Robust Glint Detection through Homography Normalization

Dan Witzner Hansen*
IT University of Copenhagen

Lars Roholm
IT University of Copenhagen

Iván García Ferreiros
IT University of Copenhagen

Abstract

A novel normalization principle for robust glint detection is presented. The method is based on geometric properties of corneal reflections and allows for simple and effective detection of glints even in the presence of several spurious and identically appearing reflections. The method is tested on both simulated and data obtained from web cameras. The proposed method is a possible direction towards making eye trackers more robust to challenging scenarios.

CR Categories:

I.4.9 [Computer Vision]: Applications I.4.4 [Pattern recognition]: Applications—Computer vision I.4.8 [Computer Vision]: Scene analysis—Tracking

Keywords: Eye tracking, Gaze, cross ratio, homography normalization corneal reflections, mobility

1 Introduction

This paper addresses the generic problem on how to make eye trackers robust in mobile situations where larger variations of light conditions occur and where flexible setups are desirable. The paper is particularly focusing on a robust method for detecting corneal reflections through geometric constraints. We show that there exist well-founded and efficient methods for localizing glints based on their relative positions.

Eye trackers are currently useful in somewhat constrained settings but most eye trackers have problems when used in situations where light conditions change. Judging on the capabilities of current eye trackers, it seems there is a lack of strong methods to handle these challenges. The reasons for eye tracking failures can largely be attributed to glint detection. Corneal reflections are the cornerstones in making remote eye trackers robust to head movements but they also constitute a point of failure when eye trackers are used in mobile and other challenging scenarios where the glint intensities may become weak, fluctuate or too many reflections appear in the images. Several eye trackers use active or multiple IR light sources to reduce the effects of light fluctuations, yet there are very few theoretical and practical results that allow eye trackers to be used in challenging light conditions. An obvious suggestion would be to avoid the dependency on the light source, but so far there are only limited results that make this a convincing direction [Hansen and Ji 2010]. Relying on corneal reflections may even cause problems in constrained settings as the reflections may fall off the corneal surface and either disappear or severely distort the shape, intensity and position relative to other system reflections. Distortion or loss of corneal reflections occur when the reflection is near the boundary between the cornea and the sclera or on the sclera itself. The distortion of the reflections are due to the different curvatures of the sclera and the cornea, while the rougher surface of the sclera can cause valid image reflections to disappear or spurious reflections to appear. The appearance (shape and intensity) of glints change as a function of the ambient light, the relative orientation of the cornea, camera and light source and luminance of the light source as well as where on the eye the light is reflected. The appearance of the reflections may change quite a bit during a session. Corneal reflections are often obtained through some function of intensities, shape of the individual reflections and relative position of reflections and the pupil [Hansen and Ji 2010]. Too many glints are found when the glint-classifier is lenient and vise versa when the classifier is too strict. For example setting thresholds on intensities too high can remove valid glints while setting it too low means too many reflections are detected. Neither shape or intensity are sufficient classification parameters for general situations. While two light sources are sufficient to determine gaze in fully calibrated settings, it would require more lights to ensure robustness for indoor/outdoor uses e.g. a simple depth change between eye and camera or a head rotation will also change the relative positions of the glints on the cornea. Patterns of higher cardinality may be used and in some cases even modulation of the lights have shown useful [Hennessy and Lawrence 2009]. Modulation of light sources are facing challenges since the light sources with regular intervals disappear and requires careful engineering to synchronize with the cameras. It is therefore cheaper and reliable that the light sources remains turned on and perhaps use multiple light sources to ensure robustness. Many reflections are in this respect preferred over too few since potential glints that have been removed can be hard to recover in the subsequent processes.

Related work is presented in section 2. Section 3.1 introduces the method of *glint normalization* and its geometric basis. Section 4 presents the detection results on both simulated data and on data obtained from a web camera. Section 6 concludes the paper with future perspectives on the method.

2 Previous work

Reflections on the corneal surface are important in making gaze estimation robust to head movements [Hansen and Ji 2010]. If eye trackers are to become robust (such as in outdoor settings) it is with current gaze estimation models imperative that reflection detection is also robust to variable working conditions. While only a few methods addressing glint detection beyond constrained indoor use, [Hennessy and Lawrence 2009; Li et al. 2007] present nice methods for glint detection using multiple light sources and geometric constraints. The method of [Hennessy and Lawrence 2009] compensates for (1) translation and some distortion, (2) few spurious reflections and (3) deletion of system glints. However, for proper operation, the light sources must be placed such that at least two valid corneal reflections are will always visible to the camera. The method also requires a unique displacements between all pairs of reflections. The method does not explicitly handle rotation or

---

*e-mail: witzner@itu.dk

changes in scale between the reference and image point patterns, but good results were presented. Good results were also presented by [Li et al. 2007]. The method uses a regular pattern of 9 light sources and matches glints located within specific squares placed in the grid. The method is capable of tracking glints when the size of the pattern remains relatively fixed. The method allows for glint detection even when the glints are located outside the corneal region.

3 Method

This section describes and formalizes an effective and generalizable method for robust glint detection for to both remote and head mounted eye trackers. The method uses a formal geometric correction scheme to remove the effect of perspective projection and head pose changes using homography normalization [Hansen et al. 2010].

The theory is derived through point light sources, but due to the mathematical basis the results can be extended to variable setups (such as lines and conics) of light sources without altering the theory. Line and conic patterns may even increase robustness even further, though not treated separately in this paper.

System light sources, L_i, form a pattern C_L that, in turn, may generate a pattern C_g of glints, g_i on the cornea. Denote a pattern with n light sources *n-pattern*. Glint detection will in the following refer to a method of detecting possible glints in the image while the purpose of the *pattern detection* is to determine the set of glints that correspond to system light sources. In some cases (e.g. homebrew eye trackers where the user is placing the light sources as most convenient) the precise shape of C_L is a priori unknown. In the following sections we assume the pattern C_L to be known, however section 5 shows that homography normalization can also be used when learning an unknown configuration C_n^* from image observations. In general settings (see figure 1) where spurious reflections are present, the main and challenging task for the pattern detection is to determine C_g from the set of all reflections $\mathcal{P}$ measurable in the image. The cardinality of $\mathcal{P}$ is usually low in indoor environments. The problem becomes significantly harder as the cardinality of $\mathcal{P}$ increases e.g. in outdoor scenarios.

3.1 Glint normalization

The general setup of an gaze tracker is shown in figure 1. The cornea is approximately spherical with a radius, R_c, about 7.8mm. It reflects light similarly to a convex mirror and has a focal point, f_c, located halfway between the corneal surface and the center of corneal curvature ($f_c = \frac{R_c}{2} \approx 3.9$ mm). Reflections on the cornea consequently appear further away than the corneal surface (a.k.a virtual reflections). In this paper the light sources, L_i, span a plane Π_l which does not necessarily coincide with the camera or screen planes.

For the sake of simplicity and without loss of generality, assume $(g_1^c \dots g_4^c)$ come from point light sources. With four light source there exist a plane Π_c (in fact a family of planes related by homographies) spanned by their reflections on the corneal surface. This plane is denoted the *normalized space* and is located close to f_c when L_i at infinity. When considering the optical laws, the corneal reflections can through physical measurements of common setups (e.g. working distance of about $50 - 60$ cm) and anthropomorphic averages be calculated to be planar up to the fourth decimal. These results were obtained through simulated experiments of an eye located in different positions in 3D space [Böhme et al. 2008] and can be generally applied in similar methods (e.g. cross ratio-based gaze estimation methods) to argue that the reflections are planar.

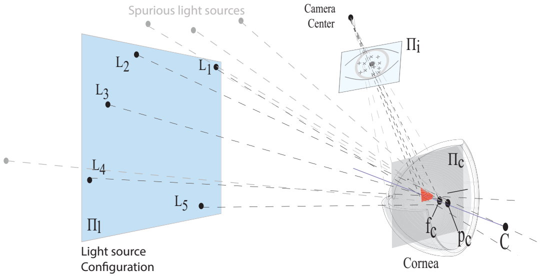

Figure 1: *Geometric setup of camera, cornea and light sources*

The quadrilateral of glints $(g_1^c \dots g_4^c)$ (denoted *reference glints*) on the cornea is consequently related to the corresponding quadrilateral $(g_1^i \dots g_4^i)$ in the image and a quadrilateral of light sources L_i e.g. through homographies H_c^i and H_c^s respectively. The homography from the image to the light sources $H_i^l = H_c^l \circ H_i^c$ via Π_c will therefore exist as long as the reflections appear on the cornea. The reflections of light sources on the corneal surface will remain planar but distorted. Due to the planarity constraint it is possible to largely remove distortions and recover C_L by mapping the pattern of a quadrilateral from the image to the normalized space through H_i^n. The normalized space may coincide with C_L, but in principle any space up to a homography to C_l may define the pattern in normalized space, C_n. Homographic mappings require at least 4 point-correspondences to be valid. Therefore, at least one additional light source is needed to ensure that a pattern can be detected (up to a four fold ambiguity). The ambiguity is removed when light sources have an ordering e.g. knowing that the top left glint in the image correspond to a light source that is also to the left and above the other light sources.

The basic principle of the proposed glint detection method (see Figure 2) is to use each combination of quadrilateral (4-pattern) of glints as reference for the mapping, $\bar{H}_I^N$ from the image to normalized space. The mapping ensures that $p_n^i = H_I^N g_i$. When all or sufficiently many p_n^i match the system pattern there will be a high probability that these correspond to system light sources C_L.

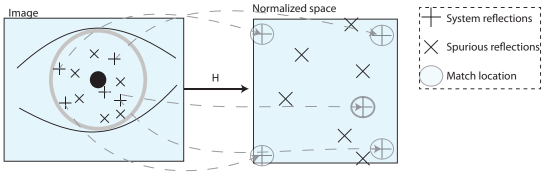

Figure 2: *The principle of glint normalization*

Detecting g_5 in normalized space may be influenced by noise and modeling errors. The location of g_5 may therefore in practice be within a small region in normalized space. That is, the point p_n^k is considered to be a system light source when it is located within a confidence region ρ of g_5. The parameter should be as small as possible to minimize the number of false positives while maintaining a high classification rate. Setting ρ too small may result in the glint not being detected while setting ρ too large will introduce false positives.

Adding robustness and downgraded strategies Homography normalization trivially extends to general n-patterns ($n > 5$) by simply matching additional reflections to the corresponding confidence regions in normalized space. Rotated variants of the 5-patterns could be used for making robust patterns of cardinality up to 8. Each additional light source provides extra information to be used for verification but it also increases the chance that only a subset of the reflections appear simultaneously e.g. when squinting and at various head and eye orientations. Obviously, partial matching may be possible when sufficiently many reflections are used in C_L. The only requirement is that each sub-pattern is unique in normalized space. Multiple light sources could therefore extend the range for which the system can be used simply by performing partial matches of the main pattern. The subsequent sections are developed around 5-patterns but extensions will also be discussed. Specific homographies (such as affine and similarity transformations) may become useful in cases where fewer system light sources are used or when a backup strategy is needed when too few glints appear on the corneal surface. The general normalization principle is not changed with downgraded strategies except using more constrained mappings than homographies. With 4-patterns, the underlying model would be affine and with 3-patterns only translations are modeled. In case the correct 5-patters has not been found through the normalization procedure, the method iteratively investigates configurations of lower cardinality. That is, by first investigating 4-patters using an affine mapping for the normalization, and if no configurations are found then 3-configurations using a similarity transformation for the normalization.

3.2 Relating normalization similar methods

Glint normalization allows for a clear characterization of the underlying assumptions and tells to which degree the pattern may vary to ensure reliable detection. For example the constraints used in [Hennessy and Lawrence 2009; Li et al. 2007] are directly explained through normalization. In Li et al. [2007] the pattern would be of cardinality 9 and placed in a regular grid. The constraints employed could be encoded in a prior by filtering the glints based on their location relative to the limbus size and location and by fixing the spatial size of the pattern before normalization. The normalization homography would be a rigid transformation. Deviations due to e.g. head movements in depths are handled through a parameter similar to the precision parameter ρ. However ρ should possibly be set larger in Li et al. [2007] as to account for scale changes.

4 Experiments with pattern detection

This section presents the results of glint detection using the pattern, C^*, shown in figure 4. The choice of C^* is not arbitrary but arguments are left out due to space constraints. The experiments were conducted on both simulated data and real data obtained from a web camera. The simulated data is generated through the open source eye tracker simulator [Böhme et al. 2008] while the experiments on real data uses a standard (Sandberg) web camera with night vision. The standard wide angle web camera lens has been replaced with a corresponding 12 mm lens to ensure the camera is placed away from the user. A GPU implementation is used to analyze the video sequences offline as to address computational demands for future high speed cameras.

Detection rates can possibly improved through stronger priors e.g. using the previous location, shape and size of the reflections, and eye location. Adding more constraints would also constrain possible situations in which the eye tracker could be used e.g. less flexible. In this paper we have purposefully chosen to use weak prior models as to investigate the properties of the normalization procedure rather than biasing the results with strong priors. Reflections are identified using connected components on binarized images. Intensity thresholds have been kept relatively low as to retrieve even weak reflections but obviously large reflections (e.g. on glasses) are removed.

Detection results on simulated data. In the simulated experiments, the eye is placed in a $20 \times 20 \times 20$ grid. The camera and light sources are located 50 cm from the eye. n_s spurious light sources obtained by uniform sampling of the 12×12 cm region around the camera. Each experiment is repeated 100 times to provide empirical means and variances. The simulated data allows the analysis to be concluded as if glint detection is made in scenarios where shape and intensities cannot be used to distinguish glints reliably e.g. when used outdoor. The results for the mean number of detections as a function of the number of spurious light sources on simulated data is shown in figure 3.

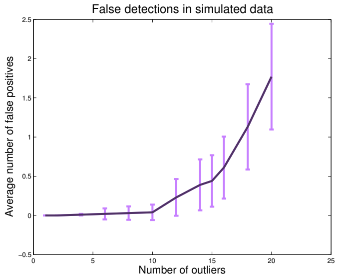

Figure 3: *Number of detected patterns as a function of spurious light sources in simulated data.*

The results show that on average the proposed glint normalization technique works well on discriminating the correct glints up to about 10 spurious identically appearing reflections. Higher detection rates should be expected when additional constraints, such as size of the pattern, are applied. In a few cases (less than 1% of the cases) the method detected two feasible solutions but there were no missed detections in any of the experiments. The mean number of the detected patterns is still close to one even though the variance increases. This makes glint normalization very robust on simulated data.

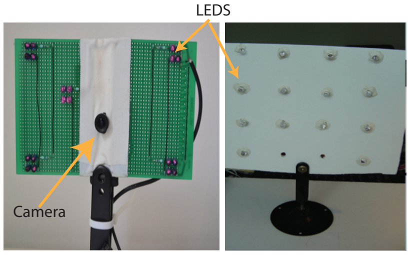

Figure 4: *(Left) Camera and light sources (Right) Board with* 18 *light sources for generating spurious reflections in real image sequences.*

Experiments with web camera data On real data obtained from a web camera the spurious reflections (assumed identical to the system light sources) are generated by the board of LED's (see figure 4).

Figure 5 shows the results of glint detection as a function of the number of spurious reflections and figure 6 shows the detection rates on camera data as the precision parameter ρ is changed. The figures shows that ρ should be around 0.05 in normalized space and that the method is robust up until 7 spurious reflections. The results furthermore indicate that it would be possible to use several (about 20×20) unique patterns of light sources in the same setup and still be able to discern these. However, using the same precision parameter ρ for other patterns may theoretically lead to different results.

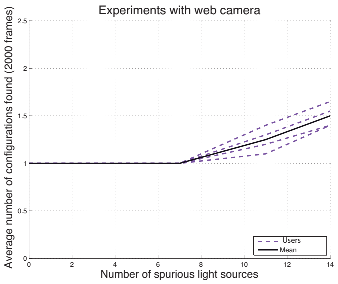

Figure 5: *Performance of glint normalization on video sequences as a function of the number of spurious reflections. The solid line is the mean of the different test cases (dashed lines)*

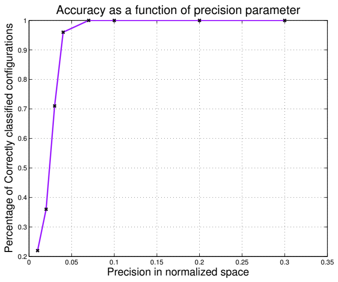

Figure 6: *Detection rates as a function on the precision parameter, ρ using web camera data.*

5 Configuration learning from cornea images

Methods to estimate gaze in uncalibrated calibrated setups exist [Hansen and Ji 2010]. An interesting direction for future eye trackers intended for mobile use is to let the eye trackers be more flexible and allow setups that suit particular needs. In this section we present a method based on the normalization procedure that allows the eye tracker to learn the current configuration, C_g, over time despite spurious random reflection occurring on the cornea. The underlying assumption is that systematic spurious reflections occur less frequently than C_n.

The configuration learning method uses glint normalization to a discretized normalized space. Evidence for each hypothetical configuration is collected by counting the number of times a given configuration appears in the discretized normalized space. The method operates in a similar way as the Hough transform and thus shares some of the same robustness properties.

Figure 7 shows the learning curve for a sequence using a moderate number of spurious reflections (randomly between 3-5 reflections). The figure show the percentage of votes that are located in the correct configuration as a function of the number of frames. In this case the learning method quite rapidly converges to the correct configuration.

6 Discussion

The paper addresses the problem of how eye trackers can be made robust and eventually used for mobile scenarios. The specific problem being addressed was mainly on how to ensure glint detection in the presence of noisy and spurious measurements. A novel normalization method based on well-founded geometric properties of corneal reflections was suggested. The same principle was used to learn a unknown configuration of light sources. Experiments on both simulated data and web camera data were presented. The method was effective up to around 10 spurious reflections without using other constraints than the pattern. With an effective pattern filtering method it is possible to relax the criteria for detecting the glints (e.g. lowering the threshold values of intensity and shape parameters) and thus make eye tracking robust in more challenging scenarios such as for mobility.

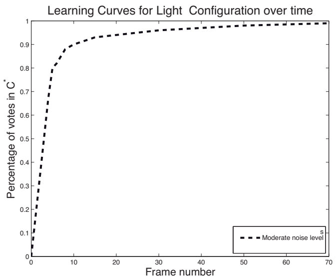

Figure 7: *Examples of learning rates using an isotropic confidence region* $\rho = 0.05$ *on web camera data*

The principle of glint normalization encompass the method presented in [Hennessy and Lawrence 2009], but with several added properties. Some of the benefits with glint normalization are that rotations and scalings (through the general perspectivity) of the light pattern are handled implicitly. The normalization principle may be used for both uncalibrated and calibrated setups, even though the full homographic model is not strictly needed for gaze estimation in calibrated setups [Hansen and Ji 2010].

The method is not restricted to point light sources but can trivially be extended to other planar shapes e.g. lines, rectangles, circles and combinations of these without altering the general principle. Shape patterns may in fact be easier to detect than points patterns since shapes may be detected even if parts of it disappears. Glint normalization and related methods are currently being pursued further to challenge the problem of eye tracking in mobile scenarios. Mobile eye tracking systems that do not require models that fundamentally differ from established eye and gaze tracking methodology indeed look viable.

References

BÖHME, M., DORR, M., GRAW, M., MARTINETZ, T., AND BARTH, E. 2008. A software framework for simulating eye trackers. In *Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2008, Savannah, Georgia, USA, March 26-28, 2008.*

HANSEN, D. W., AND JI, Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 3, 478–500.

HANSEN, D. W., AGUSTIN, J. S., AND VILLANUEVA, A. 2010. Homography normalization for robust gaze estimation in uncalibrated setups. In *ETRA*, 13–20.

HENNESSY, C., AND LAWRENCE, P. 2009. Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking. In *IEEE Transactions on Biomedical Engineering.*, vol. 56, 1891–1900.

LI, F., KOLAKOWSKI, S. M., AND PELZ, J. B. 2007. Using structured illumination to enhance video-based eye tracking. In *International Conference on Image Analysis*, 373–376.