Improved Likelihood Function in Particle-based IR Eye Tracking

Dan Witzner Hansen[1] Riad Hammoud[2] Ronald Satria[1] Jakob Sørensen[1]

Dept. of Innovation[1]
IT University, Copenhagen
DK-2300 Copenhagen
{witzner,leon98usdk,vanity}@itu.dk

Delphi Electronics and Safety[2]
M/C E110, P.O. Box 9005
Kokomo, Indiana 46904-9005
riad.hammoud@delphi.com

Abstract

In this paper we propose a log likelihood-ratio function of foreground and background models used in a particle filter to track the eye region in dark-bright pupil image sequences. This model fuses information from both dark and bright pupil images and their difference image into one model. Our enhanced tracker overcomes the issues of prior selection of static thresholds during the detection of feature observations in the bright-dark difference images. The auto-initialization process is performed using cascaded classifier trained using adaboost and adapted to IR eye images. Experiments show good performance in challenging sequences with test subjects showing large head movements and under significant light conditions.

1. Introduction

As one of the most salient features of the human face, eyes play an important role in interpreting and understanding a person's desires, needs, and emotional states. In addition, the unique geometric, photometric, and motion characteristics of the eyes also provide important visual cues for face detection, face recognition, and for facial expression understanding. Robust non-intrusive eye detection and tracking is therefore crucial for human computer interaction, attentive user interfaces, and understanding human affective states and is gaining importance outside laboratory experiments and even found in domestic appliances. For example eye tracking is used for driver fatigue and behavior [12], eye typing and in connection with rendering digital displays [2].

Direct eye detection and tracking methods search for eyes without prior information about face location, and can further be classified into passive and active methods. Passive eye detectors work on use images taken in natural scenes, without any special illumination. Good light conditions often lead to greater success and less effort on algorithm research and development. In the vision-based eye tracking methods, it is found that the use of near infrared light (reflected infrared) can be very rewarding in terms of ease and efficiency. This includes practically all stages in the eye tracker, starting from the detection, to tracking and gaze estimation. Active eye-detection and tracking methods employ special IR illumination. One of their limitations is that they are mainly applicable in controlled indoor environments as the amount of IR light in outdoor scenes may seriously disturb the image observations.

When IR light falls on the eye, part of it is reflected back, through the pupil, in a tiny ray pointing directly towards the light source. When a light source is located close to the optical axis of the camera (on-axis light), the captured image shows a bright pupil. This effect is similar to the the red-eye effect when using flashlight in photography. When a light source is located away from the camera optical axis (off-axis), the image shows a dark pupil. However, neither of these light schemes solely allow for robust results, as there are also other bright and dark objects in the scene that would generate pupil-like regions in the image.

One of the limitations of these systems is their use of thresholds when tracking the eyes. Defining thresholds can be difficult to define generically as the light conditions and head poses may influence the image observations of the eye. In this paper we propose a method based on particle filtering for tracking the eye. The method uses a new likelihood model for the image observations which avoids explicit definition of features and corresponding thresholds. For initial detection a method based on a cascaded classifier is used.

The proposed method uses an near infrared illumination (780 nm) to produce the bright pupil effect. The method is illustrated in figure 1 and consists of two parts: eye detection and eye tracking. Both methods are accomplished by simultaneously using the bright/dark pupil effects under active IR illumination and statistics of the eye appearance pattern under ambient illumination. The eye detection method

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

employs boosting methods on Haar wavelet classifiers. To meet the properties of the eye's appearance a set of additional Haar features are added compared with the original method. The eye tracking method employs particle filtering as it is capable of representing multi-modal distributions and has proven robust in clutter. In contrast to other methods we include the intensity distributions of the eye regions in conjunction with the information from the difference image in one unified model. The method is capable of tracking the eye under changing light conditions, when the bright pupil effect disappears and when people wear glasses.

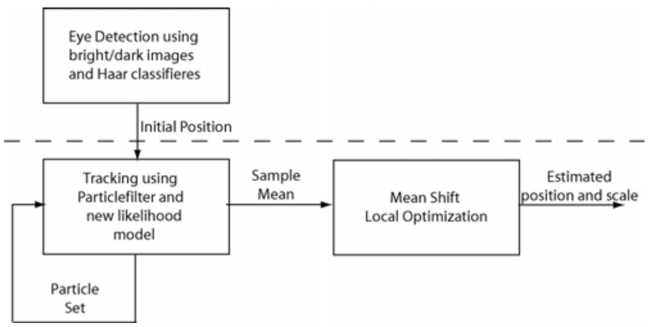

Figure 1. Detection and tracking model. Initially the eye regions are found through a cascaded Haar classifier. The eyes are tracked through particle filtering. The weighted mean of the particle set is used for initializing a local Means Shift mode optimization.

The paper is organized as follows: Section 2 provides an review of current eye tracking methods. In section 3 a detection method based on a cascaded classifier in dark and bright pupil images is presented and section 4 describes the tracking method proposed in this paper. The new eye region likelihood model for dark and bright pupil images is described in section 5. The dynamics of the tracking methods is given in section 6. The test results of using the model in a particle filtering setting is provided in section 7. We conclude the paper in section 8.

2. Current Eye Tracking Methods

Methods based on active light is the most predominate in both research [1, 8, 26, 27] and in commercial systems [22, 28, 29]. Eye detection and tracking based on the active remote IR illumination is simple and effective. The eye can be tracked efficiently by tracking the bright areas in the difference image resulting from subtracting the dark pupil image from the bright pupil image. Temporal filtering and geometric constraints are often incorporated to remove spurious candidates and to group the positive candidates. Most of these methods require distinct bright/dark pupil images to work well and thus strongly depend on the brightness and size of the pupils. Consequently, these methods are affected by factors such as occlusion from eye closure and head orientation, interferences from external illumination, and the distances of the subjects to the camera.

Ebisawa et al. [8] use a novel synchronization scheme in which the difference between images obtained from on axis and off axis light emitters are used for tracking. Kalman filtering and Mean shift tracking are more recently applied in similar approaches [35]. The success of these approaches is highly dependent on external light sources and the apparent size of the pupil. Efforts are made to focus on improving eye tracking under various light conditions. Sun light and glasses can seriously disturb the reflective properties of IR light. Methods using IR can therefore be less reliable in these situations and methods exist that address these issues [35].

Eye tracking and detection methods fall broadly within three categories, namely *deformable templates*, *appearance-based* and *feature-based* methods. Deformable template and appearance-based methods rely on building models directly on the appearance of the eye region while the feature-based methods rely on extraction of local features of the region. The latter methods are largely bottom up while template and appearance-based are generally top-down approaches. That is, feature-based methods rely on fitting the image features to the model while appearance and deformable template-based methods strive to fit the model to the image.

In general appearance models detect and track eyes based on the photometry of the eye region. A simple method to track eyes is through template-based correlation. Tracking is performed by correlation maximization of the target model in a search region. Grauman et al. [10] use background subtraction and anthropomorphic constraints to initialize a correlation-based tracker. Trackers based on template matching and stereo cameras seem to produce robust results [25]. Using fixed templates may yield good results, but the model may be a too committed model as variability is neglected. King et al [20] applies a linear principal component analysis approach to find the principal components of the training eye patches. Haro et al [14] 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. More sophisticated learning techniques are proposed in [11, 31, 35]. The support vector machines (SVM) classifier [3] is a very popular strategy and is used to train dark-pupil eye patches and non-eye patches, on gray-scale image vectors and invariant local jet, [11]. Vogelhuber at al. [31] use a Gaussian Mixture Model to learn the different aspect (variability) of the left eye and right eye patches in the local jet feature space.

Feature-based methods extract particular features such as skin-color, color distribution of the eye region. These features could be tracking the in-between eyes [19, 32], filter banks [9, 15, 30], iris location using the Hough transform [33], facial symmetries [24] Eye tracking methods committed to using explicit feature detection (such as edges) rely on thresholds. Defining thresholds can in general be difficult since light conditions and image focus change. Therefore methods relying on explicit feature detection may be vulnerable to these changes.

Deformable template-based method [5, 7, 13, 18, 21, 34] rely on a generic template which is matched to the image for example through energy minimization. In particular deformable templates [34], construct an eye model in which the eye is located through energy minimization. In the experiments it is found that the initial position of the template is critical. Another problem lies in describing the templates. Whenever analytical approximations are made to the image, the system has to be robust to variations of the template and the actual image. The deformable template-based methods seem logical and are generally accurate. They are also computationally demanding, require high contrast images and usually needs to be initialized close to the eye. While the shape and boundaries of the eye are important to model so is the texture within the regions. For example the sclera is usually white while the region of the iris is darker. Witzner et al. [13] propose a method which uses Active Appearance Models for local optimization and a Mean Shift color tracker for handling larger movements.

3. Eye Detection Through AdaBoost

Fast and accurate detection method is required to make eye tracking reach the general public. We employ a method similar to the boosted classifier method proposed by [30] for face detection, but the feature set is extended with center surround like features [23] as to direct the classification towards the problem of eye detection. Figure 2 shows the base set of simple feature classifiers used in the method.

Using the cascaded classifier for eye detection directly on the dark or bright images may result in a fair amount of false detections. We follow the direction of robust eye detection methods [11, 35] where detection is performed in two stages: 1) a set of potential eye candidates is obtained through the dark/bright difference image. 2) Based on the sparse set of candidates each candidate is examined for consistency through a cascaded classifier. The limitation of this approach is that this stage relies on thresholding.

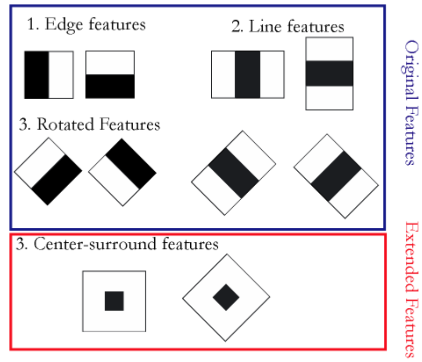

Figure 2. The set of Haar features originally proposed and the center surround feature added for eye detection.

3.1 Experimental Results

The classifier is trained from 2600 positive examples and 5000 negative examples resulting in a cascade classifier consisting of 19 stages. Figure 3 shows the candidates proposed by the cascaded classifier and using the difference image for generating the sparse candidate set. On a set of 260 images of near frontal faces the detection rate of detection both eyes simultaneously is about 80.7%. However, detection rates of single eyes (either left or right) is beyond 95%.

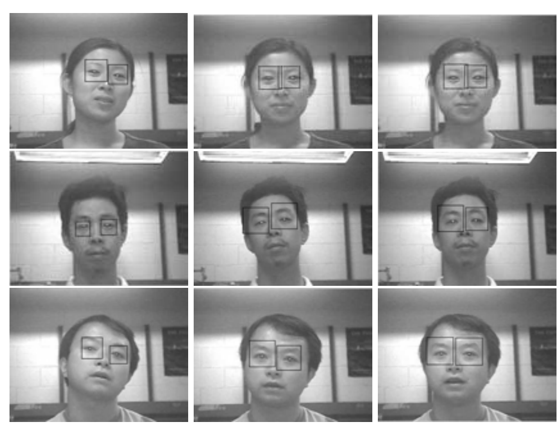

Figure 3. The result of using the difference image and a cascaded classifier for eye detection.

4. Tracking Method

The absence of bright pupils or even weak pupil intensity poses serious problems to the existing eye tracking methods using IR as they require relatively stable light conditions, users close to the camera, small out-of-plane face rotations, and open and un-occluded eyes. These conditions impose serious restrictions on both the system and the user. One reason for this lies in the frequent use of thresholds: the large difference in the dark and bright pupil images makes it tempting to use threshold values and connected components in the difference image. However, as the bright pupil effects may be reduced or eliminated due to light and head changes, appropriate threshold values may be difficult or impossible to set. Additionally, the use of threshold values may throw away useful information. Relying solely on the observations from the difference image is not necessarily sufficient as the bright pupil effects may disappear. In this case we use information from the the original bright and dark pupil images.

Continuous detection in individual frames neglects prior information from previous frames and are thus limited in scope. Even a 99 percent accurate detection system will fail every four seconds. This can be avoided by temporal filtering. The priors may in case of eye tracking include information about the history of previous positions, whether the eye is open or closed, reflectance properties and other information describing the appearance of the eye.

The proposed method is based on recursive estimation of the state variables of the iris. Given a sequence of T frames, at time t only data from the previous $t-1$ images are available. The states and measurements are represented by $\mathbf{x}_t$ and $\mathbf{y}_t$ respectively, and the previous states and measurements are represented $\underline{\mathbf{x}}_t = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$ and $\underline{\mathbf{y}}_t = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$. At time t the observation $\mathbf{y}_t$ is assumed independent of the previous state $\mathbf{x}_{t-1}$ and previous observation $\mathbf{y}_{t-1}$ given the current state $\mathbf{x}_t$.

Employing these assumptions the tracking problem can be stated as a Bayesian inference problem in the well known recursive relation [17]:

$$p(\mathbf{x}_{t+1} \mid \underline{\mathbf{y}}_{t+1}) \propto p(\mathbf{y}_t \mid \mathbf{x}_t) p(\mathbf{x}_{t+1} \mid \underline{\mathbf{y}}_t) \quad (1)$$

$$p(\mathbf{x}_{t+1} \mid \underline{\mathbf{y}}_t) = \int p(\mathbf{x}_{t+1} \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \underline{\mathbf{y}}_t) d\mathbf{x}_t \quad (2)$$

Particle filtering is used, as it allows for maintaining multiple hypotheses which in turn make it robust in clutter and capable of recovering from occlusion.

The aim of particle filtering is to approximate the filtering distribution $p(\mathbf{x}_t \mid \underline{\mathbf{y}}_t)$ by a weighted sample set $\mathbf{S}_t^N = \{(\mathbf{x}_t^{(n)}, \pi_t^{(n)})\}_{n=1}^N$, where $\mathbf{x}_t^{(n)}$ is the n^{th} instance of a state at time t with weight $\pi_t^{(n)}$. This sample set evolves into a new sample set $\mathbf{S}_{t+1}^N$, representing the posterior pdf (probability density function) $p(\mathbf{x}_{t+1} \mid \underline{\mathbf{y}}_{t+1})$ at time $t+1$. The object location in the particle filter is usually represented by the sample mean. *Factored sampling* is utilized in the CONDENSATION approach to particle filtering [17]: the samples are drawn from the prediction prior $p(\mathbf{x}_{t+1} \mid \underline{\mathbf{y}}_t)$, and sample weights are proportional to the observation likelihood $p(\mathbf{y}_t \mid \mathbf{x}_t)$. This approach is employed here. Particle filtering is particularly suitable for pupil tracking, because changes in pupil position are fast and do not follow a smooth and predictable pattern. The robustness of particle filters lies in maintaining a set of hypothesis. Generally, the larger the number of hypotheses, the better the chances to get accurate tracking results, but the slower the tracking speed. Therefore, there is a trade-off between tracking accuracy and speed.

Using particle filters in large images may require a large set of particles to sufficiently sample the spatial parameters. Adding particles to the particle set may only improve accuracy slowly, due to the sampling strategy. This added accuracy may become costly in terms of computation time. On the other hand, Mean Shift [4] is an efficient method to estimate the local mode of a distribution using a gradient-based optimization. This method combines particle filtering with the Mean Shift algorithm. In the approach particle filtering is used to obtain an estimate of the pupil position. The Mean Shift algorithm is then applied to find the local mode using the sample mean estimate from the particle filer for initialization. In this way the particle filter samples the posterior more effectively, while the Mean Shift reaches the local maximum.

5. Likelihood Model

In this section we describe our contribution of the likelihood model used in the particle filter-based eye tracker. The model uses the distribution of the eye regions in the dark and bright pupil images. The log ratio likelihood of the foreground and background models are used to avoid the use of thresholds. The use of thresholds should be minimized if possible as they may throw away useful information and can be difficult to define to be applicable in a wide variety of condition. The problem with thresholds becomes apparent for eye tracking in cases where the bright pupil disappears or when light condition change.

5.1. Likelihood of the image

The observations depend on the eye region position in the image. This means that the likelihoods computed for different locations are not comparable, as they are likelihoods of different observations. A better evaluation function is given by the likelihood of the entire image $\mathcal{I}$ given a region at location μ, as a function $f^*(\mathcal{I} \mid \mu)$ of the contour location μ.

Denote by $f_a(\mathcal{I})$ the likelihood of the image given no contour and by $f_R(\mathcal{I} \mid \mu)$ the ratio $f^*(\mathcal{I} \mid \mu)/f_a(\mathcal{I})$, then the log-likelihood of the entire image can be decomposed as follows:

$$\log f^*(\mathcal{I} \mid \mu) = \log f_a(\mathcal{I}) + \log f_R(\mathcal{I} \mid \mu) \qquad (3)$$

The first term on the right-hand side of equation 3 involves complex statistical dependencies between pixels, and is expensive to calculate as all image pixels must be inspected. Most importantly, the estimation of this term is needless as it is an additive term which is independent on the presence and location of the eye region. Consequently, in order to fit the eye model to the image, we consider only the log-likelihood ratio $\log f_R(\mathcal{I} \mid \mu)$. This derivation is fairly standard in the field of active contours, see e.g. [6]. Note that $f_R(\mathcal{I} \mid \mu)$ is the ratio between the likelihood for the hypothesis that the target is present; and the null hypothesis that the target is not present (equation 8). Hence the likelihood ratio can also be used for testing the hypothesis of the presence of a the eye region.

5.2. Eye region Model

Object Model Each pixel within a hypothesized eye region λ_j is considered. A region is defined by its location μ_j and scale Σ_j and we assume that the probability of an image coordinate $\mathbf{u}$ given the region is a Gaussian function of the distance to the mean μ:

$$g_j(\mathbf{u} \mid \lambda_j) = \frac{1}{2\pi \mid \sqrt{\Sigma_j} \mid} \exp(-\frac{1}{2}\Delta u_j^T \Sigma_j^{-1} \Delta u_j) \qquad (4)$$

where $\Delta\mathbf{u}$ is the distance of $\mathbf{u}$ from the centroid μ

Target regions can a priori have any distribution of gray levels and may change over time. This change may be due to head rotations and external disturbances from other light sources. In other words we need to represent the object region through its distribution. The Gaussian probability of the coordinates provides the basis for a kernel density estimate in which the weighting kernel reflects the importance (or the reliability) of that particular area of the object. The obvious reason for employing this kernel is that the peripheral areas are assumed to be the least reliable due to background clutter. The localization and positiveness of the kernel enables the calculation of the local mean and furthermore regularizes the distribution. The desire for differentiability is due to the computational efficiency. A differentiable kernel yield a differentiable similarity measure. The differentiability of kernels may, in turn, be used for finding the mode of the distribution, through gradient-based optimization.

Similarity of Target and Candidate Distributions The model consists of a target model $\mathbf{q}$ and a candidate $\mathbf{q}(\mathbf{u})$ evaluated in the coordinate $\mathbf{u}$. The similarity between the two distributions is expressed as measurements derived from the Bhattacharyya distance, which is defined by:

$$\xi(y) \equiv \sqrt{1 - \rho(y)} \qquad (5)$$

where

$$\rho(y) \equiv \rho[p(y), q] = \sum_{z=1}^{m} \sqrt{p_z(y) q_z} \qquad (6)$$

To make the observation model exploit both feature probabilities from both the dark and bright pupil images, the target and candidate distributions are constructed using the joint feature histogram of the both images within the region λ_i

Background Model The background model is defined to be those pixels where there is no eye present. We assume the background model only depends on the dark/bright difference image. It is well known that the the pdf of gray-level differences is well approximated by a Laplacian [16]:

$$f_L(\Delta I) = \frac{1}{Z_L} \exp\left(- \mid \frac{\Delta I}{\sigma} \mid\right) \qquad (7)$$

where ΔI is the gray level difference, σ is related to width of the laplacian function.

If there is no known object in the region λ_j, the pdf of the gray levels follows the laplacian distribution in equation 7. Assuming statistical independence between gray level differences in a hypothesized region λ_h the pdf of the observation in the absence of an eye is given by

$$f_a(\mathbf{I}) \equiv \prod_{i \in \lambda_h} f_L[\Delta I(i)] \qquad (8)$$

Note that the absence of an eye does not imply the absence of high value intensities: there can be regions within the background that exhibit similar properties as the bright and dark pupil effects. Even though the background is occluded by the eye it is still present at every location. By explicitly modeling the background thresholds become needles.

Likelihood ratio We are now in a position to formulate the likelihood ratio $f_R(\mathcal{I} \mid \mu)$ defined in equation 3. The likelihood ratio is given by equation 9:

$$p(I \mid x) = \frac{f_e}{f_a} = \frac{\xi(y)}{\prod_i f_L[\Delta I(i)]} \qquad (9)$$

Notice that the likelihood term fuses information from both dark and bright pupil images and their difference image into one model. In contrast to other eye tracking methods using dark and bright pupil images, this leads to a model which avoids explicit feature detection.

6. Eye State Model and Dynamics

The state of the eye region is modeled as a rectangle with position (x, y) and scale s. The dimensions of rectangle is fixed once initialized. The state to be estimated is therefore given by

$$X = (x, y, s)^T \tag{10}$$

Pupil movements can be very rapid from one image frame to another and thus acceleration is therefore less important to model. As no a priori knowledge of the movements is available, the dynamics is modeled as a first order auto regressive process using a time dependent Gaussian noise model: The actual dynamical model being used is therefore:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(0, \Sigma_t) \tag{11}$$

where Σ_t is the covariance matrix of the noise v_t at time t. Time dependence is included as to reflect that size changes may also change the apparent movements of the eye region. For this reason the elements in the covariance matrix corresponding to the two first components in the state model (x and y) are changed according to a linear function of the the size of the sample mean in the previous time step.

7. Experimental Results

In this section, we present experiments conducted to validate the performance of the eye tracker under different conditions.

In the all experiments the number of particles N is set to 100 with one iteration per frame. The noise parameters in the dynamical model is defined manually, but kept constant for the initial frame in each sequence. The noise term ν_t may change during tracking due to the adaptive dynamical model. Figure 4 shows the results of using the tracker under challenging scenarios with significant head poses. The bright pupil is reduced or vanishes during these sequences, but even in these cases tracking is maintained. The tracker may be slightly inaccurate when the bright pupil disappears (i.e. during eye closure). However as the pupil disappears there is little evidence of the true position.

Figure 5 shows the results of the tracker in the presence of glasses and under drastic and challenging light conditions. The accuracy of the method is validated against a manually annotated set of images. The Euclidian distance

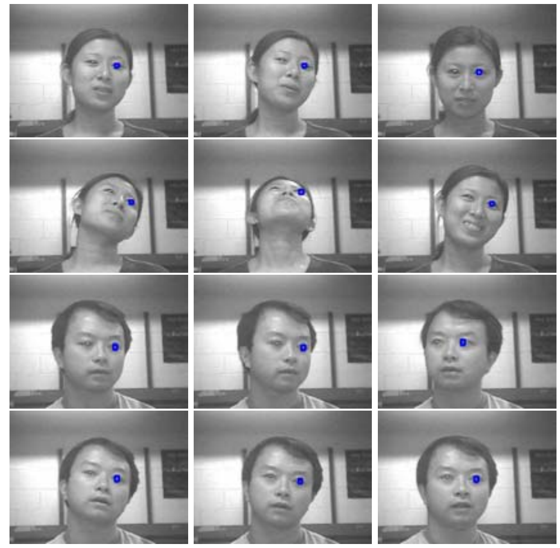

Figure 4. Tracking the left eye of the persons under significant head pose changes. The center of the eye region is indicated with a circle.

of the center of the rectangle found by the tracker and the annotated points is used for the evaluation. The accuracy of the tracker is calculated over 997 frames divided amongst 5 subjects in images of size 320×256 pixels. When neglecting the local optimization of the Mean Shift method, the accuracy of the tracker is 2.47 pixels with a variance of 1.362. Using the local optimization leads to a deviation of 1.9 pixels with a variance of 1.325 pixels and thus the use of Mean Shift optimization has improved the accuracy of the tracker. The added accuracy comes at the cost of a few additional iterations of the Mean Shift method. Notice that tracking performance could be even further improved by using additional local optimization. This may for instance be done through thresholding and other well known methods that have previously been used for eye tracking.

8. Conclusion

During the last decade, a tremendous effort has been made on developing robust and cheap eye tracking systems for various real-world applications. Most current methods use IR light and thresholding on the difference image to obtain the eye observations for detection and tracking the eyes. In this paper we also use thresholding on the difference image to obtain a set of candidates for the detection method, but we avoid explicit feature detection while tracking. This is an important distinction as it is important that tracking is

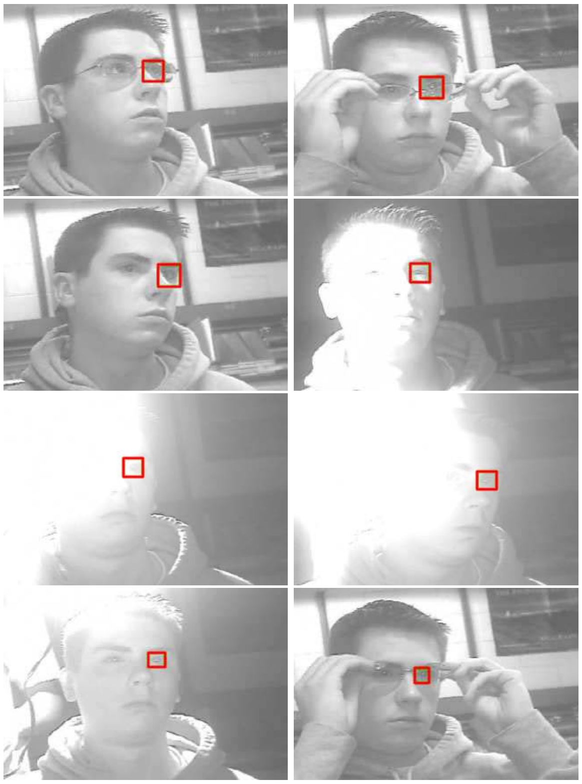

Figure 5. Tracking the eye with glasses and under challenging light conditions.

maintained even though image observations are poor. Detection can be reattempted in one of the following frames when the quality is improved. Relying on fixed thresholds in these cases seem less appropriate.

We use a set of candidate regions obtained in the difference image of the dark and bright pupil images to obtain a set of candidate regions. Then we apply a cascaded Haar-based eye-classifier on these candidate region for the final detection. Good results are obtained on frontal face images. To handle the cases where the image observations are poor we propose a likelihood model to be used in a particle filter setting to track the eye of a person. In contrast to many current methods this tracking model avoids the use of threshold. The tracker performs well on challenging sequences where the user undergoes significant head pose changes. Even though thresholding is not used for tracking it is still possible to use when applicable after tracking has been performed and in this way improve accuracy even further.

Acknowledgment

This research was partly supported by the European Network of Excellence COGAIN, Communication by Gaze Interaction, funded under the FP6/IST programme of the European Commission. The authors would like to thank Qiang Ji for providing the dataset.

References

[1] A. Amir, L. Zimet, A. Sangiovanni-Vincentelli, and S. Kao. An embedded system for an eye-detection sensor. *Computer Vision and Image Understanding*, 98(1):104–123, April 2005.

[2] Patrick Baudisch, Doug DeCarlo, Andrew T. Duchowski, and Wilson S. Geisler. Focusing on the essential: considering attention in display design. *Communications of the ACM*, 46(3):60–66, 2003.

[3] C. Burges. A tutorial on support vector machines for pattern revognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages II:142–149, 2000.

[5] T. F. Cootes and Taylor. Active shape models – 'smart snakes'. In *Proc. British Machine Vision Conf., BMVC92*, pages 266–275, 1992.

[6] J. Coughlan, A. Yuille, C. English, and D. Snow. Efficient deformable template detection and localization without user initialization. *Computer Vision and Image Understanding*, 78(3):303–319, 2000.

[7] J.Y. Deng and F. Lai. Region-based template deformation and masking for eye-feature extraction and description. *Pattern Recognition*, 30:403–419, 1997.

[8] Y. Ebisawa and S. Satoh. Effectiveness of pupil area detection technique using two light sources and image difference method. In *5th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, pages 1268–1269, 1993.

[9] I.R. Fasel, B. Fortenberry, and J.R. Movellan. A generative framework for real time object detection and classification. *Computer Vision and Image Understanding*, 98(1):182–210, April 2005.

[10] K. Grauman, M. Betke, J. Gips, and G.R. Bradski. Communication via eye blinks: Detection and duration analysis in real time. In *IEEE Computer Vision*

and Pattern Recognition (CVPR), pages I:1010–1017, 2001.

[11] Riad I. Hammoud. A robust eye position tracker based on invariant local features, eye motion and infrared-eye responses. In *SPIE Defense and Security Symposium, Automatic Target Recognition Conference, Proceedings of SPIE Vol. Nb. 5807*, 2005.

[12] Riad I. Hammoud and Dan Witzner Hansen. *Physics of the Automatic Target Recognition*, chapter Biophysics of the eye in computer vision: methods and advanced technologies. Springer Verlag, 2005.

[13] Dan Witzner Hansen, John Paulin Hansen, Mads Nielsen, Anders Sewerin Johansen, and Mikkel B. Stegmann. Eye typing using markov and active appearance models. In *IEEE Workshop on Applications on Computer Vision*, pages 132–136, 2003.

[14] A. Haro, M. Flickner, and I. Essa. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In *IEEE Conf. on Comp. Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, JUNE 2000.

[15] R. Herpers, M. Michaelis, K. Lichtenauer, and G. Sommer. Edge and keypoint detection in facial regions. In *International Conference on Automatic Face and Gesture-Recognition*, pages 212–217, 1996.

[16] J. Huang and D. Mumford. Statistics of natural images and models. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages I: 541–547, 1999.

[17] Michael Isard and Andrew Blake. Contour tracking by stochastic propagation of conditional density. In *European Conference on Computer Vision*, pages 343–356, 1996.

[18] J.P. Ivins and J. Porrill. A deformable model of the human iris for measuring small 3-dimensional eye movements. *Machine Vision and Applications*, 11(1):42–51, 1998.

[19] S. Kawato and N. Tetsutani. Detection and tracking of eyes for gaze-camera control, 2002.

[20] Irwin King and Lei Xu. Localized principal component analysis learning for face feature extraction and recognition. In *Proceedings to the Workshop on 3D Computer Vision*, pages 124–128, Shatin, Hong Kong, 1997.

[21] K.M. Lam and H. Yan. Locating and extracting the eye in human face images. *Pattern Recognition*, 29:771–779, 1996.

[22] http://www.eyegaze.com. LC Technologies INC., 2004.

[23] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, volume 1, pages 900–903, September 2002.

[24] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *PAMI*, pages 959–973, August 2003.

[25] Y. Matsumoto and A. Zelinsky. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In *International Conference on Automatic Face and Gesture Recognition*, pages 499–504, 2000.

[26] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *IVC*, 18(4):331–335, 2000.

[27] B. Noureddin, P.D. Lawrence, and C.F. Man. A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding*, 98(1):52–82, April 2005.

[28] http://www.smarteye.se. Smart Eyes A/B, 2004.

[29] http://www.tobii.se/. Tobii Technologies, 2004.

[30] P. Viola and M. Jones. Robust real-time face detection. In *International Conference on Computer Vision*, page II: 747, 2001.

[31] V. Vogelhuber and C. Schmid. Face detection based on generic local descriptors and spatial constraints. volume vol. 1, pages 1084–1087, 2000.

[32] Jie Yang, Rainer Stiefelhagen, Uwe Meier, and Alex Waibel. Robust detection of facial features by generalized symmetry. In *International Conference on Pattern Recognition*, pages I:117–120, 1992.

[33] David Young, Hilary Tunley, and Richard Samuels. Specialised hough transform and active contour methods for real-time eye tracking. Technical Report 386, School of Cognitive and Computing Sciences, University of Sussex, 1995.

[34] A. L. Yuille, P. W. Hallinan, and D.S Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.

[35] Z. Zhu and Q. Ji. Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding*, 98(1):124–154, April 2005.