Multi-Depth Cross-Calibration of Remote Eye Gaze Trackers and Stereoscopic Scene Systems

T. Kowsari[1] S.S. Beauchemin[1] M.A. Bauer[1] D. Laurendeau[2] N. Teasdale[3]

Abstract— We present a robust and accurate technique for the cross-calibration of 3D remote gaze trackers with stereoscopic scene vision systems between which no common imaging area exists. We empirically demonstrate that a multi-depth calibration approach yields remarkably superior results for obtaining 3D Point-of-Gaze (PoG) when compared with traditional methods using monocular scene cameras and co-planar eye gaze calibration points.

I. INTRODUCTION

Remote gaze trackers have been in use for various applications together with scene cameras to determine the point of gaze (PoG) of human subjects on an imaged scene. Several types of applications benefit from the use of such systems including vehicle driver training and advanced driver assistance systems, the context in which the results herein have been obtained.

The task of projecting back the 3D gaze direction onto the imaged scene requires a cross-calibration between the remote gaze tracking device and the scene. In most if not all of commercially available systems, this type of calibration is performed by requiring that test subjects fixate specific, pre-selected image points on a planar surface placed at a known distance such as on a computer screen or, by using a scene image from a monocular camera and treating it essentially as a 2D object (co-planar fixation calibration points).

Such approaches are dependable when the subject's eye center is not highly offset from the scene camera(s). In other words, because the origin of the reference system of the scene cameras and the subject's eye center approximately coincide, the projection ray of any fixated object will also approximately lie on the line of sight regardless of the depth of the object. In such cases, the calibration process may be performed correctly. Otherwise, objects with different depths along the line of sight correspond to different image locations, and must be calibrated for as such.

Our primary goal is to determine whether driver intent and driving-related actions can be predicted from qualitative and quantitative analyses of driver behavior. Toward this end, it is necessary to establish the correspondence between cephalo-ocular behavior and visual stimuli in such a way as to identify the elements in the visual field to which driver attention turns to. This type of information in turn may facilitate the task of a driving assistance system to assess whether drivers are attending to the appropriate stimuli, given traffic context [1].

A. Literature Survey

It has been known for some time now that the direction of driver gaze possesses predictive qualities in regards to the detection of driver intent and following driver maneuvers [2]. In particular, the relation between gaze direction and subsequent vehicle steering has been clearly demonstrated in several studies [3], [4], in which it was found that driver gaze generally points towards the tangent point of a road bend, and does so 0.8 to 1s before effecting steering maneuvers, providing evidence that the eyes begin processing information before the associated action has begun [4].

Hennessey and Lawrence presented a 3D PoG method which employs eye vergence to estimate the 3D position of a fixated object [5]. In their experiments, fixated objects were contained in a 0.01725 m^3 volume located in front of the subject. The reported average positional error was 3.93 cm. It constituted the first binocular system for estimating the absolute 3D coordinates of where one is looking in the 3D world.

Alternatively, Yamashiro *et al.* devised an automatic calibration to estimate the gaze of vehicle drivers by using known reference points such as the rear-view and the side mirrors of the vehicle [6]. The gaze of drivers was recorded and an Expectation-Maximization algorithm was used to cluster glances to the reference points. An automatic calibration of gaze could be achieved from the collected gaze data over time as the vehicles were driven.

In these approaches, it is assumed that the subject's eye center coincides with the origin of the reference frame of the scene cameras. When this constraint is satisfied, the depth of fixated objects does not influence the position of the gaze onto the scene images, thus avoiding parallax errors in the identification of the PoG.

The problem posed by non-coinciding frames of reference between the eye tracker and the scene camera(s) has been recently addressed by Kim *et al.* by tracking both eyes and using vergence to assist in the localization of the 3D point being fixated [7]. Their approach relies on a binocular tracking in order to estimate eye vergence and to locate the 3D PoG. However, according to Land, vergence tend to be set anywhere from 25% to 45% beyond the visually attended object. Additionally, the role of vergence in real tasks is still not well understood [4]. Bernet *et al.* also investigated this problem with monocular scene camera systems, but readily admit that a stereoscopic scene system constitutes

[1]Department of Computer Science, Western University, London, ON, N6A 5B7, Canada

[2]Département de Génie Electrique, Université Laval, Québec, QC, G1V 0A6, Canada

[3]Département de Médecine Sociale et Préventive, Université Laval, Québec, QC, G1K 7P4, Canada

Fig. 1. *Physical configuration* **a) (left):** *Remote eye-tracking system, and* **b) (right):** *RoadLAB stereoscopic vision system*

Fig. 2. *The topology of the tracker and scene reference frames, where* $\mathbf{x}_i$ *and* $\mathbf{y}_i$ *are coordinates of the fixated point in the scene,* $\mathbf{o}$ *is the reference frame of the stereo scene system,* $\mathbf{o}'$ *that of the tracker, and* $\mathbf{c}_i$ *and* $\mathbf{g}_i$ *are the eye center position and gaze vector respectively.*

the best approach to precisely locating the PoG in absolute 3D coordinates [8]. Hence we propose a method to cross-calibrate a stereo scene camera system with a remote eye gaze tracker using variable depth calibration points and compare the resulting error with a typical, monocular scene, co-planar calibration process.

B. System Configuration

Our systems consist of a remote gaze tracker with two cameras pointed toward the driver's face and a stereo system oriented toward the front of an experimental vehicle. Our aim is to determine which objects within the visual scene in front of the vehicle elicit visual responses from drivers.

Our remote gaze tracker computes several variables including gaze Euler angles, eye center location, and head position and orientation with respect to the coordinate system of the tracker, located in the middle of the stereo cameras pointed toward the driver's face. Our scene stereo system attaches to the roof of the vehicle, with its reference frame centered on the left camera. Both systems require a calibration prior to use. Figure 1 shows the configuration of the experimental vehicle.

II. Cross-Calibration Technique

The orientation of the gaze with respect to the coordinate system of the tracker is given by Euler angles describing the rotations around the X axis and the Y axis. Performing these rotations amounts to aligning the the Z axis of the tracker with the 3D direction of the gaze. The eye gaze direction defined in this way is a unit vector originating from the eye center. Figure 2 shows the relation between the gaze vector and a fixated point in the field of view. Given Euler angles θ_i and ϕ_i, the gaze unit vector is obtained as:

$$\mathbf{g}_i = R_y(\theta_i)R_x(\phi_i) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \sin(\phi_i)\cos(\theta_i) \\ -\sin(\theta_i) \\ \cos(\theta_i)\cos(\phi_i) \end{bmatrix} \quad (1)$$

A. Description of Algorithm

The objective consists of computing estimates of the rotation matrix and the translation vector between the reference frame of the scene stereo system and that of the remote eye tracker. The calibration process consists of asking the driver to fixate pre-selected points for which depth estimates are available and record the gaze vector and eye center location
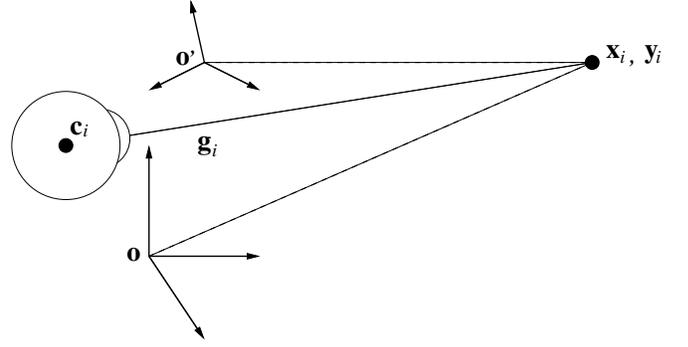
of the driver, along with the 3D position of the fixated points in the scene for a brief period (2 s) per fixated point. This data is then used to estimate the rotation matrix and the translation vector relating the reference frames.

The eye center and gaze vector, both expressed within the reference frame of the tracker, represent a 3D line passing through the fixated point which in turn is expressed in the reference frame of the stereo scene system. Let us assume that the fixated points are known in both reference frames, and find the rigid body transformation parameters that bring the points from one reference frame to the other. The relation between the fixated points and the reference frames is given by:

$$\mathbf{y}_i = R\mathbf{x}_i + \mathbf{T} \quad (2)$$

where $\mathbf{x}_i$ is the position of the ith fixated point measured in the scene reference frame, $\mathbf{y}_i$ is the position of $\mathbf{x}_i$ in the reference frame of the tracker, and R and $\mathbf{T}$ are the rotation matrix and translation vector between the reference frames.

We estimate the rigid transformation parameters following the approaches devised by Arun [9] and Challis [10], and use a confidence measure on the fixated points $\mathbf{x}_i$ based on the inverse of disparity.

The centers of mass of the fixated points in both reference frames are given by:

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^{n} w_i^2 \mathbf{x}_i}{\sum_{i=1}^{n} w_i^2} \quad \text{and} \quad \bar{\mathbf{y}} = \frac{\sum_{i=1}^{n} w_i^2 \mathbf{y}_i}{\sum_{i=1}^{n} w_i^2} \quad (3)$$

where w_i is a weight factor reflecting the reliability of the ith point, and n is the number of points ($n > 2$ [10]).

With the following substitution in variables:

$$\mathbf{x}_i' = \mathbf{x}_i - \bar{\mathbf{x}} \quad \text{and} \quad \mathbf{y}_i' = \mathbf{y}_i - \bar{\mathbf{y}}$$

a matrix can be formed as:

$$C = \frac{\sum_{i=1}^{n} w_i^2 \mathbf{y}_i' \mathbf{x}_i'^{T}}{\sum_{i=1}^{n} w_i^2} \quad (4)$$

and decomposed with SVD as

$$C = UDV^T \quad (5)$$

According to [10], setting

$$R = UV^T \qquad (6)$$

minimizes the error in the least-squares sense. Since both the reflection and the rotation matrices minimize the least-squares error, then R is either the reflection or the rotation matrix. If R is the reflection matrix, then $\det(R) = -1$ and the rotation matrix is obtained in the following way:

$$R = U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(R) \end{bmatrix} V^T \qquad (7)$$

and the translation vector is obtained as:

$$\mathbf{T} = \bar{\mathbf{y}} - R\bar{\mathbf{x}} \qquad (8)$$

The 3D fixated points are not directly known, as the eye center and the gaze direction only yield a 3D line onto which a fixated point lies. Additionally, these lines are expressed within the reference frame of the eye tracker. In order to overcome this difficulty, we assume that the transformation is already known and write:

$$d_i = \|R\mathbf{x}_i + \mathbf{T}\| \qquad (9)$$

where d_i is the distance of the ith fixated point from the origin of the reference frame of the eye tracker, and $\mathbf{x}_i$ is the 3D coordinate of the point in the scene reference frame. Hence, the fixated points can be approximated in the reference frame of the tracker as:

$$\mathbf{y}_i = d_i\mathbf{g}_i + \mathbf{c}_i \qquad (10)$$

where $\mathbf{y}_i$ is the 3D position of the fixated point in the reference frame of the tracker, with $\mathbf{g}_i$ and $\mathbf{c}_i$ defined as before. This set of fixated points allows us to approximate the rotation matrix R and the translation vector $\mathbf{T}$ iteratively. It is initially assumed that both reference frames coincide exactly, starting the iterative process with R the identity matrix and $\mathbf{T}$ a null vector. Since the reliability of the fixated points (partly) depends on their distance to the stereo system (a characterization of this error is provided in [1]), we use the stereo disparity of the fixated points, defined as the inverse of distance, to provide the weighting values w_i in (3) and (4). Algorithm 1 shows the detailed procedure for the cross-calibration.

B. Calibration Data Collection

We proceed to describe the data-gathering procedure that is used with the drivers of the experimental vehicle. For each selected calibration point the driver is asked to fixate, the gaze vector and the position of the eye center in the reference frame of the eye tracker are recorded, along with the 3D position of the calibration point in the reference frame of the stereo imaging system. We refer to these captured data elements as gaze data sets. While a minimum of three non co-planar calibration points are needed, we generally use 15 to 20 points to ensure sufficient precision in the computation of the calibration parameters. The calibration procedure is defined as follows:

Algorithm 1 *Cross Calibration Algorithm*

$R \leftarrow I$
$\mathbf{T} \leftarrow$ initial estimate
repeat
    for $i = 0 \rightarrow n$ **do**
        $d_i \leftarrow \|R\mathbf{x}_i + \mathbf{T}\|$
        $\mathbf{y}_i \leftarrow d_i\mathbf{g}_i + \mathbf{c}_i$
    end for
    $C \leftarrow \frac{\sum_{i=1}^{n} w_i^2 \mathbf{y}_i' \mathbf{x}_i'^T}{\sum_{i=1}^{n} w_i^2}$
    $(U, D, V^T) \leftarrow \text{SVD}(C)$
    $R \leftarrow UV^T$
    $R \leftarrow U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(R) \end{bmatrix} V^T$
    $\mathbf{T}' \leftarrow \mathbf{T}$
    $\mathbf{T} \leftarrow \bar{\mathbf{y}} - R\bar{\mathbf{x}}$
until $\|\mathbf{T}' - \mathbf{T}\| < \epsilon$

---

1) Salient points provided by the stereo imaging system are detected and the calibration operator selects a suitable subset of these points (a suitable subset contains 3D points that are visible to the driver and that are found at various depths in the scene). We use the `GoodFeaturesToTrack` function from the OpenCV library to provide the initial set of salient points (Figure 3a).
2) The software displays the calibration point the driver is asked to fixate and records the current gaze data set for a period of 2s (Figure 3b). A RANSAC algorithm is used in cases when the driver experiences a saccade while requested to fixate the calibration point. This ensures the rejection of the saccade gaze data from the sample.
3) When all the points have been fixated by the driver and the gaze data recorded for each point, the operator initiates the calibration stage. Once the systems are cross-calibrated the gaze of the driver is in relation with the depth map from the stereo imaging system in real-time (see Figure 3a), b) and c)).

The eye tracker provides a real-time confidence measure related to the quality of the computed gaze for each eye of the driver. During the calibration process, we compute a set of cross-calibration parameters R and $\mathbf{T}$ for each eye. Once the systems are cross-calibrated and in use, we determine in real-time which set of parameters to use based on the confidence measures provided by the eye tracker. It is possible to force the system to use a specified eye for both the calibration and the gaze projection stages in case of abnormality of one of the driver's eyes.

C. Projection of the Gaze on the Scene Image

Once the cross-calibration process has completed, the Line of Gaze (LoG) is projected onto the imaging plane of the stereo system and, when this line intersects with a valid depth estimate (which is most times), the PoG is then identified as the region around this intersection. To perform
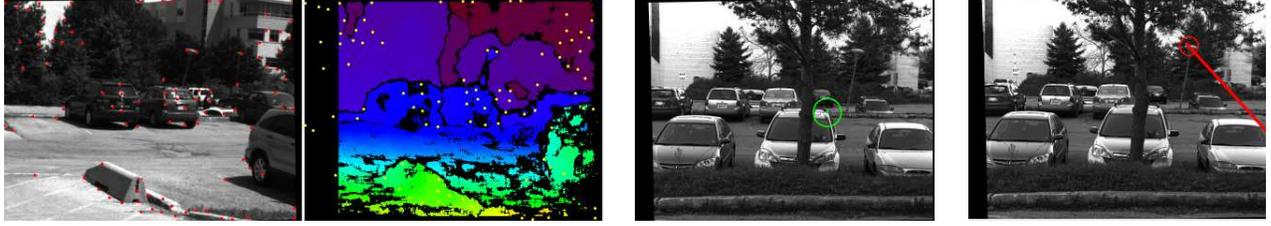
Fig. 3. *The calibration procedure* **a) (left):** *The operator selects calibration points from a set of Hessian salient points provided by OpenCV.* **b) (center):** *The driver gazes at selected points one at a time while the gaze data and depth is recorded.* **c) (right):** *Driver gaze transformed into the reference frame of the stereo imaging system and intersected with the depth-map at frame rate (30Hz).*

this projection, we first compute the 3D parameters of the LoG in the reference frame of the stereo system. The gaze vector and eye center position in the scene frame are obtained as:

$$\mathbf{g} = R^T \mathbf{g}_e \tag{11}$$

where $\mathbf{g}$ and $\mathbf{g}_e$ represent the gaze direction in the stereo imaging reference system and in that of the eye tracker, respectively, and

$$\mathbf{c} = R^T(\mathbf{c}_e - \mathbf{T}) \tag{12}$$

where $\mathbf{c}$ and $\mathbf{c}_e$ represent the eye center position in the stereo imaging system reference system and in that of the eye tracker, respectively. Then the LoG in the scene camera coordinate system becomes

$$\frac{X - c_x}{g_x} = \frac{Y - c_y}{g_y} = \frac{Z - c_z}{g_z} \tag{13}$$

where $\mathbf{c} = (c_x, c_y, c_z)^T$ and $\mathbf{g} = (g_x, g_y, g_z)^T$. Using $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ (perspective projection) yields

$$(g_x c_z - g_z c_x)y - g_x c_y = (g_y c_z - g_z c_y)x - g_y c_x \tag{14}$$

where x and y are 2D coordinates of the LoG stereo scene camera frame of reference. To obtain the LoG in image coordinates, the intrinsic calibration matrix of the stereo scene system is applied to the equation, resulting in

$$(g_x c_z - g_z c_x)\left(\frac{y' - o_y}{f_y}\right) - g_x c_y =$$
$$(g_y c_z - g_z c_y)\left(\frac{x' - o_x}{f_x}\right) - g_y c_x \tag{15}$$

where x' and y' are image coordinates of the perspective projection of the LoG. o_x, o_y, f_x, and f_y are obtained from the intrinsic calibration matrix K of the scene stereo system:

$$K = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \tag{16}$$

Then, the 2D image coordinate of the PoG is that which satisfies

$$(x'_p, y'_p)^T = \underset{(x',y')}{\operatorname{argmin}} \|Z_d - Z_l\| \tag{17}$$

where (x', y') is a pixel on the projected LoG, Z_l is its depth component, and Z_d is the corresponding depth value within the depth map. Z_d and Z_l are obtained as:

$$Z_d = \frac{Z}{W} \tag{18}$$

$$Z_l = \frac{c_z(g_z - g_x)}{g_z\left(\frac{x' - c_x}{f_x}\right) - g_x} \tag{19}$$

Here, Z and W originate from the re-projection in 3D of points $(x', y', d, 1)^T$:

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q \begin{bmatrix} x' \\ y' \\ d \\ 1 \end{bmatrix} \tag{20}$$

where d is the disparity associated with $(x', y')^T$ and Q is the re-projection matrix obtained with the `StereoRectify` function from OpenCv:

$$Q = \begin{bmatrix} 1 & 0 & 0 & -o_x \\ 0 & 1 & 0 & -o_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -T_x^{-1} & (o_x - o'_x)T_x \end{bmatrix} \tag{21}$$

As usual, $(o_x, o_y)^T$ is the principal point in the left image, and o'_x the x coordinate of that of the right image [11]. Since the correct disparity d_p is immediately available once $(x'_p, y'_p)^T$ is obtained with (17), then the 3D PoG is directly given by:

$$\mathbf{G} = (X_p, Y_p, Z_p, 1)^T = W_s^{-1}(X_s, Y_s, Z_s, W_s)^T \tag{22}$$

where

$$\begin{bmatrix} X_s \\ Y_s \\ Z_s \\ W_s \end{bmatrix} = Q \begin{bmatrix} x'_p \\ y'_p \\ d_p \\ 1 \end{bmatrix} \tag{23}$$

III. EXPERIMENTAL PROTOCOL

Two important aspects of this technique need to be evaluated. First, an empirical convergence study must be conducted[1] and second, an error analysis performed within

---

[1]Our algorithm is a straightforward extension to Arun *et al.*'s and consequently subjected to identical noiseless and noisy degenerate cases [9], justifying our decision to only study the numerical convergence rate.
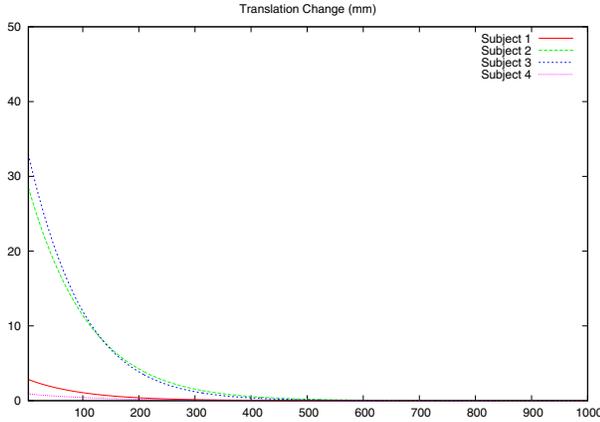
Fig. 4.   *A depiction of the convergence rate of the cross-calibration algorithm for the four test subjects.*

the conditions in which the experimental vehicle is used. We performed the convergence rate and error analysis with a group of four test drivers, composed of two males and two females, averaging 26.5 years of age. This group was composed of one Caucasian and three Middle-Eastern subjects and had no known visual problems.

A. Convergence Rate

Our study of the convergence rate begins with initializing the cross-calibration parameters R and $\mathbf{T}$. The rotation matrix is set to identity, while the translation vector is given a manually measured (an therefore approximate) vector between the centers of projection of both the eye tracker inside the vehicle and the stereo scene system on its rooftop. Figure 4 shows the progression of $\|\mathbf{T}' - \mathbf{T}\|$ toward 0 with respect to the number of iterations. As it is observed, a few hundred iterations (≈ 500) ensure adequate convergence for all test subjects. Since the algorithm is numerically simple, convergence is achieved within 1 s. Interestingly, convergence is particularly rapid for two of the four subjects. While only a conjecture, we believe this may be due to an unusual precision of the gaze of the test subjects when requested to fixate calibration points.

B. Error Analysis

In order to visually appreciate the error differences between calibration with coplanar (CoP) and non-coplanar (NcP) points, we requested one of the test drivers to fixate a number of pre-selected points on a test scene with a CoP calibration configuration (Figure 7a)) and then another set of pre-selected points on a test scene with an NcP calibration configuration (Figure 7b)). We then projected the difference of PoGs between points requested to be fixated (displayed in green) and points actually fixated (displayed in red) determined by the cross calibration parameters. These results show a significant error reduction.

The error analysis we conducted included two distinct scenarios: one for which the cross-calibration points were co-

planar (CoP), and the other for which the points experienced significant non co-planarity (NcP). The aim was to compare the effects on precision when the scene camera is monocular (and hence the calibration must proceed with forcibly co-planar image points, an assumption only valid when the centers of projection of the scene camera and the eye tracker coincide) and our technique. Figure 5 shows a typical CoP scene along with an NcP scene, each used for CoP and NcP calibrations, respectively.

In each scenario, we measured angular error for fixated points within the scene used for calibration (which we refer to as the training scene), and then within an altogether different scene (which we refer to as the test scene), using identical cross calibration parameters for both the training and test scenes.

In all cases, we performed angular error analysis by requesting test subjects to fixate pre-selected points $\mathbf{p}$ in the scene for which the 3D position is known within the error margin of the scene stereo system. For each point $\mathbf{p}$, we requested the test subject to fixate it for 2 s (using the same technique as when calibrating), recorded the gaze data set, and computed its LoG in 3D, where we measured the angle between it and the LoG of $\mathbf{p}$. This method of error evaluation comprises the stereo scene system error (characterized in [1], the eye tracker error (characterized by the manufacturer of the eye tracker[2]), and whether the test subject is accurately fixating the point (difficult to quantify).

Figure 6a) displays the angular errors obtained on a per test subject basis. The green bars represent angular errors for the test scenes and the blue bars those from the training scenes, for both co-planar (CoP) and non co-planar (NcP) calibration points. As expected, for experiments conducted with CoP calibration, the errors for the test scenes (blue bars) are significantly higher than those of the training scenes (green bars). This experimental context clearly shows the inadequacy of assuming coinciding projection centers for the scene camera and the eye tracker. In the case of experiments conducted with NcP calibration, the angular error differences between the training and test scenes are significantly smaller, empirically demonstrating the superiority of our approach. This result is also clearly observed in Figure 6b), where errors are averaged over the test subjects. The difference in angular error between CoP and NcP calibration for the test scenes is superior to $2°$, (or by a multiplicative factor just under 3).

IV. CONCLUSION

In 2009, Hennessey and Lawrence claimed to be first in devising a binocular system for estimating the 3D coordinates of where one is looking in the 3D world, by using vergence [5]. In their experiments, fixated objects were close to test subjects and contained in a 1.725 m^3 volume. They obtained an average PoG error of 3.93 cm. We devised a novel, superior method which remains precise for much larger volumes and distances by combining a binocular eye

---

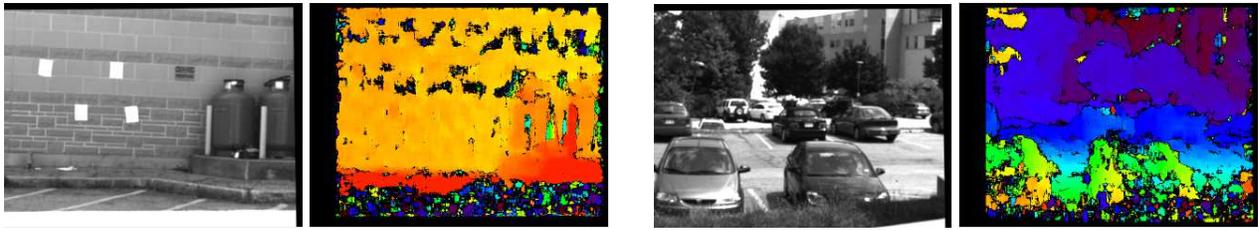[2]FaceLAB 5, from SeeingMachines Inc.

Fig. 5.   **a) (left):** *A wall and its depth map are used to perform experiments with co-planar calibration points.* **b) (right):** *A typical scene used to perform experiments with non co-planar calibration points.*
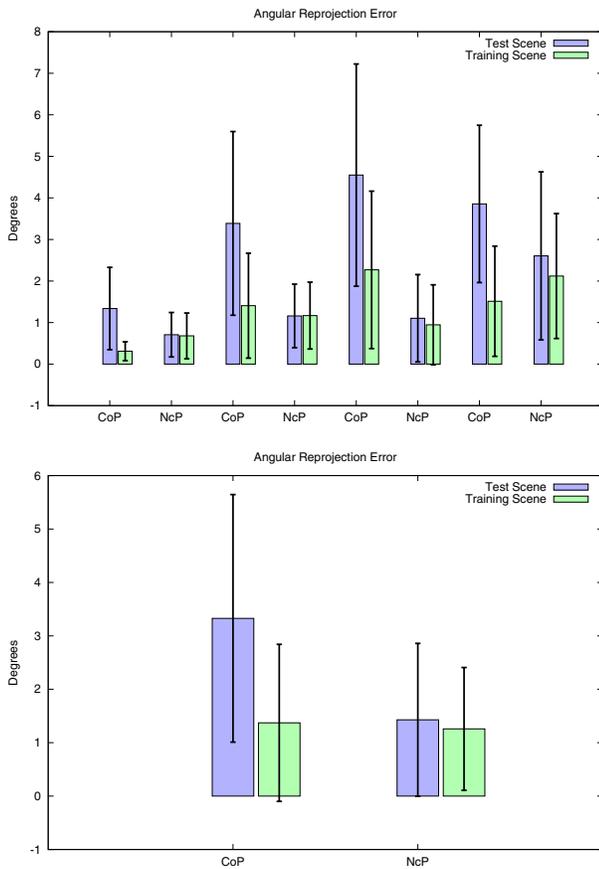

Fig. 7.   **a) (left):** *2D image errors in re-projection between points requested to be fixated and points actually fixated under a CoP calibration for a test scene* **b) (right):** *2D image errors in re-projection between points requested to be fixated and points actually fixated under an NcP calibration for a test scene*

Fig. 6.   **a) (top):** *Angular errors (with standard deviation bars) obtained with the four test subjects on test and training scenes with CoP and NcP calibrations.* **b) (bottom):** *Angular error averages (with standard deviation bars) over the test subjects obtained on test and training scenes with CoP and NcP calibrations.*

gaze tracker with a binocular scene stereo system through an innovative cross calibration procedure. Our system operates in real time (30Hz) and is installed in an operational, experimental vehicle. To our knowledge, this experimental vehicle is the first of its kind, capable of computing the absolute 3D PoG of its driver at 30Hz sufficiently precisely to conduct scientific experiments addressing ocular behavior in relation to visual stimuli.
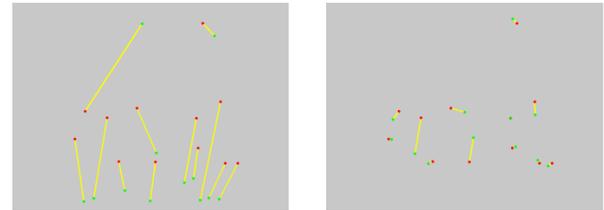
REFERENCES

[1] S. Beauchemin, M. Bauer, T. Kowsari, and J. Cho, "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 2, pp. 391–401, 2012.
[2] F. Lethaus and J. Rataj, "Do eye movements reflect driving manoeuvres?" *IET Intelligent Transportation Systems*, vol. 1, no. 3, pp. 199–204, 2007.
[3] M. Land and J. Horwood, "Which parts of the road guide steering?" *Nature*, vol. 377, pp. 339–340, 1995.
[4] M. Land, "Eye movements and the control of actions in everyday life," *Prog. in Ret. and Eye Res.*, vol. 25, pp. 296–324, 2006.
[5] G. Hennessey and P. Lawrence, "Non-contact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions," *IEEE Transactions on Biomedical Engineering*, vol. 8, no. 3, 2009.
[6] K. Yamashiro, D. Deguchi, T. Takahashi, I. Ide, H. Murase, K. Higuchi, and T. Naito, "Automatic calibration of an in-vehicle gaze tracking system using driver's typical gaze behavior," in *Intelligent Vehicles Symposium*, 2009, pp. 998–1003.
[7] M. Kim, S. Yang, and D. Kim, "Head-mounted binocular gaze detection for selective visual recognition systems," *Sensors and Actuators A: Physical*, vol. 197, pp. 29–36, 2012.
[8] S. Bernet, C. Cudel, D. Lefloch, and M. Basset, "Autocalibration-based partitioning relationship and parallax relation for head-mounted eye trackers," *Machine Vision and Applications*, vol. 24, no. 2, pp. 393–406, 2013.
[9] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.
[10] J. Challis, "A procedure for determining rigid body transformation parameters," *Journal of Biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.
[11] G. Bradski and A. Kaehler, *Learning OpenCV*. O'Reilly Media, 2008.