Heuristics for speeding up gaze estimation

Denis Leimberg, Martin Vester-Christensen, Bjarne Kjær Ersbøll and Lars Kai Hansen
Department of Informatics and Mathematical Modelling
Technical University of Denmark
denis@kultvizion.dk, martin@kultvizion.dk, be@imm.dtu.dk, lkh@imm.dtu.dk

Abstract

A deformable template method for eye tracking on full face images is presented. The strengths of the method are that it is fast and retains accuracy independently of the resolution. We compare the method with a state of the art active contour approach, showing that the heuristic method is more accurate.

1 Introduction

Gaze is very important for human communication and also plays an increasing role for human computer interaction. Gaze can play a role, e.g., in understanding the emotional state for humans [1, 2], synthesizing emotions [3], and for estimation of attentional state [7]. Specific applications include devices for the disabled, e.g., using gaze as a replacement for a computer mouse and driver awareness monitoring to improve traffic safety [5].

It has been noted that the high cost of good gaze detection devices is a major road block for broader application of gaze technology, hence, there is a strong motivation for creating systems that are simple, inexpensive, and robust [4].

Figure 1: Examples of the dataset. The region surrounding the eyes can be found in various ways. We use a head tracking algorithm[5] based on Active Appearance Models. A subimage is extracted and subsequently processed by the eye tracking algorithms.

Detection of the human eye is a difficult task due to a weak contrast between the eye and the surrounding skin. As a consequence, many existing approaches use close-up cameras to obtain high-resolution images[4]. However, this imposes restrictions on head movements. Wang et al.[8] use a two camera setup to overcome the problem.

The present paper is inspired by the line of thinking mentioned above. We focus on some of the image processing issues. In particular we propose a robust algorithm for swift eye tracking in low-resolution video images. We compare this algorithm with a proven method[4] and relate the pixel-wise error to the precision of the gaze determination.

2 Deformable Template Matching

In many existing approaches the shape of the iris is modeled as a circle. This assumption is well-motivated when the camera pose coincides with the optical axis of the eye. When the gaze is off the optical axis, the circular iris is rotated in 3D space, and appears as an ellipse in the image plane. Thus, the shape of the contour changes as a function of the gaze direction and the camera pose. The objective is then to fit an ellipse to the pupil contour, which is characterized by a darker color compared to the iris. The ellipse is parameterized,

$$\mathbf{x} = (c_x, c_y, \lambda_1, \lambda_2, \theta)\,, \tag{1}$$

where (c_x, c_y) is the ellipse centroid, λ_1 and λ_2 are the lengths of the major and minor axis respectively. θ is the orientation of the ellipse.

The pupil region P is the part of the image I spanned by the ellipse parameterized by $\mathbf{x}$. The background region B is defined as the pixels inside an ellipse, surrounding but not included in P, as seen in figure 2. When region P contains the entire object, B must be outside the object, and thus the difference in average pixel intensity is maximal. To ensure equal weighting of the two regions, they have the same area.

The pupil contour can now be estimated by minimizing the cost function,

$$\mathcal{E} = \mathrm{Av}(P) - \mathrm{Av}(B), \tag{2}$$

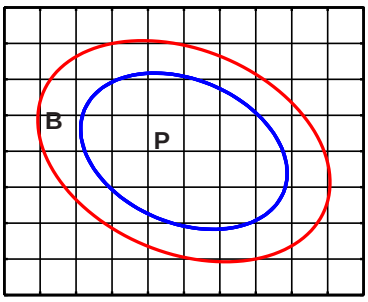

Figure 2: The deformable template model. Region P is the inner circle, and region B is the ring around it.

where $\mathrm{Av}(B)$ and $\mathrm{Av}(P)$ are the average pixel intensities of the background - in this case the iris - and pupil region respectively.

The model is deformed by Newton optimization given an appropriate starting point. Due to rapid eye movements[6], the algorithm may break down if one uses the previous state as initial guess of the current state, since the starting point may be too far from the true state. As a consequence, we use a simple 'double threshold' estimate of the pupil region as starting point.

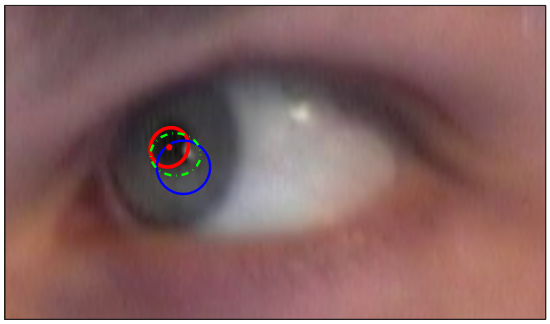

Figure 3: The blue ellipse indicates the starting point of the pupil contour. The template is iteratively deformed by an optimizer; one of the iterations is depicted in green. The red ellipse indicates the resulting estimate of the contour.

An example of the optimization of the deformable model is seen in figure 3.

2.1 Constraining the Deformation

Although a deformable template model is capable of tracking changes in the pupil shape, there are also some major drawbacks. Corneal reflections, caused by illumination, may confuse the algorithm and cause it to deform unnaturally. In the worst case the shape may grow or shrink until the algorithm collapses.

We propose to constrain the deformation of the model in the optimization step by adding a regularization term.

3 EM Contour Tracking

The iris is circular and characterized by a large contrast to the sclera. Therefore, it seems obvious to use a contour based tracker. Witzner et al.[4] describe an algorithm for tracking using active contours and particle filtering. A generative model is formulated which combines a dynamic model of state propagation and an observation model relating the contours to the image data. The current state is then found recursively by taking the sample mean of the estimated posterior probability.

The proposed method in this paper is based on [4], but extended with constraints and robust statistics.

3.1 The Dynamic Model

A dynamic model describes how the iris moves from frame to frame. Since the pupil movements are quite rapid at this time scale, the dynamics are modeled as Brownian motion (AR(1)),

$$\mathbf{x}_{t+1} = x_t + v_t, \quad v_t \sim \mathcal{N}(0, \mathbf{\Sigma}), \tag{3}$$

where $\mathbf{x}$ is the state from (1) and $\mathbf{\Sigma}$ is covariance matrix of the noise v_t.

3.2 The Observation Model

The observation model consists of two parts. A geometric component modeling the deformations of the iris by assuming a Gaussian distribution of all sample points along the contour. Secondly a texture component defining a pdf over pixel gray level differences given a contour location. Both components are joined and marginalized to produce a test of the hypothesis that there is a true contour present. The contour maximizing the combined hypotheses is chosen.

3.3 Active Contour Tracking

The tracking problem can be stated as a Bayesian inference problem by use of the recursive relation,

$$p(\mathbf{x}_{t+1}|\mathcal{M}_{t+1}) \propto p(\mathcal{M}_t|\mathbf{x}_t)p(\mathbf{x}_{t+1}|\mathcal{M}_t) \tag{4}$$

$$p(\mathbf{x}_{t+1}|\mathcal{M}_t) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathcal{M}_t)\mathrm{d}\mathbf{x}_t \tag{5}$$

where $\mathcal{M}_t$ is the observations. Particle filtering is used to estimate the optimal state in a new frame.

3.4 Constraining the Hypotheses

We propose to weigh the hypotheses through a sigmoid function. This has the effect of decreasing the evidence when the inner part of the ellipse is brighter than the surroundings. An example is depicted in figure 4. In addition, this relaxes the importance of the hypotheses along the contour around the eyelids, which improves the fit.

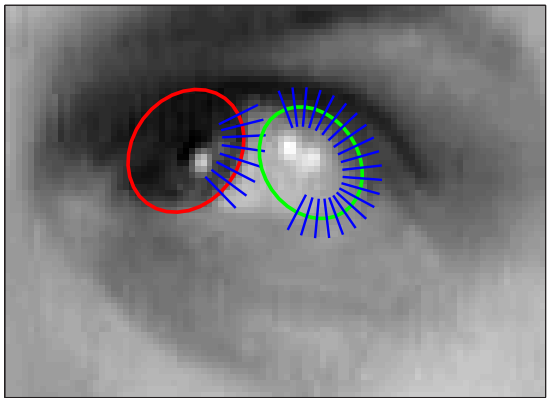

Figure 4: This figure illustrates the importance of the gray level constraint. Due to the general formulation of absolute gray level differences, the right contour has a greater likelihood, and the algorithm may thus fit to the sclera. Note the low contrast between iris and skin.

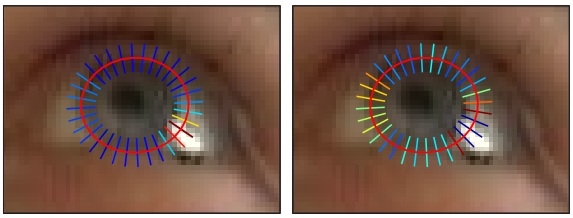

Figure 5: The relative normalized weighting of the hypotheses - Blue indicates low, while red indicates high scores. (*1*) Corneal reflections cause very distinct edges. Thus some hypotheses are weighted unreasonably high, which may confuse the algorithm. (*2*) This is solved by using robust statistics to remove outlying hypotheses.

3.5 Robust Statistics

By using robust statistics, hypotheses which obtain unreasonably high values compared to the others, are treated as outliers and therefore rejected, as seen in figure 5.

4 Results

A number of experiments have been performed with the proposed methods. We wish to investigate the importance of image resolution. Therefore the algorithms are evaluated on two datasets. One containing close up images, and one containing a down-sampled version hereof.

The algorithms estimate the center of the pupil. For each frame the error is recorded as the difference between a hand annotated ground truth and the output of the algorithms. This may lead to a biased result due to annotation error. However, this bias applies to all algorithms and a fair comparison can still be made.

Figure 6 and 7 depicts the error as a function of the number of particles used, for low resolution and high resolution images respectively. The errors for three different active contour(AC) algorithms are shows; basic, with EM refinement, with deformable template(DT) refinement. The error of the deformable template(DT) algorithm, initialized by double threshold, is inserted into the plot.

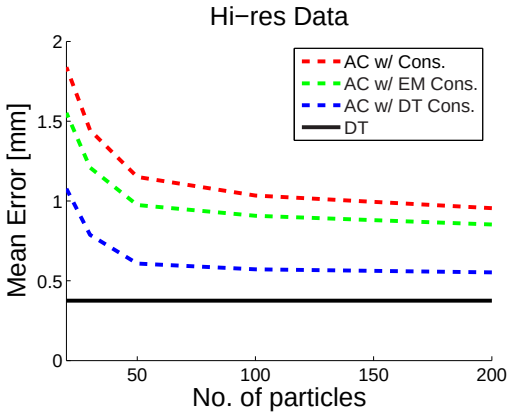

Figure 6: The error of the algorithms as a function of the number of particles for the high resolution data.

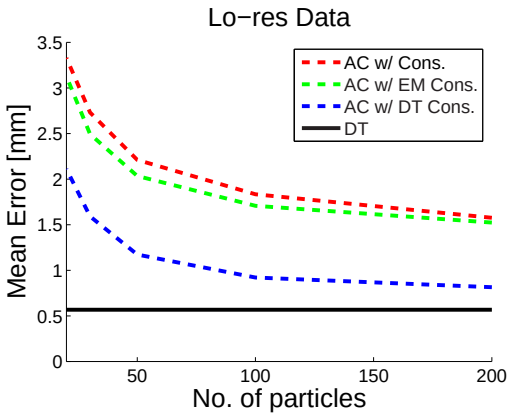

Figure 7: The error of the algorithms as a function of the number of particles for the low resolution data.

It can be seen that the proposed constraints on the active contour generally improves the accuracy of the fit. The refinement by the deformable template performs better than the EM method. The cost is an increased number of computations, which is resolution dependent. However, the deformable template method, initialized by double thresholding, is seen to outperform all active contour algorithms.

Hi-res	$E(x,y)$[mm]	$E(\theta)$	[frame/s]
AC	0.9	4.1	0.54
AC w/EM	0.8	3.7	0.49
AC w/DT	0.5	2.3	0.25
DT	0.3	1.4	2.2
Lo-res	$E(x,y)$[mm]	$E(\theta)$	[frame/s]
AC	1.5	7.3	0.57
AC w/EM	1.5	6.9	0.55
AC w/DT	0.8	3.7	0.49
DT	0.5	2.3	8.4

Table 1: Speed and precision comparison of the algorithms. The active contour uses 200 particles.

The table in figure 4 lists the mean error in accuracy in centimeters and degrees. Also listed is the computation time in frames per section of a Matlab implementation run on a 2.4Ghz PC. In general, the accuracy improves with high resolution as seen in table 4. However, the methods utilizing deformable template matching are less sensitive. The computation time for the basic active contour and EM refinement methods are independent of resolution. A significant increase in speed is noticed for the deformable template methods.

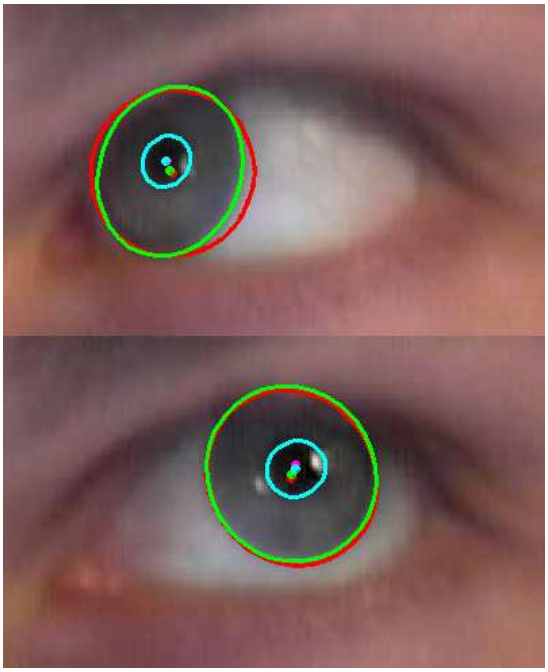

Figure 8: The resulting fit on two frames from a sequence - the red contour indicates the basic active contour, green indicates the EM refinement and the cyan indicates the deformable template initialized by the heuristic method. The top figure illustrates the benefit fitting to the pupil rather than the iris. Using robust statistic the influences from corneal reflections on the deformable template fit are ignored as depicted in the bottom image.

5 Conclusion

In this paper we have presented heuristics for improvement of the active contour method proposed by [4]. We have shown increased performance by using the prior knowledge that the iris is darker than its surroundings. This prevents the algorithm from fitting to the sclera as seen in figure 4.

Also presented is a novel approach to eye tracking based on a deformable template initialized by a simple heuristic. This enables the algorithm to overcome rapid eye movements. The active contour method handles these by broadening the state distribution and thus recovering the fit in a few frames. Furthermore, the accuracy is increased by fitting to the pupil rather than iris. This is particularly the case when a part of the iris is occluded as seen in figure 8.

It is shown that the deformable template model is accurate independent of resolution and it is very fast for low resolution images. This makes it useful for head pose independent eye tracking.

Acknowledgements

We wish to thank Hans Bruun Nielsen, Department of Informatics and Mathematical Modelling - Technical University of Denmark, for providing the optimization implementation. Additionally, we wish to thank Dan Witzner Hansen, IT-University of Copenhagen, for inspiring and insightful discussions. This research is supported by the IST Network of Excellence - Communication by Gaze Interaction (COGAIN).

References

[1] Jr. Adams, R.B. and R.E. Kleck. Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, 2003.

[2] R.B. Jr. Adams, H.L. Gordon, A.A. Baird, N. Ambady, and R.E. Kleck. Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, 300:1536–1537, 2003.

[3] Jonathan Gratch and Stacy Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. *Proceedings of the 5th International Conference on Autonomous Agents, Montreal, Canada*, June 2001.

[4] Dan W. Hansen and Arthur E. C. Pece. Iris tracking with feature free contours. In *Proc. workshop on Analysis and Modelling of Faces and Gestures: AMFG 2003*, October 2003.

[5] Takahiro Ishikawa, Simon Baker, Iain Matthews, and Takeo Kanade. Passive driver gaze tracking with active appearance models. In *Proceedings of the 11th World Congress on Intelligent Transportation Systems*, October 2004.

[6] J. Pelz, R. Canosa, J. Babcock, D. Kucharczyk, A. Silver, and D. Konno. Portable eyetracking: A study of natural eye movements, 2000.

[7] Rainer Stiefelhagen, Jie Yang, and Alex Waibel. Estimating focus of attention based on gaze and sound. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–9. ACM Press, 2001.

[8] J.G. Wang and E. Sung. Study on eye gaze estimation. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernectics*, 32(3):332–350, June 2002.