Contact-analog Information Representation in an Automotive Head-Up Display

T. Poitschke, M. Ablassmeier, and G. Rigoll*
Institute for Human-Machine Communication
Technische Universität München

S. Bardins, S. Kohlbecher, and E. Schneider†
Center for Sensorimotor Research
Ludwig-Maximilians-University Munich

Abstract

This contribution presents an approach for representing contact-analog information in an automotive Head-Up Display (HUD). Therefore, we will firstly introduce our approach for the calibration of the optical system consisting of the virtual image plane of the HUD and the drivers eyes. Afterward, we will present the used eyetracking system for adaptation of the HUD content at the current viewpoint/position of the driver. We will also present first prototypical concepts for the visualization of contact-analog HUD content and inital test results from a brief usability study.

Keywords: automotive, contact-analog, calibration, eye tracking, Head-Up Display, HUD

1 Introduction

In recent years a novel approach for information visualization – the so-called Head-Up Display (HUD) – has become more and more popular in the context of automotive displaying concepts. This output device was pioneered for fighter jets in the early 70s and later for low-flying military helicopter pilots, for whom information overload was a significant issue, and for whom changing their view to look at the aircrafts instruments could be a fatal distraction. Thus, the HUD makes it possible to project information directly into the user's visual field. Nowadays, HUDs are becoming increasingly available in production cars. Here, the information appears to hover above the engine hood, a few meters away from the drivers eyes. The first example of the HUD was used in the field of automotive transportation in 1988: in the General Motors' model Cutlass Supreme used a very simple HUD and showed only the speedometer in a single available color. Current HUDs usually offer important data, like speedometer, various status and navigation information. The advantages and disadvantages of automotive HUDs are a quite big research topic and are extensively discussed in several publications (e.g., [Gish and Staplin 1995], [Kiefer 1998], [Kiefer 2000]).
An upgrade of this device can be the contextualization of the displayed information: the HUD now becomes "contact-analog". Here, the information displayed in the windscreen merges with the real environment. Therefore, a contact-analogue HUD consists of a conventional HUD with an enlarged field of view that is combined with an eyetracker to adapt the shown content at the driver's current position. This device offers a broad spectrum for novel advanced driver information and assistance systems.

*e-mail: {poitschke,ablassmeier,rigoll}@tum.de
†e-mail: {sbardins,skohlbecher,eschneider}@nefo.med.uni-muenchen.de

ETRA 2008, Savannah, Georgia, March 26–28, 2008.

2 Background

2.1 Contact-analog information representation

So-called contact-analog HUD content is spatially related to the outside world. The image refers to some object in the background, and the driver perceives a sort of fusion between real world and virtual object.
As a result of the use of contact-analog contents in the vehicle, e.g., the following possibilities arise for a better driver support: purposeful direction of the drivers attention on accident-prone situations, marking of weaker road users, marking of road signs, night vision, fading in trace-exact navigation references, representation of driver assistance systems.
In the following, some kind of possible use of contact-analog HUDs will be shown in order to demonstrate the potentiality of this technology. The pictures (see Fig. 1) are photos elaborated with particular tools to get a realistic forecast of how such an application would look like. They are only a few examples of all the possibilities.

(a) Guidance advices: taxiway concept. (b) Concept of traffic sign visualization: stop.

Figure 1: *Examples for contact-analog displaying concepts.*

Since contact-analog HUDs integrate the displayed contents directly into the driving scene, the displaying system has to be adapted at current conditions. The position of the car on the road is to be determined accurately. Then, existing navigation data (e.g. location of traffic signs, road process etc.) can be used. Therefore, differential GPS systems, gyro-sensors, and map matching algorithms can be used. Exact line of sight data are needed to adjust the HUD at the drivers current point of view. Also, it is necessary to calibrate the displaying system at the user's optics, to merge virtual information with the environment.
However, this contribution focuses on the calibration and adaptation inside the vehicle, i.e., initial calibration of the display on the optical system of the driver, and adjustment of this calibration data to the current driver's perspective/point of view.

2.2 Mathematical background

For the calibration of the display, we use the model of the central projection for the visual system consisting of the user's eyes and the virtual displaying device. The central projection describes the transformation of a 3D point in the real world on a 2D image plane. Here, a projection ray is drawn over a fixed point in space, the projection center C, and cuts a certain plane in space. This plane does not contain C, and represents the image plane. The intersection of the ray with this plane is the picture of the 3D-point. The simplest camera model for a central projection is the so-called pinhole camera [Hartley and Zisserman 2004]. This assumes an infinite small pinhole which represents only one point. This projection center is the origin of an Euclidean coordinate system, and $Z = f$ is the image plane. One point in space with the coordinates $\vec{X} = (X,Y,Z)^T$ is transformed on that point on the image plane, where the connecting straight line between $\vec{X}$ and the projection center cuts it. If the third coordinate is neglected, one sees that

$$(X,Y,Z)^T \mapsto (fX/Z, fY/Z)^T \qquad (1)$$

is a central projection from $\mathbb{R}^3$ to $\mathbb{R}^2$. Here, the projection center is also called camera center or optical center. The perpendicular of the camera center on the image plane is the principal axis of the camera, its intersection with the image plane is the picture main point. If homogeneous coordinates are used instead of Cartesian, (1) can be simplified to a linear transformation in form of a matrix multiplication:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \qquad (2)$$

If $\vec{X}$ is defined as a point in real world represented by the 4-vector $(X,Y,Z,1)^T$, $\vec{x}$ as its image point represented by a homogeneous 3-vector, and P as the 3×4 camera or projection matrix, (2) can be written as $\vec{x} = P\vec{X}$ where P is the matrix from (2). This represents the basic camera matrix of the central projection via a pinhole camera. If

$$K = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (3)$$

is set, (2) can be arranged more clearly as

$$\vec{x} = K\,[I \mid \vec{0}]\,\vec{X}_{Kam}, \qquad (4)$$

where I is a 3×3 unit matrix, and $\vec{0} = (0,0,0)^T$. K is called calibration matrix. Here, $(X,Y,Z,1)^T$ is written as $\vec{X}_{Kam}$, in order to clarify that the camera is in the origin of a Euclidean coordinate system, and its principal axis is arranged exactly along the Z-axis. This coordinate system is called camera coordinate system. Usually, points in space are specified in the world coordinate system. These two systems are related to each other via a translation and a rotation.

$\tilde{\vec{X}}$ is an inhomogenous 3-vector that refers to a point in the world coordinate system and $\tilde{\vec{X}}_{Kam}$ the representative of the same point represented in camera coordinates. $\tilde{\vec{C}}$ is the position of the camera center in world coordinates. Then $\tilde{\vec{X}}_{Kam} = R(\tilde{\vec{X}} - \tilde{\vec{C}})$, where R is a 3×3 rotation matrix, which represents the orientation of the camera coordinate system. Using homogeneous coordinates this equation can be written as

$$\vec{X}_{Kam} = \begin{bmatrix} R & -R\tilde{\vec{C}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{\vec{C}} \\ 0 & 1 \end{bmatrix} \vec{X}. \qquad (5)$$

Applying formula (4) this is

$$\vec{x} = KR\,[I \mid -\tilde{\vec{C}}]\,\vec{X}, \qquad (6)$$

where $\vec{X}$ is now in the world coordinate system. This is the general transformation, which is given by the pinhole camera. A general pinhole camera has 9 degrees of freedom (DOF), i.e. 3 for the calibration matrix K (f, p_x, p_y), 3 for R and 3 for $\tilde{\vec{C}}$. The parameters of K are called intrinsic camera parameters. R and $\tilde{\vec{C}}$, which represent the camera rotation and position relative to the world coordinate system, are called extrinsic camera parameters. If all parameters are summarized, one receives the model of a finite projective camera:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = KR\,[I \mid -\tilde{\vec{C}}], \qquad (7)$$

where the calibration matrix K is in the form out (3). This extended model of a pinhole camera has 11 degrees of freedom: 6 extrinsic (position and orientation, contained in $\tilde{\vec{C}}$ and R) and 5 intrinsic (focal lengths in x and y-direction, skew and coordinates of the picture main point, contained in K). The projection matrix P has 12 entries. However, it is clearly defined up to an arbitrary scaling factor, that causes the 11 DOF.

3 Calibration and viewpoint adaptation

3.1 Test setup

The laboratory prototype consists of a simple model of a car with a driver's seat, a rudimentary steering wheel and a mounted Augmented Reality system (ARS) (see Fig. 2). The model's position and orientation (COS_1) are tracked by an infrared-tracking system (IRTS) based on the IRTS's world coordinate system (COS_0). A

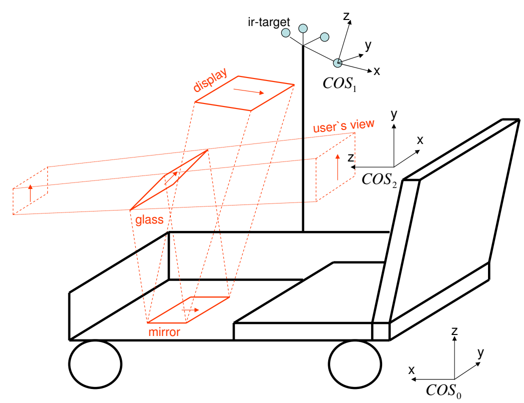

Figure 2: *Schematic overview.*

monitor mounted on top of the car's model displays virtual content, which is mirrored into the windscreen by a reflector at the bottom of the model. To create a contact-analog perception of the displayed content, the user‘s direction of sight is tracked by an eyetracking system (COS_2). Based on this direction of view, virtual content is displayed by the ARS using the monitor. Thus, moving the user‘s view results in moving virtual content.

3.2 Calibration

The visualization techniques, on which this contribution is based, originate from the research area of Augmented Reality (AR). An AR environment covers both, real objects in the user's environment and virtual objects, which exist only in a computer-internal model of the world.
Therefore, we adapted and extended an approach for the calibration of Optical-See-Through Head-Mounted Displays (HMD) (e.g.,

[Tuceryan et al. 2002]). These displays are used for the augmentation of the user's view in various scenarios, e.g. medical surgeons, mechanics, etc. We use an adapted AR-software, that implements basic functionalities (e.g., rendering and positioning of virtual content relative to defined coordinate systems) for displaying virtual content on such HMDs. Here, the projection matrix P is needed to augment the real world by emphasized or overlayed additional information.

3.2.1 Determination of the projection matrix P

The projection matrix P is computed from a number of given correspondences between the 3D-world and the 2D-image. 3D-points and their pertinent 2D-image points, resulting from the central projection, represent the simplest kind of such correspondences. Since normally over-determined linear sets of equations arise with the computation method, P cannot be determined accurately, but only an approximation. In the following, our approach to calculate the needed camera matrix P is presented.

Basic equations: Here, the assumption is made that a certain number n of point correspondences $\vec{X}_i \leftrightarrow \vec{x}_i$ between 3D-points in reality and 2D-image points is given. The 3×4 camera matrix P is wanted so that $\vec{x}_i = P\vec{X}_i$ applies for $1 \leq i \leq n$. It is to be considered that the equation contains homogeneous vectors and thus $\vec{x}_i$ and $P\vec{X}_i$ are not identical, but only have the same direction. Since two vectors, which have the same direction, result in the zero-vector in the cross product, the equation can be written as follows: $\vec{x}_i \times P\vec{X}_i = \vec{0}$ This representation facilitates the deduction of a simple linear solution for P. If the j-th row of the matrix P is denominated as $\vec{P}^{jT}$, $P\vec{X}_i$ results in

$$P\vec{X}_i = \begin{pmatrix} \vec{P}^{1T}\vec{X}_i \\ \vec{P}^{2T}\vec{X}_i \\ \vec{P}^{3T}\vec{X}_i \end{pmatrix}. \tag{8}$$

With $\vec{x}_i = (x_i, y_i, w_i)^T$, the cross product can be written as

$$\vec{x}_i \times P\vec{X}_i = \begin{pmatrix} x_i \\ y_i \\ w_i \end{pmatrix} \times \begin{pmatrix} \vec{P}^{1T}\vec{X}_i \\ \vec{P}^{2T}\vec{X}_i \\ \vec{P}^{3T}\vec{X}_i \end{pmatrix} = \begin{pmatrix} y_i\vec{P}^{3T}\vec{X}_i - w_i\vec{P}^{2T}\vec{X}_i \\ w_i\vec{P}^{1T}\vec{X}_i - x_i\vec{P}^{3T}\vec{X}_i \\ x_i\vec{P}^{2T}\vec{X}_i - y_i\vec{P}^{1T}\vec{X}_i \end{pmatrix} \tag{9}$$

As it holds that $\vec{P}^{jT}\vec{X}_i = \vec{X}_i^T\vec{P}^j$, a set of three equations with the entries of P arises for each pair of points $\vec{X}_i \leftrightarrow \vec{x}_i$. This is represented as follows:

$$\begin{bmatrix} \vec{0}^T & -w_i\vec{X}_i^T & y_i\vec{X}_i^T \\ w_i\vec{X}_i^T & \vec{0}^T & -x_i\vec{X}_i^T \\ -y_i\vec{X}_i^T & x_i\vec{X}_i^T & \vec{0}^T \end{bmatrix} \begin{pmatrix} \vec{P}^1 \\ \vec{P}^2 \\ \vec{P}^3 \end{pmatrix} = \vec{0} \tag{10}$$

Since these three equations are linearly dependent, the set of equations can be reduced to two linear independent rows:

$$\begin{bmatrix} \vec{0}^T & -w_i\vec{X}_i^T & y_i\vec{X}_i^T \\ w_i\vec{X}_i^T & \vec{0}^T & -x_i\vec{X}_i^T \end{bmatrix} \begin{pmatrix} \vec{P}^1 \\ \vec{P}^2 \\ \vec{P}^3 \end{pmatrix} = \vec{0} \tag{11}$$

These equations can be abbreviated as $A_i\vec{p} = \vec{0}$, whereby A_i is a 2×12 matrix and $\vec{p}$ is the 12-vector with the entries from P. From the given n pairs of points one receives a $2n \times 12$ matrix A through "stacking" the respective equations in (11). The projection matrix P is determined by solving the system of equations $A\vec{p} = \vec{0}$, whereby the trivial solution $\vec{p} = \vec{0}$ is not of interest.
To obtain the intrinsic and extrinsic parameters from the camera matrix P, one must convert it accordingly to the form out (7):

$$P = [M \mid -M\tilde{\vec{C}}] = K[R \mid -R\tilde{\vec{C}}] \tag{12}$$

If K and R are needed, one must divide the matrix M by means of the RQ-decomposition. The RQ-decomposition is based on the multiplication of the right of the matrix with a set of suitable rotation matrices, so that the entries underneath the main diagonals become zero. The used matrices are called Givens-rotations, and each of them represents the rotation around one of the three axes of coordinates.
The AR visualization software uses the determined intrinsic and extrinsic parameters to adapt the augmented representations to the user's virtual camera system.

3.2.2 Execution of the calibration

The main task during the calibration is the production of the pairs of 3D points (computed by the IRTS) and their appropriate 2D points, and providing these to the computer. As the view of the world, and consequently of the virtual objects in the AR environment, depends on the focus, the position and orientation of the vehicle are also tracked by the IRTS. To collect the point pairs, a calibration body is placed in front of the vehicle's HUD, such that the user sees it through the virtual image plane. Since the used camera model possesses only one projection center, the user has to look with one eye. Further, the head is assumed as immovable during the entire calibration (relative to the virtual image plane), since a head movement would change the intrinsic camera parameters and no correct solution for P could be determined.
The user points at each individual marker of the body with the cursor and presses a mouse button. The screen coordinates of the mouse cursor when clicking are noted and assigned to the appropriate world coordinates. Since the used body consists of 6 markers, but the calibration algorithm partly needs more then 6 pairs, it can be moved during the calibration procedure. Thus, other world points are transformed on the image plane.

3.3 Eyetracking for dynamic viewpoint adaptation

To adapt the representation in the HUD to different focuses of the driver, its position relative to the display has to be analyzed. Therefore, the position of the driver's eyes during the calibration procedure is determined and used as reference value for the viewpoint adaptation. Since the extrinsic values of the matrix P represent the position and orientation of the virtual camera center, these have to be adapted to the new position of the driver's optics by adding an offset computed from the new eye positions.
Our eyetracking system uses cameras with Region Of Interest readout (ROI) for the driver monitoring and eye tracking. Thus, a frame rate of 200 fps is achieved. We use an objective with 40 mm of focal length and mounted one camera at a position comparable to a position at the front column of a car. A second camera is mounted on a position comparable to the inner rear view mirror. A position nearly under the vehicle roof guarantees the lowest possible distance to the driver's eyes and thus the highest possible local accuracy with fixed sensor and optics. For a robust system setup, we use the difference image technique and different LED light sources. The pupil is illuminated by 2 different, alternating LED sources (one close to the optical axis causes the red-eye effect, the other far from the optical leaves dark pupils). Thus, the difference image only shows two bright spots at the location of the pupils.For the suppression of artifacts and noise, a filter chain consisting of a temporal low-pass, a spatial Gauss filter and various morphologic filters is applied.
After the system start and after losing the tracked pupils in a frame, the pupil search is initiated. The rough position of the pupil is determined using a dynamic threshold value and a following Canny edge detector determines its boarder. Finally, an ellipse fitting algorithm computes the coordinates of that ellipse, whose pixels have the lowest possible square error to the given edge pixels. These op-

erations are applied on the interpolated picture, whereby sub pixel accuracy can be achieved.
To compute the positions of the eyes in the 3D car coordinate system, we use a calibrated stereo camera system. For the calibration of the camera system, the calibration toolbox for MATLAB is used. An iterative algorithm based on [Zhang 2000] is then used to calculate the intrinsic and extrinsic parameters of both cameras as well as the extrinsic parameters of the stereo rig.
In each rendering phase of the AR visualization software, the computed position of the driver's eyes is added to the initial determined extrinsic calibration values (i.e., rotation and translation). Thus, the calibration is adapted on different positions of the driver and thus the center of the virtual camera.

4 Displaying concepts and evaluation results

To analyze the performance of the implemented calibration and displaying software, different concepts were implemented and evaluated in a brief study. During a test, the subjects are located in the mock-up described above. They have to accomplish a primary driving task. The driver's view through the combiner (i.e., the driving simulation and the HUD content) can be seen in Fig. 1. A special round course was modeled. This contains most diverse intersections and turning scenarios, pedestrian crossings as well as a multiplicity at traffic signs. The simulation platform provides all relevant environment data (e.g. position of traffic signs and the road) for the positioning of the HUD content.

Evaluated displaying concepts: The main idea of these concepts is to present relevant information about the primary driving task directly on the road. Fig. 3 illustrates a concept to provide guidance advices from the navigation system in a HUD. As shown in the image, the orange direction arrow appears to be actually integrated in the street. This has the following two advantages: the information is synchronized with reality, and it should not distract the driver; the information of the navigation system is placed at a known location.

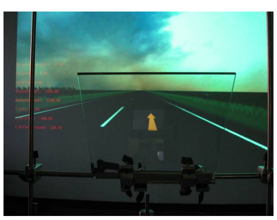

Figure 3: *Drivers view on contact-analog arrow in the HUD.*

Test procedure: The concepts were tested in separate sessions, and all participants conducted the test in the same order. In the first session, the test persons were to acclimatize them self with the used driving simulator. Afterwards, they drove the round course with the simulation of an ordinary static HUD where a symbolic arrow representation was used as reference system (HUD-REF). During the next test runs they used the contact-analog HUD concepts. For the evaluation of the concepts, we used a semantic differential, and an adapted version of the NASA TLX.

Test persons corpus: Six women and 11 men participated. The age of the participants varied between 20 and 42 years (average at approx. 31 years). 58% of the subjects knew the HUD reference system, and 35% of the subjects already had it in use. Already 50% of the subjects knew about the contact-analog extension of the HUD system, although this is not offered yet on the market.

4.1 Results

The main focus of the test series was on the evaluation of the displaying quality, which was evaluated as good to very well by nearly all of the subjects. Only for 2 persons the representation was not satisfying, since the calibration values could not be determined exactly. Thus, the HUD content could not be adapted accurately to the eye position, and the representations lay outside of the desired lane. Furthermore, the procedure of the calibration was regarded of 40% of the subjects as too pedantic and time-consuming. However, this evaluation related itself, after it was explained that this procedure would be necessary only with the first use of the HUD.
A majority of 83% of the test persons considered the contact-analog arrow representation as good or very well. According to their statements, this justifies itself, as they mentioned they already knew this layout from lane markings during several trips in real environments. Only one test person, which used the HUD-REF frequently, rated this as more intuitive compared with the contact-analog representations. The contact-analog representations were not rated as obstructing the view. 58% of the subjects did not see a danger to be diverted from traffic by the arrows. This section presents first preliminary results since the evaluation of all recorded data is still in progress.

5 Summary and outlook

In this contribution, we presented advanced visualization techniques for the automotive domain, using Augmented Reality and a contact-analog Head-Up Display. The main focus was on the calibration of the displaying device at the driver's optical system. Therefore, we presented our approach for the calibration of an automotive HUD. Using an eyetracking system, we could dynamically adapt the initial calibration of the displaying device.
Our short evaluation showed a high acceptance by the users and ideas for improvement. Future work will aim at the integration of more contact-analog content. Currently, we are also working on the implementation of a pivotable stereoscopic eyetracking system, which provides a greater and adaptable "field of view". Thus, we will be able to allow broader ranges for natural occurring head and eye movements of the driver.

References

GISH, K., AND STAPLIN, L. 1995. Human factors aspects of using head up displays in automobiles: A review of the literature. Tech. Rep. Report No. DOT HS 808 320, Washington: U.S. Department of Transportation.

HARTLEY, R., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press.

KIEFER, R. 1998. *Defining the "HUD Benefit Time Window"*. Amsterdam: Elsevier, ch. Vision in Vehicles - VI, 133–142.

KIEFER, R. 2000. *Older Drivers' Pedestrian Detection Times Surrounding Head-Up Versus Head-Down Speedometer Glances.* Amsterdam: Elsevier, ch. Vision in Vehicles - VII, 111–118.

TUCERYAN, M., GENC, Y., AND NAVAB, N. 2002. Single point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators and Virtual Environments 11* (June), 259–276.

ZHANG, Z. 2000. A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):13301334.