Adaptive Object Tracking for Improved Gaze Estimation Based on Fusion of Starburst Algorithm and Natural Features Tracking

Peter Bazanov, Toni Jarvenpaa, Martin Schrader
Nokia Research Center
{peter.bazanov, toni.jarvenpaa, martin.schrader }@nokia.com

Abstract

Gaze tracking is an important task for near-eye devices. One of the sophisticated problems in gaze tracking is tracker robustness due to high eye topology variation, lighting variations, and partial occlusions of the eye.

The proposed method overcomes this problem by employing advantages of adaptive feature based approach based on natural FERN features, strong model-based Viola Jones object detector, and fusion with Starburst algorithm.

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

I. Introduction

Robust gaze tracking is the primary task for 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 challenging task. We present fusion of STARBURST with Viola Jones object detector to achieve robustness in accuracy under high eye variation databases. Second we propose new adaptive gaze tracking algorithm that is based fusion of STARBURST with natural feature tracking and inspired by SLAM based approaches. 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.

A. 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 with corneal reflection glints. However corneal reflections features have no stable glint form, and sometimes glints are not presented. One important part of gaze estimation problem is reduced here to estimate coordinates of pupil and corneal reflection, which could further be used for tracking the actual gaze angle with proper algorithms.

The main problem is to estimate gaze pose according the marked gaze pose. [5][6]

The considered metrics could be calculation of average, std. deviation for the difference of the estimated and marked pose. This criterion is used for comparison between “Starburst” version and fusion version based on FERN features.

According some assumptions genera 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 average, std. deviation for the difference between the localized points and marked points. We used such metrics to compare STARBURST with fusion of STARBURST and Viola Jones. However this criterion is not so suitable for gaze tracking estimation based on natural features.

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

B. Related work

We can differentiate problems in eye-gaze estimation to eye model recognition and eye model localization. Eye recognition algorithms provide pattern recognition of some classes like “human eye”, “animal eye”, “asian eye”, “european eyes”. Sometimes eye regions are needed to be localized from the whole face image. The recognition algorithms can be sophisticated and can describe high-level recognition of human identity for example for the biometric security solutions.[12][13] In comparison to recognition algorithms, detection, localization methods are more related to accurate localization of the major eye features like pupil, iris, sclera. Eye class is apriori recognized and input for the localization algorithm. Eye localization algorithms are usually considered for gaze estimation. They are commonly 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 more high accuracy than the preliminary detection process [1][2][3][4] . Starburst algorithm is efficient in speed however its drawbacks are in reduced feature stability and model fitting. As an example, one implementation of the starburst algorithm was fairly stable for a gaze tracker prototype where the eye-camera was located close to the visual axis and illumination was strictly organized [5]. 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 [6].

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 [12][13] . This approach can overcome the constraints of general eye model distribution but it will be adaptive model based approach rather adaptive 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 twinking and partial occlusion is needed more robust technique. 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 marker less tracking and SLAM. It can be applied without knowing any kind of information about the object. [8][9][10][11]. We have also utilized Viola Jones detector for stable general model based technique of eye detection even in case of occlusion [7].

II. MAIN PART

A. Starburst algorithm

The main goal of the starburst algorithm is to extract pupil and corneal reflection points. The algorithm is done by noise reduction, corneal reflection detection, localization and removal of corneal reflection, edge point’s localization, pupil contour detection, ellipse fitting, model based optimization, calibration. [1]. The key idea of the algorithm is to use limited number of rays from the center to extract pupil contour. That is the algorithms is applied only for the part of whole image, meanwhile edge based approaches are applied for whole image. This Starburst optimization in edges search demonstrates benefits in speed [1] [2] [3] [4].

B. Feature –model based Starburst modification

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.

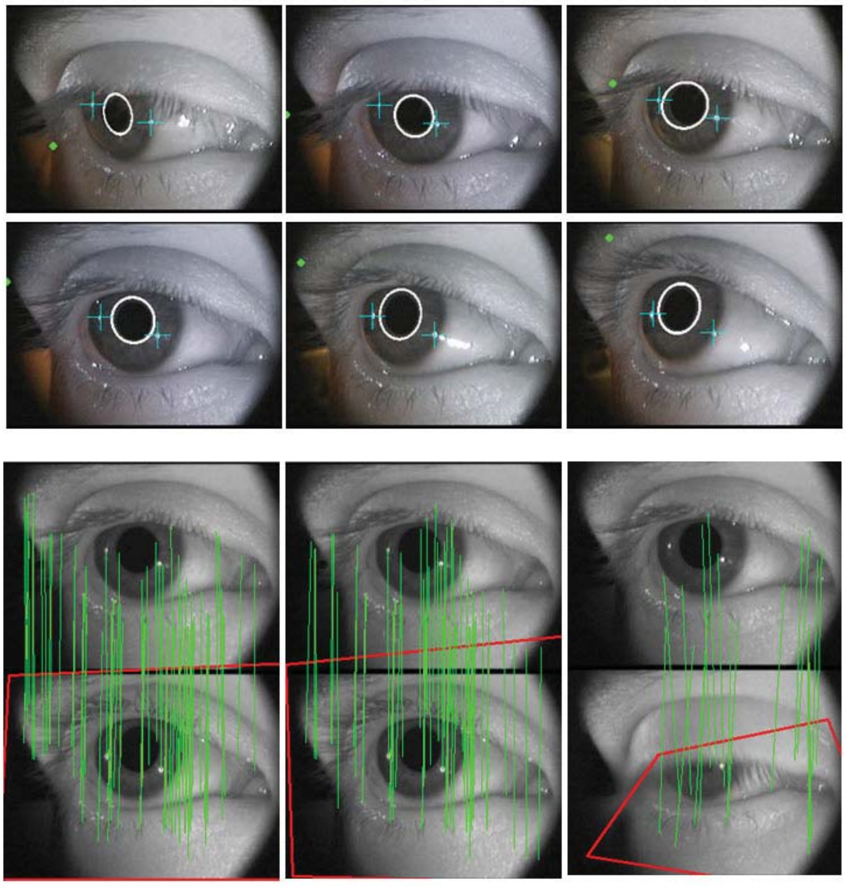

Fig. 1. Pupil contour detection, corneal reflection detection, gaze estimation.
Natural FERN features extracted from the eye and tracking

Our solution has the following steps:

1) Attentive detector. Select the start point for STARBURST algorithm. It can be useful gradient map, center of image, or even Viola Jones detector for some 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 find stable point (pupil, pupil corners or some other eye parts according 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
7) Estimate gaze tracking.

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 RANSCAC could not distinguish the nature of these edges. By application of Viola Jones algorithm we achieve robustness in removal of the outliers.

C. Tracking 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.

D. Adaptive eye tracking algorithm

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 2 minute to train the FERN decision trees[8][9] [10]. 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[9][10], calculation of FERN descriptor for the points and homograph 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 predetected features from Viola Jones or starburst scheme. Algorithm can work also without STARBURST or Viola Jones by utilizing general framework of markerless natural tracking. In our experiments two corneal features were among the FERN features. The model can be useful for the fast feature tracking and complex topology (when we cannot find pupil or corneal reflection but still need to estimate pose on the frame).

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 choose 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. 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 twinking, and different eye topology. The algorithm demonstrates higher detection rate, that is about 98,3% 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.

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.

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.

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 [9][10][11].

REFERENCES

[1] 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 .

[2] 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.

[3] Wayne J. Ryan, Damon L. Woodard, Andrew T. Duchowski, and Stan T. Birchfield Adapting Starburst for Elliptical Iris Segmentation, 2008 .

[4] Bei Yan, Xigong Zhang, Lifeng Gao,Improvement on Pupil Positioning Algorithm in Eye Tracking Technique. 2009.

[5] 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.

[6] 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.

[7] Paul Viola Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple features", 2001.

[8] Mustafa Ozuysal, Michael Calonder, Vincent Lepetit and Pascal Fua "Fast Keypoint recognition using random Ferns", 2008.

[9] Daniel Wagner, Alessandro Mulloni, Dieter Schmalstieg "Real-time detection and tracking for augmented reality on mobile phones". In Proceedings Visualization and Computer Graphics, 2010 .

[10] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, Dieter Schmalstieg "Pose Tracking from Natural Features on Mobile Phones" In Proceedings ISMAR 2008.

[11] H.Bay, T. Tuytelaars, and L.Van Gool. SURF: Speeded Up Robust Features. Proc. European Conf. Computer Vision (ECCV), 2006.

[12] J. Daugman. How Iris Recognition Works. IEEE Transactions on Circuits and Systems for Video Technology, 14(1):21–30, January 2004.

[13] J. Daugman. New Methods in Iris Recognition. IEEE Transactionson Systems, Man, and Cybernetics, Part B, 37(5):1167–1175, October 2007.