J. Schiehlen, E.D. Dickmanns: Paper No 23, Page 1
IEEE/RSJ/GI International Conference on Intelligent Robots and Systems
September 12-16, 1994, in Munich, Germany

Design and Control of a Camera Platform for Machine Vision

J. Schiehlen and E.D. Dickmanns
Universität der Bundeswehr München
Fakultät für Luft- und Raumfahrttechnik
D-85577 Neubiberg, Germany

ABSTRACT

In machine vision applications controllable vision sensors improve essentially the performance of the system. The design and realization of a camera platform is presented in this paper. Two modes of operation are important, fast gaze shifts to bring regions of interest into the field of view and smooth pursuit of a moving target. In addition, image stabilisation is needed for application of the camera platform in a vehicle. Using visual feedback in the control loop requires careful consideration of all effects like time and speed restrictions for real time control, image processing delays, interfacing the controller to image processing, smoothing the camera motion and rejection of disturbances. These control aspects are addressed in this paper.

Keywords: active vision, gaze control, camera head, machine vision

1. INTRODUCTION

The main problem of machine vision in perception of the 3D-environment is the loss of one dimension during the 2D-image aquisition. In order to recover the third dimension, two cameras with different positions can be used resulting in the stereo approach. For accurate results a great distance between the cameras is advantageous, but due to occlusions the correspondence problem related to stereo vision is a difficult one. For a small base line, correspondence is no problem, but the results are poor due to ill-conditioned measurement equations. These problems can be avoided if visual tracking is applied with an active observer. Correspondence between two successive frames is easy and an image sequence with a large number of frames following from a continous camera motion results in accurate measurements due to the great distance between the first and the last camera position.

Also in other nonlinear and ill-posed vision problems, the ability to control the vision sensor is advantageous. [Aloimonos et. al. 86] showed that low level vision tasks are simplified into low complexity algorithms using a controllable sensor system.

The goal of the human vision system is not only to perceive the whole environment, but also to perceive as much information as necessary to solve a given task. Vision is continuous and reactions have to be in real time. These requirements, different from the static image processing view, are also valid in the field of robotics [Bajcsy 88]. So, an image sequence has to be processed under real time conditions and constrained resources, the maximum information from one single image is not needed. Also it is much easier to solve a special task than the general recognition problem [Aloimonos 90][Ballard 91], because the goal can be achieved by a combination of simple robust behaviors.

In order to solve a task only some region of interest have to be analysed. This reduces the computational burden of image processing a lot, since image processing power is used optimally. An active vision system that selects regions with an exploratory behavior can avoid the current limitations of vision systems like extremly high data rates and complex data processing calculations on the whole image.

Image aquisition could be done by a spherical vision sensor covering the whole world. But this is not efficient, because the region of interest is at low resolution and difficult to analyse while most of the sensor information is not used. More effective is a vision sensor with viewing direction control. Here, the image covers only the region of interest and high resolution imaging is possible. Also, there is less optical distortion because only the area near the optical axis is used. For comparison with the biological systems, the viewing direction of the human eye is variable, too. With this ability nature developed a human retina consisting of a small area with high resolution (fovea) and large periphery with low resolution. This periphery gives humans a large field of view simultaniously with local high accuracy. Virtual high resolution is achieved by quick movements of the fovea. These advantages can be gained only with active vision systems.

In addition to such a vision sensor also appropriate behaviors are important. From physiology it is known, that there are two clearly seperate types of movements of the human eye. Firstly, rapid gaze shifts, called saccades, are used to point the gaze to the points of

interest. Secondly, during image acquisition the smooth- pursuit mode is active to track the analysed object and the optokinetic reflex and the vestibuloocular reflex to stabilize the image for compensation of the agents' motion. This has also the advantage to reduce motion blur. These simple behaviors are the fundamentals of the powerful human dynamic vision system.

First, some active vision systems known in literature are reviewed. Then, a new camera platform design especially for application in vehicle guidance is described. Further, the two layers of control loops are explained. The motor or low level (behavior level) controller with a sample time of 2 milliseconds and the interface to the special hardware will be presented. Then the high level (task level) controller with a sample time of 0.04 seconds is explained. Experimental results are given to show the performance of the integrated control system and the results are summarized.

2. STATE OF THE ART

During recent years a number of binocular camera heads have been developed inspired by biological vision. These binocular platforms have been mainly build in order to measure depth from stereo images .

The first constructions had a coupled (symmetric) vergence. At Harvard University [Clark et. al. 88] a pan/tilt head with controllable focus and aperture of each eye has been built. At the University of Pennsylvania [Bajczy 88][Krotkow et. al. 90] a head with additional two translational degree of freedom and a motor zoom was developed. These two constructions were followed by heads with vergence and version. Yorick [Murray et. al. 93] is a simple head with pan and tilt capabilities and Turnip [Mowforth 91] can also control the tilt of each eye. TRISH [Jenkin et. al. 93] uses a pan head and controls tilt and cyclotorsion of each eye. A pan/tilt head with variable baseline and focus, aperture and zoom control has been build by [Christensen 93] and a head with additional optical center correction is due to KTH [Pahlavan 92][Viéville et.al. 92]. In Bochum [Tölg 90] a head with tilt, focus, shutter (autoiris) and zoom control has been built. Also heads for use together with a manipulator have been designed. For example [Ballard 91][Brown 90] at the University of Rochester developed a head with additional tilting of each eye. [Crowley et. al. 92] from LIFIA, [Pretlove et. al. 93] from Surrey built a head with focus and aperture control.

In order to measure depth from stereo the viewing direction has to be known very precisely. So, the nontrivial calibration problem has to be solved and high resolution motion is required. A position sensor for feedback with resolution of 0.01 degree or better is required. Most of the designs described use motors with gearboxes, but the use of gearboxes leads to backlash and reduced accuracy.

Although motorized lenses with focus, aperture and zoom control are nice to have, they are rather large and heavy. The use of an electronic shutter for aperture control is preferable.

For motor control of the heads often simple PID controllers are being used that are connected to the host and vision computer via a serial RS232 interface. Because of the low bandwith of this link the overall performance has been slow. [Brown 90] presented a simple predictor tracker and the first integrated control system was given in [Murray et. al. 93].

For the tracking experiment often a white blob with a dark background has been choosen in order to get away with simple image processing. Using real scenes for image processing optical flow or brightness gradients are used, respectively. There is always the problem of correct segmentation of the object and the background.

In image processing two approaches are possible [Bajcsy 88] to get the desired qualitative information. Firstly, the data driven approach, where first the shape or depth map is estimated from image data and then objects are recognized using high level features. Secondly, the model driven recognition, where simple features of the objects are searched in the image. While the first approach is used in the references cited, the second is popular in the field of robotics. Feature based recognition is especially suited for active vision because only the areas of the features have to be analysed and the use of simple features results in a very fast processing.

In the field of robotics the first experiments have been performed using fixed stereo camera pairs resulting in image processing problems. Further, vision and robot control in an open loop fashion were investigated, but this is only suitable for slow applications. Also, research with fixed overhead cameras has been reported, but there are problems too: If the robot arm is near the object it obstructs the object and for the precise final task only low image resolution is obtained. Further, robots of the "eye in hand" type have been built, too. For example [Weiss et. al. 87] demonstrated tracking with an adaptive controller. In the field of mobile robots camera positioning systems are important as reported by[Mysliwetz et. al. 86][Schiehlen et. al. 93]. In [Mori 90] a stereo head is used.

3. HARDWARE DESCRIPTION

The aim of this research project is to design and to test a camera platform for application in vehicles. Here, typically distant objects have to be recognized and disturbances from vehicles' motion have to be suppressed. For that purpose a dynamic monocular vision system is appropriate with controllable gaze direction in two axis. The camera platform shown in Figure 1 has been built.

The outer rotation is denoted as panning or yawing, the inner rotation is called tilting or pitching, respectively. In order to supplement the high resolution in a subarea by a large field of view at the same time like in the human retina, a bifocal vision system is used. It consists of two cameras, one with a tele lens (horizontal aperture angle 15 degree) and one with a wide angle lens (horizontal aperture angle 42 degree). The tele camera is for analysing details of distant objects while the wide angle camera supervises the periphery. Due to the two different focal length no zoom is needed, resulting in a small and light design. As the

distance between the focal points of the two cameras is much smaller than the distance to the scene, the effect on the image can be neglected. In order to adapt to different brightnesses an auto-iris is used for aperture control. With new miniature CCD cameras a small size and weight of the platform is achieved. With special hardware also space-variant sensing is possible [Sandini 90], but the simple bifocal system represents a good and inexpensive compromise.

Figure 1: Two-axis camera platform

axis	pan	tilt
angle range	± 80 deg	± 10 deg
angle sensor resolution	0.011 deg	0.044 deg
max. gyroscope angular rate	± 50 deg/s	± 50 deg/s
max. motor torque	3.25 Nm	0.64 Nm

Table 1: Parameter of platform

Due to the demand of a highly dynamical platform for fast saccades and of high precission for accurate pursuit, a direct drive is most appropriate. Direct drives also have the advantage, that the camera remains inertially stabilized if there are no motor couplings. Each axis is driven by a torque motor; the absolute angles are measured by potentiometers and the inertial angular rates are obtained from piezoelectric gyroscopes. The camera is balanced in order to allow low motor power. To protect the cameras in case of a fault, soft mechanical stops are applied.

4. MOTOR CONTROLLER

The aim of the motor controller is to realize the interface between the special hardware and the image processing controller. A general protocol is used, that is independent of the platform hardware chosen.

As the kinematical and dynamical coupling between the axes can be neglected, two independend controllers are chosen.

The design of one controller allows either a good tracking performance or a good step response performance, but not both at the same time. Therefore, two separate controllers have been implemented. The pursuit motor controller is best suitable for tracking, the saccade motor controller is specialized on step commands. The saccade motor controller uses the angular position for feedback only, since the gyroscopic sensor saturates during operation.

The controller parameters are obtained by the pole placement technique [Aström et. al. 90]. For improved tracking performance additional feedforward terms are used. Both controllers have an integral component for compensation of constant disturbances resulting from gravity torques (i.e. camera cables or unbalanced installation of the camera) and Coulomb friction (resulting from the brushes of the motor). This integral component is implemented as an observer in order to minimize saturation effects resulting from the limited motor voltage. A high sampling frequency of 500 Hz has been chosen in order to be robust with respect to parameter variations and to suppress disturbances fast. Figure 2 shows the motor performance during saccades.

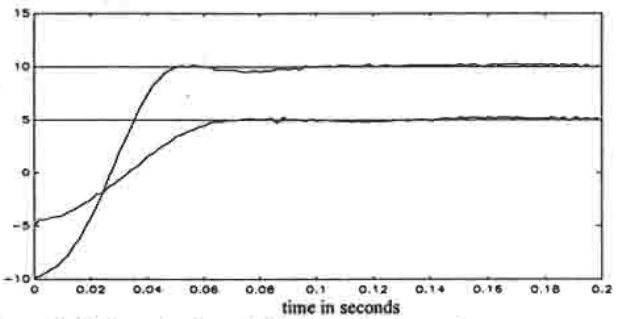

Figure 2: Fast angle steps: ±10 degree in pan axis,
±5 degree in tilt axis

In order to synchronize both controllers to video image processing, a Phase Lock Loop controller has been added. This is necessary in order to guarantee the relation of the motor sample loop to the image acquisition loop and to start saccades always at the beginning of an image acquisition cycle for saving time, because during saccades image processing makes no sense.

5. PURSUIT SYSTEM

If for the pursuit motor controller only an angle sensor is used, this sensor needs a higher resolution than the aperture angle of the picture element for good stabilisation. If the camera head is used together with a manipulator all angles have to be measured with very high resolution, too. In other applications, where it is impossible to measure all these angles like in vehicle applications, stabilisation has to be done with an inertial angular rate measurement. With the high bandwith even of low cost gyroscopes, stabilisation is achieved and the long term drift of the integrated rate signal has to be compensated by image processing.

If a visual feedback is used in the control loop, the measurements are not only noisy but also delayed. This is caused by the image acquisition and processing time required. A simple controller neglecting the delay has a very slow response in order to be stable. The only way to improve performance is to use prediction for compensation of the time delay [Wünsche 88][Murray et. al. 93]. Biological experiments have shown that humans make also an internal prediction [Robinson 88]. For implementing prediction, always a model for the object's motion is needed. If this model corresponds to the real motion, the time delay is compensated completely.

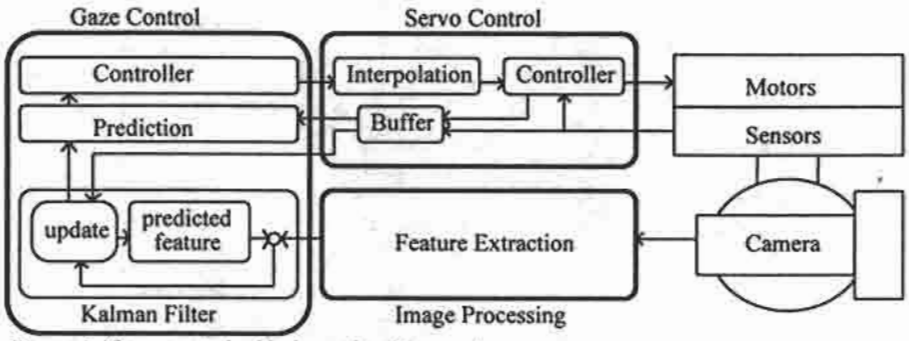

Figure 3: Gaze control with dynamic vision system

The dynamic vision concept is especially suited for the pursuit problem, since fast and reliable image processing is needed. A survey on dynamic vision can be found in [Dickmanns et. al. 88]. This approach uses spatio-temporal internal models of the real world. Comparing measured image features with predicted features from the internal model the state of real world objects can be estimated. As only simple features are searched in small areas and no storing of old images is needed, image processing at frame rate (25Hz) is possible. In addition to this, a representation with the physical states of the objects is obtained that can be directly used for gaze control.

Using dynamic vision the gaze control concept shown in Figure 3 is obtained. The image of the real world is grabbed by the camera and feature positions are extracted. The Kalman filter uses a model for object motion, camera motion, object shape and the imaging process. The difference between predicted and measured feature position is used to correct the estimated state. As the measurements are delayed, this state vector represents the delayed world state. The estimated current state is obtained by predicting the estimated delayed state using also the former platform states. This current state is used for the state feedback controller. This controller works at image sampling rate. As the motor controller needs a much higher sampling rate for disturbance suppression, an interpolation is needed. In order to get smooth motion a linear interpolation has been choosen. This cascade structure has the advantage that the motor controller immediately uses the undelayed inertial angular rate for stabilizing the platform, and the vehicle's own motion is compensated fast. For comparison, also humans have an inertial stabilising path called vestibuloocular reflex.

6. TRACKING EXPERIMENT

The aim of this experiment is to track a randomly moved object, here a red triangle. The computer hardware is a T800 transputer network. The camera image is digitized by the frame grabber (FG). The three edges of the triangle are searched using correlation of edge masks with the image in areas where the object is expected. Using these measurements the location of the object in the image is determined (IP). The control algorithm including the motor controller is implemented on the PI-transputer. The platform electronic is the interface to the hardware.

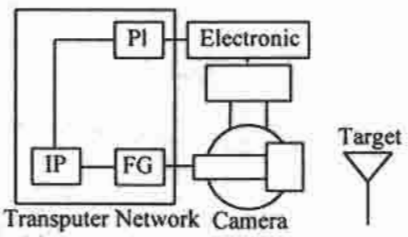

Figure 4: Tracking experimental setup

As in this case the control algorithms for the pan and tilt axis are the same only one axis is presented.

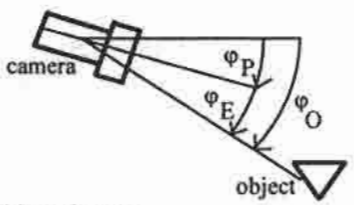

Figure 5: Definition of angles

The angle φ_P is the absolute platform angle with respect to an inertial frame, φ_O is the absolute object angle, $\varphi_E = \varphi_O - \varphi_P$ is the relative difference angle. The angle φ_E is proportional to the object position in the image plane y_{im}, measured in φ-direction.

$$\varphi_E = y_{im} \frac{s}{f} \frac{180}{\pi} \text{deg} = k\ y_{im}$$

The focal length f is 25 mm, and the size s of one pixel after A/D conversion in the frame grabber is 1/84 mm in the horizontal and 1/96 mm in vertical direction. Thus, the resolution in the pan axis is k=0.027 deg/Pixel and in the tilt axis k=0.024 deg/Pixel.

The object motion is assumed to be harmonic with the frequency of $\omega/2/\pi$, resulting in the following differential equation:

$$\dddot{\varphi}_O + \omega^2 \dot{\varphi}_O = 0$$

For vanishing ω a constant acceleration is modelled. In state space representation the following equations are obtained:

$$x = \begin{pmatrix} \ddot{\varphi}_O \\ \dot{\varphi}_O \\ \varphi_E \end{pmatrix} \quad ; \quad \dot{x} = \begin{pmatrix} 0 & -\omega^2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} x - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \dot{\varphi}_P$$

$$y_{im} = (0 \;\; 0 \;\; k)\, x$$

With the state transition matrix $\Phi(\Delta t,\omega)$ a discrete time Kalman-Filter has been built. In the steady state case, the Kalman gain is constant and the filter can be interpreted as a state feedback controller with observer. For estimation of the parameter ω a discrete Extended Kalman Filter is required. These computations contain trigonometric functions, and divisions by zero occur if ω tends to zero. In order to make the computations fast, only the first terms of a Taylor series expansion for ω^2 have been used. This also has the advantage that divisions by zero are replaced by their limiting value. Details can be found in [Schiehlen 93].

In the first experiment the object generated with a graphics workstation is moved vertically with a frequency of 0.5 Herz and an amplitude of 3.4 degree. This motion starts after 2 seconds. After the start of the motion the error increases due to the image processing time delay, but after this delay the controller detects the new motion type and the error decreases.

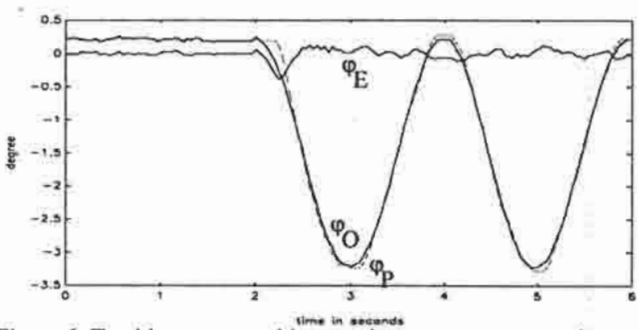

Figure 6: Tracking: φ_O = object motion, φ_P = camera motion , φ_E = tracking error

In the next experiment the object is static, but the camera platform is moved. Since this motion is measured directly with the gyroscope there is no overshoot during the start phase of the motion.

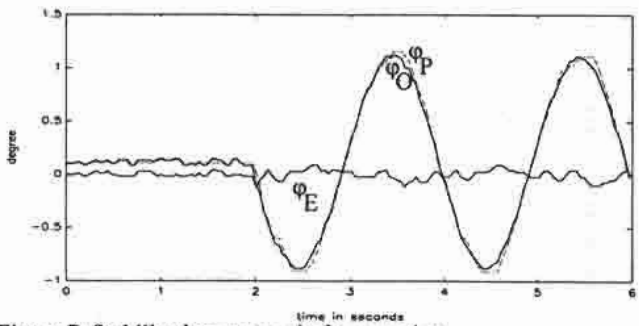

Figure 7: Stabilisation: φ_O = platform motion, φ_P = camera motion , φ_E = tracking error

In the last experiment the object is moved by hand. The motion of figure 8 lasted 4.4 seconds.

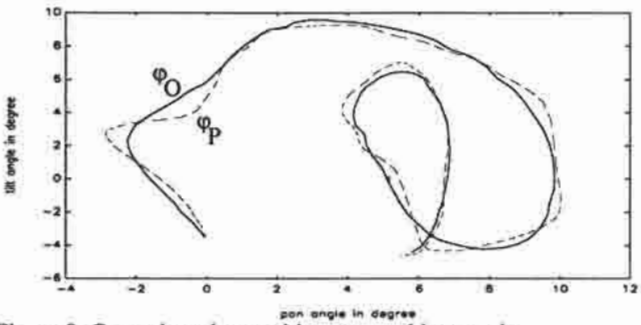

Figure 8: General motion tracking: φ_O = object motion, φ_P = camera motion

7. CONCLUSION

Visual perception is used by an autonomous agent to perceive its environment. If the vision system is active, parts of the real world can be selected for image processing, using the limited processing power optimally. For this process two complementary behaviors are needed: fixation of a point of interest and selection of new fixation points with fast gaze shifts (e.g. 20 degree in 50 milliseconds).

For application in machine vision a compact two-axis camera platform has been developed. The capability of fast gaze shifts has been demonstrated. For the tracking task an integrated control concept is given with visual feedback. It is based on dynamic vision, because feature extraction in predicted areas is most efficient and possible under real time conditions. In order to compensate the image processing time delay prediction is used. If the camera platform is implemented in a vehicle, compensation of the vehicle's own motion is necessary; this, usually, has much higher bandwith than image processing. So, fast inertial measurements are used for stabilisation and the drift is compensated with image processing . Real experiments verified that tracking of fast moving objects is possible.

REFERENCES

Aloimonos J., Weiss I., Bandyopadhyay A.: Active Vision. First International Conference on Computer Vision 1986, IEEE 1987, pp 35-54

Aloimonos J.: Purposive and Qualitative Active Vision. 10th International Conference on Pattern Recognition, June 1990. IEEE, pp 346-360

Aström K.J., Wittenmark B.: Computer Controlled Systems. Prentice-Hall Int. Inc., 1990

Bajcsy R.: Active Perception. Proceedings of the IEEE, Vol. 76, Aug. 1988, pp 996-1005

Ballard D.H.: Animate Vision. Artificial Intelligence 48. Elsevier 1991, pp 57-86

Brown C.: Prediction and Cooperation in Gaze Control. Biological Cybernectics, Springer 1990, pp 61-70

Christensen H.I.: A low-cost robot camera head. International Journal of Pattern Recognition and Artificial Intelligence, vol.7 no.1, Feb. 1993, pp 69-87

Clark J.J., Ferrier N.J.: Modal Control of an Attentive Vision System. Second International Conference on Computer Vision, Dec. 1988. IEEE, pp 514-523

Crowley J.L.,Bobet P., Mesrabi M.: Gaze Control for a Binocular Camera Head. Second European Converence on Computer Vision, May 1992. Springer, pp 588-596

Dickmanns E.D., Graefe V.: a) Dynamic monocular machine vision, b) Application of dynamic monocular machine vision. J. Machine Vision Application, Springer Nov. 1988, pp 223-261

Jenkin M., Milios E., Tsotsos J.: Design and Performance of TRISH: A binocular robot head with torsional eye movements. International Journal of Pattern Recognition and Artificial Intelligence, vol.7 no.1, Feb. 1993, pp 51-68

Krotkov E., Henriksen K., Kories R.: Stereo Ranging with Verging Cameras. IEEE Transactions on Pattern Recognition and Machine Intelligence, vol.12 no. 12, Dez. 1990, pp 1200-1205

Mori H.: Active Sensing in Stereotyped Motion. Intelligent Workshop on Intelligent Robots and Systems '90, Proc. IEEE, July 1990, pp IP-11 IP-19

Mowforth P.: Active Sensing For Mobile Robots. IEE International Conference on Control 91, Edinburgh March 1991. pp 1141-1146

Murray D.W., McLauchlan P.F., Reid I.D., Sharkey P.M.: Reactions to Peripheral Image Motion using a Head/Eye Platform. Forth International Conference on Computer Vision, May 1993. IEEE, pp 403-411

Mysliwetz B.,Dickmanns E.D.: A Vision System with Active Gaze Control for real-time Interpretation of Well Structured Dynamic Scenes. Intelligent Autonomous Systems, Dec.1986

Pahlavan K., Eklundh J.-O., Uhlin T.: Integrating Primary Ocular Processes, Second European Converence on Computer Vision, May 1992. Springer, pp 526-541

Pretlove J.R.G., Parker G.A.: The Surrey attentive robot vision system. International Journal of Pattern Recognition and Artificial Intelligence, vol.7 no.1, Feb. 1993, pp 89-107

Robinson D.A.: Why Visiomotor Systems Don't Like Negative Feedback and How They Avoid It. Vision Brain and Cooperative Computation, MIT Press 1988, pp 89-107

Sandini D.W., Dario P.: Active Vision Based on Space-Variant Sensing. Fifth International Symposium on Robotics Research, Aug. 1989. MIT-Press 1990, pp 75-83

Schiehlen J.: Regelungskonzepte für die Blickrichtungssteuerung in Fahrzeugen. Interner Bericht 1993, UniBwM/LRT/WE13/IB/93-4

Schiehlen J.,Dickmanns E.D.: Two-Axis Camera Platform for Machine Vision. AGARD Conference Proceedings 539, Oct. 1993

Tölg P., Mallot H.A.: Tracking: ein Verfahren zur Stabilisierung bewegter Objekte mit einer aktiven Kamera. In: Mustererkennung '90, DAGM, Informatik-Fachberichte Vol. 254. Springer-Verlag 1990, pp 642-649

Viéville T., Ekhlund J.-O., Pahlavan K., Uhlin T.: An example of artificial oculomotor behavior. International Symposium on Intelligent Control, Glasgow Aug 1992. IEEE, pp 348-353

Weiss L.E., Sanderson A.C., Neuman C.P.: Dynamic Sensor-Based Control of Robots with Visual Feedback. IEEE Journal of Robotics and Automation, Vol. 3, Oct. 1887, pp 404-417

Wünsche H.J: Bewegungssteuerung durch Rechnersehen. Fachberichte Messen, Steuern, Regeln Bd. 20, Springer-Verlag, Berlin 1988