Eye Tracking using Active Deformable Models

Samah Ramadan Wael Abd-Almageed Christopher E. Smith
Robotics, Artificial Intelligence and Vision Laboratory
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131, USA
saramadan, wamageed, chsmith@eece.unm.edu

Abstract

Developing a robust, non-intrusive, inexpensive eye tracker is quite a challenge in the computer vision field. This paper introduces a new eye tracking approach based on an enhanced formulation of the statistical pressure snakes. Using a head-mounted camera, to eliminate the effect of head movements, we can detect and track the eye pupil robustly. Experimental results on video sequences for normal and blinking eyes are presented with high accuracy.

1. Introduction

The need for an accurate eye tracking system is constantly growing with the growth of its applications. For instance, eye tracking is a very important building block of multi-modal Human Computer Interaction (HCI) [15], more specifically vision-based user interfaces. Human Computer Interaction systems with multi-modal interfaces uses text, speech, lip reading, eye tracking, face recognition and tracking, and gesture and handwriting recognition to interpret user intent in a user interface environment.

In a vision-based user interface, people ideally can interact with computers by pointing with their eyes [10]. The direction of the eye gaze expresses the intent of the user. This can be used for people with disabilities who cannot use a mouse or a keyboard as in command-based user interfaces. Therefore, the eye gaze of the computer user can be used as an interface for controlling the applications.

Visual attention can also be inferred from analysis of eye movements. When we scan a visual scene, our eyes alternate between rapid jumps (i.e. “saccades”) and brief stops (or “fixations”). Although little information is processed during saccadic movement (because of the fast motion of images across the retina), eye fixations enable us to focus our attention like a spotlight. Therefore, one way of exploring what people pay attention to, in any given situation, is to use a computerized eye tracking system to record their visual attention strategies (“scan paths”) and the locations and durations of their fixations. This can be used, for example, to determine the visual attention of a vehicle operator. Eye blinking rate and duration can also be used to detect and evaluate the fatigue or sleepiness of the driver.

From a psychological point of view, tracking eye movements can be used during the process of lie detection, since certain patterns of eye movements can indicate deception.

Active deformable models, more commonly known as snakes, are a rapidly evolving area of computer vision. Deformable models were first introduced as a tool for image segmentation by Kass et al. [12]. The active contour model, or snake, is defined as an energy minimizing spline. The snake's energy depends on its shape and location within the image. Statistical approaches to active deformable models were also proposed. Ivins and Porill [8] proposed a statistical pressure snake, in which a pressure force, that is a function of the statistical characteristics of the image data, is used to drive the snake.

Deformable templates, proposed by Yuille [16], were used in eye feature extraction to track both eye locations and extract their parameters. Deformable templates consist of a parameterized template of simple geometric primitives and an energy function. Deformable templates have also been used to extract the eye and mouth boundary [3][7]. Much of this prior work did not yield real time results, since the original snake formulations had a very slow convergence time. Saccades and smooth pursuit were too fast for many early snakes for online tracking.

Current non-invasive eye tracking approaches rely heavily on using IR cameras with specific IR lightning structures. They depend on obtaining a dark and a bright pupil images for each frame, which has high computational complexity [11][6][13]. Neural networks are commonly used for eye and gaze tracking [11][4]. Abd-Almageed et al. proposed a stochastic model within a Kalman filter framework for tracking an eye undergoing pursuit movements.

In this paper we use the new pressure model proposed in [2] and the new curvature formulation of [14] to track the pupil location using an inexpensive IR camera mounted to the head with no specific structure of IR lightning source.

The head-mounted camera is used to eliminate the effect of the dynamics of head movements. The camera was built using a standard black and white pinhole micro camera with an IR filter (IR passes, visible light is filtered) produced from unexposed, but developed, E-6 35mm photographic slide film. The camera, filter, and mount were built for less than $60 (U.S.). Our system provides reliable, real time eye tracking using the IR camera connected to a consumer-end personal computer for video processing.

This paper is organized as follows, Section 2 provides an overview of the deformable model used. In Section 3 we discuss the eye tracking problem. We introduce our experimental results in Section 4. Section 5 concludes the paper and suggests some future research.

2. Snakes

The active contour model, or snake, is an energy minimizing spline. The snake's energy depends on its shape and location within the image. The energy functional to be minimized is given by

$$E_{Snake} = \int_0^1 (E_{Int}(S(u)) + E_{Img}(S(u)) + E_{Ext}(S(u)))du \tag{1}$$

where E_{Int} is the internal spline energy caused by stretching and bending (tension and curvature forces), E_{Img} represents the image energy that is a measure of the attraction to the image features such as edges and E_{Ext} is a measure of external constraints (e.g. higher level understanding of the general shape or user-applied energy.) S(u) is the parametric representation of the contour in the 2D space, and is given by Equation(2).

$$S(u) = (x(u), y(u))' \quad ; \quad u = [0, 1] \tag{2}$$

In most formulations of snakes, the image must contain strong edges to drive the image force of the contour, which causes a performance degradation in weak gradient fields. Ivins and Porrill [8] proposed a statistical approach for computing image pressure forces (statistical pressure snakes), where edge energy is replaced by region energy that is a function of the statistical characteristics of the object of interest. The pressure model was given by

$$F(S) = (1 - \frac{|I(s) - \mu|}{k\sigma})(\frac{\partial S}{\partial u})^{\perp} \tag{3}$$

where S is the contour, μ and σ are the mean and the standard deviation of a user-defined seed region, respectively, and k is a user-defined parameter that represents the spread of the population. The model of Equation(3) obviously assumes a Gaussian distribution of the gray levels of the region of interest. From Equation 3, a positive pressure is applied if the distance between the image density and the mean is within $k\sigma$ and negative pressure is applied otherwise. Another issue that arises with Ivins and Porrill mode is that the model ignores the color complexity of the background.

A new pressure model was introduced in [2], where two probability density functions (*pdf*s) for the target object and background are assumed. This model assumes multi-modal *pdf*s for both the background and the object. The *pdf*s are estimated using the Expectation Maximization (EM) algorithm [5]. The pressure force is then computed using the estimated *pdf*s. The pressure model is given by

$$F(S) = (p(x|O) - p(x|B))(\frac{\partial S}{\partial u})^{\perp} \tag{4}$$

where $p(x|O)$ and $p(x|B)$ are the probability density functions of the object and the background respectively. There are two main advantages of using the model of Equation(4). First, the model does not put any assumptions on the distribution of the coloring of the object. Second, the model accounts for the coloring for the background.

A new formulation for internal energy of snakes was proposed in [14] for calculating the internal energies of the active contour models. The tension force was eliminated since it produced undesirable behaviors and it was redundant with a newly formulated curvature. The curvature force was reevaluated, under the alternative definition of a smooth curve where the third derivative is constant. This new curvature model fits our application well since a circle is a curve with a minimal (zero) energy, and we expect the pupil to be roughly circular. The new formulation has reconsidered the internal forces against the original model resulting in a more robust model.

3. Eye Tracking

The new formulations of [2] and [14], discusses before, are used as follows. Using the pressure model introduced in [2], two mixtures of Gaussian distributions ares used to model the eye pupil and the background. There are several advantages of this model. First, the hand tuning of pressure model parameters (i.e. μ, σ, and k) is eliminated. Second, the model alleviates the problem of the precise manual initialization of the placement of the snake, because of the use of object and background probabilistic models. Finally, as stated before, the model accounts for the background's *pdf*.

Using the reformulated model introduced in [14], a single internal force (curvature) is assumed, resulting in a lower computational complexity and less sensitivity to changes in the weighting forces. Eliminating the tension force also allows to snake to both expand and collapse, while in the old model, the tension force was driving the snake always to collapse.

The calibration of the system requires that the target moves his eyes in all four extreme directions. The snake then is driven by the various forces to track the eye robustly. The problem of eye blinking is overcame using two correction mechanisms. The first is an active mechanism in which the contour was reinitialized and allowed to collapse again to the pupil region. A passive approach was also taken in which the contour was left until it touches the pupil region and therefore expands using the image forces. The main advantage of the statistical snake used here is the ability of the snake to expand after collapses when the eye blinks. No quantitative differences was noticed between the two approaches.

When the snake collapses that means that the eye is probably closed. By measuring the duration of collapse and blink frequency a system can be built to infer if the target is sleepy or fatigued and an alert can be given. The frequency of blinking and the direction of eye movements can also be analyzed psychologically to implement a robust lie detector.

4. Experimental Results

To realize the system, we used an infra-red head-mounted camera with a resolution of 640 $\times$ 480 pixels, connected to a video capture device in a 1 GHZ Pentium III computer running Windows 2000. We have achieved a tracking performance of approximately 29 frame/second using a non-optimized "C" language implementation. Ivins et al. [9] used deformable models for eye tracking. They achieved a tracking rate of 2-3 second/frame (240 $\times$ 180 pixles) on an SGI Indy machine, with a hybrid MATLAB-"C/C++" implementation.

Figures 1 and 2 show the histogram for the pupil and the background respectively. The "noise" in the histograms is a result of the specularities on the eye's surface. Figures 3 and 4 show the estimated *pdf*'s of the pupil $p(x|O)$ and the background $p(x|B)$ using the EM algorithm. It is obvious from the figure that the pupil *pdf* can be modelled by a single Gaussian distribution while the background can be modelled using a mixture of two Gaussian distributions (ignoring the noise). The two estimated *pdf*'s are used to segment out the eye pupil from the background. The snake is then driven using the pressure model given by equation 4. The magnitude and direction of the pressure will be determined by the term $(p(x|O) - p(x|B))$ and the spilt of the pressure in x and y directions is determined by $(\partial S)/\partial u)^{\perp}$.

The calibration of our tracking system is very easy, which requires moving eye pupil in all four extreme posi-

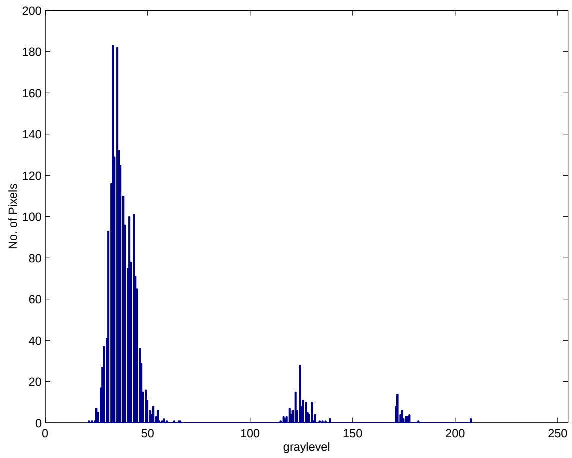

Figure 1: Histogram of the gray levels of the eye pupil.

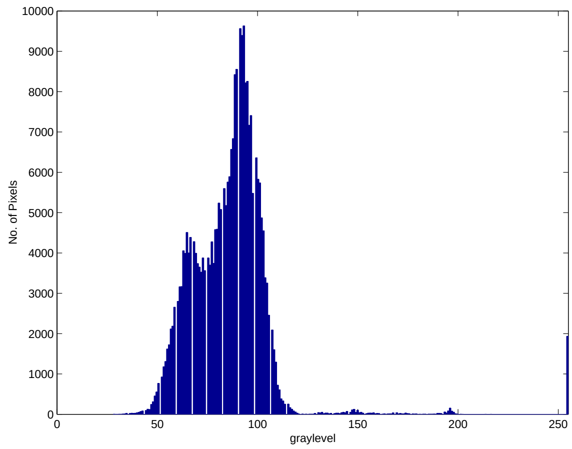

Figure 2: Histogram of the gray levels of the eye background.

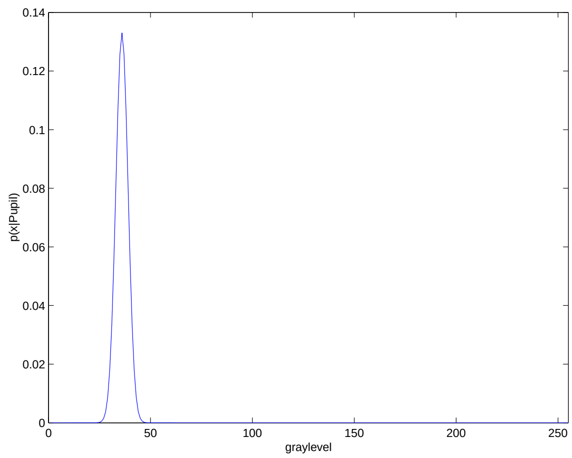

Figure 3: Estimated PDF for the gray levels of the eye pupil.

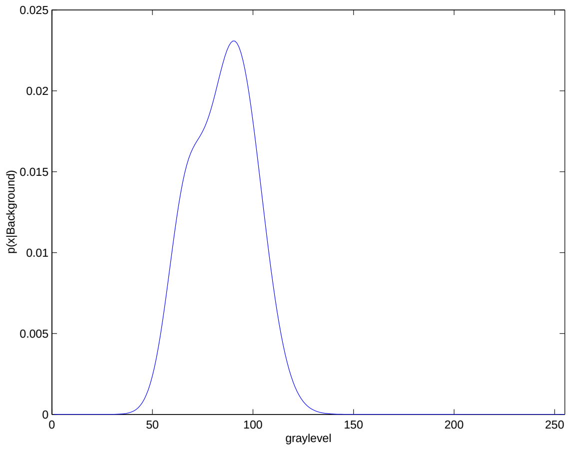

Figure 4: Estimated PDF for the for the gray levels of the eye background.

tions. After that the snake can be used to track the eye pupil robustly.

Figure 5 shows a video sequence and the tracking results using the proposed approach. A real-time robust performance was achieved. The snake was successfully able to track the eye pupil driven by the various forces including the object pressure force which was moving the snake to track the eye pupil.

Figure 6 shows the tracking results for an eye undergoing a blink. As is shown in Figure 6.b when the eye blinks the snake collapses as the object pressure force will be zero while the background pressure force will be non-zero. The loss of tracking will occur when the eye re-opens if the pupil coordinates and the snake coordinates were different as shown in Figure 6.c, 6.d. In this case the snake will lose the tracking of the eye pupil after it re-opens because the pupil and the snake are not overlapped. The action to be taken is based upon the duration of the collapse. If the snake and the pupil are overlapped again after a few frames, the snake will expand automatically as the object pressure force will capture it again as shown in Figure 6.d. If the duration of collapse was long, the snake is simply re-initialized to restore the snake forces.

To evaluate the tracking accuracy of our system we have manually measured the true position of the pupil manually and a comparison was made to the estimated snake centroid of the same frame. Figures 7 and 8 show the true pupil position compared to the estimated contour centroid. Our results show a very high accuracy of the tracking system.

5. Conclusions and Future Work

In this paper we introduced an eye tracking system using statistical pressure snakes. A new formulation of the original snakes was used with mixture models and reevaluation

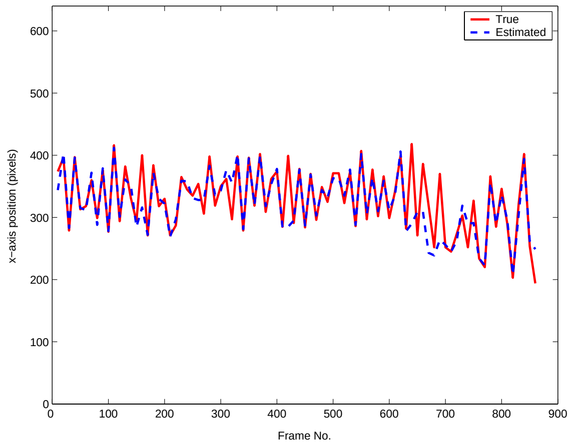

Figure 7: Tracking error in the x-axis.

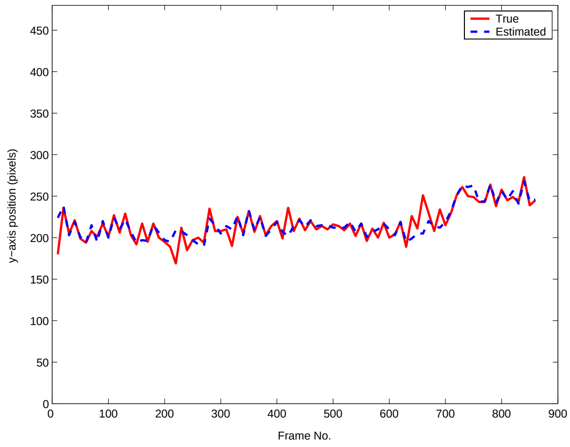

Figure 8: Tracking error in the y-axis.

of internal forces. Unlike previous eye tracking systems using snakes, our method can track smooth pursuit and saccades since our statistical snakes have lower computational overhead and since region pressure is non-zero across the entire image, unlike image gradient which has vast areas of zero or near zero gradient. The main advantage of using the new pressure model is the ability to track complex colored objects and complex colored background.

Our system tracks eye pupils using an IR head-mounted camera with no specific lighting structure. Our system can also detect eye blinking easily when the snake collapses. Using the new formulation allows the snake to both expand and collapse which added a great flexility and better performance compared to the old model. We presented the results of our tracking system using a smooth eye movement sequence as well as a sequence during which the eye undergone a blink. The performance of our tracking system

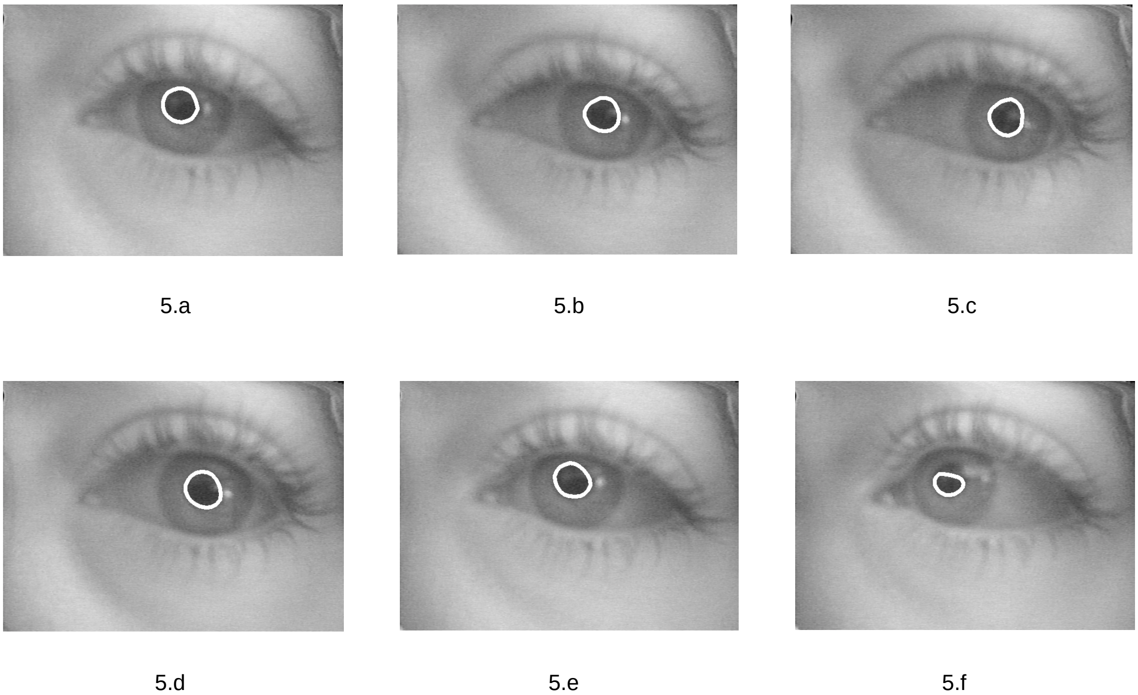

Figure 5: Tracking an eye moving smoothly.

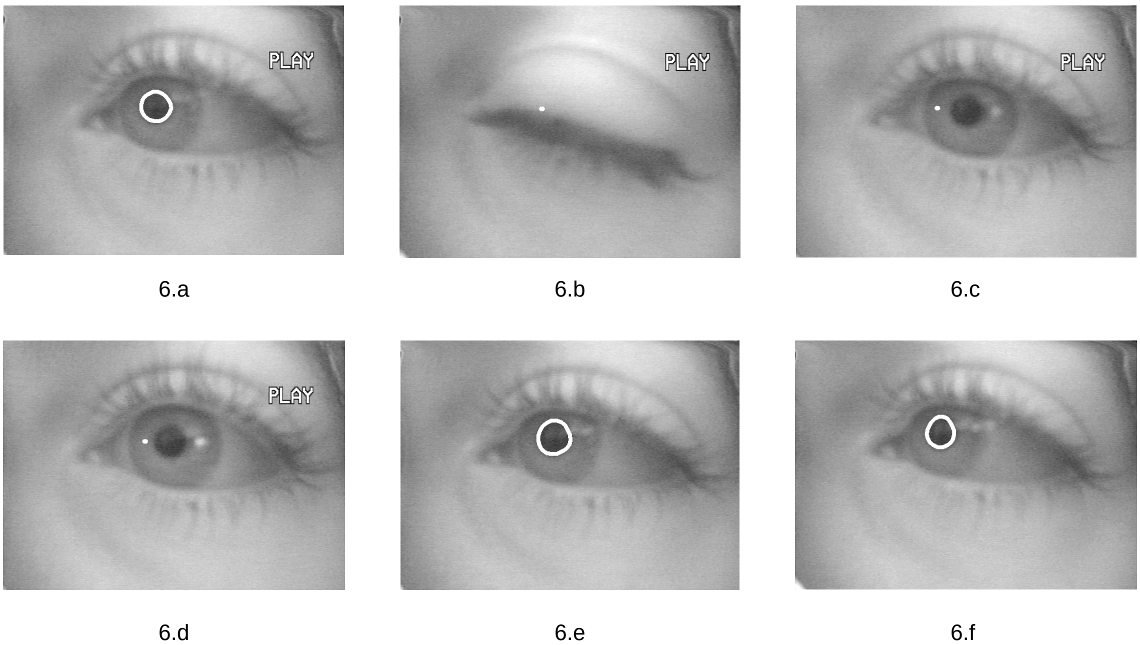

Figure 6: Tracker performance before, during and after an eye blink.

was evaluated against the real eye position data. The experiments demonstrate a high performance real-time tracking.

For future research, we want to use a Kalman filter as in [1] to obtain a more accurate estimate of the true eye position during pursuit motion of the eye given the computed position by the snake. Also, in case of tracking loss, a Kalman filter can provide a better estimate of the eye position. However, Kalman filter should be used carefully since saccades could be lost due to the "memory" of the filter. Also, we want to use Kalman filter to predict the position of the eye in the next frame and incorporate the predicted value in the motion equations of the snake.

Another area of future research is investigating the use of multiple snakes in a gaze estimation system, using a fixed camera instead of the head-mounted one. This can be used to implement a vision-based user interface.

A lie detector can be implemented using our eye pupil tracking system. From psychological point of view the pattern of eye movements can be used to infer the deception of the subject. By analyzing the estimated pupil movements we can detect lying. One potential problem in implementing such system, is the need for real data to train and evaluate the system.

References

[1] W. Abd-Almageed, M. S. Fadali, and G. Bebis. A non-intrusive kalman filter-based tracker for pursuit eye movements. In *American Control Conference, Alaska*, 2002.

[2] W. Abd-Almageed and C. E. Smith. Mixture models for dynamic statistical pressure snakes. In *IEEE International Conference on Patteren Recognition*, August 2002.

[3] G. C. andX. Li. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26(12):1739–1755, 1993.

[4] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. In *Working Notes: AAAI Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How?*, 1993.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistics Society*, B-39, 1977.

[6] A. Haro, M. Flickner, and I. Esaa. Detecting and tracking eyes by using their physiological properties, dynamics and appearance. In *IEEE CVPR*, 2000.

[7] C. Huang and C. Chen. Human facial feature extraction for face interpretation and recognition. *Pattern Recognition*, 25(12):1435–1444, 1992.

[8] J. Ivens and J. Porill. Active region models for segmenting medical images. In *Proceedings of the IEEE International Conference on Image Processing*, pages 227–31, 1994.

[9] J. Ivins, J. Porrill, and J. Frisby. A deformable model for the human iris driven by non-linear least squares minimization. In *Sixth International Conference on Image Processing and Its Applications*, 1997.

[10] R. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9(3):152–169, April 1991.

[11] Q. Ji and Z. Zhu. Eye and gaze tracking for interactive graphic display. In *Int. Symp. on Smart Graphics*, June 2002.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision*, pages 259–268, 1987.

[13] C. Morimoto, D. Knoos, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, Feb 2000.

[14] D. Perrin and C. Smith. Rethinking classical internal forces for active contour models. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2001.

[15] M. Vo and A. Waibel. Multimodal human-computer interaction. In *In Proc. Int. Symposium on Spoken Dialogue, pp. 95–101*, 1993.

[16] A. Yuille, D. S. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1989.