Stereoscopic Displays & Applications

Eye-Tracking for Autostereoscopic Displays using Web Cams

Markus Andiel, Siegbert Hentschke, Thorsten Elle, Eduard Fuchs

IPM, Universität Gesamthochschule Kassel

Wilhelmshöher Allee 71

34121 Kassel

Germany

Published as part of Proceedings of SPIE:

Stereoscopic Displays and Applications XIII

Vol. 4660

21-23.01.2002

San Jose, Ca., USA

Eye-Tracking for Autostereoscopic Displays using Web Cams

Markus Andiel, Siegbert Hentschke, Thorsten Elle, Eduard Fuchs

Universität Gesamthochschule Kassel, IPM, Wilhelmshöher Allee 71, 34121 Kassel, Germany

ABSTRACT

The tracking of observer's eyes positions in front of a 3D display is necessary to ensure a correct autostereoscopic view of position-dependent 3D images. We present a new real-time eye tracking system using two commercially available web cams detecting the observer's eyes in x, y and z direction. The entire system can be installed on a standard PC together with an autostereoscopic display. In a first process eye-candidates are detected by implementation of a fast pattern recognition. The additional use of the colour information of the web cams provides more useful information on finding eye-candidates. In a second step from eye-pairs that are nearest to a monitor world-coordinates are calculated. Signal transmission and processing delays are compensated by an adaptive predictor.

Thus the entire system is cheaper, smaller in size and it can be installed on a standard PC. In addition the tracking software can also support other applications, e.g. to set up a teleconference system in conjunction with an autostereoscopic monitor.

Keywords: Eye-position tracking, autostereoscopic display, stereo camera

1. INTRODUCTION

In recent publications we have presented several systems of autostereoscopic monitors without movable mechanical parts [1], [2]. This monitors enable an observer to see the perception of depth in a given area. Images for left and right eyes have to be placed on screen to obtain stereo viewing zones. This zones depend on to the observer's position. Thus a position detection is necessary.

Several head tracking methods exist using infra-red or ultrasonic position detectors. The main disadvantage of these head trackers is that the observer has to carry a reflector or any other kind of marker. According to Woodgate, (see [3]), this decreases the acceptance of autostereoscopic displays.

During the last years also video camera based high-tech head trackers have been developed. These trackers use special signal processing units to find the accurate observer position in real time avoiding the above mentioned disadvantage [4], [5].

But it has to be taken into account that often the costs for tracking systems exceed the costs of the system itself, thus augmenting the overall system costs in a non tolerable way.

We have developed an eye-tracking system using of the shelf web cams holding the overall costs low. By integration in our autostereoscopic systems there is the ability to generate two views of images and arrange them on screen according to the observer's position.

2. AUTOSTEREOSCOPIC DISPLAY

The "Position-adaptive Autostereoscopic Monitor (PAM)", based on location multiplex principle, is implemented on commercially available products without modifying the hardware, i.e. a TFT-monitor with digital interface and a graphic card with 3D acceleration is used, see Fig. 1. Optionally, a 3D-mouse can be added for easier controlling the perspectives of computer generated images.

To obtain the autostereoscopic 3D impression a low-cost parallax barrier or, for high-end applications, a more expensive lenticular sheet has to be placed in front of the display. Thereby, the size of the barrier can be

E-mail: andi@digi.e-technik.uni-kassel.de

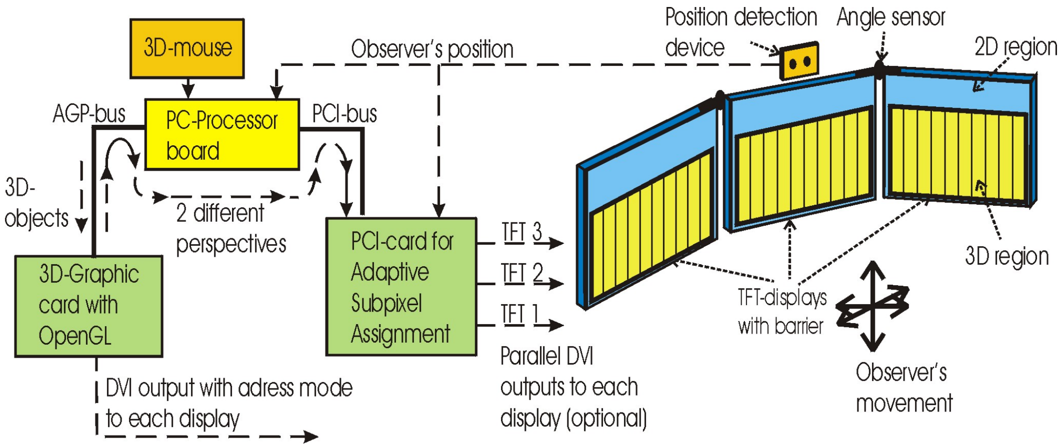

Figure 1: 3D-autostereoscopic system

defined according to the desired task. So the entire display can be covered by the barrier to have an exclusive 3D display. The display has also the possibility to cover certain areas to obtain 2D windows for convenient interaction, e.g. where the user can read texts or perform inputs.
For a detailed description of our autostereoscopic display see [1] or [2] and [6] for the multi-screen display.

The position detection device consists of our eye-tracker-system described in the following. Through the use of more than one camera a 3D-position of objects can be obtained employing well known stereo measurement methods.

3. EYE-TRACKER SYSTEM CONFIGURATION

The cameras are arranged in that way that they match the requirements for the autostereoscopic display system.

3.1. Camera Setup and Geometry

The camera setup is mainly determined by the required viewing area of an autostereoscopic monitor. This can vary with the size of the monitor. So the detection volume has the minimum and maximum viewing distance and the viewing angle to the monitor as parameters. As a third parameter the viewing zone (i.e. the angle) of the used cameras has to be taken into account. With the third parameter in mind, there are two possible camera setups depending on the direction of the optical axes:

- parallel camera setup
- toed-in camera setup

Due to greater non-overlapping area in the parallel setup, we decided for the toed-in setup, like in figure 2, by which we can achieve a greater detection volume to meet our requirements.

An object point in a monitor coordinate system can be easily described by considering the imaging process as a series of separate coordinate transforms (a detailed description can be found in [7]):

image space −>	sensor space −>	camera space −>	world space −>	monitor space
(xbi,ybi)	(xsi,ysi)	(xki,yki)	(Xw,Yw,Zw)	(Xm,Ym,Zm)

N.B. i denotes left or right camera.

For every step in the transformation process parameters of rotation and translation matrices have do be determined. The accuracy of the transformation of objects from the image space to monitor space depends on a well done calibration process.

3.2. Stereo Camera Calibration

For the camera calibration a simplified calibration method of Tsai[8] and Lenz[9] is used.

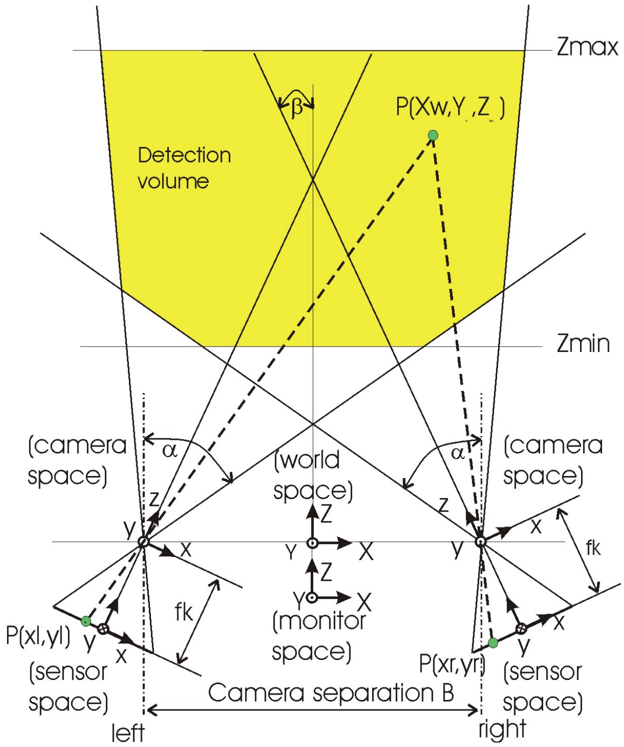

Figure 2: Stereo camera configuration

N.B.:	β	convergence angle
	α	view angle of the camera
	$Zmin, Zmax$	viewing rang
	fk	camera constant

As these calibration techniques are intended for very high precision measurements for a single camera, there are several parameters to be calibrated. Especially the lens distortion is modelled with several coefficients.
For our needs we are using a geometrical camera model described by 12 parameters: six external and six internal parameters. The lens distortion is modelled only by one parameter describing the radial lens distortion. We found that modelling the lens distortion with more than one coefficient doesn't increase accuracy and would lead to instable coefficients.

At the first step each camera was calibrated separately according to Tsai and Lenz to determine the 12 parameters. The cameras are looking to the same nonplanar calibration object placed approximately in the middle of the desired detection volume at a known distance to the monitor surface.
Then, because we want to use triangulation to calculate objects in real world coordinates, a transformation matrice from each camera coordinate system to a real world coordinate system is determined which is assumed to be in the middle of the baseline of both cameras.

But as we want to have the position of objects, i.e. in our case the observer's eyes, as a position to the monitor, the real world coordinate systems has to be transformed into a monitor coordinate system. Positioning the calibration object in known distance to the monitor (which gives the monitor space coordinates) and obtaining left and right camera coordinates of the corresponding points, a matrix equation can be solved to receive transformation and rotation matrix for transforming camera space into monitor space. So monitor space coordinates can directly be calculated from the camera space.

After the calibration all external and internal camera parameters are determined and the relationship between image coordinate system (xb, yb) and monitor coordinate system (Xm, Ym, Zm) is known.

4. IMPLEMENTATION

The eye-tracking unit consists of two web cams each connected to a USB controller . The cameras are equipped with a CCD sensor suitable for 3D-measurements. According to the camera setup described in 3.1 and to match the requirements for monitor system, the cameras are arranged about 10cm apart. The angle of convergence is ca. 5°.

Our image processing unit is a software solution, but it is designed to be implemented in the graphic board described in [6], too. The image processing unit performs several tasks:

1. Searching for eye-candidates in an Area-Of-Interest (AOI)
2. Excluding candidates until relevant eye-pair is found
3. Predict the 3D position in case no eye-pairs found (i.e. observer is blinking or to compensate the signal transmission and processing delays

A schematic control flow diagram of the entire processing is shown in Fig. 3. In the initialisation step of the system, the observer has to move into a standard position where a calibration image can be clearly seen. This known position gives the initial area-of-interest where the eyes are assumed. With this initial procedure we have the advantage of getting the observer's eye distance. So we have an exact value for disparity estimation when tracking the eyes.

After initialisation, the processing is controlled by the eye position tracking. The AOI is moved according to a predicted new eye position, so the eyes are again assumed to be in the middle of the AOI.

The position information of both eyes is linked to the interface of a 3D-program in order to adapt the appropriate viewing perspectives for each eye position. Thereby motion parallax is simulated. Also the information is used to optically address the user's eyes by the autostereoscopic 3D monitor. The location multiplex can be performed in two ways:

1. Linking the position information to a special graphic board of the autostereoscopic system. This increases speed enormous
2. Linking the position directly to the 3D-graphic program. The location multiplex is then performed by software

If the tracker fails finding eyes (e.g. eyes are closed longer than blinking or abrupt changing light conditions, etc.), it has to be restarted with the initiating procedure.

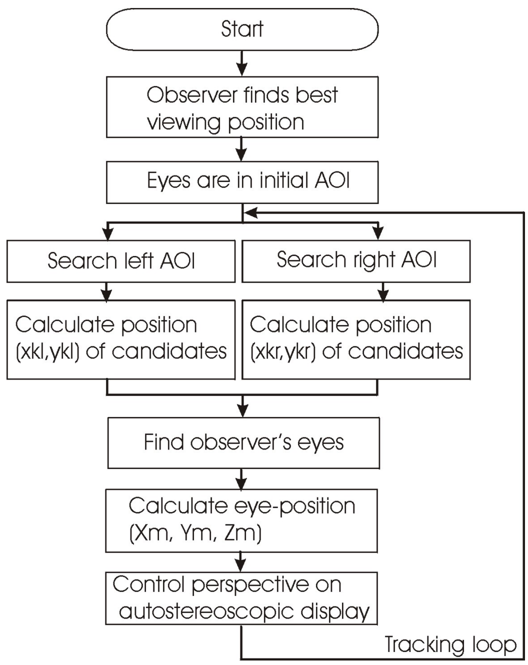

Figure 3: Control flow diagram

5. FINDING EYES IN IMAGE

Most known eye detection algorithms were designed for special purposes (like biometry, gaze detection etc.). Main disadvantages is that they need special equipment. Also they consume a lot of computing power or don't have real-time ability.

But the above mentioned criteria are mandatory for us. Considering this, we have to find a different approach for our eye-tracking.

Within the AOI for left and right eye of each camera eye features are selected and extracted. Thereby, colour information in the images helps us to obtain appropriate points by implementing a block matching technique. A reference pattern of the pupil and its surrounding, recorded with different users and distances, is compared within each AOI. Before the comparison starts, the image inside the AOI is transformed into a binary image using a window discriminator, thus reducing the image planes. If the grey value of all colour planes fits into a range representing eyes, it is marked one or the other case zero in the binary image.

The result of the comparison leads to an error ratio. The objects found during this process are expected to be eye candidates if the error ratio is below a threshold.

Objects, which don't form pairs and which are out of the tracking range, are excluded by a disparity estimation in the next step. These detected pairs which are nearest to the display are assumed to be the observer's eyes. The observer's position in real world coordinates is calculated by stereo correlation methods (see 3.1).

As stated previously, by using a restricted AOI the performance is increased. Thereby, the AOI is defined as a rectangle with 100x100 pixels in size. The rectangle size was calculated to track an observer with a maximum velocity of $0,5\frac{m}{s}$. Knowing the internal camera geometry, the resulting search area in real world covered by the AOI is approximately 10° horizontal and vertical. So it can be supposed that the observer's eyes remain inside the AOI during tracking.

Fig. 4 shows a situation where multiple persons are inside the images. But, only the eyes of the person M., closest to the monitor are tracked (marked by a white rectangle, the rectangle around it is the AOI) because person E (at left in each image) and person S (at right in each image) remain outside the search area. This holds also if multiple eyes are in one AOI. Because of disparity evaluation and the known eye distance of the tracked observer, they are excluded.

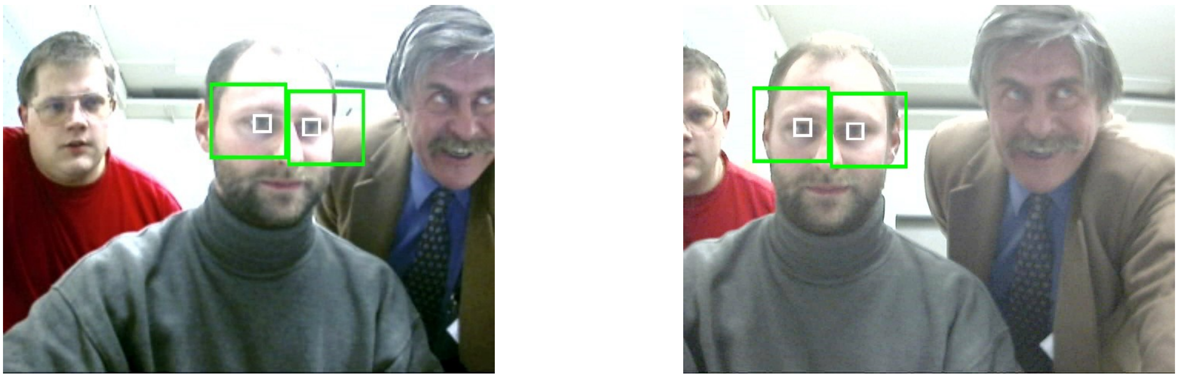

Figure 4: Detected eye-pairs taken from left and right stereo image

Continuation proofs within the prediction algorithm help to eliminate such frames which cannot lead to correct results, for example if the observer is blinking.

6. EYE-TRACKING PERFORMANCE

For the presented eye-tracking system, the following performance characteristics were achieved with respect to the used autostereoscopic display system:

distance range:	400mm ... 800mm
horizontal range:	500mm ... 700mm
max. detection error:	$\pm$1,5mm (in all directions)
head rotation:	$\pm 40^\circ$ (all directions)
measurement rate:	10Hz
camera resolution	640x480Pixel
detection certainty:	$>$ 94% (dependent on lightning)

The vertical angular detection range is determined by the vertical viewing angle of the used cameras and is sufficient for those types of autostereoscopic displays where the observer is viewing mainly in horizontal direction. So, the horizontal detection range is of greater importance and is determined by the camera geometry.

As our algorithm takes only ca. 10ms to process two images and to find appropriate eye-pairs, the measurement rate is determined by the image transmission time of the USB cameras (15Hz at 640x480 pixels) and can't be influenced by ourselves. A second bottle neck is time slot given by the operating system of the PC to perform a complete process, i.e. transmit images from two cameras and perform the eye finding.

In spite of the above mentioned restriction concerning transmission rate, a measurement rate of ca. 10Hz is realised, based on the described block matching algorithm.

An example of observer tracking can be seen in Fig. 5. Both eyes of an observer are recorded in x direction, to show detection ability. The x direction is the main direction observers tend to move. The measurement was taken at a distance of ca. 600mm. The x-axis indicate the frame number, whereas the y-axis indicate the x direction in monitor space. It can be seen that both eyes are tracked individually.

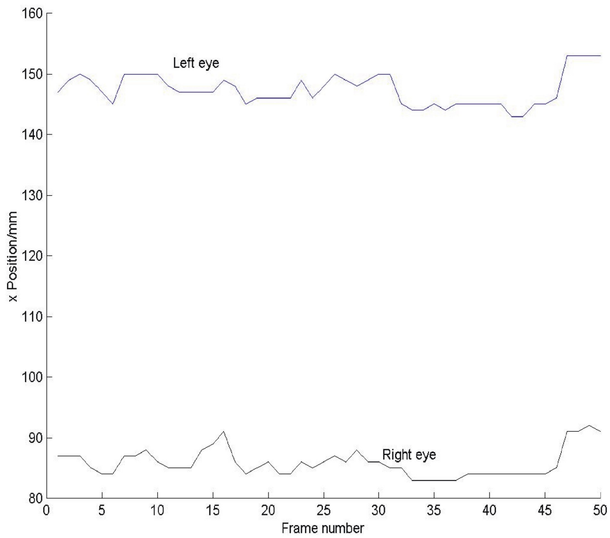

Figure 5: Observer tracking in x direction

In case the observer was blinking, the tracker has predicted the next position for the AOI to continue the search, and also the next eye positions were predicted for the following processing steps explained in 4.

Considering the results it is proofed that eye-tracker can track the eyes robust and stable in real-time.

Determining eye-pairs is complicated by light conditions which darken the eye region, whereas glasses are not crucial as long as the eyes remain completely visible and are not occluded by the frame.

7. CONCLUSION AND OUTLOOK

We have developed an eye-tracking system with web cams using specific image processing. The performance of our system is stable and reliable. By the use of of-the-shelf components we achieve a low-cost tracking system from which we assume that the acceptance of autostereoscopic displays simulating motion parallax can be augmented.

Although it is mainly intended for integration in autostereoscopic displays, of course, the tracking system supports also other applications with web cams, e.g. setting up a teleconference system in conjunction with the autostereoscopic display.

With the availability of web cams using the USB2 standard the image resolution and transmission speed of those cameras are increased. Further developments concentrate on using this standard to obtain higher detection rate and area and to handle difficult light conditions.

REFERENCES

1. M. Andiel, S. Hentschke and A. Herrfeld, "Autostereoskopische 3D-Displays," *Electronic Displays* **16**, pp. 154–158, 2001.
2. S. Hentschke, A. Herrfeld, M. Andiel et alt., "Person adaptive autostereoscopic monitor (PAM)," *IEEE, International conference on consumer electronics* **Paper WPM P2.07**, pp. 226–227, 2000.
3. G. Woodgate, J. Harrold, A. Jacobs and D. Ezra, "Flat panel autostereoscopic displays: characterisation and enhancement," *Proc. SPIE* **3957**, pp. 153–164, 2000.
4. H. Heidrich, A. Schwerdtner, A. Glatte et alt., "Eye position detection system," *Proc. SPIE* **3957**, pp. 192–197, 2000.
5. H. Imai, S. Tsujikawa and M. Imai, "Viewing point detection system using specific image processing for eye-position tracking autostereoscopic display," *Proc. SPIE* **3639**, pp. 92–98, 1999.
6. J. Börcsök, S. Hentschke, M. Andiel et alt., "Autostereoscopic adaptive multi-screen systems," *IEEE, International conference on consumer electronics* **10**, p. 10, 2001.
7. R. Klette et alt., *Computer Vision. Räumliche Information aus digitalen Bildern*, F. Vieweg u. Sohn Verlagsgesellschaft, Braunschweig, München, 1996.
8. R. Y. Tsai, "An efficient and accurate camera calibration technique for 3d machine vision," *Proc. Intl. Conference on Computer Vision and Pattern Recognition*, pp. 364–374, 1986.
9. R. Lenz, "Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standartobjektiven für hochgenaue 3D-Messungen in Echtzeit," *Informatik aktuell*, pp. 212–216, 1987.