

# Gaze Estimation for Near-Eye Display Based on Fusion of Starburst Algorithm and FERN Natural Features

Peter Bazanov, Toni Järvenpää  
Nokia Research Center  
Finland  
{peter.bazanov, toni.jarvenpaa}@nokia.com

## Abstract

In future, near-eye display devices will be commonly used by people with or without prescription glasses, with different eye topologies, and under different lighting variations. Gaze tracking is an important task for these near-eye devices. New robust gaze tracking methods should be explored for these sophisticated devices.

The proposed method overcomes the problem of adaptability to eye topology and occluded objects by employing advantages of natural FERN features tracking, strong model-based Viola Jones object detector and a fusion of these algorithms with the Starburst algorithm.

**Index Terms:** Gaze tracking, STARBURST, FERN, natural feature tracking, Viola Jones detection.

## I. INTRODUCTION

Eye tracking is the localization process of the eye object on the image and point tracking from frame to frame. Eye tracking usually is the primary task for the gaze estimation problem that is concerned in measurement either the point of gaze or the motion of an eye relative to the head. Eye tracking and gaze estimation plays an important task in the great variety of applications like human computer interaction, eye control of the computer, near eye solutions, virtual reality, augmented reality, web usability, advertising, marketing, automotive navigation, cognitive studies, improved image and video communications, face recognition.

Robustness of eye tracking technique to lighting variation, race variance, and occlusion is a challenging task. In this paper we investigate pupil detection algorithm and gaze estimation problem on challenging variations in eye databases. The primary variations are prescription glasses, secondary lighting variation, next is eye race/topology variation. We present a fusion of the STARBURST with Viola Jones object detector to achieve robustness in accuracy under high eye variation databases. Second we enhance new adaptive gaze tracking algorithm that is based on fusion of STARBURST with natural feature tracking. This adaptive approach is inspired by SLAM based approaches [1]. The proposed fusion provides robustness of pupil detection, reflection point detection and adaptability to the eye form, variance due to eye activity, pupil occlusion by the eye lids, and racial difference.

### 1.1 Problem statement

In this paper we consider the gaze estimation mainly in the infrared spectrum, even though the visible spectrum is not fully excluded. Infrared light is typically used in gaze tracking for controlling the illumination, among other things. Both infrared and visible spectrum captures ambient light reflected from the eye. We consider cases where we have ambient light from the environment and from a near-eye display. In addition, we have infrared illumination that provides us corneal reflection glints. However corneal reflections features have no stable glint form, and sometimes glints are not presented.

General camera based gaze tracking problem can be defined as by a given image, calibrated model should result in detection of the gaze direction on the scene location to where a user is fixating his gaze. The solution of the gaze estimation problem could be founded by image based and feature based methods.

In image-based approaches whole image is usually analyzed based on a preliminary taught model and gaze estimation is based on uncovering hidden dependencies in image data to the template. In feature based approaches the algorithm starts usually from the detection of the pupil position, corneal reflections then by calibrated data the algorithm should estimate the gaze position with respect to the location in space of the light source that creates the corneal reflection.

For the objective comparison we investigate the gaze estimation according the marked gaze pose [6, 7]. We have defined Euclid metrics for the difference between algorithm and marked point and defined the criterion of the quality for gaze estimation. For the tests we define the quality metrics based on calculation of average, standard deviation for the difference between marked gaze point and test gaze point. This criterion is used for comparison between “Starburst” version and fusion version based on FERN features.

According to some assumptions general gaze estimation problem can be considered as localization problem of the principal eye parts and major image characteristics such as pupil contour and corneal reflection. Using these robust points gaze can be estimated. The considered metrics are the same average, std. deviation for the difference between the localized points and marked points. We used such metrics to compare STARBURST with a fusion of STARBURST and Viola Jones.

The quality of accuracy and robustness due to eye topology and glasses should be estimated on eye marked database gathered from different users.

### 1.2 Related work

In [1] we have already estimated the gaze estimation algorithm adaptive to occlusion based on FERN features. In current work we investigate modification of this algorithm for complex sessions with glass occlusions, and variations in eye database sequences.

Eye localization algorithms are usually differentiated on feature based and image based approaches. Feature based approach is oriented to find robust invariant image features and then utilize information about the eye structure. Commonly, image intensity levels and gradients are used.

One of the best feature based approach is “Starburst”. There are also model-based enhancements of the “Starburst” - eye major model features localization, like model of pupil, or eye contour. An iterative optimization technique is applied to find accurate solution, like Active Shape Models, Snakes, or circle and ellipse fitting. However, this technique is computationally intensive and is used to provide higher accuracy than the

preliminary detection process [2-5]. Starburst algorithm is efficient in speed. However its drawbacks are in the reduced feature stability and model fitting. As an example, one implementation of the starburst algorithm is fairly stable for a gaze tracker prototype where the eye-camera is located close to the visual axis and illumination is strictly organized [6]. In a more challenging camera-geometry (camera located in an extreme angle) and illumination (not uniform, ambient illumination present) setting however, the pupil detection quality as such was not really acceptable anymore [7].

Some stable solutions can be found in iris recognition technique. Iris descriptor from recognition process can be also implemented for the simple localization process. That solution can provide calibration and adaptation for the concrete user's eye topology and guarantee some stability for the algorithm. Such kind of approach is provided by Daugman iris descriptor [13, 14]. This approach can overcome the constraints of general eye model distribution but it will be model based approach rather than feature based approach. There will be some constraints in user feature topology variation, even for the same user.

For the outdoor environment with non uniform ambient lighting, for different eyes variation, eye activity like twinkling and partial occlusion, more robust technique is needed. In this paper we propose adaptive feature based approach that utilizes natural feature tracking on FERN for gaze estimation. That kind of technique is usually useful for markerless tracking and SLAM. It can be applied without knowing any kind of information about the object [9-12]. We have also utilized Viola Jones detector for stable general model based technique of eye detection even in case of occlusion [8].

## II. MAIN PART

### 2.1. Starburst algorithm

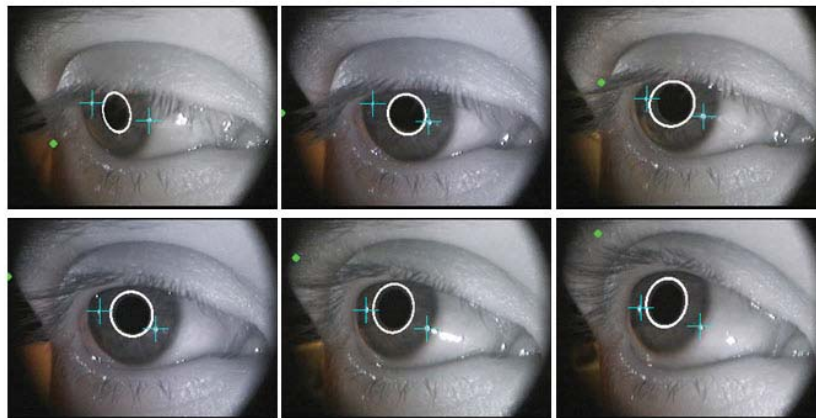
The main goal of the starburst algorithm is to extract pupil and corneal reflection points. The algorithm's major steps are as follows: image noise reduction, corneal reflection detection, localization and removal of corneal reflection, edge point's localization, pupil contour detection, ellipse fitting, model based optimization, calibration [2]. The key idea of the algorithm is to use limited number of rays from the center to extract pupil contour. That allows to apply the detection algorithm for only the part of whole image and provides optimization. Meanwhile edge based approaches are usually applied for the whole image. Major iteration steps of the algorithm are as follows:

- 1 Input: Eye image with corneal reflection removed, Best guess of pupil center
- 2 Output: Set of feature points
- 3 Procedure:
- 4 Iterate
- 5 Stage 1:
- 6 Follow rays extending from the starting point
- 7 Calculate intensity derivative at each point
- 8 If derivative > threshold then
- 9 Place feature point
- 10 Halt marching along ray
- 11 Stage 2:
- 12 For each feature point detected in Stage 1
- 13 March along rays returning towards the start point

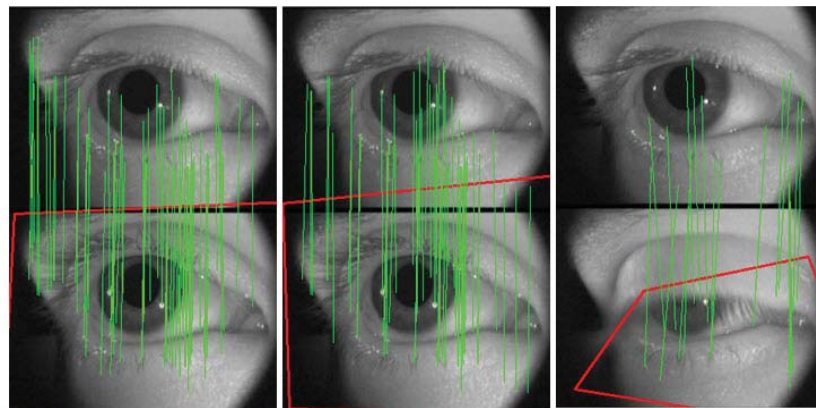
- 14 Calculate intensity derivative at each point
- 15 If derivative > threshold then
- 16 Place feature point
- 17 Halt marching along ray
- 18 Starting point = geometric center of feature points
- 19 Until starting point converges

This Starburst algorithm demonstrates optimization in edges search and benefits in speed [2-5].

2.2 Viola Jones Starburst modification



(a)



(b)



(c)

Fig. 1. Pupil contour detection, corneal reflection detection, gaze estimation. (a) Natural FERN features extracted from the eye and tracking, by red it is defined homography plane. (b) Working mobile NED and integrated gaze tracker prototype.(c)

The first Starburst problem is recognition of “eye class” , it is challenging problem for partly occluded eyes where it produces edge points that are not concerned with pupil. The second problem is that we have usually two contours of edge points; RANSAC technique cannot estimate correct ellipse in sophisticated cases. Thirdly there is no tracking mode for Starburst that is not utilized information from the last frame

We integrate feature based and strong model based approach. Fig,1 . Feature based approach is done initially to provide major hypothesis about eye pupil. It is a coarse approach that extracts many edge points. These edge points represent two types of pupil and iris contours. Starburst RANSAC could not distinguish the nature of these edges. By application of Viola Jones algorithm we achieve robustness in removal of the outliers.

Our solution has the following steps. Fig.2:

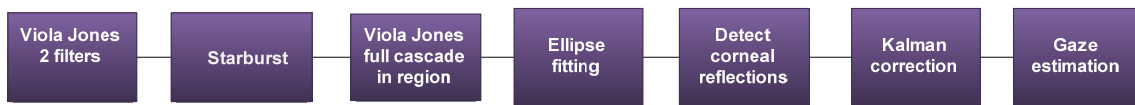


Fig.2. Fusion of Viola Jones and Starburst

- 1) Attentive detector. Select the start point for STARBURST algorithm. We have applied Viola Jones detector based on two filters (horizontal and vertical) to solve sophisticated cases like partial occluded eyes for “eye class” verification.
- 2) Starburst Edge Points localization.
- 3) Apply Viola Jones cascade in the region of interest and search through the image pyramid for the stable point with maximum score response. Founded points can be pupil, pupil corners or some other eye parts according to the marking data. Find the pupil center based on stable point and divide pupil contour points from iris contour. Remove outliers.
- 4) Starburst Ellipse fitting by RANSAC for only pupil edge points. It is used ellipse fitting by optimization of six parameters of ellipse. Additional enhancements are done by contour approximation using snakes.
- 5) Detect corneal reflection.
- 6) Add Viola Jones stable points and perform Kalman filtration for the tracking mode.
- 7) Estimate gaze tracking.

### 2.3 Tracking optimization mode for Starburst

For the tracking mode we investigate Kalman filter control of the major points coordinates. We apply Kalman filter to control the coordinates from frame to frame. The Kalman filter is applied for pupil, corneal reflection coordinates and ellipse parameters. Kalman filter is also applied to the output to control gaze points. Based on the predicted areas for the points we reduce the search of the major points. In the situation of fast movement when predicted by Kalman parameters differ more than the threshold it is performed new detection.

### 2.4. Adaptive eye tracking algorithm in case of glasses occlusion

This algorithm is based on automatic adaptive eye model construction offline for the concrete user and utilizing it in real time. For the construction of eye template it is usually needed few minutes to train the FERN decision trees [9-11]. Ferns are nonhierarchical structures where each one consists of a set of binary tests and returns the probability that a patch belongs to any one of the classes that have been learned during the training. These



responses are then combined in a Naive Bayesian way. We have taught 9 models for each eye's gaze.

We use 100 feature points for the training model and 200 feature points for test models. In the real time mode we provide corner detection by FAST algorithm [10][11], calculation of FERN descriptor for the points and homography estimation according the trained model and base pose. To achieve more robustness FERN features are applied mostly on internal eye part and some features are eliminated by pre-detected features from Viola Jones or starburst scheme. Fig.3 Algorithm can work also without STARBURST or Viola Jones by utilizing general framework of markerless natural tracking. In our experiments two corneal features are among the FERN features. The model is useful for the gaze estimation in case of glasses occurrence and complex topology or lighting variation when general algorithms of pupil and corneal reflection works unstable (Fig.4).

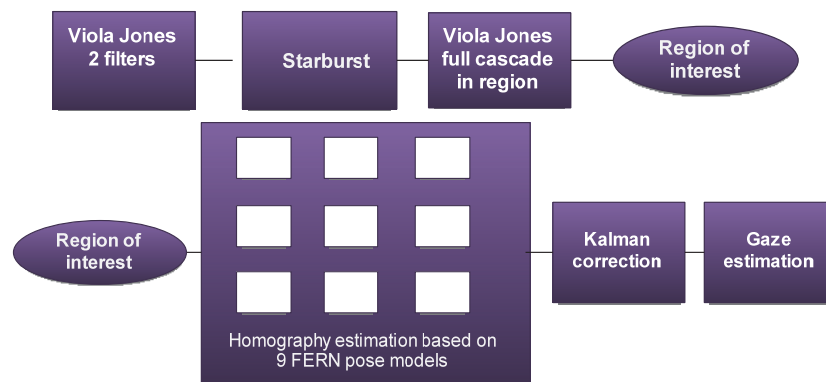


Fig. 3. Fusion of Viola Jones, Startburst, FERN natural features

First step - Application of the Viola Jones and Starburst algorithm, to generate hypothesis of eye location  
 Second step - in ROI provide matching of natural features with the predefined planar object model in right pose

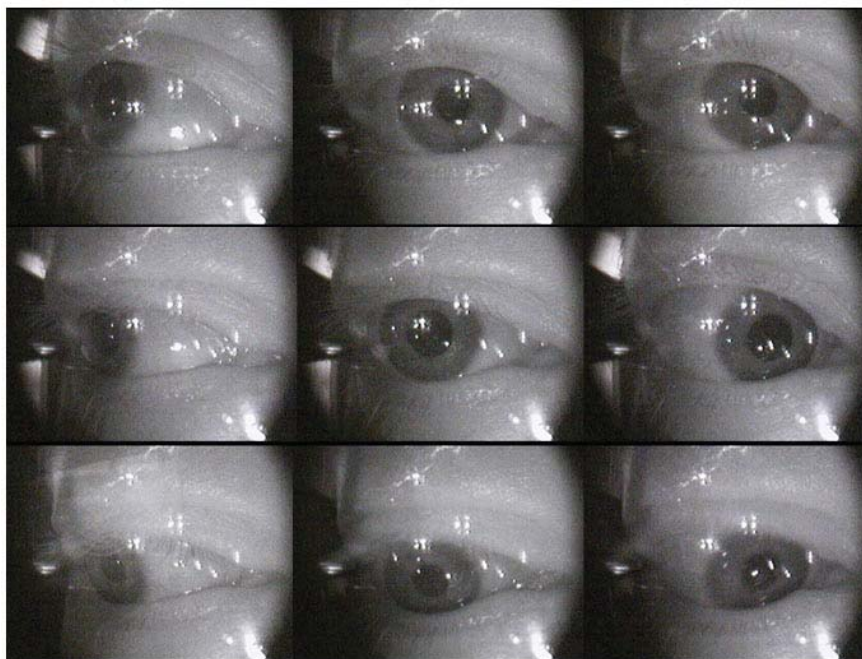


Fig.4. Database with glass occluded eyes – 9 pose to train model

### III. EXPERIMENTS

We have tested fusion version of Starburst algorithm with Viola Jones eye detector. The results are stable especially on images with partial occlusion of the pupil. We have gathered small database of 10 video eye sequences (700 image frames). On each frame there is an eye object with size of 240x140, iris with size of 95x95 and pupil with size of 45x45 pixels that is approximately contour difference is 25 pixels. General error happens in the situation when the algorithm chooses wrong edge contour (iris contour instead of pupil contour). We have set the strong and weak criterion for the algorithm accuracy.

Strong criterion is the max coordinate difference should be less than 10 pixels between marked ellipse and estimated one, otherwise it is considered misdetection. The weak criterion is the max coordinate difference should be less than 20 pixels. We have achieved the correct detection rate is about 96,6% meanwhile the Starburst algorithm results are 74,4% for the strong ellipse criterion fitting, and 88,5% for the weak criterion. Table I. For the glasses cases we achieved correct detection rate is about 60,3 % for the Starburst, 87,2% for the Viola Jones, 95,7% for the adaptive algorithm. Table II. As for the speed Viola Jones detector decreases the performance from 100 fps to 30 fps however the parameters of Viola Jones could be adjusted between speed and accuracy.

Our next model that is fusion with FERN feature tracker have provided high stability even on datasets with high occlusions, eye twinkling, and different eye topology. The algorithm demonstrates higher detection rate, that is about 98,3% and 95,7% for glasses cases while still supporting real time 15 fps on Intel Core 2 Duo P9500 2.5Ghz.

The system has been tested on small databases that are general eye movements for different users that are wearing near eye display. In future we would like to investigate results on large databases with varied eye races, partial occluded and winking, lighting variation database. We would like to estimate speed, accuracy and robustness on the databases with complex eye topology.

TABLE I  
DATABASE OF NORMAL CASES (VARIATIONS IN EYE TOPOLOGY)

Normal DB/criterion $\sigma_{XY} < 10$	STARBURST	STARBURST+ VIOLA	STARBURST+ VIOLA+ FERN
Average Performance (fps)	100	30	15
Detection Rate (%)	74,4%	96,3%	98,3%
Eye Coordinate Accuracy Sigma X (pixels)	7.25	5.7	3.4
Eye Coordinate Accuracy Sigma Y (pixels)	7.9	6.1	4.1

### IV. CONCLUSION

Feature-model based approach STARBURST shows benefits in speed. But the major drawbacks are that algorithm provides fine localization only in case the “eye” class has been already recognized, otherwise the algorithm will provide high false alarm. There are also some drawbacks in poor ellipse model based fitting technique especially when it is measured by strong criterion. Fusion of STARBURST algorithm with stable model-based Viola Jones detector demonstrates robustness due to small degradation in speed, speed

parameters are adjustable. Viola Jones detector can recognize partially occluded pupil by eye lids.

TABLE II  
 DATABASE OF GLASSES CASES (VARIATIONS IN EYE TOPOLOGY + GLASSES TOPOLOGY)

Glasses DB/criterion $\sigma_{XY} < 10$	STARBURST	STARBURST+ VIOLA	STARBURST+ VIOLA+ FERN
Average Performance (fps)	100	30	15
Detection Rate (%)	60,3%	87,2%	95,7%
Eye Coordinate Accuracy Sigma X (pixels)	8.45	6.7	4.8
Eye Coordinate Accuracy Sigma Y (pixels)	9.29	7.1	5.2

The FERN natural features are efficient instrument to construct model of the object and propose fast user enrollment and calibration. Natural features provide adaptive feature tracking on databases with high eye topology complexity. Natural feature tracking provide robustness even when general feature based and model based approaches faced with problems. Natural feature tracking is efficient in case of glass eye occlusions.

For the further exploration we would like also to consider SIFT, SURF methods whether they can be also useful for natural feature tracking on eye objects [10-12]. Glasses detection and lighting correction methods should be investigated in future.

#### REFERENCES

- [1] Peter Bazanov, Toni Jarvenpaa, Martin Schrader “Adaptive object tracking for improved gaze estimation based on fusion of starburst algorithm and natural features tracking, Fruct 9 Proceedings, p.23-27, April 2011
- [2] Dongheng Li; Winfield, D.; Parkhurst, D.J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. Proceeding of the IEEE Vision for Human Computer Interaction workshop at CVPR, 1-8 .
- [3] Dongheng Li; Derrick Parkhurst (2006). Open-Source Software for Real-Time Visible-Spectrum Eye Tracking. In Proceeding of the 2nd conference on Communication by Gaze Interaction – COGAIN 2006.
- [4] Wayne J. Ryan, Damon L. Woodard, Andrew T. Duchowski, and Stan T. Birchfield Adapting Starburst for Elliptical Iris Segmentation, 2008 .
- [5] Bei Yan, Xigong Zhang, Lifeng Gao, Improvement on Pupil Positioning Algorithm in Eye Tracking Technique. 2009.
- [6] Timo Pylvänäinen, Toni Järvenpää, Ville Nummela, Lixin Fan, “Gaze Tracking for Near to Eye Displays,” 18th International Conference on Artificial Reality and Telexistence, 5-11, 2008.
- [7] Toni Järvenpää, Pekka Äyräs, “Highly integrated near-to-eye display and gaze tracker,” Optics, Photonics, and Digital Technologies for Multimedia Applications, Proc. of SPIE Vol. 7723, 2010.
- [8] Paul Viola Michael Jones ”Rapid Object Detection using a Boosted Cascade of Simple features”, 2001.
- [9] Mustafa Ozuysal, Michael Calonder, Vincent Lepetit and Pascal Fua “Fast Keypoint recognition using random Ferns”, 2008.
- [10] Daniel Wagner, Alessandro Mulloni, Dieter Schmalstieg “Real-time detection and tracking for augmented reality on mobile phones”. In Proceedings Visualization and Computer Graphics, 2010 .
- [11] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, Dieter Schmalstieg “Pose Tracking from Natural Features on Mobile Phones” In Proceedings ISMAR 2008.
- [12] H.Bay, T. Tuytelaars, and L.Van Gool. SURF: Speeded Up Robust Features. Proc. European Conf. Computer Vision (ECCV), 2006.
- [13] J. Daugman. How Iris Recognition Works. IEEE Transactions on Circuits and Systems for Video Technology, 14(1):21–30, January 2004.
- [14] J. Daugman. New Methods in Iris Recognition. IEEE Transactionson Systems, Man, and Cybernetics, Part B, 37(5):1167–1175, October 2007.