# Detection of Head Pose and Gaze Direction for Human-Computer Interaction

Ulrich Weidenbacher, Georg Layher, Pierre Bayerl, and Heiko Neumann

University of Ulm, Germany
{Ulrich.Weidenbacher, Georg.Layher, Pierre.Bayerl,
Heiko.Neumann}@uni-ulm.de

**Abstract.** In this contribution we extend existing methods for head pose estimation and investigate the use of local image phase for gaze detection. Moreover we describe how a small database of face images with given ground truth for head pose and gaze direction was acquired. With this database we compare two different computational approaches for extracting the head pose. We demonstrate that a simple implementation of the proposed methods without extensive training sessions or calibration is sufficient to accurately detect the head pose for human-computer interaction. Furthermore, we propose how eye gaze can be extracted based on the outcome of local filter responses and the detected head pose. In all, we present a framework where different approaches are combined to a single system for extracting information about the attentional state of a person.

## 1 Introduction

### 1.1 Motivation

Interaction between humans and computers is commonly restricted to typing on a keyboard or pointing and clicking the mouse button. This type of interaction is very unnatural for humans since human-human interaction is commonly based on multi-modal interaction. For example, in a conversation auditory and visual information is important to be interpreted properly in order to react to a dialog partner. In such a conversation important visual cues can be facial expression, head pose and particularly the eye gaze for getting feedback about the attentional and mental state of a dialog partner [Emery, 2000].

### 1.2 Previous Work on Eye Gaze Estimation

One of the first applications that utilizes eye gaze as a computer interface was developed by [Hutchinson et al., 1989] where computer users could interact by directing their gaze to specific areas on the monitor. Similar to recent eyetracker applications [Eyelink, 2006] their system requires infrared light to illuminate the eye region. In general, state-of-the-art eyetracker applications are highly accurate and reliable, however most of them require complex hardware (helmet with a mounted camera) or the user has to be in a fixed position (e.g. with a chin chest). Other approaches which are not constrained by specific

E. André et al. (Eds.): PIT 2006, LNAI 4021, pp. 9–19, 2006.

hardware, referred to as 'non-intrusive' systems, have to compensate for motion affects of the user (e.g. by using tracking methods [Baluja and Pomerleau, 1994, Ji and Zhu, 2003, Zhu and Ji, 2005, Yoo and Chung, 2005]), though they still employ active sensing tools (e.g. by illuminating the eyes with infrared light). In this contribution, we concentrate on purely vision based methods ([Steifelhagen et al., 1997, Heinzmann and Zelinsky, 1998]) where eye gaze is estimated passively (i.e. without special illumination).

### 1.3 Combining Head Pose and Eye Gaze

There is a large amount of work present for the detection of head pose [Gee and Cipolla, 1994, Krüger et al., 1997, Rae and Ritter, 1998, Wang et al., 2003] and for eye gaze estimation, but there is only few work present where both information, head pose and eye gaze are combined. For example, [Matsumoto and Zelinsky, 2000] present a three stage system that combines head pose and eye gaze information to accurately estimate a person's point of attention. The particular point of their system is that they use a 3D head model and a 3D eye model to accurately detect head pose and eye gaze based on stereo vision. However, our proposed methods are based on monocular images.

### 1.4 Overview

In this contribution, we focus on methods for the estimation of head pose and eye gaze for human-computer interaction purposes. We compare two methods for the estimation of head pose. (1) a view-based approach where a small set of prototypical views of a head is used to determine the pose of a presented test head [Krüger et al., 1997] and (2) a model-based approach where geometrical information about the face is utilized to determine the head pose of a person [Gee and Cipolla, 1994]. In addition to head pose, eye gaze is an important cue for the detection of attention [Emery, 2000]. Thus, we present a novel method that uses phase information of simple biological motivated filters to determine the direction of gaze from a person.

## 2 Methods

### 2.1 Image Acquisition and Ground Truth Generation

We created a dataset of images showing different head pose / eye gaze conditions acquired from 5 subjects. The procedure of acquiring the images was as follows: we fixated a laser pointer on a tripod equipped with an angular meter. The device was used to accurately attach marks to the walls in horizontal steps of $10°$ ranging from $-90°$ to $90°$ as shown in Fig. 1a. Then, a subject had to sit on a swivel chair facing the camera. According to Fig. 1c, two clips were used to mount the laser pointer on top of the subjects heads. To calibrate the position of the laser pointer on the subjects heads the height of the chair was first adjusted so that the laser pointer was located at a level according to the marks (the height of the marks was 140 cm; see Fig. 1b). Then, the subject was told to look straight into the camera. Retaining this state, the position of the laser pointer was corrected until the laser beam spot coincided with the 0° mark. After successful calibration, images with ground truth data of different head pose/eye gaze configurations were recorded by asking the subject to align the laser pointer spot to a specific mark on the wall (head pose) while focusing another mark with the eyes (eye gaze). In this manner, 285 images from five subjects were taken with a digital camera (Casio QV-5700, maximal optical zoom to avoid perspective distortion effects).

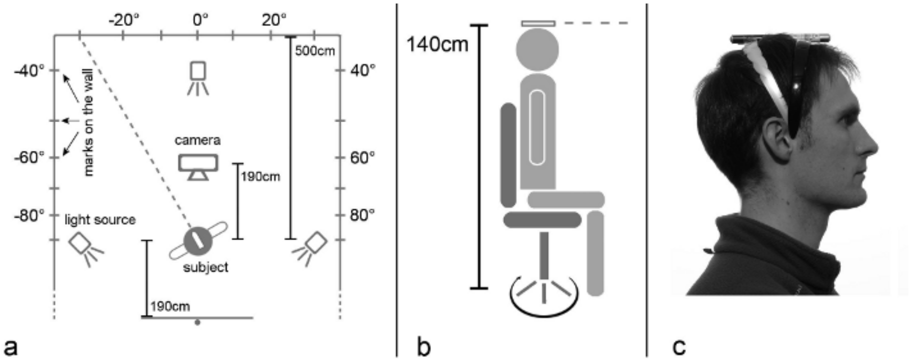

**Fig. 1.** (a) shows the setup used for image acquisition. The distance between subject and camera as well as the distance between the canvas and the subject was 190 cm. The range between subject and the opposite wall was 500 cm. Three spotlights were used to generate uniform background illumination and to prevent cast shadows. A laser pointer was mounted on a tripod equipped with an angular dimension to mark positions on the walls. These marks are later used to orient the head to a specific direction. (b) side view of a subject during image acquisition. The height of the chair was adjusted until the subject's laser pointer was located at a height of 140 cm. (c) visualizes the fixation of the laser pointer on the head of a subject.

## 2.2 Head Pose Detection

We compare two different computational approaches to determine the rotation of the head around its vertical axis (yaw or heading), namely a view-based approach and a model-based approach.

Our first method is a view-based approach which employs a simplified version of Elastic Graph Matching (EGM) proposed by [Krüger et al., 1997]. Seven images of different head poses from one individual (from $-90°$ to $90°$ in steps of $30°$ around the vertical axis) are utilized as prototype poses. On each image we manually select 10 landmarks on the face covering the nose, eyes and mouth in frontal view and also the ears in profile view (see Fig. 2a). The prototype pose

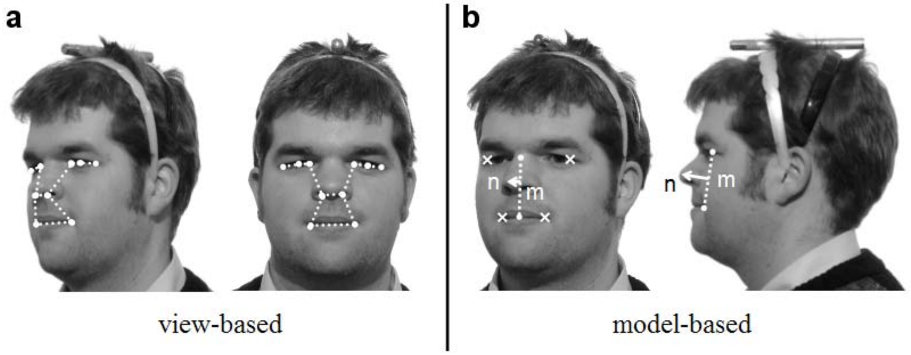

**Fig. 2.** Labeled head for the view-based approach (a) and the model-based approach (b). In the view-based approach we labeled 7 head poses from −90° to 90° in steps of 30°. Moreover, for each head pose we selected 10 facial features leading to different graphs for each pose. On each node in the graph Gabor filter responses of different orientation and scale are extracted. In the model-based approach the positions of the eye corners and the mouth corners are manually labeled to determine the symmetry axis of the head. Furthermore, the position of the nose tip is labeled manually. The projected nose length $n$ relative to the height of the face $m$ gives information about the pose of the head.

images were convolved using a family of 40 DC-free Gabor wavelets (5 frequencies x 8 orientations). On each landmark the set of 40 complex Gabor coefficients (Gabor jet) is extracted and stored together with the relative positions of the landmarks. This is done for each pose leading to 7 prototypical pose representations (bunch graphs). To detect the position and pose of a novel face, the bunch graphs are shifted over the new image while on each position in the image a similarity value is computed by a normalized cross-correlation between the Gabor responses stored in the graph and the present Gabor responses in the image. The position of the face is determined by the location in the image with the maximal correlation result. For pose estimation we consider the responses of all prototype graphs at this location. Here, we fit a quadratic function onto the prototype responses and determine the estimated head pose at the maximum of this function.

The second method employs a model-based approach for the estimation of head pose proposed by [Gee and Cipolla, 1994]. Here, we assume that localized features, namely the corners of the eyes, the mouth and the nose tip have been already detected in the image. As in the first approach we restrict the possible movements of the head to rotations around the vertical axis. Under the assumption of weak perspective projection described in [Trucco and Verri, 1998] the distance $n$ from the nose tip to the symmetry axis of the face relative to the length m (height of the face) is proportional to the sine of the head angle (see Fig. 2b). Thus, the head angle is computed by $sin^{-1}$ of the projected nose length n relative to the projected height $m$ of the face.

### 2.3 Detection of Gaze Direction

For the estimation of gaze direction we propose to employ Gabor filter responses similar to the EGM method used for pose estimation. Here, we evaluate the phase of Gabor responses [Gabor, 1946] in facial sub-regions around the eyes [Langton et al., 2000]. The idea is that the iris region is always darker than the remaining regions on the sclera (see Fig. 3; [Sinha, 2000, Langton et al., 2000]). Thus, gradual eye movements imply a gradual change of the Gabor phase. Therefore, we conclude that there exists a direct relation between Gabor phase and eye gaze dependant on the head pose.

The phase representing the optimal Gabor pattern matching the underlying image pattern at one specific location is described by Eq. 1:

$$\text{phase} = \text{atan2}(I * G_{\text{sin}}, I * G_{\text{cos}}) \tag{1}$$

where I is the input image, $*$ is the convolution operator, and $G_{\text{cos}}$ and $G_{\text{sin}}$ are the real and complex parts of the Gabor filter as follows (Eq. 2):

$$G_{\text{cos}} + iG_{\text{sin}} = \exp(i\frac{2\pi}{\lambda}x)\exp(\frac{-(x^2+y^2)}{2(0.45\lambda)^2}) \tag{2}$$

where $\lambda$ is the Gabor wave length and $x, y$ are image coordinates. Note that our Gabor filters are self-similar which means that the number of wave cycles under the Gaussian envelope function remains constant for values of $\lambda$.

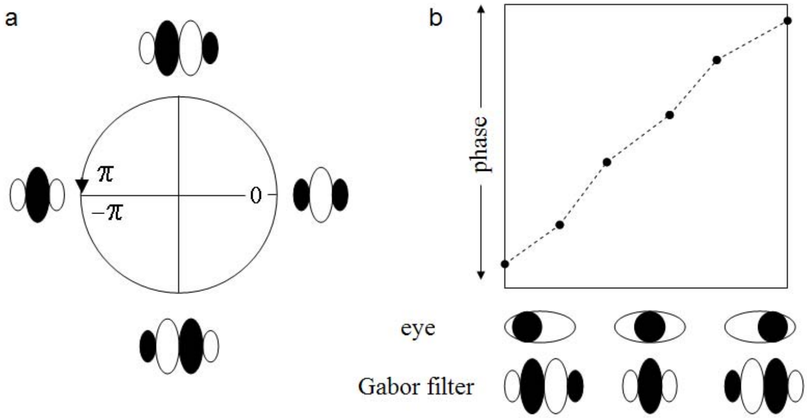

**Fig. 3.** Gabor filters consist of a wave function multiplied with a Gaussian envelope. (a) changing the phase of the wave function leads to different Gabor filter shapes. White indicates positive filter components while black stands for negative filter components (displayed are the real parts of the Gabor filter). (b) the filter that best matches the underlying eye pattern determines the phase for a specific eye gaze direction. Note that the phase is a circular measure which may lead to discontinuities when visualized in Cartesian coordinates. To avoid discontinuities the phase is shifted for visualization purposes where necessary.

For the estimation of gaze direction we generate a lookup table describing the relation between extracted image phase, gaze and head pose. This lookup table (LUT) is then utilized to map the extracted phase to the appropriate gaze direction for a given head pose[1].

## 3 Simulations

### 3.1 Head Pose Estimation

For the view-based approach the prototype bunch graphs were obtained from images of one person excluded from the test dataset as described in the previous section (see Fig. 5). For the model-based approach the length of the nose relative to the face length was measured manually for each person. Fig. 5 illustrates the estimated pose for 19 presented head poses of one person ranging from −90° to 90°. Despite that both approaches for pose estimation are rather simple in their implementation and that not much effort was put in training or calibration we obtain results which allow to determine the horizontal head orientation up to an accuracy of approximately 10°. Our experimental investigations show that 75%

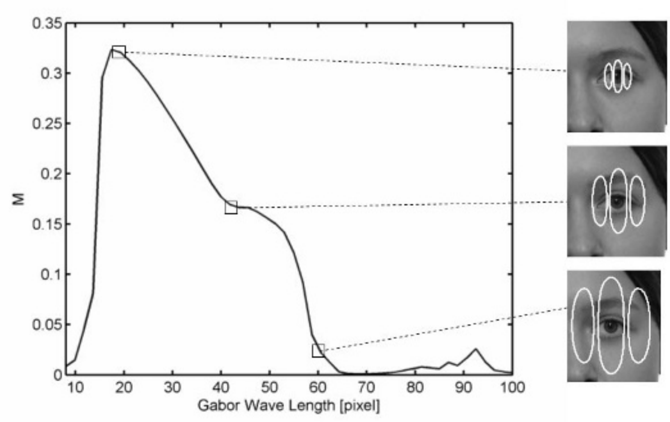

**Fig. 4.** Analysis of different Gabor filter proportions. The measure M (averaged over all heads in frontal pose) is plotted across different Gabor filter sizes. High values of M indicate phase linearity in combination with a high gradient of the phase. On the right we show three Gabor filters of different size belonging to different positions in the graph. The graph illustrates that there is an optimal Gabor filter size of about 20 pixels per cycle where slope and linearity of the phase are both high. As the filter size increases more and more adjacent parts of the eye are covered by the filter which results in a gradual drop of M towards zero. Note that the average width of the eye region within the image was about 45 pixels.

[1] This operation requires a one-to-one mapping between gaze direction and phase for each given head pose (compare section 3.2).

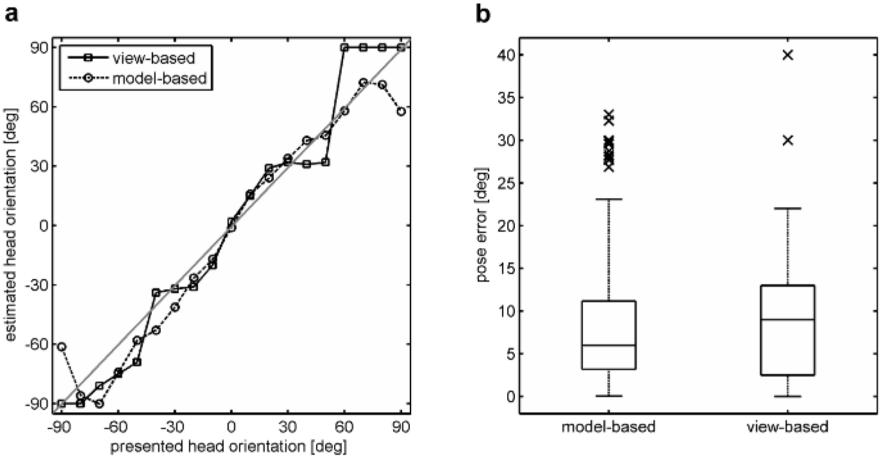

**Fig. 5.** (a) shows the estimated head pose for one head using the view-based and the model-based approach respectively. The grey line indicates optimal head pose estimation. The result demonstrates that both approaches yield good estimation results for presented head poses between −30° and 30°. Results near −90° and 90° show significantly higher errors for both approaches. (b) summarizes the distribution of the error over all sample cases for both approaches (excluding the labeled head for the view-based approach). In both cases 75% of the errors are smaller than 14°.

of all estimated head poses over all tested input images are smaller than 14° for the view-based method and smaller than 12° for the model-based approach (in accordance with the investigations in [Gee and Cipolla, 1994]).

### 3.2 Eye Gaze Estimation

**Gabor parameters.** Given the head pose and the location of the eyes it is possible to investigate the properties of the phase of Gabor responses. To determine the optimal wave length $\lambda$ of the Gabor filter we introduce a measure M that describes the quality of the Gabor filter for gaze estimation.

Fig. 6 exemplifies that a gradual change of the eye gaze direction leads to a gradual change of the Gabor phase. Therefore, we conclude that a good choice of the filter could yield a near linear dependency of the phase from the gaze direction. Moreover, a large slope of the linear dependency helps to prevent ambiguities in the mapping between phase and gaze. In other words, if the angular distance between phases is very small (i.e. low slope) then the discriminative power of the phase LUT gets lost. To investigate the phase linearity across eye gaze direction we fit a linear function to the measured phases and consider the sum of residuals as a quantitative measure for the linearity of the phase. Thus, we define M as follows (Eq. 3):

$$M = \frac{m^2}{1 + \sum r_i^2} \tag{3}$$

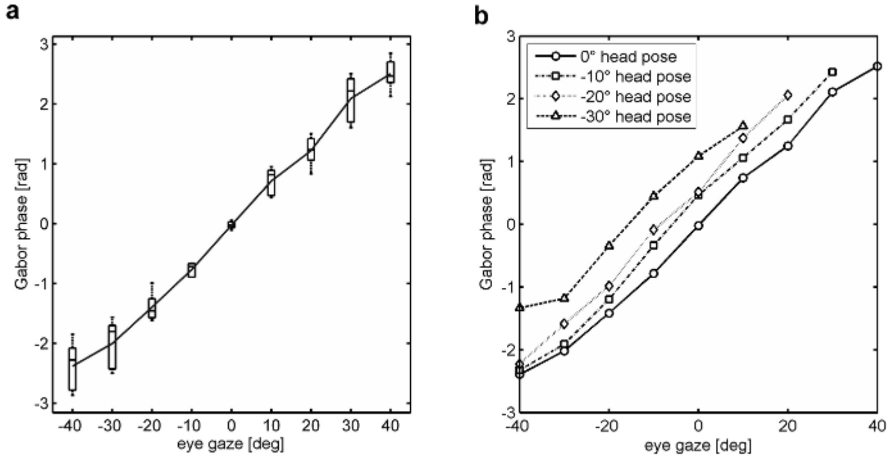

**Fig. 6.** (a) shows the average Gabor phase extracted from all heads (solid line) for frontal head poses across different eye gaze directions (gaze angles are given in world coordinates). Phase variances are smaller for central gaze directions than for peripheral gaze directions. (b) shows Gabor phases (averaged over all heads) acquired from four different head pose conditions reaching from -30 to 0 degrees. The extracted phases suggest a linear relation between gaze direction and Gabor phase. Note that we place the Gabor filters on the eye that is within the facial part of better visibility (the left eye for negative head angles and the right eye for positive head angles).

where $m$ is the gradient of the linear fit function and $r_i$ is the residual error. M should be maximal if both conditions (linearity and large slope) are present in the phase responses. In Fig. 4 we show M across different Gabor filter sizes. Small Gabor filters (smaller than the eye region) produce phase discontinuities resulting in a drop of M. Filter sizes in the proportion of the eye region lead to maximal values of M followed by a gradual decrease of M for larger filter sizes. We therefore set the size of our Gabor filters to 20 pixel per cycle for our images (corresponding to the size of the eye; see Fig. 4b).

**Gabor phase results and gaze estimation.** Fig. 6a illustrates the phase averaged over all heads in frontal head pose. The variance is very small when the eye looks straight and increases gradually when the eye looks to the left or to the right. Fig. 6b shows the extracted phase information for different gaze directions and four different head poses. As expected, the outcome suggests that for all presented head poses the phase can directly be mapped to the gaze direction.

In line with perceptual investigations [Sinha, 2000, Langton et al., 2004] our approach suggests that based on the head pose, the gaze can be determined by the local distribution of luminance values within the eyes. Consistent with experimental observation of [Gibson and Pick, 1963] the gaze direction is determined by the eye pattern (the phase) relative to the face configuration (head pose).

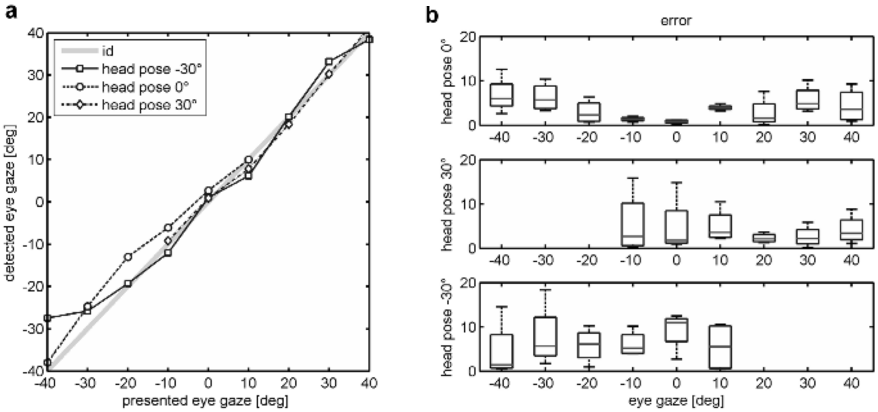

**Fig. 7.** (a) shows estimated eye gaze directions for one single head for three different head poses. The gray line indicates optimal gaze estimation. (b) shows errors across all heads for three different head poses (−30°/0°/30°). Overall the results demonstrate that in nearly all conditions 75% of the estimation errors are below 10 degrees. In frontal head pose estimation errors are specifically low for 0° gaze direction (corresponding to the mutual gaze condition).

Thus, Gabor phase responses are learned from one head by creating a simple lookup table between gaze direction, head pose, and the extracted phase. For a given head pose the gaze of a test face can now be determined by matching the detected phase to the linearly interpolated phases in the lookup table. The results are illustrated in Fig. 7a where we show the estimated eye gaze based on the learned phases for one test head. Fig. 7b shows the gaze detection errors for all test heads (excluding the head used for generating the LUT). Errors are mostly under 10 degrees and are significantly small (almost zero) for frontal head pose and straight eye gaze.

## 4 Discussion and Conclusion

In this contribution we describe how a database of images is generated showing faces from different head poses with different gaze directions. In contrast to other image databases [Phillips et al., 2000, Sim et al., 2003] our experimental setup allows to provide a ground truth for both gaze direction and head pose. We compare two approaches for head pose estimations and present how the gaze direction can be extracted from the local phase of a given gray-level image of a person's face.

For gaze detection we present a quality measurement to determine the parameters of the employed Gabor filter. Our measurement is based on the linearity and the slope of the relation between gaze direction and extracted Gabor phase. Note that ambiguities between gaze configurations can occur if the complete range of possible phases is covered by this relation caused by the circularity of the phase (e.g., $-\pi = \pi$, see Fig. 3). To take this into account we choose the parameters of the Gabor filters so that the angular distance between minimal and maximal detected Gabor phase is no less than $\frac{\pi}{8}$.

We claim that all approaches for head pose estimation as well as for gaze detection that we investigated here either utilize information that is represented in the visual system or induce perceptual effects observed in experiments with human observers. (1) The correlation of Gabor responses for head pose estimation represents a pattern matching of filter responses similar to neural responses of cells in early visual cortex [Hubel and Wiesel, 1968]. (2) The length or asymmetry of the projected nose has a direct effect on the perception of the head and gaze direction [Langton et al., 2004]. (3) The perception of the gaze direction is highly dependent on the luminance distribution of the presented face, in particular the perceived gaze is inverted if the polarity of the image is inverted [Sinha, 2000]. Furthermore, the employed filter responses proposed for gaze estimation are also expected in visual cortex [Langton et al., 2000].

Thus, we propose an extended framework in which all visual information described in this contribution are merged to determine the facial configuration concerning head pose and gaze direction. Particularly, we show how simple view or model based approaches can be utilized for determining the head pose and illustrate how gaze can be extracted in a framework for human-machine interaction.

## Acknowledgements

This work has been supported by a grant from the Ministry of Science, Research and the Arts of Baden-Württemberg (Az: 23-7532.24-13-19/1) to Heiko Neumann and Ulrich Weidenbacher.

## References

[Baluja and Pomerleau, 1994] Baluja, S. and Pomerleau, D. (1994). Non-intrusive gaze tracking using artificial neural networks. *Technical Report CMU-CS-94-102, Carnegie Mellon University.*

[Emery, 2000] Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24:581–604.

[Eyelink, 2006] Eyelink (2006). http://www.eyelinkinfo.com.

[Gabor, 1946] Gabor, D. (1946). Theory of communication. *Journal of IEE*, 93: 492–457.

[Gee and Cipolla, 1994] Gee, A. H. and Cipolla, R. (1994). Determining the gaze of faces in images. *Image and Vision Computing*, 12(10):639–647.

[Gibson and Pick, 1963] Gibson, J. J. and Pick, A. D. (1963). Perception of another persons looking behaviour. *American Journal of Psychology*, 76:386–394.

[Heinzmann and Zelinsky, 1998] Heinzmann, J. and Zelinsky, A. (1998). 3-d facial pose and gaze point estimation using a robust real-time tracking paradigma. *Intern. Conf. on Automatic Face and Gesture Recognition.*

[Hubel and Wiesel, 1968] Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Psychology*, 195:215–243.

[Hutchinson et al., 1989] Hutchinson, T., Jr., K. W., Reichert, K., and Frey, L. (1989). Human-computer interaction using eyegaze input. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:1527–1533.

[Ji and Zhu, 2003] Ji, Q. and Zhu, W. (2003). Non-intrusive eye gaze tracking for natural human computer interaction. *MMI-Interactive*, 6.

[Krüger et al., 1997] Krüger, N., Pötzsch, M., and von der Malsburg, C. (1997). Determination of face position and pose with a learned representation based on labelled graphs. *Image Vision Comput.*, 15(8):665–673.

[Langton et al., 2004] Langton, S. R., Honeyman, H., and Tessler, E. (2004). The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics*, 66(5):752–771.

[Langton et al., 2000] Langton, S. R., Watt, R., and Bruce, V. (2000). Do the eyes have it? cues to the direction of social attention. *Trends in Cognitive Science*, 4(2):50–59.

[Matsumoto and Zelinsky, 2000] Matsumoto, Y. and Zelinsky, A. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *4th Intern. Conf. on Face and Gesture Recognition*, pages 499–505.

[Phillips et al., 2000] Phillips, P., Moon, H., Rauss, P., and Rizvi, S. (2000). The feret evaluation methodology for face recognition algorythems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.

[Rae and Ritter, 1998] Rae, R. and Ritter, H. (1998). Recognition of human head orientation based on artificial neural networks. *IEEE Transaction on Neural Networks*, 9(2):257–265.

[Sim et al., 2003] Sim, S., Baker, S., and Bsat, M. (2003). The cmu pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618.

[Sinha, 2000] Sinha, P. (2000). Last but not least. heres looking at you, kid. *Perception*, 29:1005–1008.

[Steifelhagen et al., 1997] Steifelhagen, R., Yang, J., and Waibel, A. (1997). Tracking eyes and monitoring eye gaze. *Proc. of the Workshop on Perceptual User Interfaces*, pages 98–100.

[Trucco and Verri, 1998] Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.

[Wang et al., 2003] Wang, K., Wang, Y., Yin, B., and Kong, D. (2003). Face pose estimation with a knowledge based model. *IEEE Int. Conf. Neural Networks and Signal Processing*, pages 1131–1134.

[Yoo and Chung, 2005] Yoo, D. H. and Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ration under large head motion. *Computer Vision and Image Understanding*, 98:25–51.

[Zhu and Ji, 2005] Zhu, Z. and Ji, Q. (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding*, 98:124–154.