COMPETITIVE AND COOPERATIVE SOLUTIONS FOR REMOTE EYE-TRACKING

Giovanni Crisafulli, Giancarlo Iannizzotto and Francesco La Rosa
Visilab, Faculty of Engineering, University of Messina, Contrada Di Dio (S. Agata), Messina, Italy

Keywords: Eye-tracking, Image analysis, Competitive approach, Cooperative approach.

Abstract: Reliable detection and tracking of eyes is an important requirement for attentive user interfaces. In this paper, we present an innovative approach to the problem of the eye-tracking. Traditional eye-detectors, chosen for own properties, are combined by two different schemes (competitive and cooperative scheme) to improve own robustness and reliability. To illustrate our work, we introduce a proof-of-concept single camera remote eye-tracker and discuss its implementation and the obtained experimental results.

1 INTRODUCTION

Eye Tracking (ET) is the process of measuring eye positions and eye movements in a sequence of images (Ji and Zhu, 2004), (Morimoto et al., 1998b).

Specifically, detection and tracking of the iris or pupil can be used to infer the direction of interest of the human subject, this is denoted gaze (Matsumoto and Zelinsky, 2000).

By tracking the eye-gaze of a user, valuable insight may be gained into what the user is thinking of doing, resulting in more intuitive interfaces and the ability to react to the users' intentions rather than explicit commands (Jacob, 1991), (Morimoto and Mimica, 2005).

Also, gaze can play a role in understanding the emotional state for humans, syntesizing emotions, and for estimation of attentional state. Specific applications include devices for the disabled, e.g., using gaze as a replacement for a mouse and driver awareness monitoring to improve traffic safety (Coifman et al., 1998).

Early eye-tracker (and eye gaze tracker) were developed for scientific exploration in controlled environments or laboratories. Eye gaze tracking (EGT) data have been used in ophthalmology, neurology, psychology, and related areas to study oculomotor characteristics and abnormalities, and their relation to cognition and mental states.

Successful attempts are still limited to military applications and the development of interfaces for people with disabilities.

To be applied in general computer interfaces, an ideal eye tracker should be accurate, reliable, robust (should work under different conditions, such as indoors and outdoors, for people with glasses and contact lenses, etc), non-intrusive, allow for free head motion, not require calibration and to have real-time response.

Our work is concerned with the usability, the reliability and the robustness of eye-trackers for general applications. As a proof-of-concept, we propose a single camera remote eye-tracker that use two different schemes (*competitive and cooperative*) to merge (Freund and Schapire, 1997) the results of some eye-detectors. An improving of the robustness and reliability is obtained.

The paper is organized as follows. In Section 2 we describe the main ET approaches to be found in the literature. In sections 3 we describe the system we propose. In Section 4, we give details of the experimental results obtained. Finally, in Section 5, we present our conclusions and some indications of future development.

2 RELATED WORKS

Detection of the human eyes is a difficult task due to a weak contrast between the eye and the surrounding skin. As a consequence, many traditional techniques for ET and EGT are intrusive, i.e., they require some equipment to be put in physical contact with the user. These techniques include, for example, contact

Crisafulli G., Iannizzotto G. and La Rosa F. (2009).
COMPETITIVE AND COOPERATIVE SOLUTIONS FOR REMOTE EYE-TRACKING .
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 521-528
DOI: 10.5220/0001555105210528

lenses, electrodes, and head mounted devices (Morimoto and Mimica, 2005). Non-intrusive techniques (or remote techniques) are mostly vision based (Ji and Zhu, 2004), (Morimoto et al., 1998a), (Hsu et al., 2002), i.e., they use cameras to capture images of the eye. Some camera-based techniques might be somewhat intrusive if they require to be head mounted (Babcock and Pelz, 2004), (Li et al., 2005), (Li et al., 2006).

For diagnostic applications, where eye data can be recorded during a short experiment and processed later, the time required to setup the eye tracker and the discomfort that the equipment might cause do not constitute a problem. This is also true for a few interactive applications where the user has to depend heavily on the eye tracker to accomplish some task (i.e., there is little choice or no alternative device).

A remote eye tracker (RET) offers comfort of use, and easier and faster setup, allowing the user to use the system for longer periods than intrusive techniques. Although the accuracy of RETs is in general lower than intrusive ETs, they are more appropriate for use during long periods. The pupil–corneal reflection technique (Zhu and Ji, 2005), (Morimoto et al., 1998a) is commonly advertised as a remote tracking system that is robust to some head motion.

Camera-based EGT techniques rely on some properties or characteristics of the eye that can be detected and tracked by a camera or other optical or photosensitive device. Most of these techniques have the potential to be implemented in a non-intrusive way.

The limbus and the pupil are common features used for tracking (Daugman, 1993), (Haro et al., 2000). Limbus is the boundary between the sclera and the iris. Due to the contrast of these two regions, it can be easily tracked horizontally, but because the eyelids in general cover part of the iris, limbus tracking techniques have low vertical accuracy. The pupils are harder to detect and track because of the lower contrast between the pupil–iris boundary, but pupil tracking techniques have better accuracy since they are not covered by the eyelids (except during blinking).

To enhance the contrast between the pupil and the iris, many eye trackers use an infrared (IR) light source (Zhu and Ji, 2005), (Morimoto et al., 1998a) . Because IR is not visible, the light does not distract the user.

Sometimes, the IR source is placed near the optical axis of the camera. Because the camera now is able to ‘‘see’’ the light reflected from the back of the eye, similar to the red eye effect in night photography using a bright flash light, the camera sees a bright pupil instead of a regular dark pupil.

The light source can also generates a corneal reflection (CR) or glint on the cornea surface, near the pupil. This glint is used as a reference point (Li et al., 2005) in the pupil–corneal reflection technique for EGT.

Due to the use of active IR lighting, this technique works better indoors and even in the dark, but might not be appropriate outdoors, because sunlight contains IR and the pupils become smaller in bright environments.

The literature offers several techniques for detecting eyes directly (Kawato and Tetsutani, 2002a; 2002b), or as a sub-feature of the face (Hsu et al., 2002). Faces can be detected from background subtraction, skin color segmentation, geometric models (Li et al., 2005) and templates (Matsumoto and Zelinsky, 2000), artificial neural networks (Ji and Zhu, 2004), etc.

Direct methods for eye detection use spatial and temporal information to detect the location of the eyes.

Their process starts by selecting a pool of potential candidates using gradient fields and then heuristic rules and a large temporal support are used to filter erroneous pupil candidates.

The use of a support vector machine (SVM) avoids falsely identifying a bright region as a pupil: the pupil candidates are validated using SVM (Zhu and Ji, 2005) to remove spurious candidates.

Instead of using explicit geometric features such as the contours of the limbus or the pupil, an alternative approach is to treat an image as a point in a high-dimensional space. Techniques using this representation are often referred to as being appearance-based or view-based (Black and Jepson, 1998).

In (Haro et al., 2000) applies a linear principal component analysis approach to find the principal components of the training eye patches. In (King and Xu, 1997) use a probabilistic principal component analysis to model off-line the intra-class variability within the eye space and the non-eye distribution where the probability is used as a measure of confidence about the classification decision.

The method proposed in (Ji and Zhu, 2004) also does not use explicit geometric features. They describe an EGTer based on artificial neural networks (ANN). Once the eye is detected, the image of the eyes is cropped and used as input to a ANN. Training images are taken when the user is looking at a specific point on a computer monitor.

Finally, when some pupils are detected, the information about its position and velocity can be passed to a tracker module (Shi and Tomasi, 1994), (Bouguet, 1999) to enforce motion tracking stability. Kalman filtering is often used to predict pupils position in current frame, therefore greatly limiting the search space. Sometimes, Kalman filtering can be improved by mean-shift tracking (Ji and Zhu, 2004), which tracks an object based on its intensity distribution.

3 SYSTEM

In this paper we propose an innovative passive remote ET system that uses hardware off-the-shelf, whose performances are independent from the lighting conditions (natural or artificial).

Our ET system uses several ET techniques without the needs of any initial calibration. An advantage from Optical Flow (Tomasi and Kanade, 1992) has been taken.

The main contribute of our work is to obtain an accurate and robust estimate of eye-position, using some eye-detection techniques described later, according to two different schemes, called in this paper competitive and cooperative scheme.

In the first scheme, frame by frame, it's estimated the reliability of each technique and then it's chosen which one to use. In the second scheme, for each frame, all techniques contribute to the determination of the final results.

To develop the schemes, we analyzed the different performances of each eye-detection technique according to the operating conditions: lighting conditions, dynamic of the user, partial iris occlusions, and distance of the user from the camera.

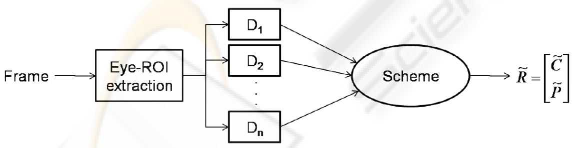

Figure 1: The main operational phases of the system.

As shown in Figure 1, each frame is elaborated by a face detector. It returns, eventually, a Region Of Interest (ROI) that encloses the face of the "observed" user. The region returned from the face detector is elaborated to define two Eye-ROI (E-ROI) that will enclose the eyes of the user. The face detection is obtained using a boosted classifier (Freund and Schapire, 1996) based on Haar-like features (Viola and Jones, 2001) and improved in (Lienhart and Maydt, 2002).

To select the E-ROI we have tried to use the method proposed in (Peng et al., 2005), but in according to the measures that we have realized, we use some parameters determined empirically.

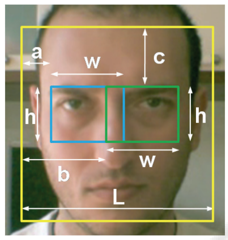

Figure 2: Eye-ROI definition.

Let L be the side length of the Face-ROI, the anchor point of the E-ROI on left will have the following coordinates *(a,c)*, referred to the top-left of Face-ROI. The E-ROI will have a width equal to w and a height equal to h. In our system, we use $a=L/6$, $c=L/3.6$, $w= L/2.8$ and $h=L/3.7$.

The same size, but with a different anchor point, is used for the other E-ROI (see Figure 2).

Also in our system, the choice of the number of detection techniques adopted is exclusively limited by the available computational resources. We use six eye-detection techniques, selected for them complementary properties (Section 4).

A brief description of the detection algorithms adopted is presented in the following section.

3.1 Eye-Detection

The **first eye detection technique** that we use is described in (Daugman, 1993) and was used, originally, as pre-processing step into an iris recognition system.

The algorithm returns the coordinates of pupil centre. Daugman asserts that the technique is optimum (it will certainly find the optimum contour) but it is computationally expensive, because the number of contours elaborated depends on the analyzed region size. This technique offers the further advantage to be robust respect to iris occlusions due to the overlap of eyelid on iris.

On the other hand, when the user wear glasses the accuracy results is considerably lower than in normal operating conditions.

The **second technique** is proposed in (Hsu et al., 2002). In this case, the detection of eyes is obtained processing the acquired frame in the YCbCr color space. The original algorithm was modified to work with E-ROI.

An EyeMap is obtained processing the E-ROI in the YCbCr space. In EyeMap, pixels around the eye are characterized by high value. Finally, the EyeMap is processed with a Dilation operator and then thresholded. After the extraction of the Connected Components (CC) in the EyeMap, the system calculates the barycentre of the one with the greatest area (and bounding box with width/height ratio between 0,6 and 1,4).

The calculated barycentre is an estimate of the eye-position for this technique.

This technique, in good and uniform lighting conditions, has proved to be robust and accurate. Performance doesn't vary significantly raising distance from camera (in a maximum distance of 60cm). Glasses reduce the performances for this technique too. False positives and/or false negatives can be returned by this technique, especially with noise and shadows.

The **third technique** process the E-ROI into HSV space.

In this case we take advantage of the high saturation in light reflection by the pupil. The pupil is characterized by low value in RGB space, and, with natural lighting conditions, it can reflect light with high saturation.

In order to extract the pupil centre the input E-ROI histogram is equalized. We use the maximum in the saturation image to threshold it with threshold value T_3. Now we extract the Connected Components (CC) as before.

The technique has shown good results and responsively in acceptable lighting conditions also when the user is moving.

The detection technique presents good results also when the user wears glasses. It can obtain detection errors with a cluttered background into E-ROI.

The **fourth technique** uses the intensity of the light reflected by the pupil. That, in fact, is a minimum of brightness into E-ROI.

Analyzing red channel we reduce the shadow presents into E-ROI. Background, eyebrows and glasses with dark frame, if enclosed into E-ROI, are the principal "obstacles" for this technique.

The **fifth technique** use two convolution masks customly designed for eye detection. These were defined using typical circular shape and characteristic gray level in the iris-sclera complex.

In our implementation we used the masks shown in eq. 1

$$K_1 = \begin{vmatrix} -1 & 2 & 2 \\ -3 & -1 & 1 \\ -1 & 2 & 2 \end{vmatrix} \quad K_2 = \begin{vmatrix} 2 & 2 & -1 \\ 1 & -1 & -3 \\ 2 & 2 & -1 \end{vmatrix} \tag{1}$$

A schematic representation of this technique is shown in Figure 3.

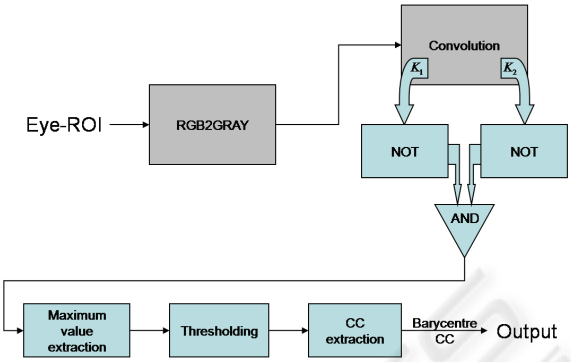

Figure 3: The main operational phases of fifth technique.

Glasses don't modify the technique's performances, also with dark frame. The detection isn't affected from shadows in region close to eye. The method proposed to find eyes doesn't guarantee the desired continuity and precision. There are cases in which the resultant mask has a single large CC, almost equal to E-ROI themselves. Instead it can't locate CC when there are excessive movements or reflections. This technique has a decrease of performances in the presence of cluttered background.

The latest and **sixth technique** used for eye detection is based on global relation between pixels of image. We take advantage of Hough Transform (Gonzalez and Woods, 2002) to extract circumferences into E-ROI previously elaborated by some mathematical morphology operators.

The circumferences that we search are the pupil and iris. Iris and pupil detection is obtained applying the Hough Transform to the edges detected by Canny's Operator (Canny, 1986).

Hough presents precision and accuracy that don't vary significantly raising distance from camera (max 60cm). On the other hand, in case of bad lighting operating conditions, Hough Transform can return high number of false positives or no response.

As for Daugman technique it needs considerable computational resources so it limits the number of techniques that can be used.

Finally, its performances decrease highly with sudden movements of the user.

3.2 Competitive and Cooperative Schemes

Now we describe the schemes that we use to merge the results obtained by the single adopted techniques. Each technique returns an estimate

($Xi=(x_i,y_i)$ with $i=1,2,\ldots,N$) of the eye position that is the barycentre of the pupil into image. This is the core of our system, and we'll try to demonstrate that cooperative and competitive schemes can return more accurate result than the described techniques.

Competitive scheme aim to determine online which technique used offers the better results. After few frames, our system will be, already, capable to establish which technique is the most reliable.

The approach's peculiarity is that we consider possible a variation of operating conditions at any time. In fact, when the operating conditions of the system change, we can use, again, the techniques that before was considered not reliable. This peculiarity is capital when, for example, lighting operating conditions and/or distance of the user from the camera change. Also the camera can change automatically some its setup (Automatic Gain Control, Auto Iris, Back Light Compensation, Exposure); we used low cost cameras.

The competitive scheme takes advantage of three principal information: the set of points returned from the techniques, the "hit" probabilities of the techniques and the estimated point (X*=(x*,y*)) obtained by the Optical Flow (OF) (Tomasi and Kanade, 1992). OF is used in our system to provide, using historical results, a prediction useful in processing of the next frame.

First of all, we define a matrix of the distances ***D*** (see eq. 2). The elements of ***D*** are the relative distances, expressed in pixel, between the points returned from the eye-detection techniques.

$$\mathbf{D} = \begin{vmatrix} 0 & d_{12} & \dots & d_{1N} \\ d_{21} & 0 & \dots & d_{2N} \\ \dots & \dots & \dots & \dots \\ d_{N1} & d_{N2} & \dots & 0 \end{vmatrix} \tag{2}$$

D is a symmetric NxN matrix (in our case N=6).

We now introduce the proximity vector ***V***:

$$V = \begin{vmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{vmatrix} \tag{3}$$

The elements v_i quantify the number of techniques that return a point near to the estimate of i^{th} technique, depending from the elements of ***D***, according to the eq. 4 and eq. 5.

$$v_i = \sum_{j=1}^{N} \delta_{ij} \tag{4}$$

with

$$\delta_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } d_{ij} < 5 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Then we define the overlap vector S:

$$S = \begin{vmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{vmatrix} \tag{6}$$

with

$$s_i = \begin{cases} \sum_{j=1}^{N} \delta_{ij} \cdot d_{ij} & \mathrm{v_{ij}} > 0 \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

s_i value will be directly proportional to the distances d_{ij} and inversely proportional to v_i. In our purpose, overlap vector is an estimate of the techniques reliability. In fact, we consider reasonable that a measure is as more reliable as more similar to the other estimates obtained.

It remains to choose the techniques that presumably have correctly responded. We consider reliable the techniques with s_i equal to the minimum value of S (s_{min}). This first analysis allows to deducing that one or more eye detection techniques can be considered reliable. This information will be used by them for a transitional period chosen equal to M frames. After the first M frames we enrich our processing considering the historical results (in our implementation M is equal to 20). To consider the historical behavior of the system, a counter (n_i) is associated to each technique and we increase his value at each hit (s_i=s_{min}). The value of the counter can be used to determine a prior probability (see eq. 8) related to the reliability of the obtained estimation. The described method is summarized in Table 1

$$p_i = \frac{n_i}{\sum_{j=1}^{N} n_j} \text{ with } i = 1,\ldots,N \tag{8}$$

Finally we sum the prior probabilities (see eq. 9 and eq. 10) of the techniques with minimum s_i:

$$p^* = \sum_i p_i \delta_i \text{ with } i = 1,\ldots,N \tag{9}$$

$$\delta_i = \begin{cases} 1 & \text{if } \mathrm{s_i} = s_{\min} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

If p^* is bigger than a prefixed threshold ($p^*>T$), then we select the point returned by most probable technique, among techniques that have s_i=s_{min}.

On the contrary case, we chose between $\boldsymbol{X}^*$ and the technique which return closest result to this.

In our experiments we chosen a threshold T equal to 0,6.

In the latest case, we preferred to take as final result one output of the eye detection method described before, only if that is distant less than K pixel (in our measures K=5) from $\boldsymbol{X}^*$.

After the first M frames:

- If $p^* \geq T$ is returned the result of the technique with maximum p_i (and with distance from $\boldsymbol{X}^*$ less than K);
- If $p^* < T$, and exist at least one $\boldsymbol{X}_i$ with distance from $\boldsymbol{X}^*$ less than K pixel, is returned the result offered from the technique that is nearest to $\boldsymbol{X}^*$;
- If $p^* < T$ and doesn't exist at least one $\boldsymbol{X}_i$ with distance from $\boldsymbol{X}^*$ less than K pixel, we discard all results and return $\boldsymbol{X}^*$;

After M frames, we also use different increasing method for the counters n_i.

In the first case, the counters will be increased for each technique with $s_i = s_{min}$.

In the second case, the counter of the successful technique will be increased.

In the third case, we won't increase any counter.

Now it is spontaneous to assign to this kind of approach the appellation of adaptive.

In fact, as more the playing techniques will cover several cases as more the system will reply precisely, robustly and with accuracy.

However in the competitive scheme the final result can depend from different techniques in different frames.

Cooperative approach, instead, take advantage from the efforts of the all techniques that give results using a particular evaluation criteria.

It's based on two operations:

- Barycentre calculus of all points (see eq. 11);
- Estimation of new barycentre under 2σ hypothesis.

$$\mathbf{B} \equiv (x_B, y_B) \equiv \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right) \tag{11}$$

Where n is the number of technique that returns a result.

After that is determined the standard deviation σ:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \left\| \mathbf{X}_i - \mathbf{B} \right\|^2}{n}} \tag{12}$$

All the points returned by the techniques that are less than 2σ distant from $\boldsymbol{B}$ will be used to calculate the new barycentre that will be the output of the cooperative scheme. We can assert that this algorithm aspires to eliminate the outliers to improve the accuracy results.

4 EXPERIMENTAL RESULTS

The system was tested during and after development by several users for a considerable number of hours in numerous environments with different external lighting conditions.

To evaluate the performance of the system in terms of accuracy and repeatability a considerable number of tests were carried out.

To produce a quantitative evaluation we compared the output of our system with a ground-truth reference obtained by manual segmentation of the video tests.

An estimation of the whole error (due to the system) can be evaluated from the comparison between the acquired coordinates and those of the manual segmentation (ground truth).

Carrying out then a statistical analysis on the measures we obtained information about the precision of the system calculating the mean error and the standard deviation of the error.

Such errors are expressed in pixel or fractions of pixel.

The measures have been realized asking 5 users to test 3 times the system using the graphic interface of the operating system and most commonly used applications.

Results obtained confirm validity of proposed solutions and allow an accuracy, precision and robustness comparing between the schemes and eye-detection techniques.

In Table 1 are summarized information related to the operating conditions of the measures, such as lighting type, number of webcam adopted, eventual worn glasses and/or shadows.

Table 1: Operating condition in experimental results.

Subject	Glasses	Webcam	Halogen lamp	Filament lamp
1	No	3		Yes
2	Yes	3		Yes
3	No	2	Yes	
4	Yes	3		
5	No	2		

As we said, we have chosen the six techniques for them complementary properties. A confirm of this is shown in Table 2 where we can see a schematic presentation of the principal properties of the techniques adopted.

Table 2: Principal properties of the techniques adopted. (a: tolerate movements, b: tolerate glasses, c: independent to the distance from camera, d: need low computational resources, e: tolerate occlusions, f: tolerate shadows)

	a	b	c	d	e	f
T1	-	n	-	n	y	-
T2	y	n	y	-	y	n
T3	y	y	n	y	-	n
T4	-	n	-	y	y	y
T5	n	y	-	-	n	y
T6	n	y	y	n	n	n

In Table 3 we present a direct comparison between the results obtained by the two schemes and the single eye-detection techniques. We indicated the competitive scheme with S1 and cooperative with S2. The single technique (T) is numbered in according to the order followed in this paper.

For each technique and scheme we report the mean error and standard deviation in pixels. We measure the response rate (%) for each technique and we show the further improvement introduced by the schemes. Response rate indicate the number of case in which a technique give a result.

Table 3: Experimental results (E: mean error; σ: standard deviation; %: response rate).

	S1	S2	T1	T2	T3	T4	T5	T6
E	6,1	6,4	8,8	6,5	8,1	6,7	12,9	11,5
σ	6,3	6,5	8,2	6,0	7,5	6,5	9,2	6,8
%	98,1	95,8	87,0	72,2	76,4	68,2	59,2	41,4

The results show that the schemes implemented reduce considerably mean error of the single techniques. The schemes also provide more continuity and robustness in offering results, contrarily to the single techniques.

5 CONCLUSIONS

In this paper, we proposed an innovative approach to the problem of the Eye-Tracking. Traditional eye-detectors, chosen for its properties, are merged by two different schemes (competitive and cooperative scheme). The described approach features high reliability and high robustness to noise and bad illumination. To illustrate our work, we introduced a proof-of-concept single camera remote eye-tracker and discussed its implementation and the obtained experimental results. More applications of the proposed approach are currently being investigated in our Lab to portable, handheld and wearable computers. At the moment, the main issues being dealt with are computational cost and power consumption reduction. Finally, we are realizing a comparative study on a number of (more) sophisticated different cooperative schemes to obtain a further improvement in accuracy and reliability.

REFERENCES

Viola, P., Jones, M., 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. In *CVPR'01, International Conference on Computer Vision And Pattern Recognition*. IEEE Computer Science Society.

Lienhart, R., Maydt, J., 2002. An Extended Set of Haar-like Features for Rapid Object Detection. In *ICIP'02, International Conference on Image Processing.* IEEE Signal Processing Society.

Freund, Y., Schapire, R., 1996. Experiments with a new boosting algorithm. In *ICML'96, International Conference on Machine Learning*. Morgan Kaufmann.

Gonzalez, R., Woods, R., 2002. *Digital Image Processing.* Prentice Hall Press, 2nd edition.

Freund, Y., Schapire, R., 1997. A Decision-Theoretic Generalization of On-Line learning and an application to boosting. *Journal of Computer and System Sciences,* Elsevier Inc.

Hsu, R., Abdel-Mottaleb, M., Jain, A., 2002. Face detection in color images. *Transactions on Pattern Analysis and Machine Intellegence.* IEEE Computer Science Society.

Coifman, B., Beymer, D., McLauchlan, P., Malik, J., 1998. A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C 6.* Elsevier Inc.

Babcock, J., Pelz, J., 2004. Building a lightweight eyetracking headgear. In *ETRA'04, International Symposium on Eye Tracking Research and Applications.* ACM Press.

Ji, Q., Zhu, Z. 2004. Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications.* Springer.

Tomasi, C., Kanade, T., 1992. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision.* Springer.

Black, M., Jepson, A., 1998. EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation. *International Journal of Computer Vision.* Springer.

Morimoto, C., Koons, D., Amir, A., Flickner M., 1998. Pupil detection and tracking using multiple light sources. *Technical Report RJ-10117, IBM Almaden Research Center.*

Morimoto, C., Koons, D., Amir, A., Flickner M., 1998. Real-time detection of eyes and faces. In *PUI'98 International Workshop on Perceptual User Interfaces.* ACM Press.

Morimoto, C., Mimica, M., 2005. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*. Elsevier Inc.

Zhu, Z., Ji, Q., 2005. Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding*. Elsevier Inc.

Jacob, R., 1991. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*. ACM Press.

Shi, J., Tomasi, C., 1994. Good features to track. In *CVPR'94, International Conference on Computer Vision And Pattern Recognition*. IEEE Computer Science Society.

Ramadan, S., Almageed, W., Smith, C., 2002. Eye tracking using active deformable models. In *ICVGIP'02, Indian Conference on Computer Vision, Graphics and Image Processing*. Springer.

Kawato, S., Tetsutani, N., 2002. Real-time detection of between-the-eyes with a circle frequency filter. In *ACCV'02, Asian Conference on Computer Vision*. Springer.

Kawato, S., Tetsutani, N., 2002. Detection and tracking of eyes for gaze-camera control. In *VI'02 International. Conference on Vision Interface*. Springer.

Canny, J., 1986. A Computational Approach to Edge Detection, *Transactions on Pattern Analysis and Machine Intellegence*. IEEE Computer Science Society.

Bouguet, J., 1999. Pyramidal Implementation of the Lucas Kanade Feature Tracker. *Intel Corporation Microprocessor Research Labs.*

Daugman, J., 1993. High confidence visual recognition of persons by a test of statistical independence. *Transactions on Pattern Analysis and Machine Intellegence.* IEEE Computer Science Society.

Haro, A., Flickner, M., Essa, I, 2000. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In *CVPR'00, International Conference on Computer Vision And Pattern Recognition*. IEEE Computer Science Society.

Peng, K., Chen, L., Ruan, S., Kukharev, G., 2005. A Robust Algorithm for Eye Detection on Gray Intensity Face without Spectacles. *Journal of Computer Science & Technology*. Springer.

Vester-Christensen, M., Leimberg, D., Ersbøll, B., Hansen, L., 2005. Heuristics For Speeding Up Gaze Estimation. In *SSBA'05, Svenska Symposium i Bildanalys*.

Li, D., Babcock, J., Parkhurst, D., 2006. OpenEyes: A low-cost head-mounted eye-tracking solution. In ETRA'06, *International Symposium on Eye Tracking Research and Applications*. ACM Press.

Li, D., Winfield, D., Parkhurst, D., 2005. Starburst: A Hybrid Algorithm For Video-Based Eye Tracking Combining Feature-Based And Model-Based Approaches. In *CVPR'05, International Conference on Computer Vision And Pattern Recognition*. IEEE Computer Science Society.

Matsumoto, Y., Zelinsky., A., 2000. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In *AFGR'00, International Conference on Automatic Face and Gesture Recognition.* IEEE Computer Science Society.

King, I., Xu, L., 1997. Localized Principal Component Analysis Learning for Face Feature Extraction and Recognition. *Workshop on 3D Computer Vision.* Chinese University of Hong Kong.