CITR-TR-2 ISSN 1171-7483
TAMAKI-TR-12 March 1997

A Modular 10-DOF Vision System for High-Resolution Active Stereo

Karsten Schlüns[*], Winfried Fellenz[**] ,
Andreas Koschan[**], and Matthias Teschner[***]

Abstract

We present a low-cost active vision system with ten degrees of freedom which has been built from off-the-shelf parts. To obtain high resolution depth information of fixated objects in the scene a general purpose calibration procedure is proposed which estimates intrinsic and extrinsic camera parameters including the vergence axes of both cameras. To produce enhanced dense depth maps a hierarchical block matching procedure is presented which employs color information. To simplify the development of controlling strategies for the head a modular hierarchy is proposed that distributes various tasks among different levels employing basic capabilities of the components of the head.

* The University of Auckland, Tamaki Campus, Computing and Information Technology Research, Computer Vision Unit
** Technical University of Berlin, Institute for Technical Computer Science, Computer Vision Group
*** University of Erlangen-Nuremberg, Telecommunications Institute

This work was funded by the Deutsche Forschungsgemeinschaft (DFG).

1 Introduction

Biological and engineering active vision systems share mechanisms which allow to accomplish visual tasks by varying view parameters like direction of gaze, vergence, focus, zoom and aperture. By adjusting these gaze parameters in a three dimensional world, the process of image acquisition can be voluntarily controlled, thereby constraining the acquired views. Fundamental advantages of actively controlling extrinsic camera parameters are the reduced computational burden for transmitting and processing visual information [BalOzc88] allowing the real-time computation of a depth map, the cooperation of visual behaviors like focus, stereo and vergence to overcome the limitations of a single fixed sensor [KroBaj93][AhuAbb93], and the use of dynamic fixations for real-time tracking of a moving object employing visual feedback [OlsCoo90][Pah et al93][Bra et al94].

In this contribution a further advantage of active vision strategies will be outlined which allows us to obtain high resolution depth maps in the immediate surround of the stereo head using active fovealization of a visual target. The principal processing scheme is outlined in Fig 1 showing a top view of the stereo head for different common viewing angles. For a fixed baseline and wide viewing angles the parallel orientation of both cameras suffices to compute a raw depth map of the scene (1). If the viewing angle of both cameras is narrowed without appropriate adjustment of the vergence or version angles (2) near objects are no longer visible in both images making them not fusable. The convergent fovealization of an object with both cameras (3) now allows the extraction of a high resolution depth map if the version angles are known. However, this enhancement in depth resolution is only limited by the zooming capabilities of the cameras and does not require a higher resolution of the sampling array.

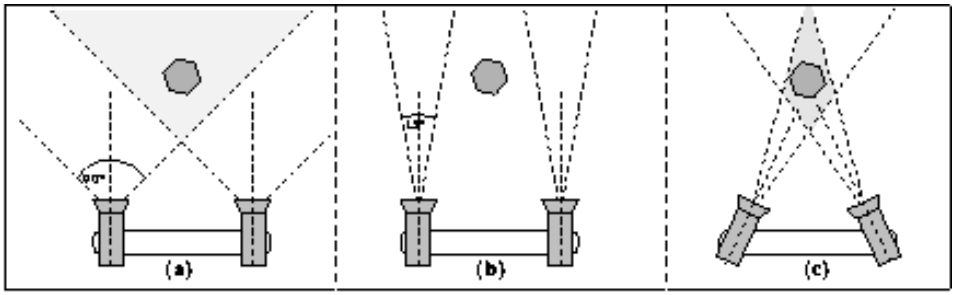


Figure 1: Convergent fovealization for high resolution sampling.

The assembled stereo head and its mechanical hardware will be presented first, followed by a description of the calibration procedure for the intrinsic and extrinsic camera parameters. A fast chromatic stereo procedure to obtain high resolution depth information will be presented next followed by the proposition of a hierarchical controlling scheme which distributes various control loops and capabilities to different levels in a modular fashion.

2 Mechanical Hardware

Apart from financial limitations, our main design objectives for the active stereo head were a hardware controlling scheme which uses a common interface protocol, a

modular design of the head using off-the-shelf parts which can be assembled without much effort, and sufficient degrees of freedom to accomplish various visual tasks. The controlling hierarchy should be modular, exploiting the intrinsic competencies of the low level modules by higher level controlling strategies. The general design philosophy was to introduce biological processing principles to engineering design, thereby acknowledging the efficient information filtering of biological visual systems fulfilling behavioral tasks in a complex environment.

The camera which is shown in Fig. 2 has been build up from the following off-the-shelf parts:

- a Pan-Tilt Unit from Directed Perception (PTU-46-17.5) with angular resolution of 3.086 arc minute and maximum speed of 150°/second,
- two small PAL color video cameras from Sony (EVI-311) with integrated 8x-zoom (f=5.9 (44.3°-horiz) to f=47.2 (5.8°-horiz), F1.4), Aperture, Focus, and 1/3" CCD-chip (752(H) x 582(V)) controllable by a RS-232 Interface Board (IF-51),
- two Vergence Stepper Motors/Indexer from IMS/SCT which have angular resolution of 50.800 counts per revolution at a torque of 0.7 Nm and optical encoders with a resolution of 500 counts per revolution.

The cameras are mounted directly to the motor shafts to minimize slippage using self-made connectors which allow to adjust the optical center to the axes. The parts were assembled using a light-weight vergence platform, build from an aluminum profile which is mounted directly on the pan-tilt unit. Additional modifications of the CCD-cameras were necessary to integrate connector sockets for RS-232, video and voltage supply.

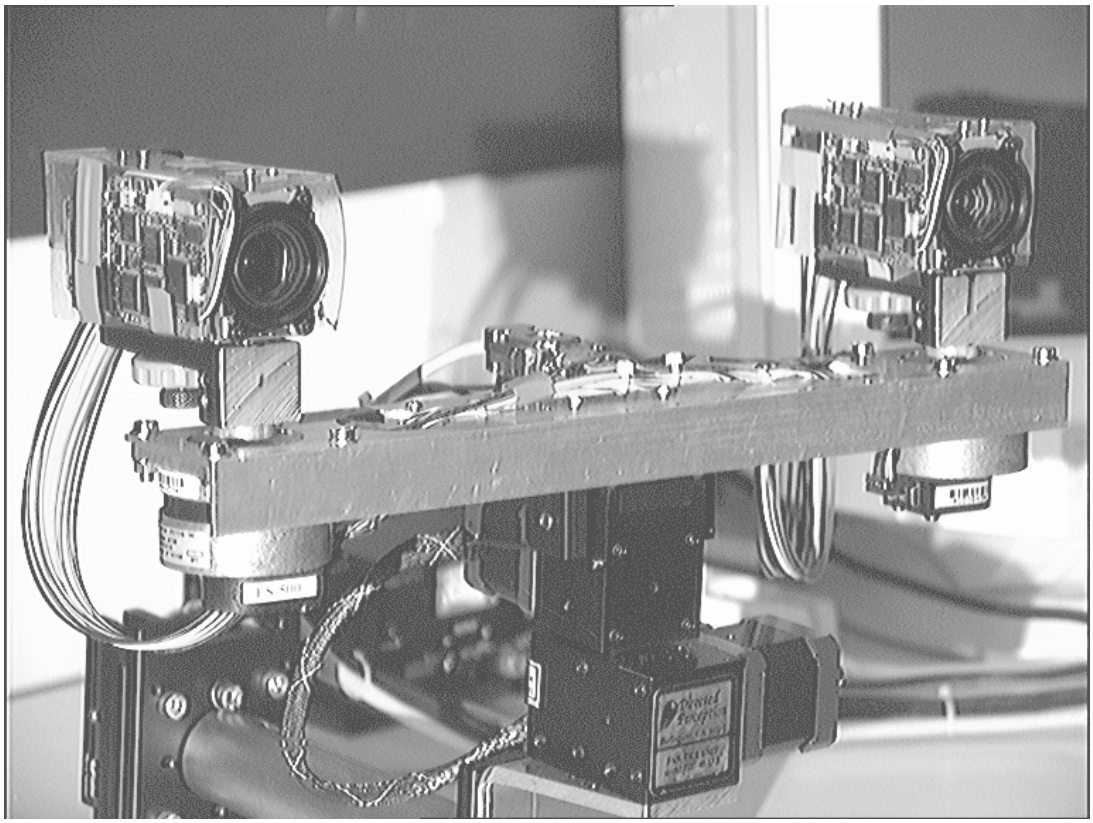


Figure 2: The Active Vision System TUBI at TU Berlin.

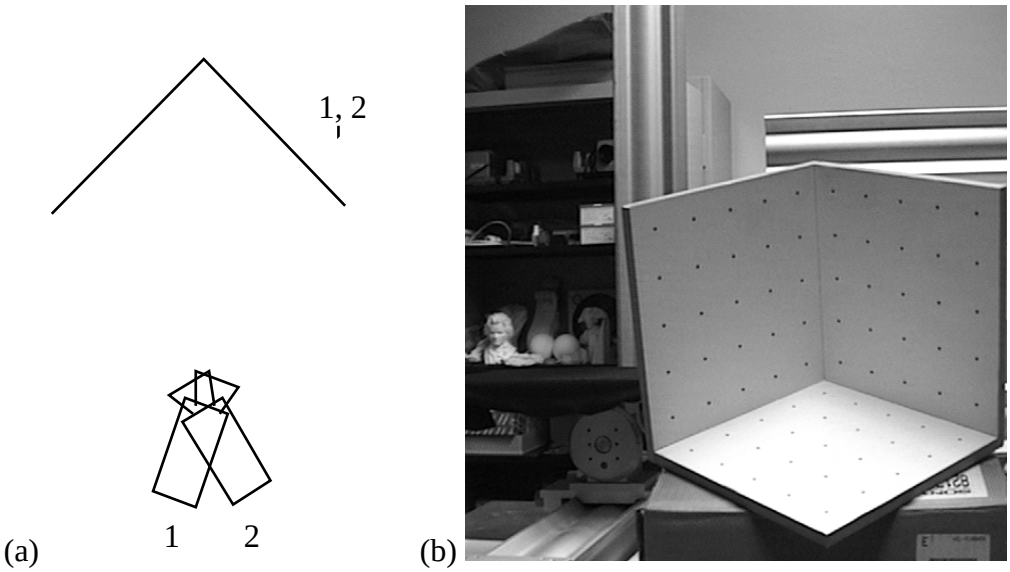


Figure 3: (a) Rotation of the camera, (b) calibration object.

The video data is collected by two Matrox Meteor frame-grabbers on a Pentium computer using routines under the Linux operating system. To communicate with the head via RS-232 protocol an internal interface-board is used to control all five serial lines of the head asynchronously from the host. The complete setup including the color cameras, two micro-steppers with indexers, the pan-tilt unit and two frame-grabbers costs about $10.000 with reasonable own work on the Head.

3 Camera Calibration for an Active Stereo System

For an active stereo system it is not only important to estimate the intrinsic and extrinsic parameters of the cameras. Additionally, it is essential to know the vergence axes of the cameras to simplify the rectification process. Generally, for every single camera the vergence axis (axis of rotation) is not located in the optical center of the camera. In this approach a general purpose camera calibration method is applied that was proposed by Tsai [Tsa86]. The algorithm estimates all of the required camera parameters.

3.1 Estimating the vergence axis

The vergence axis of each camera is estimated by employing a general purpose 3-D calibration procedure. Among the many existing calibration methods, the Tsai method [Tsa86] has some advantages. It has in possession a known behavior and a well-known stability, since several research groups with independent realizations extensively analyzed this approach. For calculating a vergence axis, a sequence of at least two images of a calibration object is taken, see Fig. 3(a). We use a concave cube as calibration object, as shown in Fig. 3(b).

In the following description we use only two of the five coordinate systems introduced by Tsai, for simplification reasons, namely the camera coordinate system (X_k, Y_k, Z_k) and the world coordinate system (X_w, Y_w, Z_w). The equation

$$P_k = \mathrm{R} \cdot P_w + \mathrm{T}, \text{ with } \mathrm{R} \in \Re^{3\times 3}, \ \mathrm{T} \in \Re^3$$

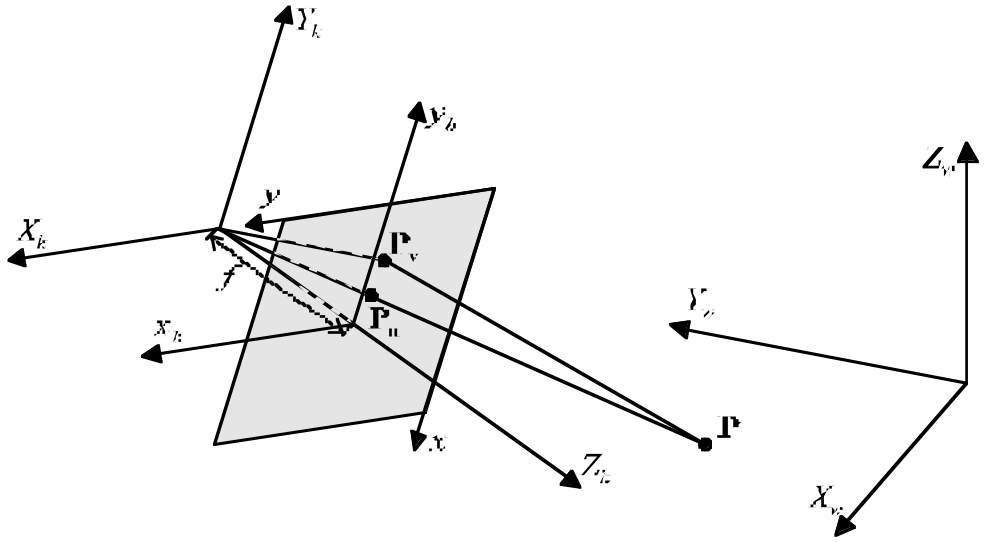


Figure 4: Camera model proposed by Tsai.

describes the relation between (X_k, Y_k, Z_k) and (X_w, Y_w, Z_w), see Fig. 4. Then, the transformation of the two 3D coordinate systems is simply

$$P_{k1} = \mathbf{R}_1 \cdot P_w + \mathrm{T}_1, \text{ with } \mathbf{R}_1 \in \Re^{3\times3}, \ \mathrm{T}_1 \in \Re^3 \text{ and}$$

$$P_{k2} = \mathbf{R}_2 \cdot P_w + \mathrm{T}_2, \text{ with } \mathbf{R}_2 \in \Re^{3\times3}, \ \mathrm{T}_2 \in \Re^3.$$

Using homogenous coordinates each of these transformations can be merged to one 4×4 matrix:

$$P_{k1} = \mathbf{M}_1 \cdot P_{k1} \text{ and } P_{k2} = \mathbf{M}_2 \cdot P_{k2} \ .$$

Now, the relation between both camera coordinate system points becomes

$$P_{k1} = \mathbf{M}_1 \cdot \mathbf{M}_2^{-1} \cdot P_{k2} \ .$$

The matrix product describes the general movement of the (rotating) camera. Furthermore, there exists a priori knowledge about the camera movement. Using a mechanical pre-calibration it can be assumed, that the camera rotates about an axis parallel to the image plane (projection plane, *XY*-plane) and aligned with the *Y*-axis of the (X_k, Y_k, Z_k) coordinate system, see Fig. 5.

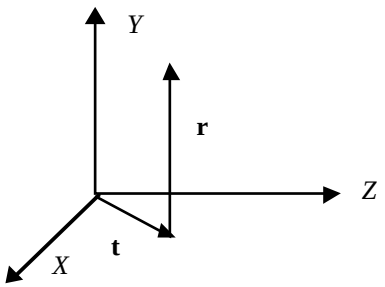


Figure 5: Model of the vergence axis.

The task is to estimate t_x and t_z which are unequal to zero, due to focus adjustment and camera shift. It is possible to describe the movement of the camera in a different way by using the model mentioned above. The movement consists of three steps:

1. Translating the vergence axis to bring it into conformity with the *Y*-axis:

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

2. The rotation itself:

$$\mathbf{A}_2 = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\alpha & 0 & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

3. Inverting of step 1:

$$\mathbf{A}_3 = \begin{pmatrix} 1 & 0 & 0 & -t_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The matrix $\mathbf{A} = \mathbf{A}_3 \cdot \mathbf{A}_2 \cdot \mathbf{A}_1$ is now an alternative representation of the movement of the camera. In addition, it is known that $\mathbf{M} = \mathbf{A}$.

$$\mathbf{A} = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha & -t_x + t_x\cos\alpha + t_z\sin\alpha \\ 0 & 1 & 0 & 0 \\ -\sin\alpha & 0 & \cos\alpha & -t_x\sin\alpha - t_z + t_z\cos\alpha \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

As shown before, **M** is completely known by applying the camera calibration method twice. For the estimation of t_x and t_z two components are chosen from the matrices to form an equation system (m_{ij} is an element of **M**).

$$\begin{aligned} -t_x + t_x\cos\alpha + t_z\sin\alpha &= m_{03} \\ -t_x\sin\alpha - t_z + t_z\cos\alpha &= m_{23} \end{aligned}.$$

Comparing further matrix elements we get

$$\cos\alpha = m_{00} = m_{22} \quad \text{and} \quad \sin\alpha = m_{02} = -m_{20}.$$

Now

$$t_x = \frac{-m_{03} + m_{03}m_{00} - m_{23}m_{02}}{2 - 2m_{00}} \text{ and } t_z = \frac{-m_{23} + m_{23}m_{00} + m_{03}m_{02}}{2 - 2m_{00}}$$

are solutions for t_x and t_z.

In that way the location of the vergence axes and the rotation angle are estimated. If the results of the camera calibration are not reliable it is possible to replace cos α and sin α using more than only one matrix element. Herewith errors are reduced that are related to numerical instabilities.

First results of the rectification procedure suggest that the estimation of the vergence axes is very reliable. Unfortunately, because of the differences between a real camera and the assumed camera model it is not possible to get the precision of the proposed method directly. Therefore, the following theoretical error propagation was analyzed.

It is assumed that the camera calibration is influenced by noise. The error is modeled by adding Gaussian noise to the transformation matrix **M**. Fig. 6 shows that erroneous camera calibration has only little effect if t_x and t_z are small. For that reason it is useful to apply the approach more than once.

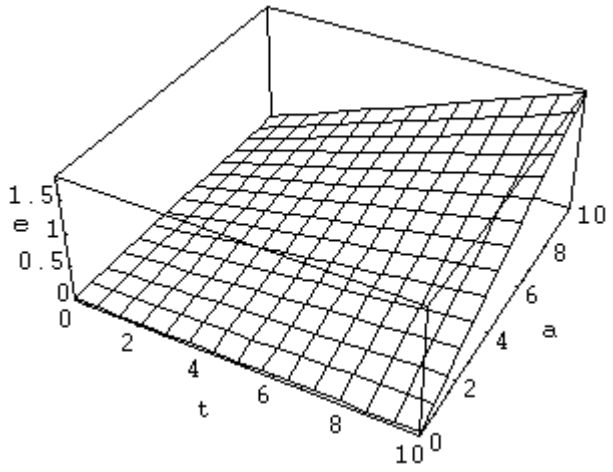


Figure 6: Error analysis, a: noise added to each matrix element of **M** (in percentage), t: t_x and t_z in mm, ($t_x = t_z$ to simplify the analysis), e: resulting error of estimating t.

4 Stereo analysis using chromatic Block Matching

A fast stereo algorithm that produces dense depth maps is needed to obtain depth information with high resolution for an active vision system. Most of the existing stereo techniques are either fast and produce sparse depth maps or they compute dense depth maps and are very time consuming. In an earlier investigation [Kos et al96], we found a hierarchical implementation of a Block Matching technique using color information to be very suitable for obtaining fast and precise dense depths maps. The use of color information enhances the quality of the results by 25 to 30 percent compared to the quality of the results obtained when using only intensity information. Thus, we employ color information to produce dense depth maps of higher quality.

A hierarchical implementation of the algorithm in an image pyramid enables an efficient realization of the matching process. However, our implementation on a single processor does not reach real-time demands at the moment. Currently, we are only approaching real-time requirements when using images of size 256×256 pixels and a parallel implementation on 10 processors [KosRod95]. Further investigations are necessary to speed up the algorithm. The principle of the chromatic Block Matching algorithm and its hierarchical implementation is outlined below.

Block Matching is based on a similarity check between two equal sized blocks ($n \times m$-matrices) in the left and the right image (area-based stereo). The mean square error *MSE* between the pixel values inside the respective blocks defines a measure for the similarity of two blocks. We propose to employ an approximation of the Euclidean distance to measure color differences. The left color image F_L and the right color image F_R may be represented in the RGB color space as $F_L(i,j) = \big(R_L(i,j), G_L(i,j), B_L(i,j)\big)$ and $F_R(i,j) = \big(R_R(i,j), G_R(i,j), B_R(i,j)\big)$. The *MSE* is defined with $n = m = 2k+1$ as

$$MSE_{color}(x,y,\Delta) = \frac{1}{n \cdot m} \sum_{i=-k}^{k} \sum_{j=-k}^{k} (\left| R_R(\mathrm{x}+i,y+j) - R_L(\mathrm{x}+i+\Delta,y+j) \right|^2 + \left| G_R(\mathrm{x}+i,y+j) - G_L(\mathrm{x}+i+\Delta,y+j) \right|^2 + \left| B_R(\mathrm{x}+i,y+j) - B_L(\mathrm{x}+i+\Delta,y+j) \right|^2),$$

where Δ is an offset describing the difference $(x_L - x_R)$ between the column positions in the left and in the right image. The block (of size $n \times m$) is shifted pixel by pixel inside the search area. Using standard stereo geometry the epipolar lines match the image lines. The disparity D of two blocks in both images is defined by the horizontal distance, showing the minimum mean square error. Furthermore, the search area in the right image is limited by a predefined maximum disparity $d_{\max}$:

$$D = \min_{|\Delta| \le d_{\max}} \{ MSE_{color}(x,y,\Delta) \}.$$

Block disparities are median filtered to avoid outliers. A dense disparity map is generated when applying a pixel selection technique to every pixel in the image. Afterwards, median filtering is applied to pixel disparities (see [Kos93] for further details).

Reducing the search space for the disparities could be a solution to the problem. This could be obtained by a very restrictive use of the continuity constraint proposed in [Mar82]. It produces a smoothed depth map where fine structures can not be represented. Discontinuities in depth that are typical for object edges will be smoothed. Thus, any segmentation may fail. This disadvantage can be solved using a pyramid model.

4.1 Hierarchical Block Matching using image pyramids

The idea of using pyramid models in image analysis was introduced by Tanimoto and Pavlidis [TanPav75] as a solution to edge detection. One important property of the pyramid model is that it is computationally extremely efficient [Kro96]. We enhanced the chromatic Block Matching algorithm in robustness and in time efficiency by using a quad pyramid. Each level is obtained by a reduction of factor 4 in resolution from the next lower level. The values for the pixels are obtained by calculating the mean value in each color channel.

The disparities $D(s+1)$ at level $(s+1)$ can be derived from the disparities $D(s)$ of the preceding level (s) by applying a modified block matching algorithm to the image of level $(s+1)$. The search space for the disparity of each block at level $(s+1)$ is derived from the disparity of the corresponding block at level (s) by a tolerance factor D_T. (see Fig. 7) This parameter defines the width D_Δ of the reduced search space $[D_{MIN}, D_{MAX}]$ and controls the smoothness of the disparity map.

$$D_\Delta(s) = 2^{(s-1)} \cdot D_T ,$$

$$D_{MIN}(s) = \begin{cases} D(0) - D_\Delta(s) & \text{for } s = 1 \\ D_{MIN}(s-1) - D_\Delta(s-1) & \text{for } s > 1 \end{cases},$$

$$D_{MAX}(s) = \begin{cases} D(0) - D_\Delta(s) & \text{for } s = 1 \\ D_{MAX}(s-1) - D_\Delta(s-1) & \text{for } s > 1 \end{cases}.$$

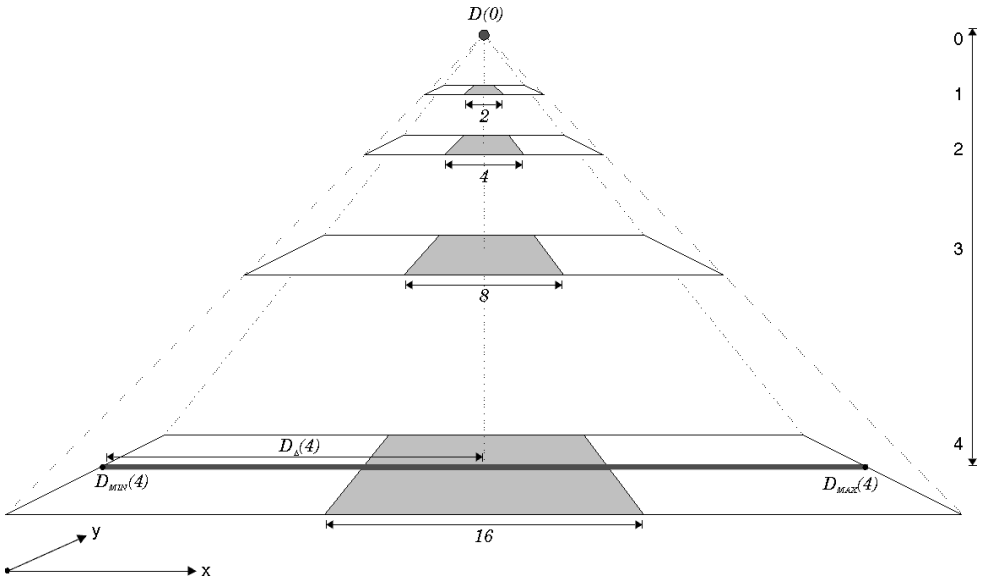


Figure 7: Definition of the search space with a tolerance factor D_T = 3.0.

When choosing a small value for the tolerance factor D_T, the difference between the final disparities and the average disparity found at level 0 will be very small. This is equivalent to a small variation of disparities over the whole image. A larger tolerance factor will cause a bigger search space and the influence of the computed disparities in the preceding levels will decline.

This hierarchical method is more robust than the non-hierarchical one. Additionally, matching the blocks at one level is still possible in parallel, because the blocks within one level are matched independently.

5 The Control Hierarchy

To simplify the implementation of control algorithms for the head we choose a modular and hierarchical controlling scheme with a common interface protocol between each layer. The control level hierarchy which is depicted in Fig. 8 includes low level occulomotoric tasks like focusing and verging and higher level tasks like stereopsis and focus-of-attention. Each of the modules in the control hierarchy can be compared to specific visual competencies which can be used at the next higher level.

(a) Level 0 incorporates the low-level occulomotoric control for the basic functional units of the system. Each of the mechanical and optical modules (pan-tilt, vergence, focus) has its own parameter control which can be exchanged by a more sophisticated controlling strategy on a higher level. The small color CCD-cameras employ an integrated auto focus and aperture control which operates fast and accurately for all of the performed visual tasks. The vergence system consisting of two high-resolution micro stepper systems can be controlled feed forward using optical encoders, or closed-loop using a scale-space disparity reduction strategy. Both axes of the pan-tilt unit can be positioned with high resolution to point the vergence platform to a predefined direction in space.

The intermediate level 1 implements kinematic control strategies for the fast and accurate positioning of the gaze using optimized acceleration and movement profiles. Level 2, which deals with ballistic and feedback eye movements, comprises modules

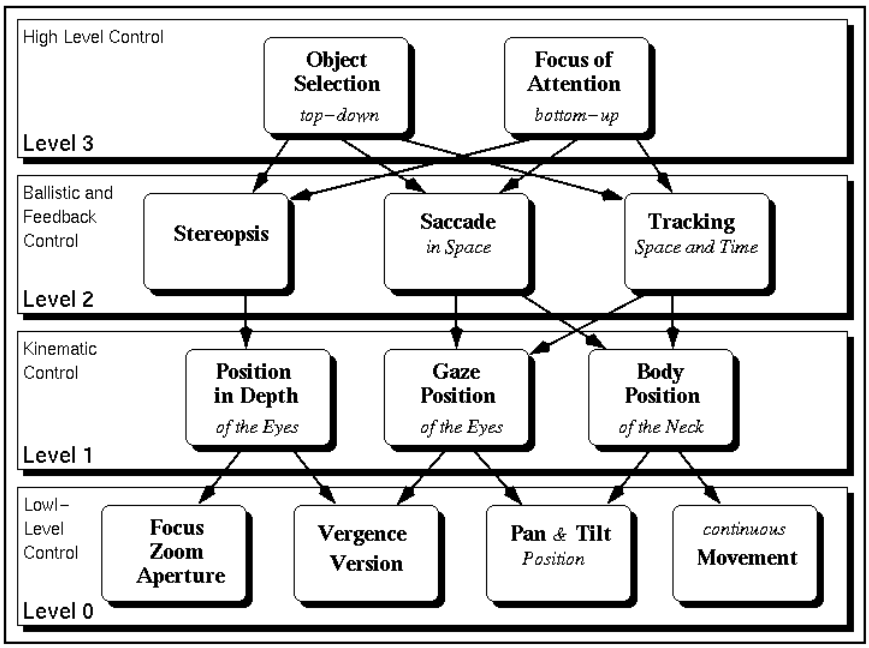


Figure 8: Hierarchical, modular controlling scheme for an active vision system.

for target fovealization in space and time. The capabilities at this level include stereopsis, target tracking and the generation of a saccadic command. Level 3 is engaged with the data-driven extraction of semantic units for a higher level recognition process by an attention mechanism based on preattentively grouped visual information [FellHar96] or the top-down generation of an object signal which can be sequentially matched to objects in the scene.

6 Results

Although not all visual capabilities have been integrated into one system, the performance of its parts can be studied. A requirement for tracking objects in real-time is the knowledge of their position in depth.

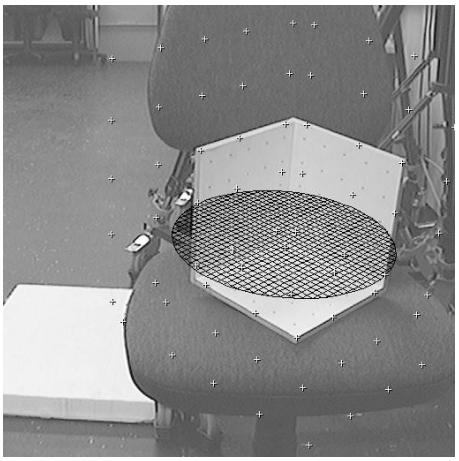
(a)  (b) 

Figure 9: Backprojection of 75 virtual points into the scene using their known depth values (a) in the left image and (b) in the right image.

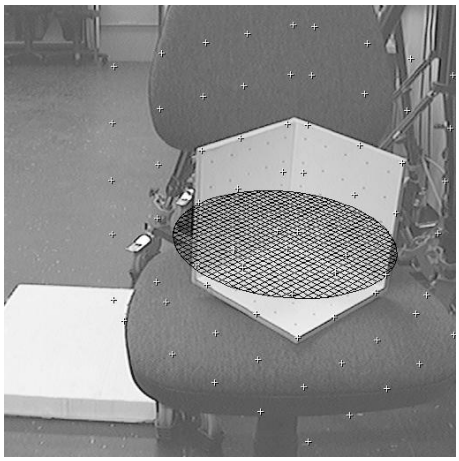


Figure 10: Backprojection of the ideal image points.

The problem of finding the point in space which corresponds to the image points u_i and v_i in both images is commonly known as triangulation. If both calibration matrices C and C' are known, it is possible in principle to calculate the intersection of the two rays in space corresponding to the image points. In practice, these lines are not guaranteed to cross caused by discretization error and noise [HarStu95]. In the presented system the calibration matrices are calculated using the calibration method after Tsai [Tsai86] and the triangulation of both image points is computed using the pseudoinverse of the resulting matrix equations. Fig. 9(a) and 9(b) shows the result of backprojecting the image points of 75 virtual points around the calibration object into the scene using the known depth of the points.

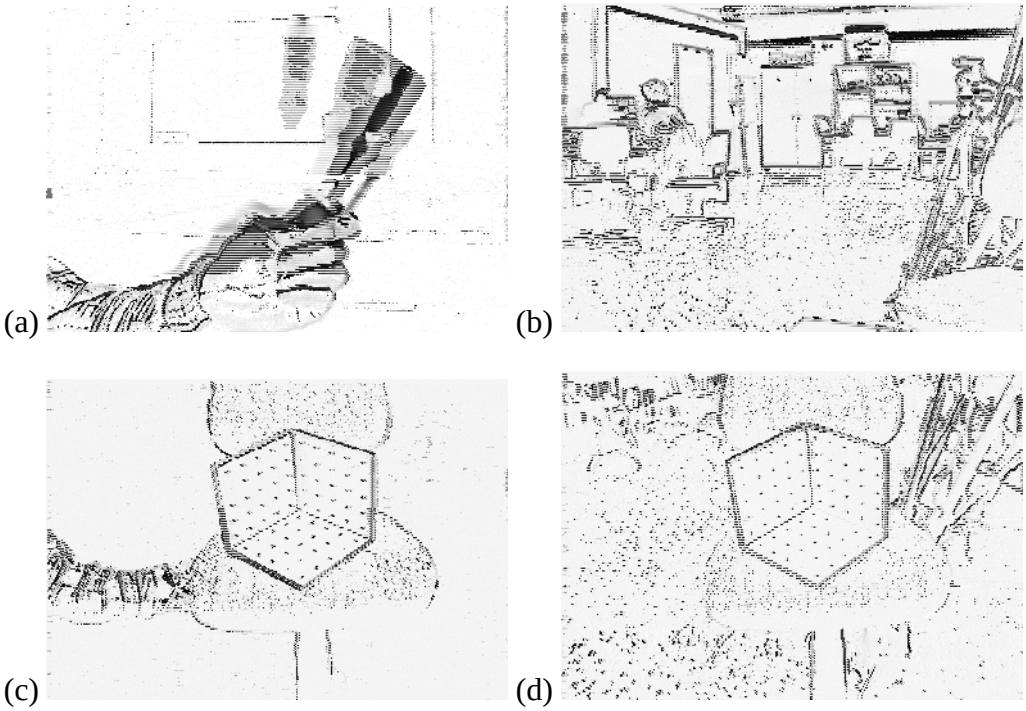


Figure 11: (a) Moving hand, (b) temporal subtraction of two successive frames, (c) object movement and (d) camera movement for calibration.

Figure 10 shows the result of backprojecting the ideal image points of both images using the triangulation method. Closer inspection of the marked points reveals small offsets of the outer marks. Table 1 (see Appendix) shows the camera parameters. Both calibration matrices with parameters are also given in the Appendix.

A simple way to detect motion in the scene is by computing the optical flow of two successive images. Fig. 11(a) shows the result of temporal differentiating two images of a moving hand with a screwdriver. Although the maximum lies on the moving object, spurious aftereffects of the scanning process are visible.

A computationally cheap alternative to get the most conspicuous edge information in the image is depicted in Fig. 11(b), showing the response of temporal subtracting two successive images during a small vergence movement. Fig. 11(c) and 11(d) illustrate the difference between moving the object and moving the camera to get temporal edges and points on a calibration object for on-line calibration. Although the background noise is lower when only the object is moved, the detected calibration pattern are similar in both cases.

7 Conclusion

We have presented a low-cost stereo active-vision head which uses a modular cooperating control strategy. By distributing capabilities of the system in a hierarchical manner we have reduced the complexity of controlling the multiple degrees of freedom of the system. Furthermore, a computationally efficient technique for calibrating the vergence axes of an active vision system has been introduced. A modular control hierarchy has been outlined for several visual tasks (e.g. stereo vision, tracking, object recognition, etc.). Additional tests and research activities are necessary for a more detailed investigation of the techniques. Further results will be presented soon.

In summary, we should like to emphasize that an active vision system always allows to improve the resolution of depth information. Therefore, we believe that precise results can be efficiently obtained when combining hierarchical chromatic Block Matching with active fovealization.

Acknowledgment

This work was funded by the Deutsche Forschungsgemeinschaft (DFG).

References

[AhuAbb93] Ahuja, N., Abbot, A. L.: *Active Stereo: integrating disparity, vergence, focus, aperture, and calibration for surface estimation*, IEEE Trans. PAMI, Vol. 15(10), pp. 1007-1029, 1993.

[BalOzc88] Ballard, D. H., Ozcandarli, A.: *Eye fixation and early vision: kinetic depth*, Proc. ICCV´88, pp. 524-531, 1988.

[Bra et al94] K.J. Bradshaw, P.F. McLauchlan, I.D. Reid, and D.W. Murray, *Saccade and pursuit on an active head-eye platform*, Image and Vision Computing 12, pp. 155-163, 1994

[FelHar96] Fellenz, W. A., Hartmann, G.: *Preattentive grouping and attentive selection for early visual computation*, Proc. 13th Int. Conf. on Pattern Recognition ICPR´96, Vienna, Austria, Vol. IV, pp. 340-345, 1996.

[HarStu95] R. I. Hartley and P. Sturm, *Triangulation*, Proc. Computer Analysis of Images and Patterns, CAIP´95, pp. 190-197, 1995

[Kos93] Koschan, A.: *Dense stereo correspondence using polychromatic block matching*, Proc. of the 5th Int. Conf. on Computer Analysis of Images and Patterns CAIP'93, D. Chetverikov, W. Kropatsch (eds.), Budapest, Hungary, pp. 538-542, 1993.

[KosRod95] Koschan, A., Rodehorst, V.: *Towards real-time stereo employing parallel algorithms for edge-based and dense stereo matching*, Proc. of the IEEE Workshop on Computer Architectures for Machine Perception CAMP´95, Como, Italy, pp. 234-241, 1995.

[Kos et al96] Koschan, A., Rodehorst, V., Spiller, K.: *Color stereo vision using hierarchical block matching and active color illumination*, Proc. 13th Int. Conf. on Pattern Recognition ICPR´96, Vienna, Austria, Vol. I, pp. 835-839, 1996.

[Kro96] Kropatsch, W.G.: *Properties of pyramidal representations*, Computing Suppl. 11, pp. 99-111 (1996).

[KroBaj93] Krotkov, E. and Bajcsy, R.: *Active vision for reliable ranging: cooperating focus, stereo, and vergence*, Int. J. of Computer Vision, Vol. 11(2), pp. 187-203, 1993.

[Mar82] Marr, D.: *Vision - A Computational Investigation Into the Human Representation and Processing of Visual Information*, New York: Freeman & Co 1982.

[OlsCoo90] Olson, T. J., Coombs, D. J.: *Real-time vergence control for binocular robots*, Proc. IUW´90, pp. 881-888, 1990.

[Pah et al93] Pahlavan, K., Uhlin, T, Eklundh J.-O.: *Dynamic fixation*, Proc. ICCV´93, pp. 412-419, 1993.

[TanPav75] Tanimoto, S., Pavlidis, T.: *A hierarchical data structure for picture processing*, Computer Graphics and Image Processing 4, pp. 104-119, 1975.

[Tsa86] Tsai, R. Y.: *An efficient and accurate camera calibration technique for 3d machine vision*, Proc. of CVPR´86, Miami Beach, USA, pp. 364-374, 1986.

Appendix

The parameters of the employed cameras are given in Table 1.

Size of sensor	X = 8.8 mm, Y = 6.6 mm
Number of sensor points	X = 786 sels, Y = 581 sels
Distance of sensor points	X = 11.196 um/sel, Y = 11.360 um/sel
Number of points	X = 512 pix
Number of image points	X = 512 pix, Y = 512 pix
Distance of image points	X = 11.196 um/pix*
Center of image	X = 256.0 pix, Y = 256.0 pix

Table 1: Camera parameters.

The optimized extrinsic parameters of the left camera with respect to the world coordinate system (X_w, Y_w, Z_w) are as follows. The translation vector (data given in mm) is

$$(X_L,\ Y_L,\ Z_L) = (-139.13,\ -7.52,\ 1253.38).$$

The rotation matrix was calculated to

$$\mathrm{Rot}_L = \begin{pmatrix} 0.669 & -0.743 & 0.019 \\ -0.348 & -0.292 & 0.891 \\ -0.656 & -0.603 & -0.454 \end{pmatrix}.$$

Using the Yaw-Pitch-Roll notation the rotational parameters are

X (*Yaw*) = -126.99°, *Y* (*Pitch*) = 41.02°, *Z* (*Roll*) = -27.49°.

The optimized intrinsic parameters for this camera are

focus *f* = 11.29 mm (44.91 mm using standard 35mm-camera),

horizontal scaling factor s_x = 0.958 and

lens distortion κ_1 = -1.161906e-03 1/mm^2 resp. κ_2 = 2.358010e-04 1/mm^2.

The orientation of the *XY*-plane with respect to the world coordinate system is

$$(X,\ Y,\ Z) = (0.019,\ 0.891,\ -0.454)$$

and *d* = 1274.016. The pose and orientation of the left camera is given by

$$(X,\ Y,\ Z) = (912.95,\ 649.94,\ 578.34)$$

and

$$(X,\ Y,\ Z) = (-0.656,\ -0.603,\ -0.454).$$

The optimized extrinsic parameters of the right camera with respect to the world coordinate system (X_w, Y_w, Z_w) are as follows. The translation vector (data given in mm) is

$$(X_R,\ Y_R,\ Z_R) = (80.85,\ -11.51,\ 1274.56).$$

The rotation matrix was calculated to

$$\mathrm{Rot}_R = \begin{pmatrix} 0.707 & -0.705 & 0.053 \\ -0.357 & -0.284 & 0.890 \\ -0.612 & -0.649 & -0.453 \end{pmatrix}.$$

Using the Yaw-Pitch-Roll notation the rotational parameters are

X (*Yaw*) = -124.91°, *Y* (*Pitch*) = 37.68°, *Z* (*Roll*) = -26.80°.

The optimized intrinsic parameters for this camera are

focus *f* = 11.36 mm (45.20 mm using standard 35mm-camera),

horizontal scaling factor s_x = 0.966 and

lens distortion κ_1 = 8.903591e-05 1/mm^2 resp. κ_2 = 2.985636e-04 1/mm^2.

The orientation of the *XY*-plane with respect to the world coordinate system is

$$(X,\ Y,\ Z) = (0.054,\ 0.890,\ -0.453)$$

and *d* = 1287.63. The pose and orientation of the right camera is given by

$$(X,\ Y,\ Z) = (718.69,\ 880.35,\ 582.79)$$

and

$$(X,\ Y,\ Z) = (-0.612,\ -0.649,\ -0.453).$$