Wearable Gaze Trackers: Mapping Visual Attention in 3D

Rasmus R. Jensen[1], Jonathan D. Stets[1](✉), Seidi Suurmets[2], Jesper Clement[2], and Henrik Aanæs[1]

[1] Technical University of Denmark, Kongens Lyngby, Denmark
stet@dtu.dk
[2] Copenhagen Business School, Frederiksberg, Denmark

Abstract. The study of visual attention in humans relates to a wide range of areas such as: psychology, cognition, usability, and marketing. These studies have been limited to fixed setups with respondents sitting in front of a monitor mounted with a gaze tracking device. The introduction of wearable mobile gaze trackers allows respondents to move freely in any real world 3D environment, removing the previous restrictions.

In this paper we propose a novel approach for processing visual attention of respondents using mobile wearable gaze trackers in a 3D environment. The pipeline consists of 3 steps: modeling the 3D area-of-interest, positioning the gaze tracker in 3D space, and 3D mapping of visual attention.

The approach is general, but as a case study we created 3D heat maps of respondents visiting supermarket shelves as well as finding their in-store movement relative to these shelves. The method allows for analysis across multiple respondents and to distinguish between phases of in-store orientation (far away) and product recognition/selection (up close) based on distance to shelves.

1 Introduction

The study of human visual attention relates to a wide range of areas such as: psychology, cognition, useability, and marketing. In order to directly study this in various settings, eye tracking has become a standard method. A common way of visualizing and analysing gaze data is using Areas Of Interest (AOI) and attentional heat maps [13]. The heat maps represent the spatial distribution of eye movement throughout the AOI and can often be used for quantitative analysis. The most common method of visualizing heat maps is using a Gaussian based solution. Here, four parameters are used to determine the appearance of the heat map: the width of the basic construct, the use of fixations vs. raw data, whether accounting for duration of fixation and the mapping color altitude form [3]. For many years, mapping visual attention as heat maps has been limited to static setups with respondents sitting in front of a screen mounted with a stationary calibrated gaze tracker. Such a setup can accurately map the visual attention as a heat map of what is projected on the screen, but obviously limits

P. Sharma and F.M. Bianchi (Eds.): SCIA 2017, Part I, LNCS 10269, pp. 66–76, 2017.
DOI: 10.1007/978-3-319-59126-1_6

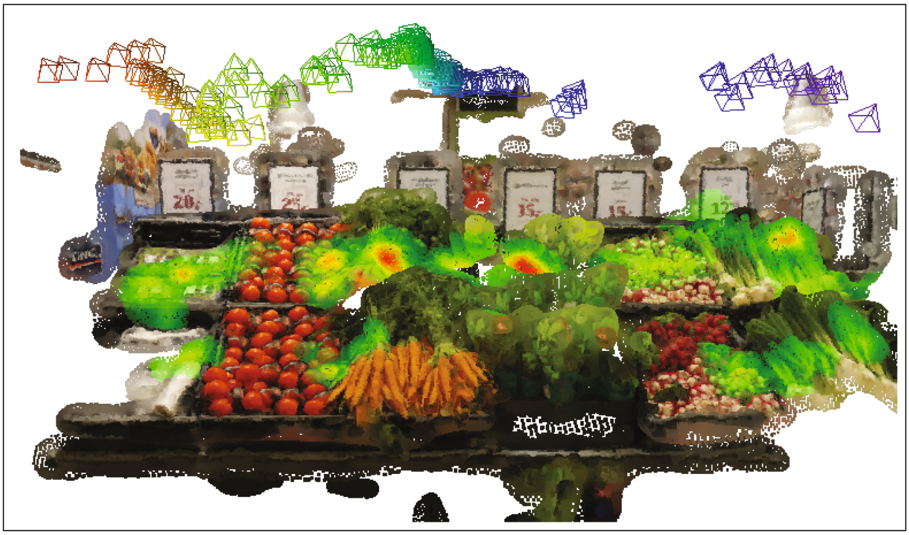


Fig. 1. Supermarket vegetables shown as a 3D model with heatmap and respondent viewing points.

the visual attention to a 2D surface. The recent introduction of mobile wearable gaze trackers (Fig. 2) enables data collection in about any real-world environment. On mobile wearable eye-trackers, the scene is recorded using a front facing camera, and gaze data collected from eye tracking cameras can be projected onto this video. Despite the potential of introducing recordings of three dimensional scenes, common for both the stationary and mobile wearable eye-tracker is that ultimately the data is still recorded and analysed in 2D.

Mapping visual attention data recorded in a 3D space to a 2D heatmap is not straightforward. A simple approach is to find the best homographic correspondence between a reference image and a given frame from the eye-tracker, and then map the gaze according to this homography [4,12]. Figure 3 shows common errors in mapping using a homography relative to the actual mapping onto a 3D AOI. We argue that gaze collected in 3D mapped onto a 2D reference image using a homography will always be limited as a result of incorrect mappings.

Fig. 2. Tobii Pro Glasses 2 [12]. A wearable gaze tracker that tracks a respondents eye movements using IR cameras, while also recording the environment with a front facing video camera.

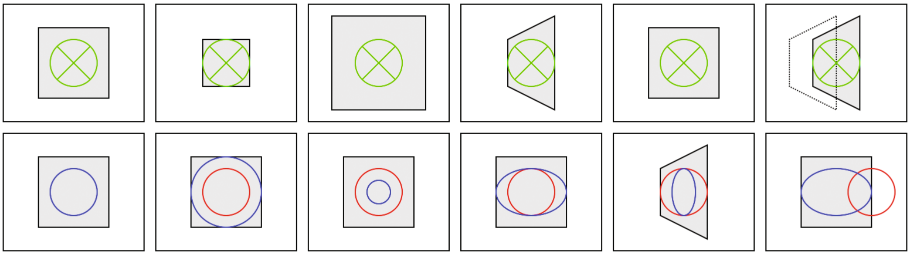


Fig. 3. This figure shows the common errors relating to mapping gaze. The top row shows a respondents viewpoint of an AOI with the gaze point in green, while the second row shows a reference view of an AOI with gaze mapped as homography mapping in red and mapping according to 3D structure in blue. First column shows mapping, when the viewpoint of the respondent and reference coincide. In this case, homography mapping and mapping according to 3D structure will be identical and perfectly overlapping. Column 2 and 3 show the mapping, when respondent is closer to or further from the AOI. Using homography mapping, the gaze point does not change along with the movement of the respondent. Row 4 and 5 show how the mapped gaze changes shape according to changes in viewpoint for the 3D mapping, while homography mapping does not change accordingly. The final column shows the error, when the homography is offset from the plane of the actual viewpoint, which introduces parallax error. (Color figure online)

We propose a solution to these problems and limitations by modelling an AOI in 3D as a reference for mapping gaze data. The reference model is reconstructed from photographs of the AOI to establish a good base for image feature matching and a high quality model mesh. We demonstrate a fully automatic pipeline for generating a 3D attention heat map, and furthermore the possibility of calculating the respondent viewing points as shown in Fig. 1. Our pipeline enables spatial filtering, positioning and orientation relative to the selected AOI, as well as correlation of multi-respondent data. We use supermarket shelves as a case study, but our pipeline is not limited to this setup. Our method requires a standard digital camera to capture images of the reference model, and a wearable gaze tracker with a front facing camera, such as the one shown in Fig. 2, for recording the scene and gaze data.

There are a number of recent studies that addresses the need to move mapping of visual attention to 3D. [11] introduces the potential of measuring 3D gaze coordinates from head-mounted gaze trackers, and [9] proposes visualisation of 3D gaze data on to virtual computer generated models. A method similar to our pipeline is described in [10], which demonstrates the use of a Microsoft Kinect to create a 3D reference model. Our method differs by using images to create a more dense point cloud, which also enables us to backproject the heat map to a traditional 2D visualization for comparison.

2 Data

We have collected data in both a real world supermarket and using a mock-up supermarket shelf in our lab. Reference data of the AOIs have been captured using a digital mirrorless camera: a Panasonic GH4 with a 12 mm lens (24 mm in 35 mm equivalent). To collect respondent data we have used the Tobii Pro Glasses 2 wearable gaze tracker [12] (Fig. 2), which collects the respondents view using a front facing video camera, while also recording the respondent gaze direction using 4 infrared cameras facing the eyes. Both cameras were calibrated using a standard checkerboard approach [16]. Data was collected of four in-store product sections in a supermarket: wine, vegetables, flour and cereal, as well as a mock-up of the cereal section in our lab. We used the digital camera to capture sets of reference images to cover the desired AOIs (12–20 images of each AOI). Gaze and video data were collected of respondents visiting the given sections (16 sets), visiting the store but acquired to get cereal (4 sets), and finally, presented for a mock-up of the cereal section in the lab (6 sets). All gaze data samples are raw, so no fixation filtering has been applied [3].

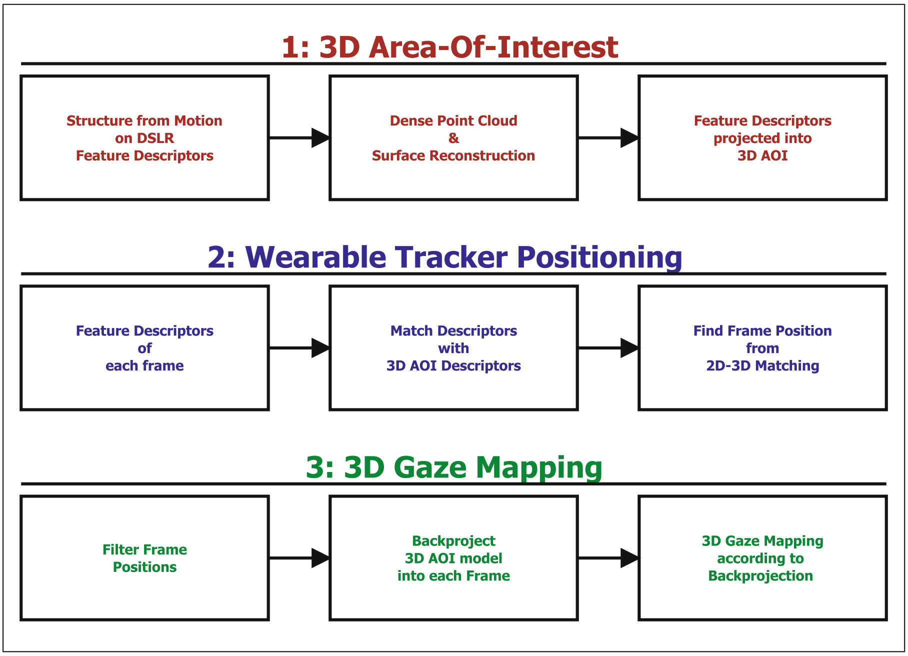

1: 3D Area-Of-Interest
Structure from Motion on DSLR Feature Descriptors
Dense Point Cloud & Surface Reconstruction
Feature Descriptors projected into 3D AOI
2: Wearable Tracker Positioning
Feature Descriptors of each frame
Match Descriptors with 3D AOI Descriptors
Find Frame Position from 2D-3D Matching
3: 3D Gaze Mapping
Filter Frame Positions
Backproject 3D AOI model into each Frame
3D Gaze Mapping according to Backprojection

Fig. 4. The 3 steps in our proposed pipeline to construct 3D gaze mapping: Modelling of an Area-Of-Interest, Eye-tracker frame positioning, and finally the gaze mapping.

3 Method

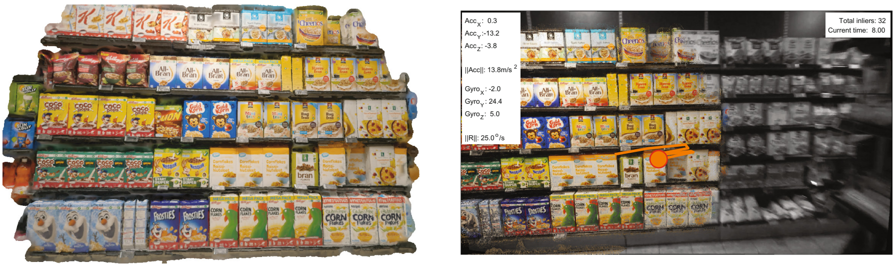
In order to map gaze data onto a 3D AOI, we propose a pipeline consisting of three parts (Fig. 4): construction of the 3D AOI reference model, localization of the wearable gaze tracker frames relative to the reference model, and finally gaze mapping onto the AOI as a heat map.

3.1 Modelling a 3D Area-Of-Interest

The 3D AOI reference model is built using a series of images of the AOI. This task is divided further into three steps (Fig. 4). First, we use structure from motion to find the spatial camera positions and a sparse point cloud representation. We have opted for a structure from motion (SfM) [6] implementation, which requires a sequence of images followed by an image rectification based on the parameters obtained from the camera calibration. SIFT descriptors [7,15] are found in each image and sequentially matched across the sequence of images in an iterative fashion. Images with sufficient feature matches are included, while the extrinsic camera parameters are estimated and refined using bundle adjustment [14].

Given the estimated extrinsic camera parameters, we move onto dense point cloud estimation using the patch-expansion approach to multiview stereopsis proposed by Furukawa and Ponce [2]. This method robustly produces dense representations from which a surfaces are reconstructed using Poisson surface reconstruction by Kazhdan et al. [5]. A 3D modelled AOI from the cereal section in a supermarket is shown in Fig. 5(a).

As a preparation step for the localization of the wearable gaze tracker later in the pipeline, we use backprojection with depth management of the 3D AOI to project the model into each reference image. This is done in order to project 2D SIFT descriptors [7] into the 3D space, allowing the 2D descriptors between each frame from the gaze tracker and 3D AOI to be compared.



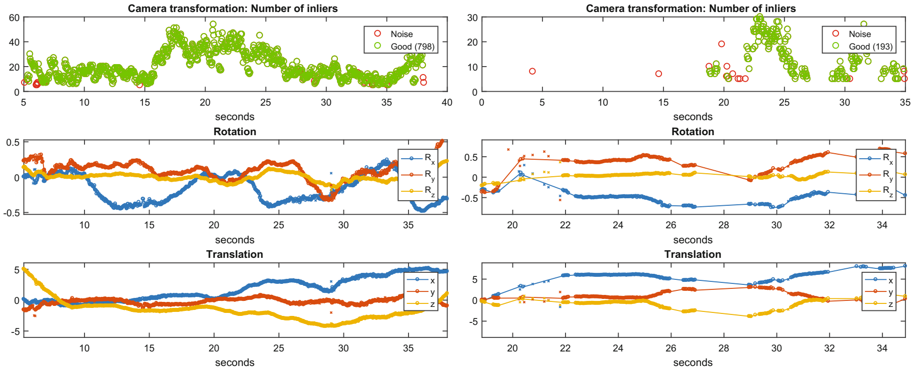
(a) 3D Area-Of-Interrest reference model.

(b) 3D AOI backprojected onto a gaze tracker frame.

Fig. 5. The 3D AOI in (a) is backprojected onto an undistorted gaze tracker frame and the gaze point with trace from previous frames (b). The frame is shown in black and white, while the projection is shown in color.

3.2 Wearable Gaze Tracker Frame Localization

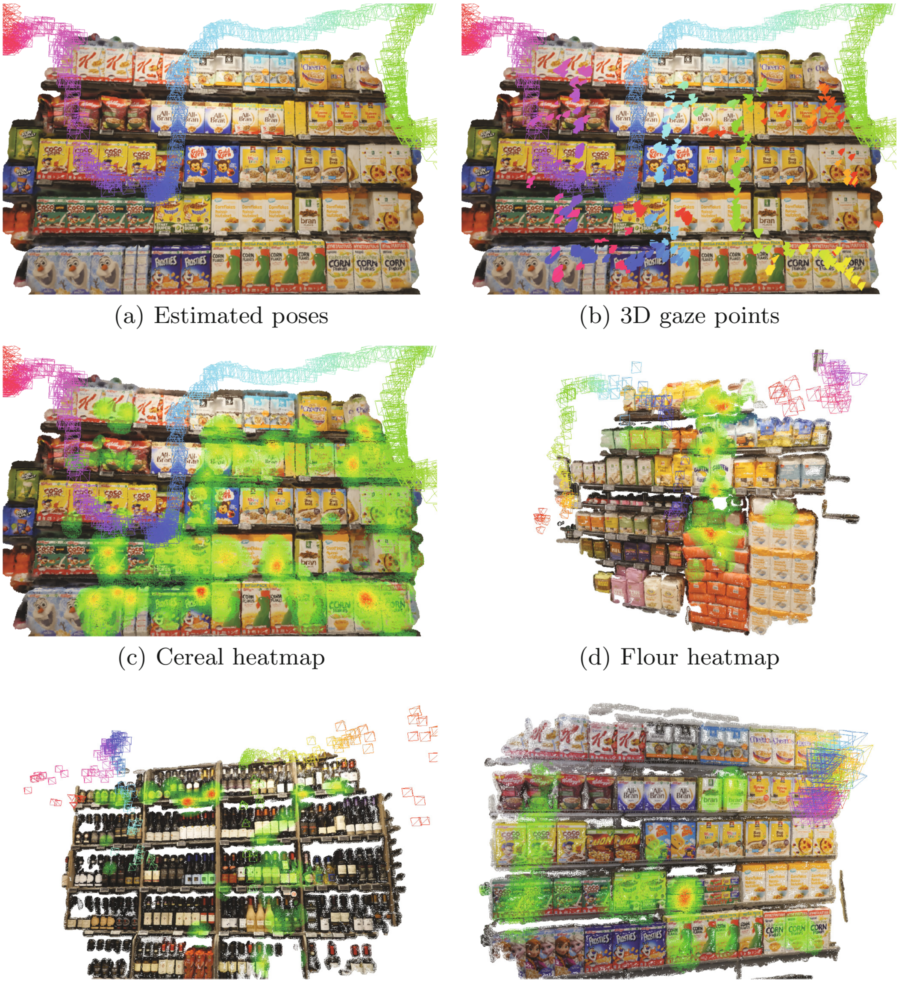
In order to correctly map gaze data on the 3D AOI, each frame from the gaze tracker has to be positioned relative to the AOI (if visible). The SIFT descriptors in each frame are matched with the reference 2D descriptors projected into 3D space, when constructing the 3D AOI. These correspondences are sent to a 2D to 3D pose solver, which finds the best fit camera pose using a RANSAC approach discarding outliers [1]. Given sufficient corresponding points, the solver will return the correct camera pose relative to the AOI. Without sufficient good matches the solver either fails or returns a false camera pose. Since a given frame might not cover any part of the AOI, the resulting matching consists of either a lot of true positive or a few false positive matches. Figure 5(b) shows the 3D AOI backprojected into a frame from the wearable tracker using the estimated camera pose. This backprojection is an immediate sanity check, showing the correctness of the pose estimation. Incorrect pose estimates tends to be very inconsistent from one frame to the next. To speed things up we have used the above approach to find the pose in keyframes, which are followed by frames, where the correspondence points are tracked using optical flow [8] (1 keyframe followed by 5 optical flow frames). This is substantially faster than finding and matching features in each frame. The pose solver is initialized with the pose from the previous frame, which along with the optical flow gives timewise consistency in the pose estimation.



(a) Cereal shelves

(b) Flour shelves

Fig. 6. Estimated respondent poses from visits to the cereal and flour shelves with timestamp. First row of plot: the framewise number of inliers (y-axis) in the camera positioning solver. Green points are included as reliable and red points are considered noise. Noise points are filtered out based on spatial and rotational inconsitency. Second and third row of plots are the rotation and translation with inliers shown as a connected graph and the few outliers as single points. (Color figure online)



(a) Estimated poses

(b) 3D gaze points

(c) Cereal heatmap

(d) Flour heatmap

(e) Wine heatmap

(f) Mock-up heatmap

Fig. 7. Poses, gaze points and heatmaps obtained from the data of the 5 sections included in our study.

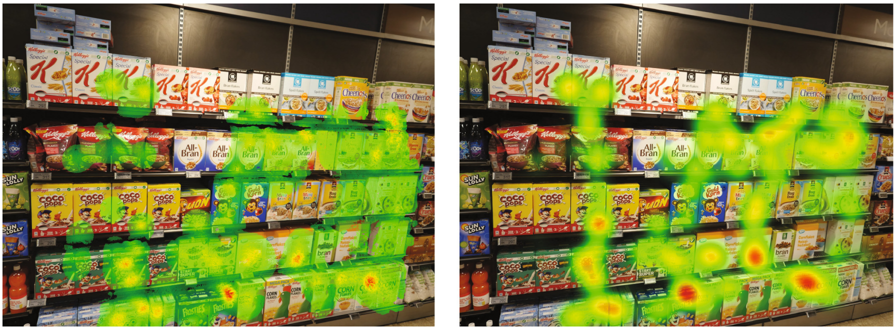
3.3 Mapping Gaze Data

The pose estimation is unbiased and may results in a few faulty poses. We consider these noise and use the following approach to filter them from the good poses. Correct pose estimates between consecutive frames is assumed to have small variation, while incorrect poses are very inconsistent. This inconsistency is used to identify and discard faulty poses. In Fig. 6 the good pose estimates is shown as a connected graphs with discarded poses as outlying points. The number of inliers returned from the 2D to 3D pose solver is a good estimator of correctness, but thresholding this number is not as robust as filtering the pose.

A respondent moving in front of the AOI is shown in Fig. 7(a). Using the good poses, the gaze can be mapped onto the 3D AOI model creating a 3D heat map as seen in Fig. 7(b). The gaze intersection with the 3D model is found using backprojection with depth management into the current frame, which is significantly faster than calculating the intersection between the line of sight and the 3D model.

A similar approach is taken, when creating the heatmap. Here a predetermined symmetric 2D Gaussian function with center at the gaze coordinate is added to a sum map of Gaussians in 3D space. Using a Gaussian serves both the fact that sight is not an infinitely small point, while also incorporating some uncertainty in the gaze estimates. Discussions about the size of the Gaussian, and whether the raw gaze data or fixation filtered data should be used is beyond the scope of this work. The resulting heatmaps visualized on the 3D AOIs can be seen in Figs. 1 and 7(c) to (e).

One benefit worth noticing is, that the approach of mapping Gaussian to the backprojection of the AOI allows for a normalization of the contribution from each gaze point. It also addresses the problems shown in Fig. 3. When a respondent is close, the covering of the Gaussian gaze point of the 3D AOI will be small with a locally high intensity. Respondents far away will cover a larger area in the 3D AOI, which will result in less locally intensive mapping. It also handles change in perspective, while effectively shaping the Gaussian according to the viewpoint without introducing parallax error. Since the sum of Gaussian gaze points is done on a 3D model, the heatmap can be projected into any frame or reference image. The backprojection of the a heatmap is shown in Fig. 8(a) as an overlay to the original image.



(a) 3D heatmap backprojected into reference image

(b) 2D heatmap from iMotions software

Fig. 8. Heatmaps based on 3D gaze mapping and 2D gaze mapping. For the 3D mapping the heatmap has been backprojected into the DSLR frame used for the 2D mapping.

4 Results

The core of our presented pipeline is the ability to correctly find the pose of the wearable gaze tracker relative to the 3D AOI in a given frame. Validating this after filtering puts each frame in one of four categories:

True positive correctly detecting the AOI.
True negative correctly not detecting the AOI.
False positive incorrect detection of the AOI.
False negative incorrectly not detecting the AOI.

Reviewing the output videos with 3D AOI overlay backprojected as presented in Fig. 5(b) is an easy way to quickly asses the quality of the AOI detection. Such a review shows non or only a very few false positives, but some false negatives. Since the gaze tracker has a very small sensor, the sensor struggles with low indoor light, which results in both frames with motion blur from head movement and rolling shutter. In the supermarket setting, these frames provide the vast majority of false negatives and one could debate, whether they are actually false negatives. Occlusion from people or other shelves can also cause false negatives. Reviewing both the frame positions as a graph in Fig. 6 or the resulting spatial positions in Fig. 7(a) are can also provide quick qualitative verifications in addition to reviewing a video with backprojected 3D AOI.

We have reprojected the heatmap into a reference image, which has also been applied homography gaze mapping using the iMotions 6.2 software [4] and the results are shown in Fig. 8. The heatmaps are both based on raw data samples but using different techniques: 3D mapping and homography mapping respectively. This means they cannot be compared directly, however there have clear similarities of the path pattern and duration of attention.

5 Conclusion

We have successfully created 3D AOIs and heat maps for respondents visiting the five sections in our data set: vegetables, cereal, flour, wine, and cereal mockup. Our proposed pipeline does away with the problems relating to mapping gaze using a homography. The proposed pipeline is fully automatic and runs at ~2 fps using a combination of Matlab, mex, vlfeat and OpenCV. A full C++ implementation will provide further speed up, but the scope was to demonstrate a feasible pipeline, which allows researchers to spend time only on the results, once processed. Concerns such as visual attention mapped based on fixation vs. raw data, size of the Gaussian gaze point, and cross respondent analysis has not been evaluated. We found that our pipeline works well in in-store settings, since store products tend to have very distinct image features. However, settings with only repeating image features, such as frames with only the same product present, complicates the feature matching. This is often the case, when the respondent is very near a product shelf. On the other hand, detection works well in the case where the respondent is inspecting the shelves at an arm length distance, which in many cases is the important frames for generating heat maps. Our approach provides a fully automatic method of mapping gaze data and positioning of the respondent relative to the AOI, thus adding another dimension to the resulting data.

Acknowledgements. This work has been funded by the Innovation Fund (Denmark) and carried out in collaboration with iMotions. We would like to thank both for the support and collaboration.

References

1. Bradski, G.: The OpenCV Library. Dr. Dobb's J. Softw. Tools, Article id 2236121 (2000). http://code.opencv.org/projects/opencv/wiki/CiteOpenCV
2. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. **1**(1), 1–14 (2007). ISSN 01628828. http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.161
3. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. OUP, Oxford (2011)
4. iMotions. iMotions biometric research platform (2017). https://imotions.com/
5. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the Fourth Eurographics Son Geometry Processing (SGP 2006), pp. 61–70, Aire-la-Ville, Switzerland. Eurographics Association (2006). ISBN 3-905673-36-3
6. Koenderink, J., Van Doorn, A.: Affine structure from motion. Optical Soc. Am. A **8**(2), 337–385 (1991). doi:10.1364/JOSAA.8.000377, ISSN 1084–7529
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
8. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision (1981). ISSN 17486815, http://www.ic.unicamp.br/rocha/teaching/2013s1/mc851/aulas/additional-material-lucas-kanade-tracker.pdf
9. Maurus, M., Hammer, J.H., Beyerer, J.: Realistic heatmap visualization for interactive analysis of 3d gaze data. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 295–298. ACM (2014)
10. Paletta, L., Santner, K., Fritz, G., Mayer, H., Schrammel, J.: 3d attention: measurement of visual saliency using eye tracking glasses. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, pp. 199–204. ACM (2013)
11. Pfeiffer, T.: Measuring and visualizing attention in space with 3d attention volumes. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 29–36. ACM (2012)
12. Tobii Pro. Tobii pro glasses 2 (2017). http://www.tobiipro.com/product-listing/tobii-pro-glasses-2/
13. Purucker, C., Landwehr, J.R., Sprott, D.E., Herrmann, A.: Clustered insights: improving eye tracking data analysis using scan statistics. In: Psychological Considerations on Car Designs-An Investigation of Behavioral and Perceptual Aspects Using Eye Tracking and Cross-Cultural Studies (2012)
14. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment — a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) IWVA 1999. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000). doi:10.1007/3-540-44480-7_21, http://www.springerlink.com/content/plvcrq5bx753a2tn

15. Vedaldi, A., Fulkerson, B.: Vlfeat. In: Proceedings of the International Conference on Multimedia (MM 2010), vol. 3, no. 1, p. 1469 (2010). doi:10.1145/1873951.1874249, http://dl.acm.org/citation.cfm?doid=1873951.1874249
16. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, no. c, p. 7 (1999). doi:10.1109/ICCV.1999.791289, ISSN 01628828