Accurate 3D eye tracking for multi viewpoint systems

Article · December 1997

Source: CiteSeer

CITATIONS
3

READS
145

3 authors, including:

André Redert
Rodotti
46 PUBLICATIONS **827** CITATIONS
SEE PROFILE

Emile Hendriks
Delft University of Technology
146 PUBLICATIONS **2,137** CITATIONS
SEE PROFILE

Accurate 3D eye tracking for multi viewpoint systems

André Redert, Joost-Jan van Klaveren, Emile Hendriks

Information Theory Group, Department of Electrical Engineering
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
phone +31 15 278 6269, fax +31 15 278 1843
email {andre, joostjan, emile}@it.et.tudelft.nl
http://www-it.et.tudelft.nl

ABSTRACT

In this paper we present a system for accurate 3D tracking of pupils of human eyes. In multi viewpoint video systems, this information is needed to present 3D scenes correctly to the viewer on a stereo display.

The system input is a stereo image sequence, taken from the viewer by cameras at the sides of the display. The system is based on four steps: global eyes detection, generation of eye feature candidates, selection of the best combinations leading to 2D pupil positions and stereopsis assisted by camera calibration leading to 3D pupil positions.

Experimental results show that the system obtains accurate 3D pupil positions with an error in the order of 2 mm.

1. INTRODUCTION

In 3D video communications, one of the most promising techniques is the multi viewpoint system [2,6,11] shown in Figure 1. The system provides a sensation of depth and motion parallax based on the position and motion of the viewer. At the presentation side new images are synthesized, based on the actual position of the viewer. The multi viewpoint system requires a head tracker that measures the 3D viewer position with respect to the stereo display.

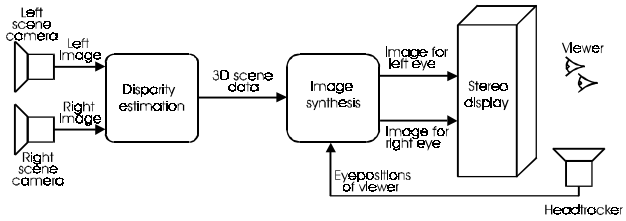

Figure 1: A multi viewpoint 3D system

Other applications that need head tracking are for instance model based coding [1,3] and face recognition [5]. Head trackers developed for these applications supply 2D positions in image coordinates, so they are not suited for the multi viewpoint system.

For complexity reasons, multi viewpoint systems often provide intermediate viewpoints only [2,10]. In this case the head tracker has to measure only the x position of the viewer [4]. For a full multi viewpoint system that provides any viewpoint [6], we need a head tracker that measures the exact 3D position of the viewer.

Although the image synthesis part needs the position of the *pupils* of the viewer's eyes [6], the term *head* tracker is widely adopted. In the remainder of this paper both head and eye tracking mean tracking of the pupil.

In this paper we will present an accurate 3D tracking system for the pupils of the viewer. The system is based on recording the viewer with a stereo camera, followed by four processing steps. The steps include global localisation of the eyes, generation of eye feature candidates, selection of the best combinations and stereopsis to obtain the 3D pupil coordinates.

In section 2 the system is described. Section 3 gives our experimental results. Finally in section 4 we give conclusions and recommendations for further research.

2. HEAD TRACKING SYSTEM

In this section we describe our head tracking system as shown in Figure 2. The system input is a stereo image sequence of the viewer, obtained by a stereo camera. At the output the estimated 3D locations of the left and right eye of the viewer are available, relative to the stereo display in the multi viewpoint system.

The tracking system is based on four steps. First the head and eyes are localised globally. Then a number of eye feature candidates is generated. After that the best combinations of candidates are selected and the pupils of both eyes are localised in each tracker camera image. In the final step 3D pupil coordinates are obtained from the stereo 2D pair of coordinates by stereopsis. To obtain a reliable stereopsis, we calibrate the tracker cameras.

In a two way video system, the same cameras can be used for recording the scene and for head tracking. In this case a tracker camera disparity field is available that can be used to increase the accuracy of the steps in the head tracker.

In the next sections we will describe the camera setup, disparity estimator and the four head tracker steps.

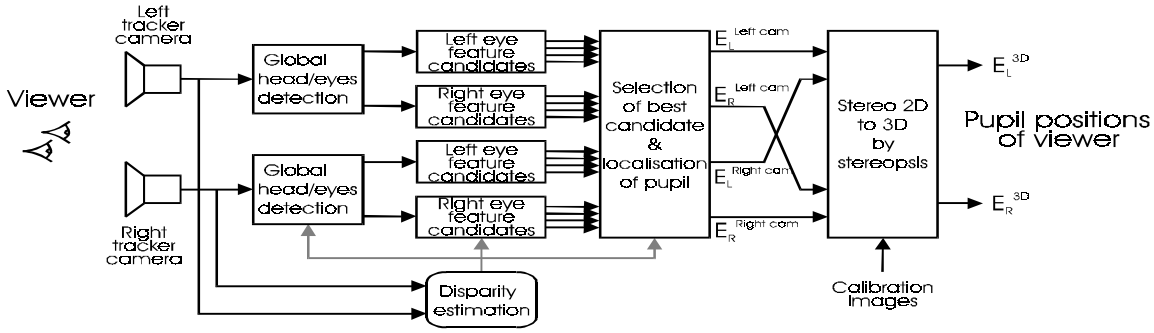

Figure 2: The head tracking system

2.1 Camera setup

Figure 3 shows the camera setup. The tracker cameras are placed on both sides of the stereo display. In this way the scene cameras can be used as tracker cameras in bidirectional multi viewpoint applications such as teleconferencing and videophony.

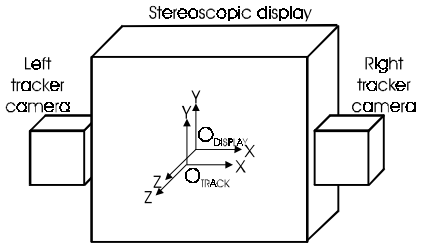

Figure 3: The camera setup

The display reference frame is centered in the display and the tracker reference frame is centered in between the camera optical centres. The head tracker will measure the pupil positions in the tracker camera reference frame, while the multi viewpoint system requires pupil positions in the display reference frame. So we need to know the relation between the two reference frames.

Currently we align the two reference frames manually. It is a challenge to obtain an accurate relationship between the reference frames since the display itself is not visible in the tracker camera images.

2.2 Disparity estimation

As shown in Figure 2, a disparity field between the left and right tracker camera images is estimated and fed into the first three stages in the head tracker system. In this way the similarities in left and right tracker images can be used to enhance the performance of these steps.

In the multi viewpoint system a disparity field is estimated for the scene camera images. If the scene cameras are used as tracker cameras, the available scene disparity field can serve as tracker disparity field.

We used the disparity estimator described in [7]. Other estimators can be used as well, provided they perform well in the area around the viewer's eyes.

2.3 STEP 1: Detection and global localisation

For the detection and global localisation of the eyes we use the automatic technique described in [5]. This system uses a background image to segment the image into foreground/background. Then the head and eyes are detected globally.

We apply the algorithm both on the left and the right image to obtain estimates of the 2D eye positions in the left and right image separately. The tracker disparity field can be used to increase the robustness of the detection, which is still under investigation, and to increase the accuracy. The latter is not necessary since step 2 needs only a very rough eye position.

2.4 STEP 2: Generation of eye feature candidates

For the generation of eye feature candidates, we use the very promising neural network approach described in [8]. This method can track a left eye in a monoscopic image sequence. The approach searches for four eye features (left, right, top, bottom) in a manually defined window around the left eye. A post processing step searches for the combination of features that best matches with the shape of a left eye.

In our system, the search window is provided by the global localisation step. The post processing in [8] is not present in step 2, to allow for a new post processor that can make use of stereo correspondences in step 3.

We run the algorithm separately on both left and right sequences. To obtain the feature candidates of the right eye, the image data in the window around the right eye is mirrored. This is necessary since the neural network is trained on left eyes.

The output of step 2 consists of 16 lists, one for each tracker camera, each viewer eye and each feature. Each list contains a number of possible image locations for that particular eye feature and a probability value.

2.5 STEP 3: Accurate localisation of the pupils

For the accurate localisation of the pupils we select the best combinations of features in the 16 lists from step 2. We choose the combination method equal to the

monoscopic post processing described in [8], performed on both the left and right image features.

After the best four eye features are found for each tracker camera image and each viewer eye, we average the positions of the left and right features to obtain the estimated 2D position of the pupil. We assume that the pupils are centered in the eye.

Step 3 provides four 2D measurements: the locations of both the left and right eye pupil in the images of both the left and right tracker cameras.

2.6 STEP 4: From stereo 2D to 3D coordinates

In step 4 the 2D measurements from step 3 are transformed into the final 3D positions by stereopsis. To obtain a reliable result, we calibrate the tracker cameras following the method described in [9].

The calibration method has as input (at least) two stereo calibration images, that contain a specific calibration object. This calibration method provides us with measurement coordinates in the reference frame of the calibration object. We use an extra post processing step that transforms the measurements to the tracker reference frame. The transform is based on the camera parameters that result from the calibration.

3. EXPERIMENTAL RESULTS

We recorded a new stereo test sequence, the ANNET sequence. Figure 4 shows the first left and right frames. The sequence is accompanied by calibration images. For the experiments we used each fifth of the first 191 frames of the sequence (1,6,11, etc.).

Figure 4: Stereo test sequence ANNET

Figure 5 shows an enlarged part of the first frame including the estimated eye features and pupils.

Figure 5: Best feature candidates (•) and pupils (×)

Figure 6 shows the obtained 3D coordinates of the left and right pupils. Subjectively, they are in accordance with the motion in the image sequence. At the start of the sequence the person is moving very little, at the end she starts laughing.

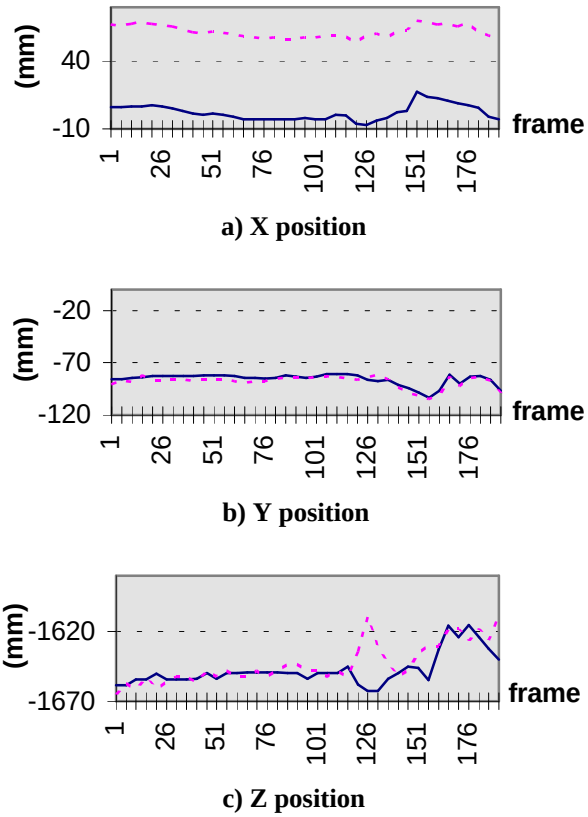

Figure 6: Measured 3D positions of the left pupil (solid line) and the right pupil (dotted line)

Since we do not have a reference tracker, we used the following two procedures to estimate the error in the found 3D pupil positions.

In the first method we estimate the error in the 3D pupil positions based on a known output error of step 3 and a linearisation of the relation between input 2D and output 3D coordinates of step 4. The found eye features in step 3 are within 2 pixels of the true positions [8]. The linearisation of step 4 is a very complex, position dependent task to perform analytically. Therefore we performed it numerically for each found 3D eye position. The linearisation results in 12 numbers: the derivatives of the 3D eye position XYZ with respect to the 2D left XY pupil and 2D right XY pupil positions.

For the eye positions in the ANNET sequence we found that the linearisation is accurate for deviations of up to five pixels in the 2D domain. Averaging over the sequence and the four possible 2D errors (left/right image, XY position), the resulting 3D errors are 0.72

mm/pixel (X), 0.68 mm/pixel (Y) and 3.63 mm/pixel (Z). So given the 2 pixel errors in step 3 we obtain a pupil tracking accuracy of 1.4 mm (X), 1.4 mm (Y) and 7.2 mm (Z).

In the second method, we checked if the estimated 3D distance between left and right eye is invariant as is to be expected. Figure 7 shows the estimated distance between the left and right pupils.

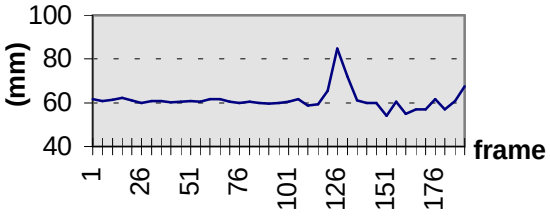

Figure 7: Measured 3D distance left/right pupils

Centred at frames 126 there is an outlier, caused by an error in the right eye position. At this point in the image sequence the right eye becomes partly occluded by the head.

Averaging over all frames we found an eye distance of 61.2 mm and a standard deviation of 4.8 mm. Excluding the outlier this results in a distance of 60.2 mm and a deviation of 2.2 mm. Dividing this over the two estimated pupils we obtain an error of 3.4 mm (any 3D direction) for all frames and 1.6 mm excluding the outlier. This is in accordance with the first error estimation method.

4. CONCLUSIONS AND FURTHER RESEARCH

A head tracking system has been designed for application in multi viewpoint 3D video systems. The objective is the accurate 3D localisation of the viewer's left and right pupils.

Experimental results show that the system obtains accurate 3D pupil positions with an error in the order of 2 mm.

In future research we would like to incorporate a calibration procedure for the automatic alignment of the tracker and display reference frames. In this procedure we expect to be able to calibrate the slightly curved display shape as well.

ACKNOWLEDGEMENT

This work was done in the framework of the European ACTS project PANORAMA. One of the major goals is the realisation of a real time multi viewpoint system in hardware.

REFERENCES

[1] K. Aizawa, H. Harashima and T. Saito, "Model-based analysis-synthesis image coding (MBASIC) system for a person's face", Signal processing: *Image Communication* 1, 1989, pp. 139-152

[2] B. Chupeau and P. Salmon, "Synthesis of intermediate pictures for autostereoscopic multiview displays", in Proceedings Workshop on HDTV '94, Turin, Italy, 1994

[3] A. Eleftheriadis and A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates", Signal Processing: *Image Communication* No. 7, 1995, pp. 231-248

[4] C. Maggioni and J. Liu, "Headposition sensor", internal report AC092/SIE/SN5/DS/I/007/b1 of the European ACTS project PANORAMA, september 1996

[5] A.V. Nefian, M. Khosravi and M.H. Hayes, "Real-time detection of human faces in uncontrolled environments", Proceedings of SPIE conference on Visual Communications and Image Processing, Vol. 3024, San Jose, California, USA, 1997, pp. 211-219

[6] P.A. Redert, E.A. Hendriks and J. Biemond, "Synthesis of multi viewpoint images at non-intermediate positions", in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, 1997, pp. (to be published)

[7] P.A. Redert, E.A. Hendriks, C.J. Tsai and A. Katsaggelos, "Disparity estimation with modelling of occlusion and object orientation", submitted to VCIP '98, to be held in San Jose, California, USA, 1997

[8] M.J.T. Reinders, R.W.C. Koch and J.J. Gerbrands, "Locating facial features in image sequences using neural networks", in Proceedings of the *Second International Conference on Automatic Face and Gesture Recognition*, Killington, USA, 1997, pp. 230-235

[9] H.J. Woltring, "Simultaneous Multiframe Analytical Calibration (S.M.A.C.) by recourse to oblique observations of planar control distributions*", SPIE Vol. 166 Applications of Human Biostereometrics (NATO)*, 1978, pp. 124-135

[10] J.S. McVeigh, M.W. Siegel and A.G. Jordan, "Intermediate view synthesis considering occluded and ambiguously referenced image regions", *Signal Processing: Image Communication* 9, 1996, pp. 21-28

[11] European ACTS AC092 PANORAMA project proposal, september 1995