A Literature Review for Eye Tracking and Face detection

1 Introduction

This project is trying to setup a reliable machine vision driven motor wheel. Since the technology for parts like motor control, communication between PC and micro controller and GUI development are relatively mature, this review focus on the eye location and tracking as well as face detection algorithms.

As one of the most vital features of face characteristics, eye motion play an quite important role in expressing a person's desires and needs, cognitive processes, emotional states, and interpersonal relations. [1] Although enormous researches have been carried out on eye motion and face detection in many fields such as biometric security, human-computer interaction, challenges are still blocking the way to implement it easily in practise. Challenges of eye detection and face detection are caused by illumination changing, in-plane rotation, out-plane rotation appearance changing and occlusion.

As shown in [2], eye location may be implemented in 5 ways:

1. Shape detection, an approach that try to match iris, pupil and shape of eyes.
2. Feature detection, which captures characteristics of human eye to locate features including limbs, pupil and cornea reflections.
3. Photometric appearance detection. This approach detect eyes directly using its photometric result characterized by the colour or some filter response of the eye and its surroundings.
4. Combination approach. This approach combines several different methods together in one system.
5. Other way. Such as symmetry operators, temporal information.

In addition to eye location, eye tracking is also reviewed in this report. Since eye tracking is one of object tracking issues, this review focused on object tracking to discuss advantages and disadvantages of currently available approaches.

Generally there are two classes of object tacking algorithms. They are generative and discriminative respectively. Generative algorithms build a model for the object tracked and use this model to the region in an image that has minimum differences. Discriminative algorithms pose the tracking problem as a binary classification task in order to find the decision boundary for separating the target object from the background [3]. Eye tracking has been studied in many fields. Zhu and JI illustrate an eye tracking system using IR illumination [3]. Villanuev et al. also showed an eye tracking system to identify gaze direction with extra hardware [4]. As shown in [5], face detection can be carried out first prior to eye tracking to improve the performance. Hansen et al. present a new approach that may perform in large information spaces with noisy inputs.

Face detection is the third issue reviewed. It may be used as an alternative method to improve or even substitute eye tracking and eye location in later phases of this project.

Researches on face detection have made a series of significant achievement, such as colour-based detection, neural-based detection and feature-based detection. Considering time complexity, accuracy and reliability, this report reviewed a machine learning algorithm proposed by Viola and Joes as described in [6]. This approach distinguished by three key contributions. The first is the introduction of a new image representation called the “Integral Image” which allows the features used by our de-tector to be computed very quickly. The second is a learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers . The third contribution is a method for combining increasingly more complex classifiers in a “cascade” which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions.

2 Discussions

2.1 Eyes detection and tracking

In [7] the Hough transform is used to locate eyes by detecting shape of iris or pupil. Problem is that it requires explicit feature detection. As shown in [8], Perez et al. used thresholds of image intensities to detect and evaluate the centre of pupil. The eye camera follows the head movements maintaining the pupil centred in the image. When a tracking error is produced, the image from a camera with a wider field of view is used to locate the eye and quickly recover the tracking process. Four infrared light sources, synchronised with the shutter of the eye camera, are used to produce corneal glints. Its special shape has been exploited to allow the optimisation of the image processing algorithms developed for this system. Special care has been taken in limiting the illumination power and working way below the dangerous levels. After a calibration procedure, the line of gaze is determined by using the pupil-glint vector. The glints are validated using the iris outline with the purpose of avoiding the glint distortion due to the changes in the curvature on the ocular globe. The proposed algorithms determine the pupil centre with sub-pixel resolution, minimising the measurement error in the pupil-glint vector. The problem with this approach is that it is quite sensitive to illumination changing.

Yuille et al. proposed a deformable template model. This approach describe the feature of interest by a parameterized template. An energy function is defined which links edges, peaks, and valleys on the image intensity to corresponding properties of the template. The template then interacts dynamically with the image by altering its parameter values to minimize the energy function, thereby deforming itself to find the best fit. The final parameter values can be used as descriptors for the feature. [9] The only problem with this method is that it is unable to handle the situation when eyes are closed. Even so Yuille et al’s model can be an option for the later phase of this project.

Neural networking is one important approach for eye detection. The simplest way to use a neural network is to scan an input window of i \* j pixels across the image. Input each colour value into the neural network as shown in Figure 1.This network will then be trained to give a higher value when the input pixels are cantered on the eye. When the entire image has been scanned the region that has highest value will be returned as output.

Problem with this approach is the high time consumption .Also it requires a quite large number of inputs and training set to cover all kinds of eye appearances. Besides, the colour values are sensitive to lighting conditions, orientation and scale.

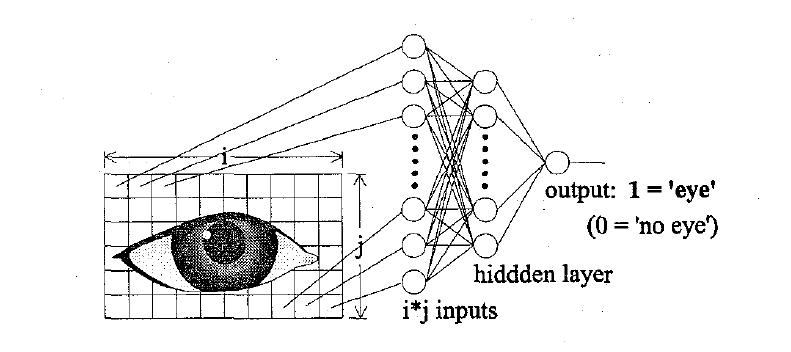


Figure 1.

To make this method faster, Vincet et al used a low resolution image first to coarsely detect the eyes. Only the most possible regions are scanned using high resolution image. This approach reduces the space down to 4% of the original size. The disadvantage is that a low resolution image is needed, which implies filtering and decimation operations. Moreover, two separate networks have to be trained.

A better approach is presented in [10] which automatically locate facial features such as eyes, nose and mouth in image sequences using a neural network approach. It is shown that by modelling the feature sought as a structural assembly of micro-features, and by using a probabilistic interpretation of neural outputs, it is possible to construct a location system that is more robust than a location system which uses the feature as a single entity. With this micro-feature approach, not only the position of the features can be found, but also the shape of the features. The problem with this approach is that it just detects frontal view face only and needs long calculation time.

2.2 Face detection

This part focus on Viola’s Adaboost paper [6]. This paper has three parts. The first is “Integral Image”. The second is learning algorithm based on AdaBoost. The third is method that used multiple classifiers in a “cascade”, which can discard background regions quickly while spending more computation time on promising parts of images.

2.2.1 Integral Image

Integral Image is an intermediate representation for the image which can be computed very rapidly. [11]The following equation shows how to build it pixel by pixel:

Where ii(x,y) is the integral image and i(x’,y’) is the original image. Using the following the recurrences:

s(x,y) = s(x,y-1) + i(x,y)

ii(x,y) = ii(x-1,y)+ s(x,y)

The integral image can be computed in one pass over the original image.

2.2.2 Learning Discussion

Many selection process can be found in [12] for a review. In [13], Papageorgiou et al proposed a scheme for feature selection based on feature variance. They demonstrated good results selecting 37 features out of a total 1734 features. In Viola’s paper, an algorithm based on Boost is given, the below is a simplified version:

1. Prepare positive and negative samples;
2. Initialize weights;
3. Normalize the weights,
4. Train a classifier hj, such that
5. Choose the classifier with the lowest error.
6. Update the weights.

The final Strong classifier is

2.2.3 Classifiers Cascade

As shown in Figure 2, a series of classifiers are applied to every sub-window. The initial classifier eliminates a large number of negative examples with very little processing. Subsequent layers eliminate additional negatives but require additional computation. After several stages of processing the number of sub-windows have been reduced radically. Further processing can take any form such as additional stages of the cascade (as in our detection system) or an alternative detection system.

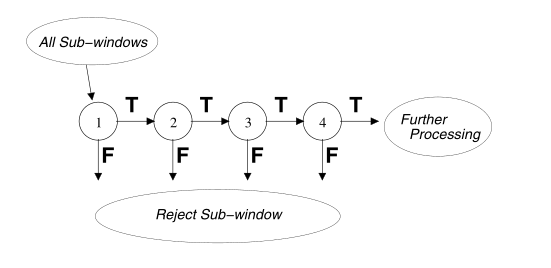


Figure 2

Summary

This report reviewed the image processing algorithms that can be used to control motor wheel. For eyes location and eyes tracking, this report reviewed the currently available approaches as well as advantages and disadvantages. For face detection, this review focused on an algorithm based on machine learning.

References

[1] Underwood G. Cognitive processes in eye guidance. Oxford: Oxford University Press; 2005.

[2] Hansen DW, Ji Q. In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans Pattern Anal Mach Intell 2010;32(3): 478–500.

[3] Zhang K, Zhang L, Yang MH. Real-time compressive tracking. In: A. Fitzgibbon, S.Lazebnik, P.Perona, Y.Sato, C.Schmid(eds.), Computer vision–ECCV. Berlin, Heidelberg: Springer;2012.p.864–77.(b)ZhuZ, JiQ .Eyeand gaze tracking for interactivegraphicdisplay.MachVisiAppl2004;15(3):139–48.

[4] Villanueva A, Cabeza R, Porta S. Gaze tracking system model based on physical parameters. IntJ Pattern Recognit Artif Intell2007;21(5):855–77.

[5] Chen YS, Su CH,Chen JH, Hung YP, FuhCS, Chen CS. Video-based eye tracking for autostereoscopic displays. Opt Eng 2001; 40(12):2726–34.

[6] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Conference on computer vision and pattern recognition, vol.1; 2001.p.511–518.

[7] Young D, Tunley H, Samuels R. Specialised hough transform and active contour methods for real-time eye tracking. Brighton: University of Sussex, Cognitive & Computing Science; 1995.

[8] Peréz A, Cordoba ML, Garcia A, Méndez R, Munoz ML, PedrazaJL. etal. A precise eye-gaze detection and tracking system. In: Proceeding soft the 11th international conferenceincentral Europeon computergraphics, visualization and computervision; 2003.p.105–8.

[9] Yuille AL, Hallinan PW, Cohen DS. Feature extraction from faces using deformable templates. IntJ Comput Vis 1992;8(2):99–111.

[10] Reinders MJ, Koch RWC, GerbrandsJJ. Locating facial features in image sequences using neural networks. In: Proceeding soft hesecond international conference on automatic face and gesture recognition; 1996.p.230–5.

[11] F. Crow. Summed-area tables for texture mapping. In Proceedings of SIGGRAPH, volume 18(3), pages 207–212,1984.

[12] Andrew Webb. Statistical Pattern Recognition. Oxford University Press, New York, 1999.

[13] Robert E. Schapire, Yoav Freund, Peter Bartlett, and Wee Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. In Proceedings of the Fourteenth International Conference on Machine Learning, 1997.