https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-055

Pupil Detection and Tracking for AR 3D under Various Circumstances

Dongwoo Kang, Jingu Heo, Byongmin Kang, and Dongkyung Nam; Samsung Advanced Institute of Technology; Suwon-si, South Korea

Abstract

Accurate pupil position detection and tracking of a driver plays an important role and is a prerequisite for augmented reality 3-dimensional head-up display (AR 3D HUD) systems of a vehicle. This paper aims to develop a robust, automated algorithm and system for detection and tracking of pupil centers of a user using a single visual camera and near-infrared (NIR) leds under dynamic driving circumstances. Our proposed pupil tracker consists of eye-nose detection, keypoint alignment, and tracking with NIR led on/off control depending on illumination conditions. Eye-nose detection, which utilizes an error reinforcement learning method for selecting best learning DB, generates facial sub-region boxes including eyes and nose. The error reinforcement learning method uses only a small fraction (less than 5%) of the training DB, while improving the detection rate. Based on the detected region, eye-nose alignment including pupil centers are then processed by Supervised Descent Method (SDM) with Scale-invariant Feature Transform (SIFT). Then, the pupil centers are tracked with Support Vector Machine (SVM) classifier and SIFT feature based tracking checker, which guarantees the aligned results contains pupil centers. This can be considered a different feature space and a strong classifier used in the eye-nose detection stage, such as Local Binary Pattern (LBP) and a set of weak classifiers (Adaboosting). However, the proposed pupil tracker cannot be applied under low light conditions. To achieve this, we add NIR leds and S/W functions to control the intensity of the NIR. After recognizing the contents of the pupil image, such as low light conditions, wearing eyeglasses or sunglasses, corresponding designated aligners are applied. We achieve fairly high detection rate (98%) and precise eye alignment (average error 2mm) even with challenging conditions, such as day and nighttime driving conditions.

Introduction

Recently, we have been witnessed an increasing popularity of head-up display (HUD), being deployed in automotive business. Automotive HUD shows information directly in the line of sight via a windshield or a combiner. It is believed that HUD reduces driver distraction and increases driver safety and convenience, while enhancing the driver's experience with a layer of information right on the street. However, 2D HUD, showing information in a virtual plane, has many limitations. One of the main limiting factors of 2D HUD is that Augmented Reality (AR) information is not overlapped in the real world, causing additional distractions and visual mismatches. AR 3D HUD can enrich HUD experience providing the drivers with real depth of AR images aligned with the street and road objects. For example, drivers can obtain the important information such as indicator arrows aligned with the road for navigation without looking at the secondary display. The pupil-tracking-based 3D method allows users to seamlessly enjoy 3D displays without glasses by overcoming the viewing position restriction, which are more suitable for automotive driving conditions. Accurate 3D pupil-position tracking plays an important role in clearly separating the left and right pupils of the viewers in autostereoscopic 3D display systems [1].

To date, many studies of eye tracking have focused on gaze tracking, which can be obtained by calculating the vector between the pupil center and the corneal reflection. Most methods use bright pupil and corneal reflections with near-infrared (NIR) light sources, where the IR camera and light sources should be located on specific locations each . The goal of gaze tracking is to estimate the viewer's looking direction, while our pupil tracking aims to estimate the 3D position of the pupil centers without relying on corneal reflections. Therefore, the two should be considered a complete different technology. Pupil centers can be calculated from conventional facial detection/tracking points. However, detection and alignments are still challenging under dynamic illumination conditions, especially low light environments.

In this work, we describe an improved automatic pupil center tracking algorithm and system for AR3DHUD, which should provide drivers stable AR3D image/graphics under dynamic illumination conditions, such as night driving. Instead of using the whole facial area, we only utilizes eye-nose regions, since it has been proven to be reliable under expression while improving the detection rate and accuracy. Our proposed pupil tracker takes only 4ms@3.5GHz GPU in tracking mode and detects visual/IR images with high success rate even with challenging driving conditions.

Method

The basic components of our proposed pupil tracker processes input images from a camera with the following three main steps, as described below: (1) eye-nose detection from a camera image, (2) eye-nose alignment of 11 landmark points and (3) tracking at every camera frame with tracker checker (Figure. 1).

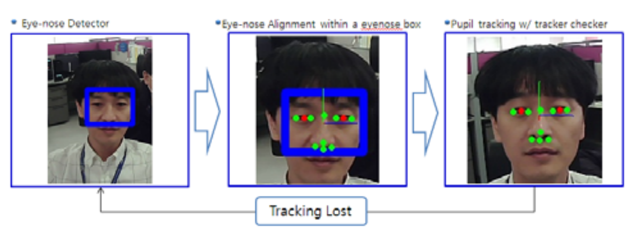

Figure 1. Overview of pupil tracker: (1) eye-nose detector (2) eye-nose alignment (3) tracking

The overall flowchart of the proposed system and algorithm is given in Figure. 2. The detector is not executed while tracing succeeds. It focuses more on pupil center alignments, preventing pupil centers from wrong alignment, less being affected by other parts alignment of face under various illumination conditions.

The proposed algorithm works fairly reliable under normal lighting conditions (100 ~10,000lux) with 99.4% detection rate and 1mm accuracy, obtained from our own collected data. However, under low light conditions, significant drop in performance has been observed. In order to handle this challenge, we design a switchable visual/IR camera. After removing the IR cut filter and adding a NIR light source, the camera switches between visual and IR modes. Especially, an in-painting algorithm was applied on eye glasses reflection area of NIR DB images for both training and testing as pre-processing. This is a necessary step to improve robustness with eye glasses under NIR.

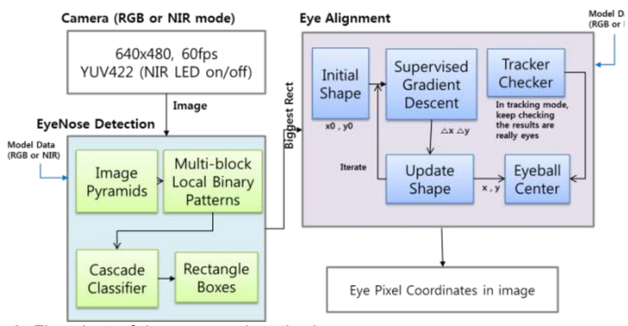

Figure 2. Flowchart of the proposed method

Unlike other object detection algorithms, eye-nose detection utilizes a novel approach, called an error reinforcement learning method. The error reinforcement learning method uses only a small fraction (less than 5%) of the training DB, while improving the detection rate. It can be analogous for human's brain developing stage, since humans can learn many rules while making mistakes and transfer those knowledges to other decision rules without having actual experience on them. In the full paper, we plan to explain in detail about the error reinforcement learning method. From the detected eye-nose region, the pupil center coordinates are extracted by facial shape alignments using supervised descent method (SDM)[2] and scale-invariant feature transform (SFIT).

Another novel idea of our proposed work is the use of a tracker check. The tracker checker, which guarantees the aligned results contains pupil centers, plays an important role in order to maintain in a tracking mode. The tracker checker consists of SIFT features and SVM classifiers. This can be considered a different feature space and a strong classifier used in the eye-nose detection stage, such as Local Binary Pattern (LBP) and a set of weak classifiers. This way, we expect more reliable tracker checker will be achieved.

However, the proposed pupil tracker cannot be applied under low light conditions. To achieve this, we add NIR leds and S/W functions to control the intensity of the leds. Based on the contents of the pupil image, such as low light conditions, wearing eyeglasses or sunglasses, corresponding designated aligners are applied. Specifically, sunglass and eyeglasses can be handled with designated aligners, accordingly. This step is illustrated in Figure.3.

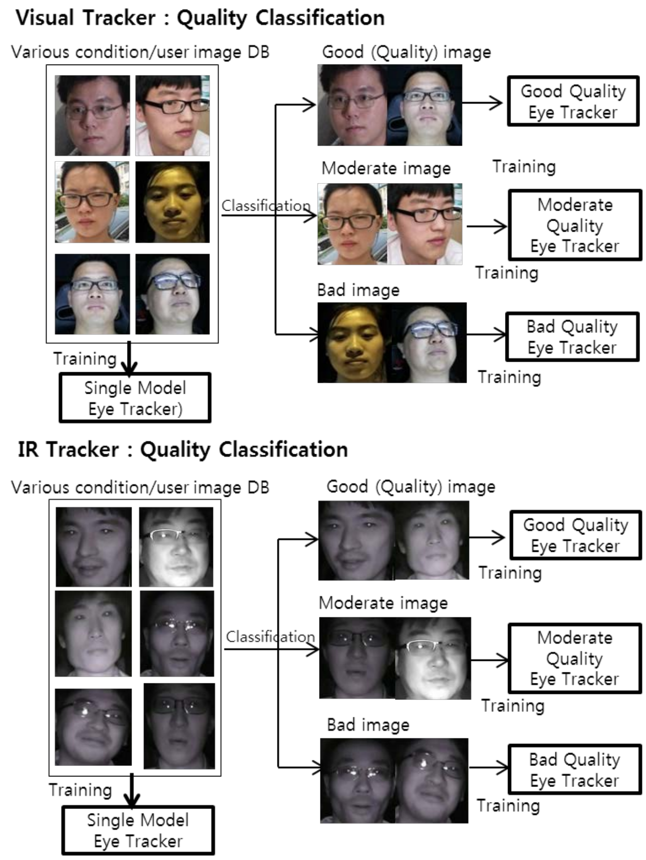

Figure 3. Visual single aligner model vs. Contents-aware quality specific aligner model (up). IR single aligner model vs. Contents-aware quality specific aligner model (bottom)

Experimental Results

Detector and tracker models were trained separately with eye-nose boxes and 11 eye-nose landmark points on images each. The ground truth of eye-nose boxes and landmark points on all image datasets were assessed by a person. The RGB detector for general usage was trained with 660,000 RGB image dataset, and the IR detector for low illumination conditions was trained with 60,000 NIR image dataset. For the tracking and tracking checker, we trained 60,000 RGB images for RGB and 60,000 NIR images for IR (Table 1).

We tested our algorithm on both 100,000 RGB images and 30,000 NIR images, not used in training, which were captured in indoor areas and real driving circumstances outside (Figure. 4). 70,000 RGB face images were captured in the office, where various illumination conditions were applied from dark to bright. 30,000 RGB face images and 30,000 NIR face images were recorded as drivers or passengers during the daytime, sunset time and night time.

The proposed algorithm ran successfully (Figure. 4) on all the 130,000 face images, in real time with an execution time of <6ms for tracking on a standard 2.5GHz personal computer running windows 7. Detection was executed only when tracking lost, with execution time <16ms (>60fps).

The detection testing accuracy was 99% on RGB images and 98% on NIR images. The alignment testing precision was 1mm on RGB images and 2mm on NIR images. The alignment precision was measured as the average distance between the ground truth and the tracked pupil center, where 65mm inter-pupillary distance (IPD) was used for physical distance conversion. NIR detector and aligner performance were lower than RGB mode due to NIR eye-glasses reflection. Overall performance of the pupil tracker has 98.5% detector accuracy and average eye precision error was 1.5mm compared to manual human marking (Table 2).

Table 1. Training DB for RGB detector/tracker and NIR detector/tracker

		Training DB
RGB Images (general illumination)	Detector DB	660,000
	Tracking DB	60,000
Infrared (NIR) Images (low illumination)	Detector DB	60,000
	Tracking DB	60,000

Table 2. Testing DB and algorithm performance of RGB detector/tracker and NIR detector/tracker

		Testing DB	Accuracy (detection)	Precision (Average Error)
RGB Images (general illumination)	Detector	100,000	99%	N/A
	Tracking	100,000	N/A	1mm
Infrared (NIR) Images (low illumination)	Detector	30,000	98%	N/A
	Tracking	30,000	N/A	2mm

Figure 4. Driver pupil tracking on RGB (left) and NIR (right). Red dots indicates the tracked center of the pupils

Discussion

A novel machine learning technique is proposed for the robust detection and tracking of pupil center under various illumination conditions. The algorithm was validated on indoor environment and real-driving image datasets and showed 98.5% detection accuracy and 1.5mm average pupil precision error in total, superior to that achieved by previous standard machine learning techniques.

The algorithm is different from previous conventional machine learning techniques. According to different image conditions, different eye detectors and trackers were applied, with functions to control the intensity of the NIR leds. Pre-trained different detector and trackers were selected depending on the contests of the pupil images, including low light conditions. Especially, even sunglass and eyeglasses can be handled with designated aligners, accordingly. When compared to the single RGB detector and aligner under low light circumstances, the detector accuracy decreased by 25% and the pupil tracking precision average error increased by 2mm When the single RGB detector and tracker was used for NIR images, the detector accuracy decreased by 19% and the pupil tracking average error increased by 2mm (Table 3).

Many studies attempted automatic face detection and facial landmark point alignment including deep learning based techniques [ref]. A performance comparison between our proposed algorithm and a fast deep learning detector technique SqueezeDet (57.2fps) [3] is shown in Table 4. Our content-aware pupil tracking algorithm achieves higher detection accuracy, lower pupil precision error, and higher speed at both the high and low light conditions. Also, the

Table 3. Algorithm performance when a single RGB detector/tracker used at low illumination circumstances. The table shows the conventional RGB detector/tracker accuracy and precision on RGB image data and NIR image data.

		Testing DB (low light 5lux)	Accuracy (detection)	Precision (Average Error)
RGB Images (general illumination)	RGB Detector	100,000	74.44%	N/A
	RGB Tracker	100,000	N/A	3mm
Infrared (NIR) Images (low illumination)	RGB Detector	30,000	80.36%	N/A
	RGB Tracking	30,000	N/A	3mm

Table 4. Eye detector performance comparison between SqueezeDet and the proposed algorithm.

Testing	SqueezeDet	Proposed detector
NIR DB (low light)	Driver DB 8059	Driver DB 8059
Accuracy	35.1%	98%
RGB DB (normal light)	16463	16463
Accuracy	99.5%	100%
RGB DB (eye-glasses)	2308	2308장
Accuracy	89.31%	99%

References

[1] Park, J., et al. "Light field rendering of multi-view contents for high density light field 3D display", SID Symposium Digest of Technical Papers, Vancouver, Canada, May 2013, 44, pp. 667–670

[2] Xiong, et al. "Supervised descent method and its applications to face alignment." Computer Vision and Pattern Recognition (CVPR)

[3] Bichen Wu et al., "SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017, pp. 129-137

Author Biography

Dongwoo Kang received his BS in Electrical Engineering from Seoul National University (SNU) (2007) and his PhD in Electrical Engineering from University of Southern California (USC) (2013). Since then he has worked in Samsung Advanced Institute of Technology in Suwon, South Korea. His work has focused on the development of image processing and computer vision application.