IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
21-23 September 2009, Rende (Cosenza), Italy

IR Based Pupil Tracking Using Optimized Particle Filter

Ognian Boumbarov [1], Stanislav Panev [2], Strahil Sokolov [3], Vladimir Kanchev [4]
Technical University of Sofia, Faculty of Telecommunications
8, Kliment Ohridski Str.
1000 Sofia, Bulgaria
[1] Assoc. Prof., olb@tu-sofia.bg
[2] PhD Student, s_panev@tu-sofia.bg
[3] PhD Student, s_sokolov@tu-sofia.bg
[4] PhD Student, v_kanchev@tu-sofia.bg

***Abstract* - In this article an algorithm is proposed for estimation of eye-pupil contour parameters in an image, by means of a particle filter and artificial infrared lighting. A description is given of the initial assumptions that must be made when the algorithm is used. Then the used system model is described, as well as the principles of evolving the state of the system in time. Also the method is depicted, for estimation of the hypotheses of eye-pupil presence in the analyzed image. This is necessary for the application of particle filtering in the algorithm. In addition, it is necessary to introduce a means for reduction of the number of hypotheses. We have also shown and analyzed the experimental results from the proposed algorithm.**

***Keywords* - eye, pupil, tracking, particle filtering, mean-shift**

I. Introduction

Nowadays there is a large numbers of systems for eye tracking available, both in commercial and in laboratory environment. Many of them are based on contact lenses, electrodes, specialized hardware, and infrared emitters. Such systems could easily be separated into three groups, according to the approach used: methods, using the electric potential of the human skin [1], methods that involve contact lenses [2] and methods that involve image analysis.

In general image based approaches are divided into two groups – appearance-based and model-based.

Appearance-based approaches directly treat an eye image as a high dimensional feature. Baluja and Pomerleau use a neural network to learn a mapping function between eye images and gaze points (display coordinates) using 2,000 training samples [3]. Xu *et al.* proposed a similar neural network-based method that uses more (3,000) training samples [4]. Tan *et al.* take a local interpolation approach to estimate unknown gaze point from 252 relatively sparse samples [5]. Recently, Williams *et al.* proposed a novel regression method called S3GP (Sparse, Semi-Supervised Gaussian Process), and applied it to the gaze estimation task with partially labeled (16 of 80) training samples [6]. Sugano *et al.* proposed appearance-based gaze estimation system based on an online learning approach [7]. Appearance-based approaches can make the system less restrictive, and can also be very robust even when used with relatively low-resolution cameras.

Model-based approaches use an explicit geometric model of the eye, and estimate its gaze direction using geometric eye features. For example one typical feature is the pupil glint vector [8, 9], the relative position of the pupil center and the specular reflection of a light source. Model-based approaches typically need to precisely locate small features on the eye using a high-resolution image and often require additional light sources but can be very accurate.

Several approaches exist for eye-tracking by using images. Some of them are largely dependent on active light-sources, such as infrared emitters [10]. Other approaches, on the other hand, do not use the information from an explicit light-source, but rather use information for improving the quality of the image. Still other approaches even exclude the usage of active lighting, but rather rely on natural light-sources.

In general, the pupil of the human eye has an elliptic shape and relatively large contrast in comparison to the iris. This is the reason why an algorithm for tracking could be used, based on contour estimation.

The rest of this article is organized as follows: in Section II an approach is described for contour estimation of the human eye's pupil, in an image obtained under infrared lighting. In order to locate the eye-pupil's contour in the image, an algorithm is used based on the so called active-contour. In particular, this article describes a method that uses particle filtering for generating of a certain number of contours of the pupil that can vary in: shape, ratio between the major and minor axes, and direction. As greater number of candidate contours are used, as greater accuracy we will achieve but at expense of longer calculation time. In section III an optimization for particle filter is applied by using EM and Mean-Shift methods along the measurement lines of the active contour, which reduces the requirements for lager particle number. In section IV we present our experimental results. In section V we conclude and discuss the future development of our algorithm.

II. Algorithm Description

Assumptions

To accomplish all the derivations, it is necessary to introduce some initial assumptions, which the algorithm will be based on. And they are as follows:

- Grey level differences (GLDs) between pixels along a measurement line are statistically independent;
- Intensity values of nearby pixels are correlated in the following manner: either they both belong to the object being tracked, or they both belong to the background: thus, a priori statistic dependencies between nearby pixels is assumed;
- There is no correlation between pixel values if they are on opposite sides of the object boundary;
- The shape of the contour is a subject of random local variability, which means that marginalization over local deformations is required for a Bayesian estimate of contour parameters.

Pupil Model

Geometrically the contour of the pupil is modeled as ellipse (Fig. 1). Thus the result of tracking is a five-dimensional vector with the following parameters:

$$\vec{x}=\left[c_x,\ c_y,\ \lambda_1,\ \lambda_2,\ \theta\right], \qquad (1)$$

where $\left(c_x,\ c_y\right)$ is the center of the pupil, λ_1 and λ_2 are the major and minor axes, and θ is the angle of the major axis with respect to the vertical coordinate. This vector represents the state of the system.

The pupil movements can be very rapid therefore they are modeled as Brownian motions (AR(1)). Thus the evolution of the state sequence is modeled [11]

$$\vec{x}_{t+1}=\vec{x}_t+\vec{v}_t,\quad \vec{v}_t\sim N\left(0,\ \Sigma_t\right), \qquad (2)$$

where Σ_t is the time dependent covariance matrix of the noise.

Estimating the Hypotheses for Contour Presence

The pdf of the GLDs is well approximated by a generalized Laplacian [12]:

$$p_L\left(\Delta M\right)=\frac{1}{Z_L}exp\left(-\left|\frac{\Delta M}{L}\right|^{\beta}\right), \qquad (3)$$

where ΔM is the GLD, β is a parameter approximately equal to 0.5, Z_L is a normalization constant and L is a parameter which depends on the distance between the two sampled image locations.

It is assumed that the geometric object model is subjected to random Gaussian deformation at each sample point on the contour and the prior pdf of deformations is defined:

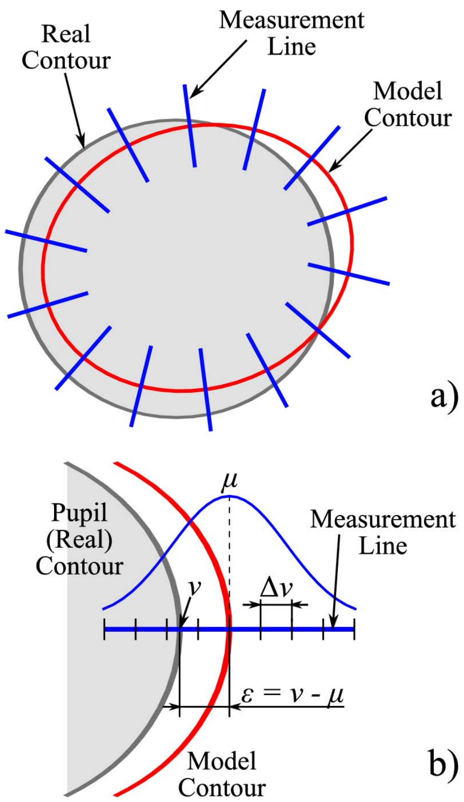

Fig. 1. Observation model.

$$p_D\left(\varepsilon\right) = \frac{1}{Z_D}exp\left(\frac{-\varepsilon^2}{2\sigma^2}\right), \qquad (4)$$

where Z_D is a normalization constant.

Using the definitions (3), (4), applying marginalization over them [13] and choosing $\beta = 0{,}5$, the point evaluation function becomes:

$$h\left(M|\mu\right) = h_0+log\sum_j exp\left[\sqrt{\frac{\left|\Delta M\left(j\Delta\nu\right)\right|}{L}}-\frac{\varepsilon_j^2}{2\sigma^2}\right], \qquad (5)$$

where $h_0 = log\frac{Z_L}{m}-log\frac{Z_D}{\Delta\nu}$, m is the number of grey levels, $\Delta\nu$ is the distance between two sampling points along a measurement line, $\varepsilon_j = j\Delta\nu-\mu$ is the distance from μ to the current point of the measurement line.

Particle Filtering

Let's suppose that the previous states of the system are known $X_{t-1} = \left\{x_i,\ i=0:t-1\right\}$. Then from the theory of Markov's chains [14]:

$$p\left(x_t|X_{t-1}\right) = p\left(x_t|x_{t-1}\right), \qquad (6)$$

which means that the current system state x_t depends only on the previous system state x_{t-1}. Therefore the particle filtering represents a first order autoregressive process. The set of measurements till the current moment is marked as $Y_t = \{y_k,\ k=1:t\}$. Let N_{ML} be the number of the measurement lines which belong to a single contour. Then for all the measurement lines of that contour is made and evaluation if there is an edge along them or not, i.e. for all the measurement lines the testing function is calculated $h(M_k^{(i)}|\mu)$. The result are summed and normalized, after which the new weight coefficients are determined:

$$w_t^{(i)} = p\left(Y_t^{(i)}|x_t^{(i)}\right) \equiv p\left(M_t^{(i)}|x_t^{(i)}\right) = \frac{\sum_{k=1}^{N_{ML}} h\left(M_k^{(i)}|\mu^{(i)}\right)}{\sum_{i=1}^{N}\sum_{k=1}^{N_{ML}} h\left(M_k^{(i)}|\mu^{(i)}\right)} \quad (7)$$

The new system state should be estimated, as follows:

$$x_t = \hat{x}_t = \sum_{i=1}^{N} w_t^{(i)} x_t^{(i)}, \quad (8)$$

which is used for following optimization by EM and Mean-Shift.

Constraining the Hypotheses

The proposed active contour method may fail due to various non-ideal conditions in the image. The reflections from cornea and sclera, caused by the artificial lightening, can cause that. That's why it is necessary to be implemented constraints for the hypothesis. This can be achieved if some anatomical characteristics of the human eye are taken into consideration. For example it is known that the average intensity of the iris in IR spectrum is greater than the average intensity of the pupil. Therefore the inner part of the contour area will be darker than the area outside of it, i.e. the direction of the contour edge is known in advance. In order to make an estimation it is necessary to calculate the average intensity of pixels lying on the inner for the contour part of the measurement line and the average intensity of the pixels lying on the outer part of the measurement line. This way after calculating the difference between the two average intensities is estimated which region is darker and which is lighter. And if the result doesn't correspond to the anatomical facts this hypothesis will be ignored.

III. Optimization

As has been mentioned above, the optimization stage is required to reduce the number of needed particles and to increase the accuracy and efficiency of the tracking. Here this is achieved by using Expectation-Maximization (EM) algorithm combined with the information for GLDs in natural images [12] and Mean-Shift algorithm.

Optimization by EM algorithm

On Fig. 2a is shown a part of the model contour (the red curve) after particle filtration and one of the measurement lines (blue dotted). As can be seen the model contour doesn't fit perfectly to the real pupil contour. The red dot shows where exactly is the real contour, e.i. where the model contour should be placed after the optimization. The task now is to find that exact location (the red dot) and to measure the distance between it and the center of the measurement line, which will give us an impression about how far is the model from the target. Two approaches have been tested. The first one relies on natural image statistics [12] and the second one relies on Mean-shift filtering along the measurement line [15].

On Fig. 2b the probability function can be seen along the measurement line in the left, which shows what the probability is that the real contour exists at the position. It is calculated by using the statistics of the GLDs and (3). The exact position of the contour is estimated by finding the center of the mass of that function:

$$\hat{v} = \sum_j p_j\, j\, \Delta v \quad (9)$$

Another approach to measure the exact position of the real contour over a ML is using the Mean-Shift smoothing filtering along a vector [15]. In the current situation the image intensity profile along a ML represents the vector values. The Mean-Shift smoothing is suitable as it underlines the border between homogeneous areas in the image.

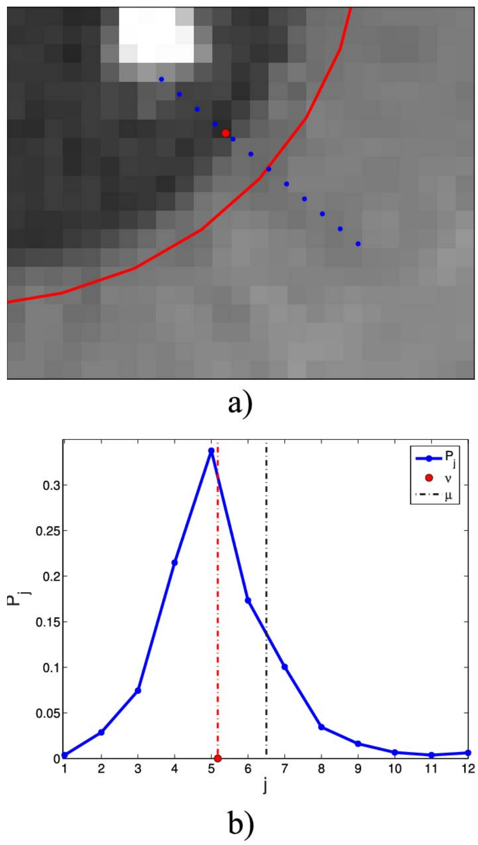

Fig. 2: Tracking optimization by EM.

To illustrate this approach let $\{x_i, I_i\}_{i=1,\dots,N}$ and $\{x_i^*, I_i^*\}_{i=1,\dots,N}$ be the 2-dimensional and filtered N image points in the spatial-range domain. Specifically the algorithm consists of 3 steps:

For each $i=1,\dots,N$

1. Initialize $k=1$ and $(x_i^{*k}, I_i^{*k})=(x_i, I_i)$
2. Compute

$$x_i^{*k+1}=\frac{\sum_{j=1}^{M} x_j e^{\frac{-(x_i^{*k}-x_j)^2}{2\sigma_x^2}} e^{\frac{-(I_i^{*k}-I_j)^2}{2\sigma_I^2}}}{\sum_{j=1}^{M} e^{\frac{-(x_i^{*k}-x_j)^2}{2\sigma_x^2}} e^{\frac{-(I_i^{*k}-I_j)^2}{2\sigma_I^2}}}$$
$$I_i^{*k+1}=\frac{\sum_{j=1}^{M} I_j e^{\frac{-(x_i^{*k}-x_j)^2}{2\sigma_x^2}} e^{\frac{-(I_i^{*k}-I_j)^2}{2\sigma_I^2}}}{\sum_{j=1}^{M} e^{\frac{-(x_i^{*k}-x_j)^2}{2\sigma_x^2}} e^{\frac{-(I_i^{*k}-I_j)^2}{2\sigma_I^2}}} \quad (10)$$

   until the displacement of the spatial points, x_i are small, i.e. $|x_i^{*k+1}-x_i^{*k}|<T$

3. Assign $(x_i^*, I_i^*)=(x_i, I_i^*)$

where σ_x and σ_I the Gaussian spatial and range kernel size, respectively. Observer that in the last step of the procedure the original spatial locations, namely x_i's are assigned with the smoothed intensity values.

After the smoothed image profile is calculated it is easy to find the first derivative of it. After that comparing the result with a predefined threshold will give us the exact position of the pupil contour boundary (Fig. 3).

On Fig. 3 is shown an example of the previously explained technique by using the same image profile as in the example with natural image statistics (Fig. 2a).

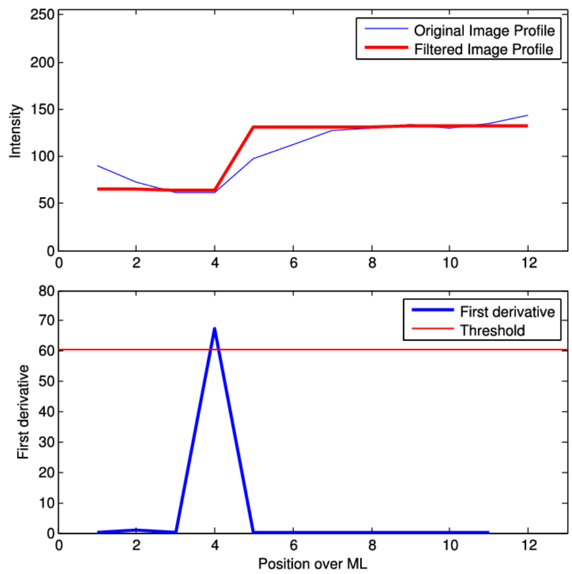

Fig. 3: Applying Mean-Shift Filter over an Image Profile.

Optimizing the ML length

Some more optimization can be done by controlling the length of the measurement lines (Fig. 4). While the tracked eye is standing still or make slight movements, the position of the model contour will be very close to the real contour. So there won't be need to spend time to process all the pixels over the full length of the MLs. That's why it is appropriate to reduce that length by a predefined number of pixels.

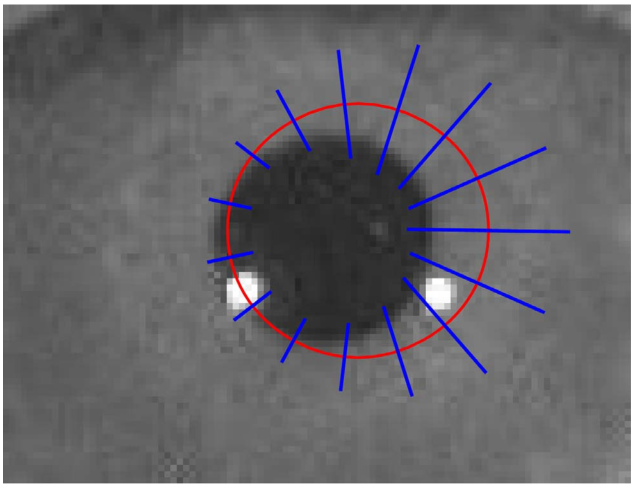

Fig. 4. Optimizing ML length.

In case that no boundary is found in the reduced limits, then the length will be increased. That's how the ML length will be adapted based on the current situation, thus the processing time will be decreased.

IV. Experimental Results

As a demonstration the algorithm was simulated in Matlab environment and after that a C application was developed which implements the algorithm in real-time. On Fig. 5 are shown few sequential moments of it in action.

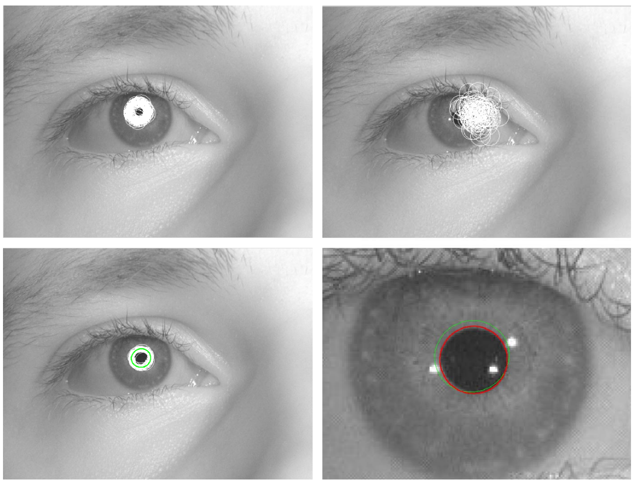

Fig. 5. Experimental Testing.

On the right picture on the second row of Fig. 5 can be seen the result after particle filtering (the green ellipse) and the result after its optimization (the red ellipse).

In Table I the dependence is shown of the mean deviation of the actual value of the coordinates of the pupil center and radius as a function of the number of the particles N before the optimization procedure. As can be seen from the table, there is a trend for the error to decrease as the number of the particles is increased. Moreover this trend is not so remarkable if $N > 100$, which means that it is not necessary more than 100 particles are used, because this won't lead to much better accuracy but instead it will lead to longer time for processing. That's why the number of particles in the experiments was set to $N = 100$.

In Table II the results are shown after the two optimization stages. As can be seen, the Mean-Shift optimization gives much better results then the statistical approach.

Table I
RESULTS FROM THE PARTICLE FILTER TRACKER

Number of Particles	Mean deviation of center estimates (pixels)	Mean deviation of radius estimates (pixels)
20	31	13
40	11	5
60	7	3
80	4.5	2
100	2.5	2
200	2	2
300	2	2

V. CONCLUSION

In this paper a new approach to IR-based eye-pupil tracking was presented. First, the parameters of the pupil are estimated and a particle filter is applied, after which the tracking procedure was described using EM and Mean-Shift algorithms. Our approach for tracking and optimization provides high near real-time performance and is suitable for applications in video surveillance, moving object control, etc.

In our future work we intend to make our approach faster and implement tracking of multiple eye-pupils within an image.

ACKNOWLEDGMENT

This work was sponsored by the Ministry of Education and Science of Bulgaria, National Foundation "Science and Research", Ukrainian-Bulgarian R&D joint project, NSF Grant DO02-41/2008 – "Human Biometric Identification in Video Surveillance Systems".

Table II
ACCURACY AFTER THE OPTIMIZATION PROCEDURE

Statistical Optimization		Mean-Shift Optimization	
Mean deviation of center estimates (pixels)	Mean deviation of radius estimates (pixels)	Mean deviation of center estimates (pixels)	Mean deviation of radius estimates (pixels)
1.36	1.75	0.45	1.43

REFERENCES

[1] J. Gips and P. Olivieri, "EagleEyes: An Eye Control System for Persons with Disabilities," *The Eleventh International Conference on Technology and Persons with Disabilities*, 1996.

[2] L. Young and D. Sheena, "Methods & Designs: Survey of Eye Movement Recording Methods," *Behavioral Research Methods & Instrumentation 7(5)*, 1975, pp. 397-429.

[3] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," *Advances in Neural Information Processing Systems (NIPS) 6,* 1994, pp. 753–760.

[4] L. Q. Xu, D. Machin, and P. Sheppard, "A novel approach to real-time non-intrusive gaze finding," *In: Proceedings of the British Machine Vision Conference*, 1998, pp. 428–437.

[5] K. H. Tan, D.J. Kriegman, and N. Ahuja, "Appearance-based eye gaze estimation," *In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV 2002),* 2002, pp. 191–195.

[6] O. Williams, A. Blake, and R. Cipolla, "Sparse and semi-supervised visual mapping with the S3GP," *Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 230–237.

[7] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike, "An Incremental Learning Method for Unconstrained Gaze Estimation," *Proceedings of the 10th European Conference on Computer Vision: Part III*, 2008, pp. 656-667.

[8] T. E. Hutchinson, K. P. White Jr., W.N. Martin, K.C. Reichert, and L.A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man and Cybernetics 19(6)*, 1989, pp. 1527–1534.

[9] R.J. Jacob, "What you look at is what you get: eye movement-based interaction techniques," *In: Proceedings of the SIGCHI conference on Human factors in computing systems*, 1990, pp. 11–18.

[10] C. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources," *IVC18(4)*, 2000, pp. 331-335.

[11] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, 2001, pp. 100-117.

[12] J. Huang and D. Mumford, "Statistics of Natural Images and Models," *IEEE Computer Vision and Pattern Recognition (VCPR)*, 1999, pp. 541-547.

[13] A. Pece and A. Worrall, "Tracking with the EM Contour Algorithm," *European Conference on Computer Vision*, 2002, pp. 3-17.

[14] M. Istard and A. Blake, "CONDENSATION – Conditional Density Propagation for Visual Tracking," *Int. J. Computer Vision*, 1998.

[15] H. Tek, D. Comaniciu, and J. Williams, "Vessel Detection by Mean Shift Based Ray Propagation," *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001, pp. 228-235.