2013 IEEE International Conference on Computer Vision

Calibration-Free Gaze Estimation Using Human Gaze Patterns

Fares Alnajar Theo Gevers Roberto Valenti Sennay Ghebreab
University of Amsterdam
Amsterdam, The Netherlands
{f.alnajar,th.gevers,r.valenti,s.ghebreab}@uva.nl

Abstract

We present a novel method to auto-calibrate gaze estimators based on gaze patterns obtained from other viewers. Our method is based on the observation that the gaze patterns of humans are indicative of where a new viewer will look at [12]. When a new viewer is looking at a stimulus, we first estimate a topology of gaze points (initial gaze points). Next, these points are transformed so that they match the gaze patterns of other humans to find the correct gaze points.

In a flexible uncalibrated setup with a web camera and no chin rest, the proposed method was tested on ten subjects and ten images. The method estimates the gaze points after looking at a stimulus for a few seconds with an average accuracy of 4.3°*. Although the reported performance is lower than what could be achieved with dedicated hardware or calibrated setup, the proposed method still provides a sufficient accuracy to trace the viewer attention. This is promising considering the fact that auto-calibration is done in a flexible setup , without the use of a chin rest, and based only on a few seconds of gaze initialization data. To the best of our knowledge, this is the first work to use human gaze patterns in order to auto-calibrate gaze estimators.*

1. Introduction

Gaze estimation is the process of determining where a person is looking at in a predefined plane. It is important for many applications such as human-computer interaction, marketing and advertisement [1], and human behavior analysis. The applications go beyond that to help disabled users (e.g. eye typing) [2].

In general, gaze estimation methods fall into two categories: 1) appearance-based methods [5, 6, 7] and 2) 3D-eye-model-based methods [8, 9, 10, 14]. The former extracts features from images of the eyes and map them to points on the gaze plane (i.e. gaze points). The latter tries to construct a 3D model of the eye and estimates the visual axis. The intersection of the axis and the gaze plane determines the gaze point. Regardless of the gaze estimation method, a calibration procedure is needed to set some parameters. The calibration can be camera-based (estimating the camera parameters), geometric calibration (estimating the relations between the scene components like the camera, the gaze plane, and the user), personal calibration (determining the angle between visual and optical axes), or gaze mapping correlation [11]. An extensive overview of the different approaches of gaze estimation can be found in [11].

3D-eye models require special equipment like cameras with multiple light sources and infrared. The costs and the strict requirements for their use (infrared, for example, is not reliable when used outdoors) limit their widespread applicability. On the other hand, appearance-based approaches are less accurate than 3D-eye-models and less invariant to head pose changes. Yet, low-cost cameras are common and sufficient for appearance-based approaches which makes them suitable for applications where high accuracy is not required. Consider for example an application of people looking at advertisements for marketing research. Asking each participant to buy dedicated cameras or to do the experiment in the lab is time and money consuming, while low-cost cameras are integrated in almost every laptop or tablet nowadays. Appearance-based methods are more suitable in such a situation.

Besides the choice of the recording equipment, the approach allows for a certain level of flexibility of the setup and the calibration. During calibration, users are usually asked to fixate their gaze on certain points while images of their eyes are captured. This procedure is cumbersome and sometimes impractical. In case of, for example, tracing costumers attention in shops, estimating the gaze points or regions should be done passively. Hence, some approaches suggest methods to reduce the number of calibration points. However, in case of passive gaze estimation, the calibration should be done completely automatically without a calibration procedure enforced on the user.

Some recent studies focus on visual saliency information

1550-5499/13 $31.00 © 2013 IEEE
DOI 10.1109/ICCV.2013.24

in images and videos to avoid applying active human calibration. Sugano et al. [4, 3] treat saliency maps extracted from videos as probability distributions for gaze points. Gaussian process regression is used to learn the mapping between the images of the eyes and the gaze points. Chen and Ji [14] use 3D models of the eye and incrementally estimate the angle between the visual and the optical axes by combining the image saliency with the 3D model. The reason behind the use of saliency is that people look at salient regions with higher probability than other regions. However, as shown in [12], the computational saliency models do not frequently match the actual human saccades (Figure 1). In this paper, we claim that the gaze patterns of several viewers provide important cues for the auto-calibration of new viewers. This is based on the assumption that humans produce similar gaze patterns when they look at a stimulus. The assumption is supported by Judd et al. [12], where the authors show that fixation locations of several humans are strongly indicative, in general, of where a new viewer will look at. To the best of our knowledge, our work is the first to use human gaze patterns in order to auto-calibrate gaze estimators.

We present a novel approach to auto-calibrate gaze estimators based on the similarity of human gaze patterns. In addition, we make use of the topology of the gaze points. Consider, in a fully uncalibrated setting, a person who follows a stimulus from left to right. It would be difficult to indicate where the gaze points are on the gaze plane. However, their relative locations can still be inferred and used for auto-calibration. In a fully uncalibrated setting, when a new subject looks at a stimulus, *initial gaze points* are inferred. Then, a transformation is computed to map the initial gaze points to *match* the gaze patterns of other users. In this way, we use all the initial gaze points to match the human gaze patterns instead of using each gaze point at the time. Consequently, the transformed points represent the auto-calibrated estimated gaze points.

The rest of the paper is organized as follows: the proposed method is explained is Section 2. Next, we describe the experimental setup and evaluation in Section 3. The results are discussed in Section 4. Finally, the conclusions are given in Section 5.

2. Calibration-free gaze estimation using human gaze patterns

We build upon the observation that gaze patterns of individuals are similar for a certain stimulus [12]. Although, there is no guarantee that people always look at the same regions, human gaze patterns provide important cues about the locations of the gaze points of a new observer. The pipeline of the proposed method is as follows: when a new user is looking at a stimulus, the initial gaze points are computed first. Then, a transformation is inferred which maps

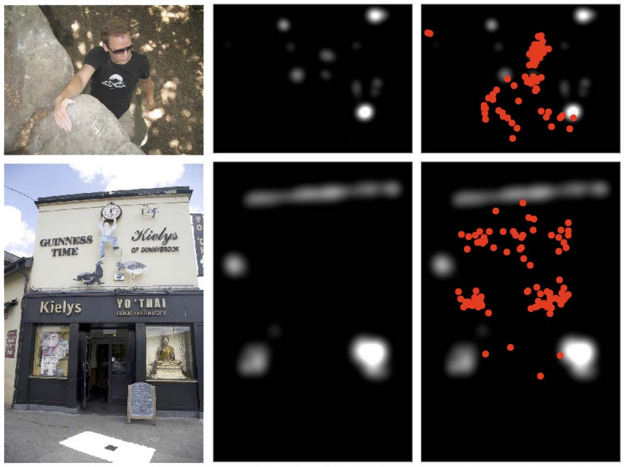

Figure 1. (Taken from [12]). Examples where saliency models do not match the human fixations. Bright spots indicate the saliency model predictions and the red dots refer to the human gaze points.

the initial gaze points to gaze patterns of other individuals. Here, we consider a transformation with translation and scaling (per dimension). Other transformations like rotation or shearing might provide better mapping. Yet, for simplicity, we focus on translation and scaling which are the most common transformations for gaze estimation. Figure 2 illustrates the pipeline.

2.1. Initial gaze points

The final gaze points should eventually match the human gaze patterns. However, we need to start from an *initial* estimation of the gaze points. Hereafter, we present two methods to achieve this: estimation of initial gaze points from eye templates and estimation based on 2D-manifold.

2.1.1 Eye templates

In this approach, the eye images of a person are captured (templates) while fixating the eyes on points on a gaze plane. The images of the eyes of a new user are captured and compared with the template eye images. The idea is to reconstruct the eye image at hand based on the eye image templates. Note that here the eye templates are captured once from a single subject. When a new subject uses the gaze estimator, his or her eye images are compared with the already-collected eye templates. This is different from the traditional calibration-based gaze estimator where the eye templates are captured and stored for each subject. This process can be performed at the raw intensity level or at the feature level. We will refer to both eye image representations as feature vectors. Consider $\{\mathbf{t}_i\}$ to be the template feature vectors, and $\{p_i\}$ denotes the corresponding

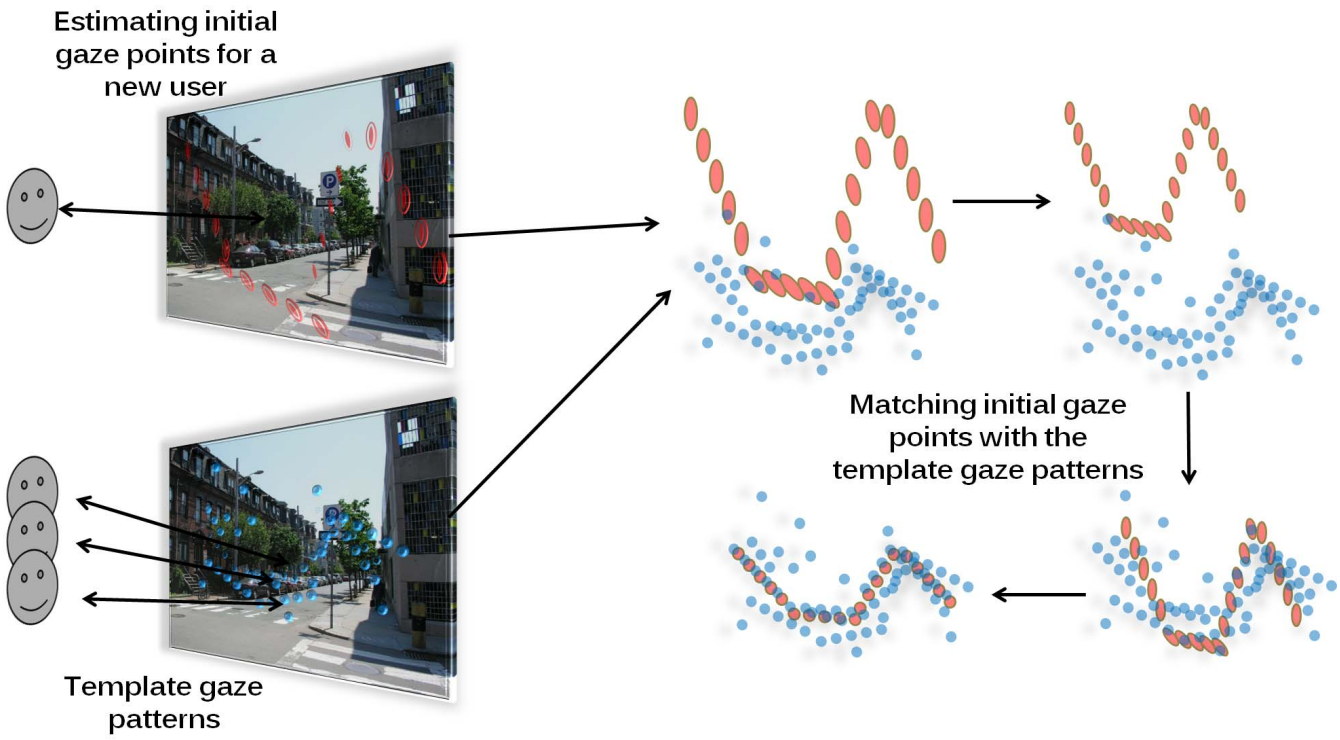

Figure 2. Graphical illustration of the proposed method. Template gaze patterns refer to the gaze points of other individuals for the same gaze plane (display). When a new user looks at the stimulus, his or her initial gaze points are first estimated which preserves the relative locations between the gaze points. These points are transformed so that they match the template gaze patterns.

gaze points. Furthermore, $\{w_i\}$ corresponds to the computed weights to reconstruct a new eye image feature vector $\hat{\mathbf{t}}$:

$$\hat{\mathbf{t}} = \sum_i w_i \mathbf{t}_i \qquad s.t. \qquad \sum_i w_i = 1. \qquad (1)$$

Then the corresponding gaze point $\hat{p}$ for $\hat{\mathbf{t}}$ is calculated as follows:

$$\hat{p} = \sum_i w_i p_i. \qquad (2)$$

To find the weights $\{w_i\}$ values, Tan et al. [13] suggest to first select a subset of $\{\mathbf{t}_i\}$ where the first and the second neighbors of the sample are selected for training. The weight values are then computed as in [15]. Lu et al. [6] select only the direct neighbors as a training subset. Here, we select only the direct neighbors as in [6].

For a new user in a different unknown scene setup, the initial gaze points will be incorrect (without calibration). However, the relative locations between the gaze points are preserved.

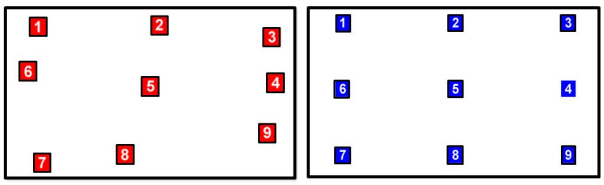

Figure 3. The projection of features of 9 eye images on a 2-D manifold (red, left) and the positions of the corresponding gaze points on the gaze plane (blue, right). The 2D manifold is computed using 800 eye images corresponding to various locations on the gaze plane.

2.1.2 2D manifold

Lu et al. [6] find that the template eye features correspond to a 2D manifold while retaining most of the important information. The reason is that the eyes move, in the appearance-based representation, in two degrees of freedom. Figure 3 shows the projection of features of 9 eye images on a 2D manifold and their corresponding 9 gaze points on the gaze plane. It can be derived that the feature projections preserve the relative locations of the corresponding gaze points.

The 2D manifold can be obtained by projecting the tem-

plate features on the first two principal components. However, the locations on the 2D manifold might be interchanged, transposed, or rotated when compared with the corresponding gaze points. For example, when the eyes move mainly vertically, the first principal component represents the pupil changes on the Y dimension and the second principal component represents the X dimension. Hence, the projected locations need to be transposed. As this step is performed once offline, the projected locations are checked once and transformed to match the corresponding gaze points locations. As in the eye templates method, this procedure is followed once with a single (template) subject. When a new user looks at a stimulus, the eye features are projected on the offline-learned 2D manifold and the projected values are treated as initial gaze points.

The previous two methods (eye templates and 2D manifold) provide a way to find the initial gaze points. In the next section we explain how to map these points to match the template (human) gaze patterns.

2.2. Gaze points mapping

Judd et al. [12] show that the fixation points of several humans correspond strongly with the gaze points of a new user. We aim to exploit this observation to perform calibration without any active user participation. To this end, we transform the initial (uncalibrated) gaze points so they match the template gaze patterns for a stimulus. By applying the aforementioned transformation, we aim to transfer the gaze points to their correct positions without explicit calibration. We present two different methods to find the best transformation. Let $\mathbb{P} = \{\mathbf{p^1}, \mathbf{p^2}, ...\mathbf{p^M}\}$ denotes the gaze patterns of M users (hereafter, we call them *template gaze patterns*) where $\mathbf{p}^u = \{p_1^u, p_2^u, ...p_{Su}^u\}$ consists of the gaze points of user u, and Su is different for each user. Let $\mathbf{p} = \{p_1, p_2, ...p_S\}$ be the initial gaze point set for a new user. The following two methods aim to transform $\mathbf{p}$ so it can match the template gaze patterns $\mathbb{P}$.

2.2.1 K-closest points

This methods tries to find the best mapping which minimizes the sum of distances of each point $p_j \in \mathbf{p}$ to its K closest neighbors of $\mathbb{P}$. Consider Φ is the set of all mappings. The method tries to find a mapping $\bar{\phi} \in \Phi$ which satisfies:

$$\bar{\phi} = \arg\min_{\phi} \Gamma(\mathbf{p}, \mathbb{P}, \phi), \tag{3}$$

where:

$$\Gamma(\mathbf{p}, \mathbb{P}, \phi) = \sum_{j=1}^{S} \sum_{k=1}^{K} \|\phi(p_j) - N(p_j, \mathbb{P}, k)\|. \tag{4}$$

$N(p_j, \mathbb{P}, k)$ is the k closest point from $\mathbb{P}$ to p_j. $\bar{\phi}$ is the computed mapping and $\bar{\mathbf{p}} = \bar{\phi}(\mathbf{p})$ represents the auto-calibrated gaze points. Note that we try to match the initial gaze points with all the gaze patterns in $\mathbb{P}$ simultaneously. To find $\bar{\mathbf{p}}$ and $\bar{\phi}$, a greedy approach is taken. At each step, we apply translation in eight directions with different scales. Then, we adopt the translation and scale which gives the best outcome according to 4. If none of the explored transformations is better than the current one, we reduce the translation step. The process is repeated until no better transformation is found i.e. reaching a local minimum. Since our matching measure is biased to smaller scales of the initial gaze points, the minimum scale is set to the average scale of the gaze patterns. To improve the search efficiency, we set the scale and the location of the initial gaze points to the average scale and location of the template gaze patterns.

2.2.2 Mixture model

To find the best mapping, this method models the fixations of the template gaze patterns $\mathbb{P}$ by a Gaussian mixture and transforms the initial gaze points to maximize the probability density function of the transformed points. While looking at a stimulus, viewers tend to fixate on some regions. The concept is that the means of the mixture model components are fit to represent the fixation centers while the covariance matrices represent the size of the fixations. Specifically, the method searches for a mapping $\bar{\phi} \in \Phi$ so that:

$$\bar{\phi} = \arg\max_{\phi} \sum_{j=1}^{S} pdf(\phi(p_j)), \tag{5}$$

where:

$$pdf(p) = \sum_{k=1}^{K} \pi_k \mathcal{N}(p|\mu_k, \Sigma_k). \tag{6}$$

K is the number of model components, π_k is the mixing coefficient of the k_{th} Gaussian component $\mathcal{N}(p|\mu_k, \Sigma_k)$ with μ_k mean and Σ_k covariance matrix. Finding $\bar{\phi}$ is done by the same greedy method described in 2.2.1.

3. Experimental results

In this section, we describe the experimental setup and the data used to evaluate the performance of our method. The first ten images of the eye tracking dataset of Judd et al. [12] are used as stimuli (Figure 4). The dataset has the advantage of containing the eye tracking data of 15 subjects for 1003 images collected from Flickr and LabelMe [18]. Hence, we can use this data as template gaze patterns. The dataset contains landscape and portrait images. The images

Figure 4. The 10 images used as stimuli in our experiments. The images show landscapes and street views where multiple objects are present in the scene.

have a resolution of 1024×768. The images contain multiple objects and they do not necessarily contain faces or objects centered in the middle of the image, which represents a realistic stimuli set.

For obtaining the ground truth, the Tobii T60XL gaze estimator [16] is used. It uses four infra red diodes mounted at the bottom of a 24 inch display with a resolution of 1920×1200 pixels. The reported accuracy of the gaze estimator is less than 1°.

The aim of the scene setup is to allow the subjects to look at the stimuli without hard constrains e.g. using a chin rest or sitting at a fixed distance from the stimuli. To collect the eye images, a web camera is mounted above the screen to record the subject. The eye image resolution is around 60×30. The coordinates and direction of the camera is unknown with regard to the gaze plane and can change for each new subject. Ten subjects were asked to sit where they wanted but within the allowed range of the Tobii system. The subject's distance from the display ranged from 55 to 75 cm. No chin rest is used in the experiments so the heads of the subjects may move during the experiment.

The subjects were asked to look at each image for three seconds followed by one second of showing a gray image. The recording of each subject is saved and later analyzed to estimate the gaze points. We follow Lu et al. [6] approach to extract the images of the eyes. For each one of the ten stimuli, the first corresponding web camera frame is taken as an input by the landmarker [17] to detect the eye corners. In [3], the eye corners are detected using the OMRON OKAO vision library. To detect the eye corners for the subsequent frames, we apply template matching using the eye corners of the first frame (for each stimulus) as templates. The eye images are then cropped from the corner and resized to 70×35. Histogram equalization is later applied to alleviate the illumination changes.

3.1. Results on artificially distorted data

Our assumption is that a collection of gaze patterns of individuals can be used to automatically infer the calibration for the gaze estimation of a new user. In this section, we validate the assumption on artificially distorted data. We use the eye tracking dataset in [12] and apply a distortion in the subject fixations. The distorted fixations are considered as a simulation of the initial (uncalibrated) gaze points. For each stimulus, we apply a random translation and scaling to the fixation set of each subject. Then, the methods in Sections 2.2.1 and 2.2.2 are used to transform the distorted gaze points to their correct locations. The first 30 images in the dataset are used in this experiment. For each image, we tested the subjects with 10 or more fixations. We discarded the images where the number of active subjects (10 or more fixations) was less than 6 to ensure sufficient gaze patterns. Using the K-closest points, the mean accuracy across all images is 2.9°, while the accuracy is 4.7° using the mixture model fitting (the scene setup details can be found in [12]). The same procedure is applied on the ground truth gaze points obtained from our collected data. For this dataset, the K-closest points and mixture model fitting obtained accuracies of 2° and 2.7° respectively. The results show the validity of the proposed methods to bring the distorted (uncalibrated) gaze points closer to their correct locations for different sets of template gaze patterns. Regarding the parameter setting, we set K in the K-closest points method to 3 and the number of Gaussian components

to 5. We examined different values of K and components number and the performance difference was not significant.

3.2. Results on the real data

The previous section shows how artificially distorted gaze points can be transformed to their correct locations with sufficient accuracy using the K-closest points. In this section, we use the aforementioned collected data to automatically calibrate the gaze estimator and find the gaze points from the videos acquired from the web camera. We apply the two proposed methods (Sections 2.1.1 and 2.1.2) to find the initial gaze points. For the eye templates method, 25 eye templates were captured while a person was fixating his eyes at 25 points on a 21.5 inch display. This process is followed once for a single (template) subject. So reconstructing an eye image of a new subject from the eye templates will not be ideal due to the changes in eye appearance between the template subject and the other subjects. However, we assume that it still gives a good representation of the topology of the gaze points. As in [6] we divide the eye image into a 5x3 grid and sum up the intensity of the pixel inside each grid cell. The resulting 15 values constitute the feature vector of the eye image.

Regarding the 2D manifold method, a template subject was asked to look at random points on the screen while his face was video recorded. The eye images are cropped and their feature vectors are computed as previously explained. Then, the feature vectors are projected on the first two principal components to constitute a 2D-manifold. The eye images of a new subject (while looking at a stimulus) are cropped, and then the feature vectors are extracted and projected on the same manifold to determine their relative locations. The distances between the initial gaze points are much larger than the actual corresponding gaze points. Yet, this will not affect the results as the initial gaze points will be scaled while finding the mapping to match the initial gaze points with the template gaze patterns.

We select the gaze template patterns in two ways: First, we use the fixation points provided in the eye tracking dataset [12]. Second, the ground truth of our collected data (via the Tobii gaze estimator) is used. In this case, for each subject, we consider the gaze points of the other subjects as template gaze patterns. The K-closest points and fitting the mixture model methods are applied to the initial gaze points. Table 1 shows the results.

The results show that the K-closest points method achieves higher accuracy than using the mixture model while 2D manifold outperforms eye templates for both template gaze pattern sets. The best accuracy (4.3°) is obtained using K-closest points and 2D manifold. Table 2 details the results per subject/stimulus. Figure 5 shows the results for the first four images with subject 3.

Regarding the template gaze patterns, the accuracies are similar for both sets with a slight improvement using the gaze patterns from [12] dataset. The template gaze pattern sets were collected in two different experiments on two different groups of subjects. This is interesting as it shows the general similarity of gaze patterns and hence suggests the validity of using them in auto-calibration regardless of the viewers. The gaze estimation accuracies vary for different subjects. The relatively lower accuracies for some subjects might be either due to errors in estimating the initial gaze points, i.e. because of eye appearance variations with the template subject eye templates which leads to incorrect initialization, or because of the gaze behavior of the subjects and its variation with the template gaze patterns.

The stimuli set contains landscapes and street views images, which makes the auto-calibration more challenging than images with clearly salient objects where humans usually focus on. Yet, the reported accuracy (4.3°) and the results in Figure 5 show the validity of our approach.

3.3. Comparison with other methods

We compare out method with other state-of-the-art auto-calibration approaches. The recent work of Chen and Ji [14] uses a single camera with multiple infrared lights to reconstruct the 3D eye model while using the saliency to estimate the angle between the visual and optical axes. The authors reported less than 3° accuracy using five images and five subjects. Clearly, the comparison with this method is not feasible as the authors use different equipment to reconstruct an accurate 3D eye model.

Sugano et al. [3] adopt an appearance-based gaze estimator and use visual saliency for auto-calibration. The authors reported accuracy of 3.5°. However, their experimental setup differ from ours in the following aspects: First, a chin rest is used in [3] to fixate the head during the experiment while the subjects in our experiment do not use any tool to fixate their heads. Second, the authors in [3] ask the subjects to look at a number of 30-second videos for training (5-20 videos), while in our method the subject needs to look at a single image for 3 seconds. Images contain less cues than videos in which moving objects attract the viewers attention. However, experimenting on still images is more natural and requiring motion in the scene limits the applicability of the gaze estimator. Finally, Sugano et al. analyze the performance variations with respect to different number of training videos. When training on 5 videos (each lasts 30 seconds), the average accuracy is about 5.2° (the exact accuracy is not reported as the results are plotted on a graph). While our method achieves an average accuracy of 4.3° by looking at a single image for 3 seconds.

4. Discussion

Although our method cannot obtain the accuracy of dedicated hardware or calibrated setup, it still provides sufficient

Table 1. Accuracies over different methods and template gaze pattern sets. KCP denotes K-closest points method, GMM refers to Gaussian mixture model fitting. The best accuracy is yielded using 2D manifold and K-closest points.

	Template Gaze Patterns from [12]		Template Gaze Patterns from our Data	
	KCP	GMM	KCP	GMM
Eye Templates	4.7°	5.1°	5.0°	5.3°
2D Manifold	**4.3**$^\circ$	4.9°	4.6°	4.9°

Table 2. Accuracies of the gaze estimation auto-calibrated using K-closest points and 2D manifold. The accuracies are shown per subject/stimulus.

	Stim. 1	**Stim. 2**	**Stim. 3**	**Stim. 4**	**Stim. 5**	**Stim. 6**	**Stim. 7**	**Stim. 8**	**Stim. 9**	**Stim. 10**	**Average**
Subject 1	5.6°	3.1°	2.4°	2.9°	7.2°	5.2°	4.4°	6.9°	6.6°	4.7°	4.9°
Subject 2	4.5°	2.1°	3.5°	2.2°	4.2°	3.5°	4.3°	6.2°	5.8°	5.0°	4.1°
Subject 3	4.7°	2.8°	1.8°	2.3°	3.6°	3.6°	3.2°	5.1°	5.2°	6.9°	3.9°
Subject 4	4.9°	2.3°	2.0°	2.7°	2.3°	2.2°	3.7°	6.5°	5.4°	6.9°	3.9°
Subject 5	3.6°	3.0°	3.5°	5.2°	5.2°	5.3°	4.9°	5.7°	5.2°	4.3°	4.6°
Subject 6	4.2°	3.3°	1.3°	2.9°	3.3°	3.4°	4.4°	5.3°	6.3°	6.0°	4.0°
Subject 7	4.7°	3.6°	3.0°	3.1°	3.5°	4.7°	5.2°	6.4°	7.8°	6.3°	4.8°
Subject 8	3.6°	3.0°	3.5°	5.2°	5.2°	5.3°	4.9°	5.7°	5.2°	4.3°	4.6°
Subject 9	4.1°	2.5°	2.2°	3.8°	4.4°	3.6°	4.9°	6.5°	5.8°	4.4°	4.2°
Subject 10	4.3°	3.2°	3.8°	4.2°	3.4°	4.8°	4.6°	6.1°	6.7°	4.9°	4.6°

accuracy to predict the areas of attention. This is especially important for tasks where gaze estimation is required with no active participation from the user and using off-the-shelf hardware. In this work, we try to simulate a flexible setup and use low-cost publicly available web cameras. There is a trend nowadays to use eye gaze estimation for electronic consumer relationship marketing which aims to employ information technology to understand and fulfill consumers needs. These applications usually collect the data passively without user active participation. Our method is suitable for such applications. Tracing consumers attention when shopping in malls or when exploring advertisements on their laptops are examples of use.

The presented method still has a couple of limitations. Significant head movements are not addressed here. Practical gaze estimators should be invariant to such head pose changes. The method assumes that the template gaze patterns are already available which might not be always the case. Our future research work is to make use of the initial gaze points of the subsequent subjects to gradually auto-calibrate the gaze estimator and to combine the saliency information with the template gaze patterns.

5. Conclusion

We presented a novel method to auto-calibrate gaze estimators in an uncalibrated setup. Based on the observation that humans produce similar gaze patterns when looking at a stimulus, we use the gaze patterns of individuals to estimate the gaze points for new viewers without active calibration.

The proposed method was tested in a flexible setup using a web camera without a chin rest. To estimate the gaze points, the viewer needs to look at an image for only 3 seconds without any explicit participation in the calibration. Evaluated on 10 subjects and 10 images showing landscapes and street views, the proposed method achieves an accuracy of 4.3°. To the best of our knowledge, this is the first work to use human gaze patterns in order to auto-calibrate gaze estimators.

Acknowledgment

This research is supported by the Dutch national program COMMIT.

References

[1] K. Smith, S.O. Ba, J. Odobez, and D. Gatica-Perez. Tracking the visual focus of attention for a varying number of wandering people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 30 Issue 7, p. 1212-1229, 2008.

[2] P. Majaranta and K.-J. Rih. Twenty years of eye typing: systems and design issues. *Symposium on Eye Tracking Research and Applications (ETRA)*, p. 15-22 , 2002.

[3] Y. Sugano, Y. Matsushita, and Y. Sato. Appearance-based gaze estimation using visual saliency. *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, Volume 35 Issue 2, p. 329-341, 2013.

[4] Y. Sugano, Y. Matsushita, and Y.Sato. Calibration-free gaze sensing using saliency maps. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2667-2674, 2010.

[5] D.W. Hansen, J.P. Hansen, M. Nielsen, A.S. Johansen, and M.B. Stegmann. Eye typing using Markov and active appearance models. *Sixth IEEE Workshop on Applications of Computer Vision*, p. 132-136 , 2002.

[6] L. Feng, Y. Sugano, T. Okabe, and Y. Sato. Inferring human gaze from appearance via adaptive linear regression. *IEEE International Conference on Computer Vision (ICCV)*, p. 153-160, 2011.

[7] R. Valenti and T. Gevers. Accurate eye center location through invariant isocentric patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 34 Issue 9, p. 1785-1798, 2012.

[8] A. Villanueva, R. Cabeza, and S. Porta. Eye tracking: pupil orientation geometrical modeling. *Image and Vision Computing*, Volume 24 Issue 7, p. 663679, 2006.

[9] E.D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, Volume 53 Issue 6, p. 1124-1133, 2006.

[10] E. D. Guestrin and M. Eizenman. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. *Symposium on Eye Tracking Research and Applications (ETRA)*, p. 267-274, 2008.

[11] D.W Hansen and Q Ji. In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 32 Issue 3, p. 478-500, 2010.

[12] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. *IEEE International Conference on Computer Vision (ICCV)*, p. 2106-2113, 2009.

[13] K. Tan, D. Kriegman, and N. Ahuja. Appearance-based eye gaze estimation. *Applications of Computer Vision*, p. 191-195, 2002.

[14] J. Chen and Q. Ji. Probabilistic gaze estimation without active personal calibration. *IEEE International Conference on Computer Vision (ICCV)*, p. 609-616, 2011.

[15] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, Volume 229, p. 2323-2326, 2000.

[16] Tobii Technology: http://www.tobii.com/.

[17] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2879-2886, 2012.

[18] B. Russell, A. Torralba, K. Murphy, and W. Freeman. Labelme: a database and web-based tool for image annotation.. *MIT AI Lab Memo AIM-2005-025, MIT CSAIL*, 2005.

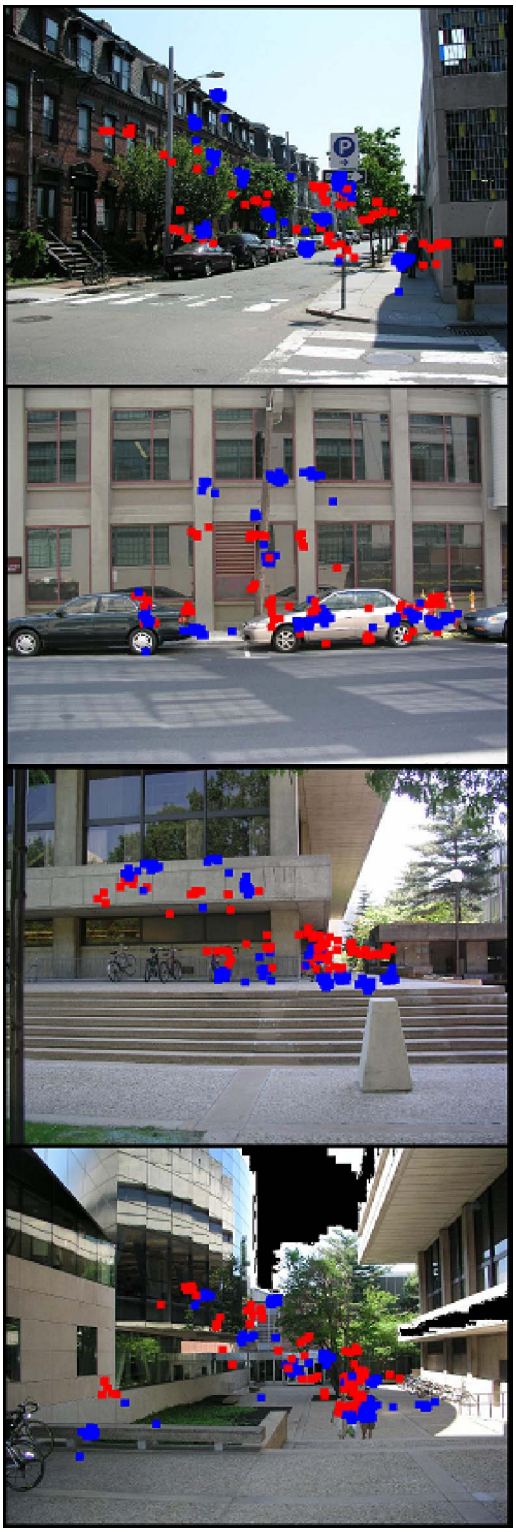

Figure 5. Gaze estimation results for the first four images with subject 3. The red traces represent the estimated gaze points while the blue traces represent the ground truth obtained from the Tobii gaze estimator. The results are achieved using 2D-manifold and K-closest points.