute mechanism a sample interval less than 33ms was needed to achieve satisfactory plant identification. For manipulator control Paul[75] suggests that the sample rate should be at least 15 times the link structural frequency. Since the highest sample frequency achievable with standard cameras and image processing hardware is 60Hz, the IBVS structure is not currently practical for visual servoing. The so called dynamic look and move structure, Figure 3, is more

suitable for control of 6DOF manipulators, by combining high-bandwidth joint level control in conjunction with a lower rate visual position control loop. Such a structure is used by Feddema,[33] and others.[25,44,52]

Feddema extends the work of Weiss in many important ways, particularly by experimentation.,[33,34,32] and cites the difficulties of Weiss's approach as

- the assumption that vision update interval, T, is constant, and
- that $T \geq 33$ms which is greater than the sub millisecond period needed to control robot dynamics.

Due to the low speed feature extraction achievable (every 70ms) an explicit trajectory generator operating in feature space is used, rather than the pure control loop approach of Weiss. Feature velocities from the trajectory generator are resolved to manipulator configuration space for individual closed-loop joint PID control.

Feddema[34,33] describes a 4DOF servoing experiment where the target was a gasket containing a number of circular holes. Binary image processing was used, and Fourier descriptors of perimeter chain codes were used to describe each hole feature. From the two most unique circles, 4 features are derived; X and Y coordinates of the midpoint between the two circles, the angle of the midpoint line with respect to image coordinates, and the distance between circle centers. It is possible to write the feature Jacobian in terms of these features, that is ${}^{f}\mathbf{J}_c(\underline{f})$, though this is not generally the case. The experimental system could track the gasket, moving on a turntable at up to 1rad/s. The actual position lags the desired position, and 'some oscillation' is reported due to time delays in the closed-loop system.

A similar experimental setup,[33] used the centroid coordinates of three gasket holes as features. This more typical case does not allow the Jacobian to be formulated directly in terms of the measured features. Two approaches to evaluating the Jacobian are described. Firstly[32] the desired pose is used to compute the Jacobian, which is then kept constant. This is satisfactory as long as the initial pose is close to that desired. Secondly,[32,101] the pose is explicitly solved using photogrammetric techniques and used to compute the Jacobian. Simulation of 6DOF image based servoing[33] required determination of pose at each step to update the Jacobian. This appears more involved than pure position based servoing.

Rives *et al.*[77,19] describe an approach that computes the camera velocity screw as a functions of feature values. Based on the task function approach, the task is defined as the problem of minimizing $\| e({}^{c}\underline{x}_{t6}(t)) \|$. For visual servoing the task function is written in terms of image features $\underline{f}$ which in turn are a function of robot pose .

$$e(\underline{x}_{t6}(t)) = \mathbf{C}\left(\underline{f}(\underline{x}_{t6}(t)) - \underline{f}_d\right) \tag{8}$$

$\mathbf{C}$ is chosen as $\mathbf{C} = \mathbf{L}^{T+}$ to ensure convergence: where $\mathbf{L}^T = \frac{\partial \underline{f}}{\partial {}^{c}\underline{x}_{t6}(t)}$ is referred to as the interaction matrix, and $^+$ denotes the generalized inverse. As previously it is

necessary to know the model of the interaction matrix for the visual features selected and the cases of point clusters, lines and circles are derived. Experimental results for robot positioning using four point features are presented.[19]

Frequently the feature Jacobian can be formulated in terms of features plus depth. Hashimoto *et al.*[44] estimates depth explicitly based on analysis of features. Papanikolopoulos[74] estimates depth of each feature point in a cluster using an adaptive control scheme. Rives *et al.*[77,19] set the desired distance rather than update or estimate it continuously.

Feddema describes an algorithm[32] to select which three of the seven measurable features gives best control. Features are selected so as to achieve a balance between controllability and sensitivity with respect to changing features. The generalized inverse of the feature Jacobian[44,52,77] allows more than 3 features to be used, and has been shown to increase robustness, particularly with respect to singularities.[52]

Jang *et al.*[53] introduce the concepts of augmented image space (AIS) and transformed feature space (TFS). AIS is a 3D space whose coordinates are image plane coordinates plus distance from camera, determined from motion stereo. In a similar way to Cartesian space, trajectories may be specified in AIS. A Jacobian may be formulated to map differential changes from AIS to Cartesian space, and then to manipulator joint space. The TFS approach appears to be very similar to the image-based servoing approach of Weiss and Feddema.

Bowman and Forrest[17] describe how small changes in image plane coordinates can be used to determine differential change in Cartesian camera position, which is used for visual servoing a small robot. No feature Jacobian is required, but the camera calibration matrix is needed.

Most of the above approaches require analytic formulation of the feature Jacobian given knowledge of the target model. This process could be automated, but there is attraction in the idea of a system that can 'learn' the non-linear relationship automatically, as originally envisaged by Sanderson and Weiss.

Skaar *et al.*[85] describes the example of a 1DOF robot catching a ball. By observing visual cues such as the ball, the arm's pivot point, and another point on the arm, the interception task can be specified, even if the relationship between camera and arm is not known a priori. This is then extended to a multi-DOF robot, where cues on each link and the payload are observed. After a number of trajectories, the system 'learns' the relationship between image-plane motion, and joint-space motion, effectively estimating a feature Jacobian. Tendick *et al.*[93] describe the use of a vision system to close the position loop on a remote slave arm with no joint position sensors. A fixed camera observes markers on the arm's links, and a numerical optimization is performed to determine the robot's pose.

Miller[67] presents a generalized learning algorithm based on the CMAC structure proposed by Albus[5] for complex or multi-sensor systems. The CMAC structure is table driven, indexed by sensor output to determine the system command. The modified CMAC is indexed by sensor output as well as the desired goal state. Experimental

results are given for control of a 3DOF robot with a hand-held camera. More than 100 trials were required for training, and good positioning and tracking capability were demonstrated. Artificial neural techniques can also be used to learn the non-linear relationships between features and manipulator joint angles, as discussed by Kuperstein,[59] Hashimoto[43] and Mel.[66]

6 Implementational issues

Progress in visual servoing is related to technological advances in diverse areas including sensors, image processing and robot control. This section summarizes some of the implementational details reported in the literature.

6.1 Video standards

Since the television and surveillance industries dominate the manufacture and consumption of video equipment, most cameras used for machine vision work conform to broadcast video standards. These standards incorporate interlacing, an artifact of technological history, introduced to reduce human perception of flicker on TV screens. An interlaced video signal comprises pairs of sequential half-vertical-resolution fields, displaced vertically by one line. High frame rate and non-interlaced cameras are available, but are expensive due to low demand, and require specialized digitization hardware.

The two fields comprising one frame are exposed at different times, separated by one field time. Thus full frame images of rapidly moving targets can be difficult to interpret. Field-rate processing offers advantages of reduced blur, and higher sample rate, 50Hz for CCIR or 60Hz for RS170. To be rigorous when using field rate data, camera calibrations should be made for each field due to the vertical offset, but in practice this is rarely done.

6.2 Cameras

The earliest reports used vidicon, or thermionic tube, image sensors. These devices had a number of undesirable characteristics including large size and weight, lack of robustness and image stability, and memory effect.[26]

Since the mid 1980s most researchers have used some form of solid state camera, either NMOS, CCD or CID. The only reference to color vision for visual servoing is the fruit picking robot,[42] where color is used to differentiate fruit from the leaves. Given the real-time constraints and with current technology, any advantages color vision may offer is offset by the increased cost and the processing requirements of three times the monochrome data rate. The following discussion will be limited to 2D, or area sensors, though line-scan sensors have been used for visual servoing.[100]

All solid state sensors comprise a rectangular array of photosites, and each site accumulates a charge related to the time integrated incident illumination at that point. The three major types of solid state area imaging sensor are;

1. CCD (Charge Coupled Device). A CCD sensor contains a transport mechanism to shift the charge from the photosite to the output amplifier. During the read-out phase, the charge packets are kept in either the vertical transport registers, for interline transfer devices, or in an on-chip framestore, for frame transfer devices. Since packets of charge must be moved about the chip, there are paths by which excess charge from high illumination can 'leak' out, manifesting itself as 'smearing' or 'blooming'.

2. NMOS, or photodiode. An array of photosites in which each site is sequentially accessed (in raster order) for a destructive read of accumulated charge.

3. CID (Charge Injection Device). A CID sensor is very similar to the NMOS sensor except that the charge at the photosite can be read non-destructively. Charge injection clears the accumulated charge, and may be inhibited, allowing for some control over exposure time.

The most significant difference between the CCD and other sensors, is that the CCD sensor samples all photosites simultaneously, when the photosite charge is transferred to the transport registers. With the other sensor types, each pixel is exposed over the field-time prior to its being read out. This means that a pixel at the top of the frame is exposed over a substantially different time interval to a pixel in the middle of the frame. This can present a problem in scenes with rapidly moving targets and is discussed by Andersson.[11] Andersen *et al.*[9] discusses the analogous sampling problem for a vidicon image sensor, and proposes a modified Z-transform approach.

In the discussion above it is assumed that the photosites are being charged, or integrating, for one whole field time. When high relative motion exists between camera and scene, this long integration, or exposure time, will cause image blur. Motion blur tends to spread light over many pixels which may cause the intensity to fall below the threshold,[11,25] so that the system loses 'sight' of the target, leading to 'rough' motion. A conventional film camera uses a mechanical shutter to expose the film for a very short period of time relative to the scene dynamics. Electronic shuttering can be achieved on CCD sensors by discharging the photosites shortly before the end of field, and only the charge integrated over the short remaining period is transferred to the transport registers. The amount of accumulated charge is reduced with exposure time, and this in turn reduces the signal to noise ratio of the image.

6.3 Lenses

Lenses can introduce a number of geometric distortions to the image, the most significant of which is radial distortion where points are displaced along radial lines in the

image. This effect is worse near the edges of the image. Photogrammetric approaches in particular must model this distortion and correct for it. In general, image-based visual servo systems are robust in the presence of lens distortion.

Maintaining focus over a wide range of camera object distances is also difficult; a large depth of field is desirable, or a lens with servo controlled focus could be employed. Large depth of field is achieved by using a small aperture, but at the expense of light falling on the sensor. The requirement for large depth of field and short exposure time to eliminate motion blur both call for increased ambient illumination, or a sensitive sensor devices.

6.3.1 Camera location

Cameras can be either fixed or mounted on the robot's end-effector. All reported stereo-vision systems use fixed cameras, although there is no reason a stereo-camera cannot be mounted on the end-effector, apart from practical considerations such as payload limitation, or lack of robustness of the camera system. The benefits of an end-effector-mounted camera include the ability to avoid occlusion, resolve ambiguity and increase accuracy, by directing its attention.

Zhang *et al.*[109] propose that prediction is always required for grasping since an overhead camera will be obscured by the gripper, and a gripper mounted camera will be out of focus during the final phase of part acquisition. A camera can be mounted remotely and a fibre optic bundle used to carry the image from near the end-effector.[91] Given the small size and cost of modern CCD cameras this approach is not particularly advantageous.

6.4 Image processing

The enormous amount of data produced by vision sensors (typically 6Mbyte/s) and technological limitations in processing that data rapidly, has meant that vision has not been widely exploited as a sensing technology when high measurement rates are required. A vision sensor's output data rate, for example, is several orders of magnitude greater than that of a force sensor for the same sample rate. Thus special purpose hardware is needed for the early stages of image processing in order to reduce the data rate to something manageable by a conventional computer.

The real-time systems described in the literature generally perform minimal image processing, and the scenes are contrived to be simple to interpret, such as a single white target on a black background.[11,78] Image processing consists of thresholding[11,78,34,25] followed by centroid determination. Selection of a threshold is a practical issue that must be addressed, and many techniques have been suggested.[103]

General scenes have too much 'clutter' and are difficult to interpret at video rates. Harrell[42] describes a vision system which uses software to classify pixels by color so as to segment citrus fruit from the surrounding leaves. Allen[7,6] uses optical flow calculation to extract only moving targets in the scene, thus eliminating stationary

background detail. The Horn and Schunk optical flow algorithm[47] is implemented on a NIST PIPE[58] processor for each camera in the stereo pair. The stereo flow fields are thresholded, the centroid of the motion energy regions determined, and then triangulation gives the 3D position of the moving body. Haynes[18] also uses a PIPE processor for stereo vision. Sequential frame differencing or background subtraction, are proposed to eliminate static background detail. These approaches are only appropriate if the camera is stationary, but Dickmanns[29] suggests that the use of foveation and feature tracking can be used in the general case of moving targets and observer.

Datacube processing modules have also been used for video-rate visual servoing,[25,27] and active vision camera control.[21]

6.5 Feature extraction

A very important operation in most visual servoing systems is determining the coordinate of an image feature point, frequently the centroid of a region. The centroid can be determined to sub-pixel accuracy, even in a binary image, and for a circle that accuracy is shown to increase with region radius.[9] The calculation of centroid can be achieved by software, or with specialized moment generation hardware.

The centroid of a region can also be determined using multispectral spatial, or pyramid, decomposition of the image.[10] This reduces the complexity of the search problem, by allowing the computer to localize the region in a coarse image, and then refine the estimate by searching progressively higher resolution images.

Software computation of image moments is one to two orders of magnitude slower than specialized hardware. However the computation time can be greatly reduced if only a small image window is processed, whose location is predicted from previous centroid calculations.[40,102,78,34,29,42] The task of locating features in sequential scenes is relatively easy, since there will be only small changes from one scene to the next,[29,71] and total scene interpretation is not required. This is the principle of verification vision proposed by Bolles[15] in which the system has considerable prior knowledge of the scene, and the goal is verify and refine the location of one or more features in the scene. Determining the initial location of features requires the entire image to be searched, but this is needed only once.

Dickmanns[29] and Inoue[50] have built multiprocessor systems where each processor is dedicated to tracking a single feature within the image. More recently, Inoue[51] has demonstrated the used of a specialized VLSI motion estimation device for fast feature tracking. Papanikolopoulos *et al.*[73] uses a sum-of-squared differences approach to match features between consecutive frames. Features are chosen on the basis of a confidence measure computed from the feature window, and the search is performed in software. The TRIAX system[12] is an extremely high-performance multiprocessor system for low latency six-dimensional object tracking. It can determine the pose of a cube by searching short check lines normal to the expected edges of the cube.

When the software feature search is limited to only a small window into the

image, it becomes important to know the expected position of the feature in image. This is the target tracking problem, the use of a filtering process to generate target state estimates and predictions based on noisy observations of the target's position and dynamic model of the motion of the target. Target maneuvers are generated by acceleration controls unknown to the tracker. Kalata[57] introduces tracking filters, and discusses the similarities to Kalman filtering. Approaches using such tracking filters,[7] Kalman filters,[102,29] AR (auto regressive) or ARX (auto regressive with exogenous inputs) models[33,48] have been described. Papanikolopoulos *et al.*[73] uses an ARMAX model, and considers tracking as the design of a second-order controller of image plane position. The distance to the target is assumed constant and a number of different controllers such as PI, pole-assignment and LQG are investigated. When the distance is unknown or time varying, adaptive control may be used.[72] The prediction used for search window placement can also be used to overcome latency in the vision system and robot controller.

6.6 Robot control/communications

A commonly cited problem is the low communication bandwidth between the vision system and the robot, typically a serial link. Such a link limits bandwidth and introduces latency into the closed loop system. Higher communication bandwidth, such as available via ethernet, has been used for stable closed-loop control at 60Hz.[24] A significant part of the problem is that off-the-shelf robot systems have low bandwidth communications facilities. This limits the researcher to using the communications facilities provided, or embarking on the more difficult path of reverse engineering the system or implementing a custom controller.

A commonly cited robot is the Unimate Puma 560, which is widely used in research laboratories. The VAL-II controller has an 'ALTER' facility, allowing trajectory modifications to be received via a serial communications channel, at a sample rate of 28ms. Tate[92] identified the transfer function between this input and the manipulator motion, and found the dominant dynamic characteristic below 10Hz was a time delay of 0.1s. Delays in the system can be overcome to some extent by introducing predictive controllers.

Bukowski *et al.*[18] describes a novel analog interconnect between a PC and the Unimate controller, so as to overcome the inherent latency associated with the VAL-II ALTER facility.

A few described systems[7,48,24] use RCCL,[45] or similar[25] to connect a 'foreign controller' to the Puma 560. This provides direct application access to the Unimate joint servo controllers, at high sample rate, and with reduced latency. Hashimoto *et al.*[44] describes a visual servo system with a Puma 560 controlled by a transputer network. Wilson[102] also uses a transputer network, but to control a small CRS robot.

Little research has been reported in the area of languages for visual task specification. Geshke[40] describes the Robot Servo System (RSS) software which facilitates

describing applications based on sensory input. The software controls robot position, orientation, force and torque independently, and specifications for control of each may be given. The programming facilities are demonstrated with applications for vision based peg in hole insertion, and crank turning. Adsit[1] discusses the use of a table-driven state machine to control a visually-servoed fruit-picking robot. The state machine was seen to be advantageous in coping with the variety of error conditions possible in such an unstructured work environment. Haynes *et al.*[18] describe a set of library routines for a PC connected to a Unimate robot controller and stereo-camera system to facilitate hand-eye programming.

7 Conclusion

This paper has presented, for the first time, a comprehensive summary of research results in the use of visual information to control robot manipulators and related mechanisms. Research results have been discussed and compared in terms of their historical context, algorithmic approach, and implementation. A large bibliography has been provided which also includes important papers from the elemental disciplines upon which visual servoing is based.

Visual servoing is a promising approach to many automation tasks, particularly where the object's location is ill-defined or time varying. Much of the pioneering work in tracking, assembly, grasping, and 3D control was done in the 1970s. The biggest problems remain technological, primarily rapid and robust image feature extraction, and a high-bandwidth path to the robot controller. As computing technology continues to improve, performance/price almost doubling annually, many of these remaining problems will be solved. The problem of determining how the robot should move using visual information is solved for all practical purposes, and the computational costs of the various proposed approaches are comparable, and readily achieved. The image-based technique has been demonstrated to work well, but the advantages have proved illusory, and the problem of Jacobian update in the general case remains. The cited disadvantages of the position-based approach have been ameliorated by recent research; camera calibration, and photogrammetric solutions can now be computed in a few milliseconds even using iteration. Image-based servoing as originally proposed by Weiss has not yet been demonstrated experimentally for 3DOF or 6DOF motion, and may require significant advances in sensor and processing technology before it is feasible. The usefulness of controlling manipulator kinematics and dynamics this way is however open to question.

Active 3D sensors based on structured lighting are now compact and fast enough to use for visual servoing. Feature processing, whether by position-based or image-based techniques is now sufficiently fast. Image feature extraction, by software or specialized hardware, is capable of 60Hz operation with binary scenes, and the limiting factor is now the sensor frame rate. Clearly more robust scene interpretation is required, if the systems are to move out of environments lined with black velvet. Optical flow

based approaches show promise, and have been demonstrated at 60Hz with specialized processing hardware

The problem of communications bandwidth is common to all off-the-shelf robot systems, and the alternatives are either considerable reverse engineering, or implementing a custom controller and robot. However as the visual sample rate is increased, and system latencies are reduced, dynamics problems in closed-loop control will come to the fore.

Before concluding it is appropriate to address the question of why there are so few, if any, visually servoed robots in industry. The situation is somewhat similar to that of robotic force control, which has seen considerable research, and is demonstrable in many laboratories worldwide. The explanation may involve the following points.

- Within a manufacturing environment, tasks that are amenable to sensor based robot control, can in general be achieved with lower cycle time, lower cost and greater reliability by re-engineering the problem. This is both a reflection of the perceptions of industry, and an indictment of state-of-the-art robot controllers. However as robotic technology moves into less structured manufacturing environments, those currently considered too hard or not amenable to re-engineering, then the advantages of sensate robots will become apparent.

- Machine vision is a notoriously difficult technology to apply robustly in industry. The use of mobile, end-effector-mounted cameras would tend to exacerbate lighting and other problems.

8 References

1. P.D. Adsit. *Real-Time Intelligent Control of a Vision-Servoed Fruit-Picking Robot.* PhD thesis, University of Florida, 1989.

2. J.K. Aggarwal and N. Nandhakumar. On the Computation of Motion From Sequences of Images – a Review. *Proc. IEEE*, 76(8), pp. 917–935, August 1988.

3. G.J. Agin. Calibration and use of a light stripe range sensor mounted on the hand of a robot. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 680–685, 1985.

4. R.S. Ahluwalia and L.M. Fogwell. A modular approach to visual servoing. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 943–950, 1986.

5. J. S. Albus. *Brains, behavior and robotics.* Byte Books, 1981.

6. P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman. Real-time visual servoing. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 1850–1856, 1992.

7. P. K. Allen, B. Yoshimi, and A. Timcenko. Real-time visual servoing. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 851–856, 1991.

8. J. Aloimonos, I. Weiss, and A. Badyopadhyay. Active vision. *Int. J. Computer Vision*, 1, pp. 333–356, January 1988.

9. N.A. Andersen, O. Ravn, and A.T. Sorensen. Real-time vision based control of servomechanical systems. *2nd.Int.Symp. Experimental Robotics*, June 1991.

10. C. H. Anderson, P. J. Burt, and G. S. van der Wal. Change detection and tracking using pyramid transform techniques. In *Proceeding of SPIE*, volume 579, pp. 72–78, Cambridge, Mass., September 1985. SPIE.

11. R.L. Andersson. *Real Time Expert System to Control a Robot Ping-Pong Player*. PhD thesis, University of Pennsylvania, June 1987.

12. R.L. Andersson. A low-latency 60Hz stereo vision system for real-time visual control. *Proc. 5th Int.Symp. on Intelligent Control*, pp. 165–170, 1990.

13. R. Bajcsy. Active perception. *Proc. IEEE*, 76(8), pp. 996–1005, August 1988.

14. P.J. Besl. Active, optical range imaging sensors. *Machine Vision and Applications*, pp. 127–152, 1988.

15. R. C. Bolles. Verification vision for programmable assembly. In *Proc 5th International Joint Conference on Artificial Intelligence*, pp. 569–575, Cambridge, MA, 1977.

16. R.C. Bolles and R.A. Cain. Recognizing and locating partially visible objects: the local-feature-focus method. *International Journal of Robotics Research*, 1(3), pp. 57–, 1982.

17. M.E. Bowman and A.K. Forrest. Visual detection of differential movement: Applications to robotics. *Robotica*, 6, pp. 7–12, 1988.

18. R. Bukowski, L.S. Haynes, Z. Geng, N. Coleman, A. Santucci, K. Lam, A. Paz, R. May, and M. DeVito. Robot hand-eye coordination rapid prototyping environment. In *Proc. ISIR*, pp. 16.15 to 16.28, October 1991.

19. F. Chaumette, P. Rives, and B. Espiau. Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2248–2253, 1991.

20. W. F. Clocksin, J. S. E. Bromley, P. G. Davey, A. R. Vidler, and C. G. Morgan. An implementation of model-based visual feedback for robot arc welding of thin sheet steel. *International Journal of Robotics Research*, 4(1), pp. 13–26, Spring 1985.

21. D.J. Coombs and C.M. Brown. Cooperative gaze holding in binocular vision. *IEEE Control Systems Magazine*, 11(4), pp. 24–33, June 1991.

22. P. I. Corke and M. C. Good. Dynamic effects in high-performance visual servoing. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 1838–1843, May 1992.

23. P.I. Corke. Real-time image feature analysis for robot visual servoing. In *Proc. DICTA-91*, Melbourne, December 1991.

24. P.I. Corke and R.P. Paul. Video-rate visual servoing for robots. Technical Report MS-CIS-89-18, GRASP Lab, University of Pennsylvania, February 1989.

25. P.I. Corke and R.P. Paul. Video-rate visual servoing for robots. In V. Hayward and O. Khatib, editors, *Experimental Robotics 1*, pp. 429–451. Springer Verlag, 1989.

26. P. Y. Coulon and M. Nougaret. Use of a TV camera system in closed-loop position control of mechnisms. In Alan Pugh, editor, *International Trends in Manufacturing Technology ROBOT VISION*, pp. 117–127. IFS Publications, 1983.

27. M. L. Cyros. Datacube at the space shuttle's launch pad. *Datacube World Review*, 2(5), pp. 1–3, September 1988.

28. E.D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1, pp. 241–261, 1988.

29. E.D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1, pp. 223–240, 1988.

30. J. Dietrich, G. Hirzinger, B. Gombert, and J. Schott. On a unified concept for a new generation of light-weight robots. In V. Hayward and O. Khatib, editors, *Experimental Robotics 1*, pp. 287–295. Springer Verlag, 1989.

31. K. A. Dzialo and R. J. Schalkoff. Control implications in tracking moving objects using time-varying perspective-projective imagery. *IEEE Trans. Industrial Electronics*, IE-33(3), pp. 247–253, August 1986.

32. J. T. Feddema, C. S. G. Lee, and O. R. Mitchell. Weighted selection of image features for resolved rate visual feedback control. *IEEE Trans. Robotics and Automation*, 7(1), pp. 31–47, February 1991.

33. J.T. Feddema. *Real Time Visual Feedback Control for Hand-Eye Coordinated Robotic Systems*. PhD thesis, Purdue University, 1989.

34. J.T. Feddema and O.R. Mitchell. Vision-guided servoing with feature-based trajectory generation. *IEEE Trans. Robotics and Automation*, 5(5), pp. 691–700, October 1989.

35. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp. 381–395, June 1981.

36. S. Ganapathy. Camera location determination problem. Technical Memorandom 11358-841102-20-TM, AT&T Bell Laboratories, November 1984.

37. S. Ganapathy. Decomposition of transformation matrices for robot vision. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 130–139, 1984.

38. S. Ganapathy. Real-time motion tracking using a single camera. Technical Memorandom 11358-841105-21-TM, AT&T Bell Laboratories, November 1984.

39. C. Geschke. A robot task using visual tracking. *Robotics Today*, pp. 39–43, Winter 1981.

40. C. C. Geschke. A system for programming and controlling sensor-based robot manipulators. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-5(1), pp. 1–7, January 1983.

41. A.L. Gilbert, M.K. Giles, G.M. Flachs, R.B. Rogers, and H.U. Yee. A real-time video tracking system. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2(1), January 1980.

42. R. C. Harrell, D. C. Slaughter, and P. D. Adsit. A fruit-tracking system for robotic harvesting. *Machine Vision and Applications*, pp. 69–80, 1989.

43. H. Hashimoto, T. Kubota, W-C. Lo, and F. Harashima. A control scheme of visual servo control of robotic manipulators using artificial neural network. In *Proc. IEEE Int.Conf. Control and Applications*, pp. TA–3–6, Jerusalem, 1989.

44. K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Manipulator control with image-based visual servo. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2267–2272, 1991.

45. V. Hayward and R. P. Paul. Robot manipulator control under unix - RCCL: a Robot Control C Library. *International Journal of Robotics Research*, 5 No.4, pp. 94–111, 1986.

46. J. Hill and W. T. Park. Real time control of a robot with a mobile camera. *9th International Symposium on Industrial Robots*, pp. 233–246, March 1979.

47. B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17, pp. 185–203, 1981.

48. N. Houshangi. Control of a robotic manipulator to grasp a moving target using vision. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 604–609, 1990.

49. Y. Hung, P.S. Yeh, and D. Harwood. Passive ranging to known planar point sets. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 80–85, 1985.

50. H. Inoue, T. Tachikawa, and M. Inaba. Robot vision server. In *Proc. 20th ISIR*, pp. 195–202, 1989.

51. H. Inoue, T. Tachikawa, and M. Inaba. Robot vision system with a correlation chip for real-time tracking, optical flow, and depth map generation. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 1621–1626, 1992.

52. W. Jang and Z. Bien. Feature-based visual servoing of an eye-in-hand robot with improved tracking performance. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2254–2260, 1991.

53. W. Jang, K. Kim, M. Chung, and Z. Bien. Concepts of augmented image space and transformed feature space for efficient visual servoing of an "eye-in-hand robot". *Robotica*, 9, pp. 203–212, 1991.

54. R. A. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2), pp. 122–139, March 1983.

55. M. Kabuka, J. Desoto, and J. Miranda. Robot vision tracking system. *IEEE Trans. Industrial Electronics*, 35(1), pp. 40–51, February 1988.

56. M. Kabuka, E. McVey, and P. Shironoshita. An adaptive approach to video tracking. *IEEE Trans. Robotics and Automation*, 4(2), pp. 228–236, April 1988.

57. P. R. Kalata. The tracking index: a generalized parameter for $\alpha-\beta$ and $\alpha-\beta-\gamma$ target trackers. *IEEE Trans. Aerospace and Electronic Systems*, AES-20(2), pp. 174–182, March 1984.

58. E.W. Kent, M.O. Shneier, and R. Lumia. PIPE - Pipelined Image Processing Engine. *J. Parallel and Distributed Computing*, 2, pp. 50–7, December 1991.

59. M. Kuperstein. Generalized neural model for adaptive sensory-motor control of single postures. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 140–143, 1988.

60. T. Lazano-Perez, J.L. Jones, E. Mazer, P.A. O'Donnell, and W. Eric L. Grimson. Handey: a robot system that recognizes, plans, and manipulates. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 843–849, 1987.

61. R.K. Lenz and R.Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3-d machine vision metrology. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10(5), pp. 713–720, September 1988.

62. Z. Lin, V. Zeman, and R. V. Patel. On-line robot trajectory planning for catching a moving object. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 1726–1731, 1989.

63. A. G. Makhlin. Stability and sensitivity of servo vision systems. *Proc 5th Int Conf on Robot Vision and Sensory Controls - RoViSeC 5*, pp. 79–89, October 1985.

64. H.A. Martins, J.R. Birk, and R.B. Kelley. Camera models based on data from two calibration planes. *Computer Graphics and Image Processing*, 17, pp. 173–180, 1980.

65. H.G. McCain. A hierarchically controlled, sensory interactive robot in the automated manufacturing research facility. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 931–939, 1985.

66. B.W. Mel. *Connectionist Robot Motion Planning.* Academic Press, 1990.

67. W.T. Miller. Sensor-based control of robotic manipulators using a general learning algorithm. *IEEE Trans. Robotics and Automation*, 3(2), pp. 157–165, April 1987.

68. F. Miyazaki and S. Aritmoto. Sensory feedback for robot manipulators. *Journal of Robotic Systems*, 2(1), pp. 53–71, 1985.

69. J. Mochizuki, M. Takahashi, and S. Hata. Unpositioned workpieces handling robot with visual and force sensors. *IEEE Trans. Industrial Electronics*, 34(1), pp. 1–4, February 1987.

70. S. Negahdaripour and J. Fox. Undersea optical stationkeeping: Improved methods. *Journal of Robotic Systems*, 8(3), pp. 319–338, 1991.

71. R. Nevatia. Depth measurement by motion stereo. *Computer Graphics and Image Processing*, 5, pp. 203–214, 1976.

72. N. Papanikolopoulos, P.K. Khosla, and T. Kanade. Adaptive robot visual tracking. In *Proc. American Control Conference*, pp. 962–967, 1991.

73. N. Papanikolopoulos, P.K. Khosla, and T. Kanade. Vision and control techniques for robotic visual tracking. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 857–864, 1991.

74. N.P. Papanikolopoulos and P.K. Khosla. Shared and traded telerobotic visual control. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 878–885, 1992.

75. R. P. Paul. *Robot Manipulators: Mathematics, Programming, and Control.* MIT Press, Cambridge, Massachusetts, 1981.

76. R. E. Prajoux. Visual tracking. In D. Nitzan etal, editor, *Machine intelligence research applied to industrial automation*, pp. 17–37. SRI International, August 1979.

77. P. Rives, F. Chaumette, and B. Espiau. Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing. In V. Hayward and O. Khatib, editors, *Experimental Robotics 1*, pp. 412–428. Springer Verlag, 1989.

78. A.A. Rizzi and D.E. Koditschek. Preliminary experiments in spatial robot juggling. *2nd.Int.Symp. Experimental Robotics*, June 1991.

79. C. Rosen et al. Machine intelligence research applied to industrial automation. sixth report. Technical report, SRI International, 1976.

80. C. Rosen et al. Machine intelligence research applied to industrial automation. eigth report. Technical report, SRI International, 1978.

81. C. Samson, B. Espiau, and M. Le Borgne. *Robot Control: the Task Function Approach.* Oxford, 1990.

82. A. C. Sanderson and L. E. Weiss. Image-based visual servo control using relational graph error signals. *Proc. IEEE*, pp. 1074–1077, 1980.

83. S. Sawano, J. Ikeda, N. Utsumi, H. Kiba, Y. Ohtani, and A. Kikuchi. A sealing robot system with visual seam tracking. In *Proc. Int. Conf. on Advanced Robotics*, pp. 351–8, Tokyo, September 1983. Japan Ind. Robot Assoc, Tokyo, Japan.

84. Y. Shirai and H. Inoue. Guiding a robot by visual feedback in assembling tasks. *Pattern Recognition*, 5, pp. 99–108, 1973.

85. S.B. Skaar, W.H. Brockman, and R. Hanson. Camera-space manipulation. *International Journal of Robotics Research*, 6(4), pp. 20–32, 1987.

86. G. Skofteland and G. Hirzinger. Computing position and orientation of a freeflying polyhedron from 3d data. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 150–155, 1991.

87. M.V. Srinivasan, M. Lehrer, S.W. Zhang, and G.A. Horridge. How honeybees measure their distance from objects of unknown size. *J. Comp. Physiol. A*, 165, pp. 605–613, 1989.

88. G. Stange, M.V. Srinivasan, and J. Dalczynski. Rangefinder based on intensity gradient measurement. *Applied Optics*, 30(13), pp. 1695–1700, May 1991.

89. T. M. Strat. Recovering the camera parameters from a transformation matrix. In *IU Workshop*, 1984.

90. I. E. Sutherland. Three-dimensional data input by tablet. *Proc. IEEE*, 62(4), pp. 453–461, April 1974.

91. K. Tani, M. Abe, K. Tanie, and T. Ohno. High precision manipulator with visual sense. In *Proc. ISIR*, pp. 561–568, 1977.

92. A. R. Tate. Closed loop force control for a robotic grinding system. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachsetts, 1986.

93. F. Tendick, J. Voichick, G. Tharp, and L. Stark. A supervisory telerobotic control system using model-based vision feedback. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2280–2285, 1991.

94. R.Y. Tsai. A versatile camera calibration technique for high accuracy 3-d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Trans. Robotics and Automation*, 3(4), pp. 323–344, August 1987.

95. R.Y. Tsai and R.K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Trans. Robotics and Automation*, 5(3), pp. 345–358, June 1989.

96. S. Venkatesan and C. Archibald. Realtime tracking in five degrees of freedom using two wrist-mounted laser range finders. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2004–2010, 1990.

97. D. Vernon and M. Tistarelli. Using camera motion to estimate range for robotic parts manipulation. *IEEE Trans. Robotics and Automation*, 6(5), pp. 509–521, October 1990.

98. A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(5), pp. 490–498, May 1989.

99. J. Wang and W. J. Wilson. Three-d relative position and orientation estimation using kalman filter for robot control. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2638–2645, 1992.

100. T.E. Webber and R.L. Hollis. A vision based correlator to actively damp vibrations of a coarse-fine manipulator. RC 14147 (63381), IBM T.J. Watson Research Center, October 1988.

101. L.E. Weiss. *Dynamic Visual Servo Control of Robots: an Adaptive Image-Based Approach*. PhD thesis, Carnegie-Mellon University, 1984.

102. D. B. Westmore and W. J. Wilson. Direct dynamic control of a robot using an end-point mounted camera and kalman filter position estimation. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 2376–2384, 1991.

103. Joan S. Weszka. A survey of threshold selection techniques. *Computer Graphics and Image Processing*, 7, pp. 259–265, 1978.

104. D.E. Whitney and D. M. Gorinevskii. The mathematics of coordinated control of prosthetic arms and manipulators. *ASME Journal of Dynamic Systems, Measurement and Control*, 20(4), pp. 303–309, 1972.

105. P.R. Wolf. *Elements of Photogrammetry*. McGraw-Hill, 1974.

106. Y. Yakimovsky and R. Cunningham. A system for extracting three-dimensional measurements from a stereo pair of tv cameras. *Computer Graphics and Image Processing*, 7, pp. 195–210, 1978.

107. J.S.-C. Yuan. A general photogrammetric method for determining object position and orientation. *IEEE Trans. Robotics and Automation*, 5(2), pp. 129–142, April 1989.

108. J.S-C. Yuan, F.H.N. Keung, and R.A. MacDonald. Telerobotic tracker. EP 0 323 681 A1, European Patent Office, Filed 1988.

109. D. B. Zhang, L. Van Gool, and A. Oosterlinck. Stochastic predictive control of robot tracking systems with dynamic visual feedback. In *Proc. IEEE Int.Conf. Robotics and Automation*, pp. 610–615, 1990.

110. N. Zuech and R.K. Miller. *Machine Vision*. Fairmont Press, 1987.