A COMPREHENSIVE HEAD POSE AND GAZE DATABASE

U. Weidenbacher*, G. Layher*, P.-M. Strauss[†] and H. Neumann*

*Institute of Neural Information Processing,
[†]Institute of Information Technology,
University of Ulm, Germany
{Ulrich.Weidenbacher,Georg.Layher,Petra-Maria.Strauss,Heiko.Neumann}@uni-ulm.de

Abstract

Within the past decade, many computational approaches have been developed to estimate gaze directions of persons based on their facial appearance. Most researchers used common face datasets with only a limited representation of different head poses to train and verify their algorithms. Moreover, in most datasets, faces have neither a defined gaze direction, nor do they incorporate different combinations of eye gaze and head pose. Therefore, we recorded an extended dataset of 20 subjects including faces in various combinations of head pose and eye gaze leading to a total amount of 2220 colour images (111 per subject). The images were produced under controlled conditions, i.e. we used a technique to make sure that the subjects adjust their head and eyes appropriately. Furthermore, all images are manually labelled with landmarks indicating important features of the face. Finally, we evaluate the dataset with two computational methods for head pose and gaze estimation respectively.

1 Introduction

1.1 Motivation

Over the last decade, more and more researchers put their focus on the investigation of how to reliably estimate the view directions of persons attending to certain objects, features or locations [4,6,11]. This question seems to be very crucial for all fields of research that are involved in interactions between humans and machines. For instance, in a dialogue scenario it would be very informative to know if a person is attending to a specific target. In the development of new methods for gaze estimation, it is essential to have a database of images that covers a wide range of pose and gaze combinations that can be utilized for testing and training known and newly developed algorithms.

1.2 Scenario

We are currently developing a multimodal dialogue system [10,11] in which we want to deploy gaze direction detection. The dialogue system is designed as an independent dialogue partner in a multi user scenario. It passively observes the conversation between two human users and captures the relevant conversational context. The system should take the initiative and get meaningfully involved in the communication process when it is required by the conversational situation. One of the dialogue partners (the one situated in front of the computer screen) is the system's main interaction partner. When this dialogue partner attends to the system either explicitly by addressing it directly or implicitly by looking at it, the system gets involved in the conversation, it also takes itself back when the task has been solved.

In order to thoroughly investigate new algorithms that can detect the attentional state of a person from camera images we need a suitable database for training and testing purposes.

1.3 Existing Databases and Overview

Currently, there exist several face databases with a large number of subjects that cover at least a sparse set of head poses [7]. Moreover some databases are available, that capture different combinations of pose, illumination [3] or with mimic expression [5,8].

However, there seems to be no database, where combinations of eye gaze and head pose are rendered in a systematic fashion. Therefore, we recorded a large set of face images from 20 different subjects covering horizontal poses as well as vertical pose conditions. Furthermore, for each pose we recorded nine different eye gaze directions. The subjects vary in gender, age, colour of the skin and some of them wear glasses. In contrast to other approaches where multiple cameras are used, our recording technique requires only one camera to be present. Additionally, we manually added landmarks to each image providing information about the exact locations of pupils, eye corners, nose tip, and mouth corners. Finally we show some evaluations of computational methods with input images from the database.

Figure 1: Studio setup showing the arrangement of camera, light sources, canvas and subject. The subject was equipped with a laser pointer on his head. With this equipment, subjects were able to control their head pose by directing the laser pointer dot to specific marks on the wall.

2 Studio Setup and Image Acquisition

We employed an ordinary lecture room, a digital camera (Casio Exilim 850), three spotlights (800 W each), a white canvas (200x170 cm) and a tripod for our recordings (see Fig. 1). Furthermore, we used a laser pointer technique to control specific pose and gaze conditions of the subjects. More specifically, we mounted a laser pointer on the head of each subject so that they could monitor a red dot appearing on the wall to precisely adjust their head pose. Additionally, we utilized a laser water-level on a tripod equipped with an angular dimension to attach marks on the wall, representing both vertical and horizontal angles. Subjects were now able to move their head until the laser beam points to a specific mark on the wall. As soon as this position was reached, we told the subjects to subsequently fixate the marks with their eyes while maintaining their head (and the laser point) in the same position.

Following this scheme, we acquired for each head pose angle nine different gaze conditions (-40° to 40° in steps of 10° relative to the prevailing head pose). Head poses were captured for horizontal head rotations in the range 0° to 90° in steps of 10° (see Figure 2a and Figure 4). Likewise, vertical head poses were recorded for 0°, 30°, and 60° azimuth and for -20° and +20° elevation. Each pose was acquired in combination with different vertical gaze deviations (-20°, 0°, +20°, relative to head pose, see Figure 2b and Figure 3). Table 1 contains a list of all viewing conditions used for acquisition of the database.

Figure 2: Schematic drawing showing the ranges of head pose and eye gaze for horizontal (a) and for vertical rotations (b).

For initial calibration of 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 vertical 0° marks. 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 dot coincided with the horizontal (and vertical) 0° mark.

vertical pose	horizontal pose	horizontal gaze (absolute)	vertical gaze (absolute)
0	0	-40...40	0
0	10	-30...50	0
0	20	-20...60	0
0	30	-10...70	0
0	40	0...80	0
0	50	10...90	0
0	60	20...100	0
0	70	30...110	0
0	80	40...120	0
0	90	50...130	0
-20	0	0	-20/0
0	0	0	-20/20
20	0	0	0/20
-20	30	30	-20/0
0	30	30	-20/20
20	30	30	0/20
-20	60	60	-20/0
0	60	60	-20/20
20	60	60	0/20

Table 1: The table shows which combinations of head pose and gaze are available in the database. Three dots means in steps of 10°. Note that the gaze angles are in absolute coordinates (corresponding to the labels on the wall).

3 Image Representation

In order to avoid the problem of capturing eye blinks we acquired three images for each pose/gaze combination. Then we manually selected the one with best quality from these three pictures. The raw resolution of the colour images is 1600x1200 pixels. The images were originally stored in jpeg format (fine quality) on the camera, but we changed the format afterwards to png. We assume that for

most applications, a downsized version of the images would be sufficient, therefore possible artefacts induced by the jpeg compression should not lead to negative effects during image processing. The filename of each image contains information about horizontal and vertical head orientation, horizontal and vertical gaze direction
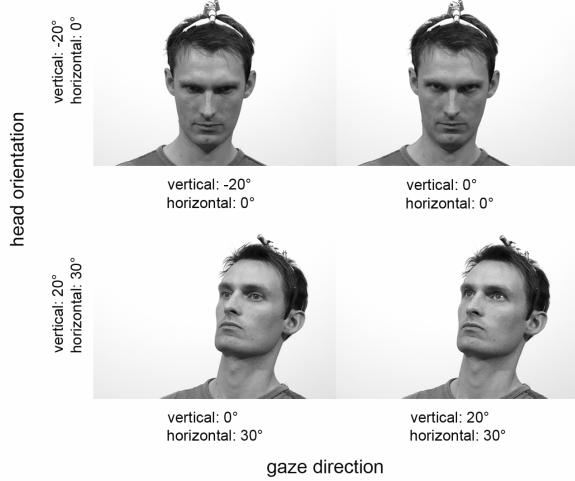

Figure 3: Example images of different gaze and head pose conditions. Note that there are significant changes in the eye patterns even for vertical gaze shifts.

4 Landmark Labelling

For many applications it is useful to have knowledge about the locations of specific features of the face. For instance, some algorithms try to estimate the head orientation based on the geometrical configuration of facial feature locations [2]. If no method is available that automatically extracts facial feature positions, manual
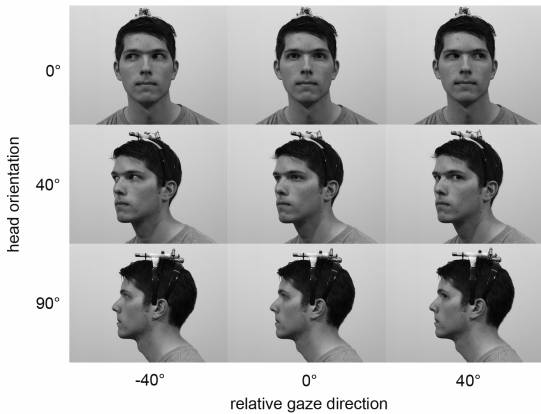
Figure 4: Example images from one subject showing horizontal variations of head pose and gaze direction. Note that a laser pointer is attached to the head to control the head pose.

labelling can be used instead. The labels are attached to important landmarks of the face such as eyes, nose, and mouth which can be used to train and test algorithms for facial feature detection.
Thus, we manually labelled each image with 9 landmarks covering the two corners of each eye, the pupils, the nose tip and the two mouth corners (see Fig. 5). Positions that are invisible due to occlusion are labelled with n/a. The

label positions are represented in ASCII file format and in Matlab file format for each image.
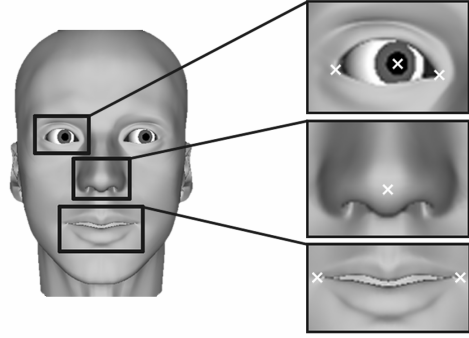
Figure 5: We show an artificial head to illustrate the label positions. Labels were manually attached to 9 positions in the face. For each eye we labelled the eye corners and the centre of the pupil. Furthermore, the locations of the mouth corners and the nose-tip were labelled.

5 Evaluation of Computational Methods

5.1 Head Pose Estimation

We used the database to evaluate a method for head pose estimation based on the pictorial structures model proposed by Felzenswalb [1]. The model builds up a graph representation by connecting important landmarks of faces. To extract the appearance in the area around the landmarks contrast information over different orientations is extracted by using Gabor filters responses. To model the spatial configuration of facial landmarks, distances between the nose and the other face parts are analysed. The connections between face parts are flexible and can be thought as spring-like connections. Spring tensions depend on the variance of feature distances extracted from the dataset.
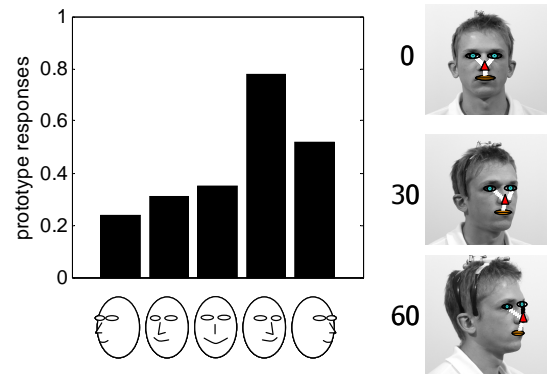

Figure 6: Model graphs, extracted for prototypical head orientations. Face part distances are used to model the spatial configuration of features. Gabor patterns model the appearance of each feature.

We extract three prototypes from different head pose conditions that lead to different face graphs (Fig. 6). To estimate the position and pose of a new face image all prototypes graphs are matched with the input image (for details about the matching procedure see [1]). This returns for each prototype a response which can be interpreted as a confidence measure (Fig. 6). Finally, the pose of the

prototype with maximal response corresponds to the estimated head pose. Note that at this stage, it is also possible to use an interpolation scheme between prototype responses to get more precise pose estimates [12].

5.2 Eye Gaze Estimation

To estimate gaze directions we employed Gabor filter responses similar to the EGM method used for pose estimation. Here, we evaluate the phase of vertically oriented Gabor responses in facial sub-regions around the eyes [12]. The idea is that the iris region is always darker than the remaining regions on the sclera [9]. Thus, gradual eye movements imply a gradual change of the Gabor phase (see Fig. 7). We tested the method on five frontal heads from the data base. The results in Fig. 8 illustrate that there is a linear dependency between gaze direction and phase.
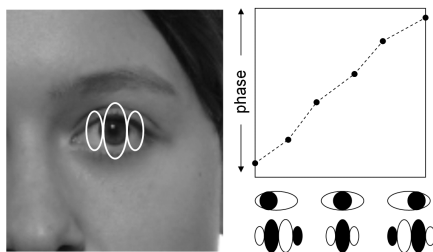
Figure 7: Illustration of Gabor filter size used to extract phase responses around the eye region (left). The graph shows, that gradual shifts of Gaze correspond to a linear shift of the Gabor phase. The correspondence can be used to estimate gaze direction from Gabor phase responses.
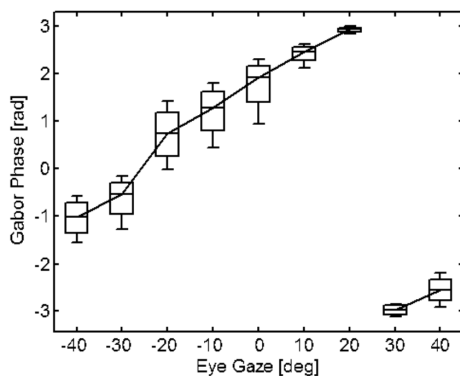

Figure 8: Median Gabor phase (solid line) extracted from five heads for frontal head pose across different eye gaze directions. Boxes represent upper and lower quartile, whiskers plotted at minimal and maximal phase value. Note that the phase is a circular measure which leads to discontinuities when visualized in Cartesian coordinates.

6 Conclusions and Outlook

We have presented a comprehensive database of faces covering large variations of head pose and eye gaze. We also described a simple technique to acquire images of faces under controlled pose and gaze conditions. We further briefly presented an example scenario for which such a database could be very useful. We pointed out that

there exist only few algorithms that deal with the estimation of gaze and the estimation of head pose. To develop new algorithms in this field of research, it is therefore essential to build up such databases. Finally, we employed the database to evaluate two algorithms respectively for head pose estimation and for gaze estimation.

Our plan is to make the database open for public access by building up a webpage with a short description of the database and a download option. The database will be made available open for public access. Download instructions and citation can be found on the website "http://www.informatik.uni-ulm.de/ni/mitarbeiter/UWeidenbacher".

Acknowledgements

References

[1]   P.F. Felzenswalb, Pictorial Structures for Object Recognition. *Int'l. J. Computer Vision,* 61(1), 55-79, 2005.

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

[3]   A.S. Georghiades, P.N. Belhumeur,d D.J. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *Proc. 4th IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2000.

[4]   T. E. Hutchinson , K. P. White Jr, W. N. Martin, K. C. Reichert, L. A. Frey. Human-Computer Int'leraction Using Eye-Gaze Input, *IEEE Trans. on Systems, Man, and Cybernetics,* 19(6), 1527-1534, 1998.

[5]   T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proc. 4th IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 2000.

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

[7]   P.J. Philips, H. Moon, P. Rauss, and S.A. Rizvi. The FERET evaluation methodology for face-recognition algorithms. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 22(10), 1090-1104, 2000.

[8]   T. Sim, S. Baker, M. Bsat. The CMU Pose, Illumination, and Expression Database, *IEEE Trans. PAMI*, 25(12), 1615 – 1618, 2003.

[9]   P. Sinha. Last But Not Least. Heres Looking at You, Kid, *Perception,* 29, 1005-1008, 2000.

[10]   P.-M. Strauss, H. Hoffmann, W. Minker, H. Neumann, G. Palm, S. Scherer, F. Schwenker, H. Traue, W. Walter, U. Weidenbacher. Wizard-of-Oz Data Collection for Perception and Interaction in Multi-User Environments, *Int'l. Conf. on Language Resources and Evaluation (LREC)*, 2006.

[11]   P.-M. Strauss. A SLDS for Perception and Interaction in Multi-User Environment, 2nd *Int'l Conf. on Intelligent Environments*, 171-174, 2006.

[12]   U. Weidenbacher, G. Layher, P. Bayerl, H. Neumann. Detection of Head and Gaze Direction for Human-Computer Interaction*, Proc. Perception and Interactive Technologies*, LNAI 4021, pp. 9-19, 2006.