Meticulously Detailed Eye Model and Its Application to Analysis of Facial Image*

Tsuyoshi Moriyama
Keio University
moriyama@ozawa.ics.keio.ac.jp

Jing Xiao
Carnegie Mellon University
jxiao@cs.cmu.edu

Jeffrey F. Cohn
University of Pittsburgh
jeffcohn@pitt.edu

Takeo Kanade
Carnegie Mellon University
tk@cs.cmu.edu

Abstract − *We propose a system that is capable of detailed analysis of eye region images including position of the iris, degree of eyelid opening, and shape and texture of the eyelid. The system is based on a generative eye model that defines fine structures and motions of eye. The structure parameters represent structural individuality of the eye, including size and color of the iris, width and boldness of the double-fold eyelid, width of the bulge below the eye and width of the illumination reflection on the bulge. The motion parameters represent movements of the eye, including up-down position of the upper and lower eyelids and 2D position of the iris. The system first registers the eye model to the input in a particular frame and individualizes the model by adjusting the structure parameters. Then, it tracks motion of the eye by estimating the motion parameters across the entire image sequence. Combined with image stabilization to compensate the head motion, the registration and motion recovery of the eye are guaranteed to be robust.*

Keywords: computer vision, face analysis, facial expression analysis, generative eye model, motion tracking, texture modeling, gradient descent.

1 Introduction

In facial image analysis for expression and identity recognition, eyes are particularly important [1][2]. Gaze tracking plays a significant role in human-computer interaction and eye analysis provides strong biometrics for face recognition [3]. In behavioral science, Facial Action Coding System (FACS [4]), the *de facto* standard for coding facial muscle activities, defines many action units (AUs) for eye.

Automated analysis of facial images has found eyes yet to be a difficult target [5][6][7]. The difficulty comes from the diversity in the appearance of eyes: they are from both structural individuality and motion of eyes as
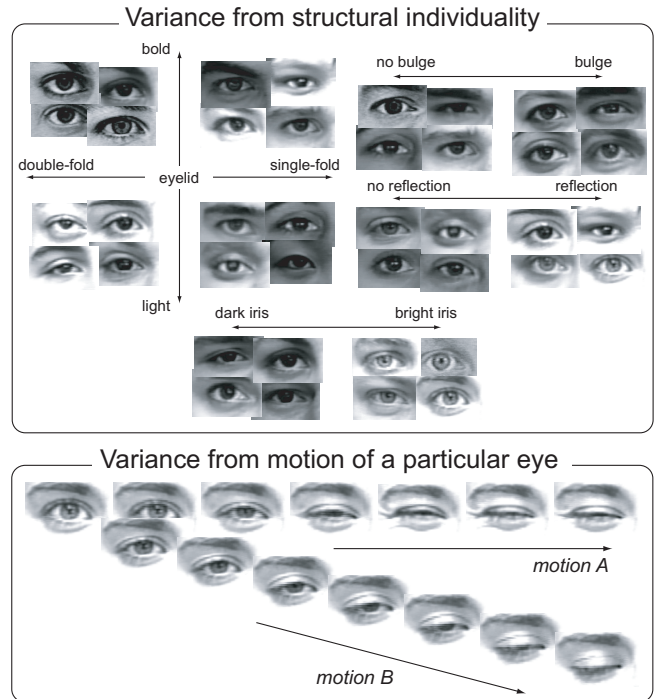
Figure 1: Diversity in the appearance of eye images

shown in Fig.1. Past studies have not been able to represent this diversity. For example, a couple of parabolic curves and a circle have been used for eye models in past studies, but they are compromised by double-fold eyelids causing inaccurate tracking of motion. More detailed models are necessary for perfect representation of eye images.

We propose a meticulously detailed generative eye model and an eye motion tracking system that exploits the model. The model parameterizes both the structural individuality and the motion of eyes. Structural individuality is represented by size and color of the iris, width and boldness of the double-fold eyelid, width of the bulge below the eye, width of the illumination reflec-
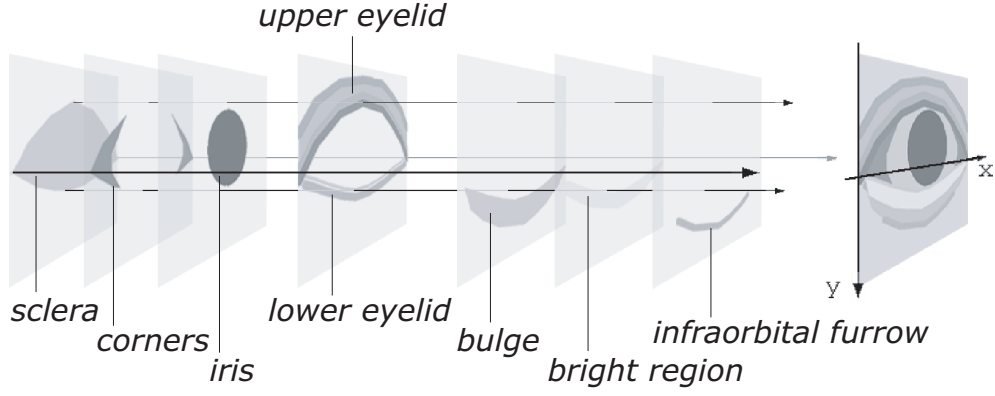
Figure 2: Multi-layered 2D eye region model

Table 1: Detailed description of our eye region model.

parts	sub-parts	model	parameters		picture
			structure **s**	motion **m**	
Upper eyelid	lower boundary	curve1	d_u : distance between curve1 and 2 I_{r1} : intensity of region1 w_{c2} : line width of curve2 (I_{r1}, w_{c2} : controlled by 'boldness' parameter f)	ν_{height} : degree of curve1 raising ν_{skew} : degree of horizontal skew of curve1	
	furrow	curve2 (region1)			
Lower eyelid	upper boundary	curve3	d_b : distance between curve3 and 4 d_r : distance between curve3 and 6	λ_{height} : degree of curve3 raising I_{c3} : intensity of curve3 I_{c5} : intensity of curve5 (I_{c3}, I_{c5} : $subject to$ λ_{height})	
	bulge below the eye	region2 (curve4)			
	infraorbital furrow	curve5 (∈ curve4)			
	reflection on the bulge	region3 (curve6)			
Sclera		region4		(upper boundary) ≡ curve1 (lower boundary) ≡ curve3 I_{r4} : intensity of region4 (I_{r4} : subject to ν_{height})	
Corners		region5 (curve7) region6 (curve8)		(upper boundary) ∈ curve1 (lower boundary) ∈ curve3	
Iris		region7	r_i : radius of region7 I_{r7} : intensity of region7	η_x, η_y : position of the center of region7	

Table 2: Appearance changes controlled by structure parameters $\mathbf{s}$.

parameter	0.0	0.5	1.0
d_u			
f			
d_b			
d_r			
r_i			
I_{r7}			

Table 3: Appearance changes controlled by motion parameters $\mathbf{m}$.

parameter	0.0	0.5	1.0
ν_{height}			
ν_{skew}			
λ_{height}			
η_x			
η_y			

tion on the bulge and furrow below the bulge. Motion is represented by up-down positions of the upper and lower eyelids and 2D position of the iris. Combined with the image stabilization with respect to head motion, the system estimates motion of the eye accurately together with structural individuality of the eye.

2 Eye region model

We exploit a 2D parameterized generative model which consists of multiple components corresponding to the anatomy of an eye. The components include white region around *iris* (*sclera*), dark regions near left and right corners, an *iris*, upper and lower eyelids, bulge below the eye, bright region on the bulge, and a furrow below the bulge (*infraorbital furrow*). The model for each component is rendered in a separate rectangle layer and overlayed together to make an eye region as illustrated in Fig.2. Pixels that render a component in a layer have color intensities otherwise transparent, so that a color pixel in a lower layer appears in the eye region model when all of the upper layers are transparent at the same location. For example, the eyelid layer has two curves to represent upper and lower eyelids, where the region above the upper curve and that below the lower curve are filled with skin color while the region between those curves (*palpebral fissure*) is transparent. The *iris* layer has a disk to represent an *iris*. When the eyelid layer is superimposed over the *iris* layer, only the portion of the disk between the eyelid curves appears in the eye region model while the rest is occluded by the skin pixels. When the upper curve in the eyelid layer comes down to represent half-closed eye, more portion of the disk in the *iris* layer is occluded.

Table 1 shows eye components represented in our multi-layered eye region model and their parameters. Curves and regions are realized by polygonal lines formed by predefined vertices, and all the graphics are rendered using Microsoft Foundation Class Library 6.0.

We call parameters d_u, f, d_b, d_r, r_i, and I_{r7} the *structure parameters* (denoted $\mathbf{s}$) that define static and structural detail of the eye region model, whereas we call parameters ν_{height}, ν_{skew}, λ_{height}, η_x, and η_y the *motion parameters* (denoted $\mathbf{m}$) that define dynamic detail of the model. $T(\mathbf{x}; \mathbf{s}, \mathbf{m})$ represents the eye region model using $\mathbf{s}$ and $\mathbf{m}$, where $\mathbf{x}$ is a vector containing pixels in the model coordinate system. Table 2 and Table 3 show appearance changes of the eye region model due to the changes of $\mathbf{s}$ and $\mathbf{m}$, respectively.

3 Model based eye image analysis

The input image sequence contains facial behaviors of a subject. Facial behaviors usually accompany spontaneous head motions. The appearance changes of facial images thus comprise both rigid 3D head motions and non-rigid facial expressions. Decoupling these two components is realized by recovering the 3D head pose across the image sequence and accordingly warping the faces to a canonical head pose (e.g. frontal and upright). We call the warped images the stabilized images that are supposed to include only appearance changes due to facial expressions and use them in eye image analysis. Fig. 3 shows a schematic overview of the whole process including the head motion stabilization. The system first registers the eye region model to the input in the initial frame and individualizes the model by adjusting the structure parameters $\mathbf{s}$ (Table 1). Then, it tracks motion of the eye by estimating the motion parameters $\mathbf{m}$ across the entire image sequence.
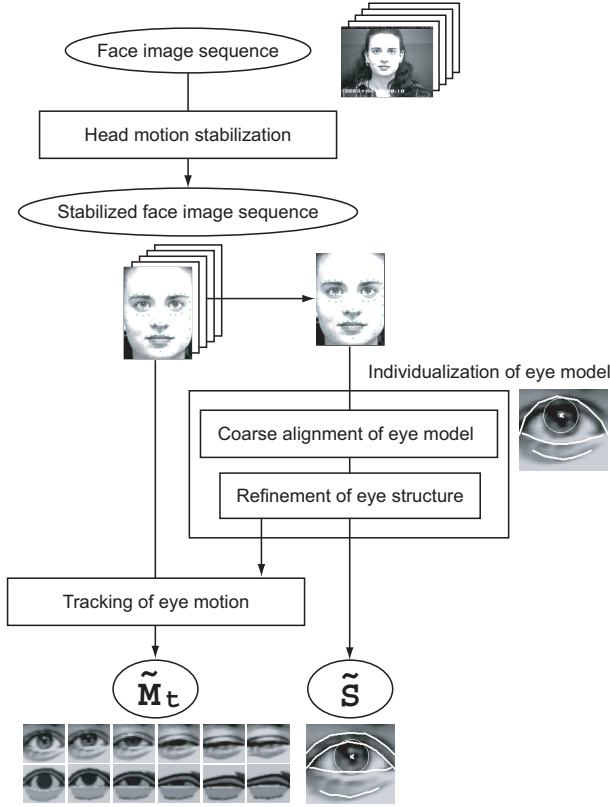
Figure 3: Schematic overview of model based eye image analysis

3.1 Head motion stabilization

We use a 3D head tracker that is based on a cylindrical head model [8]. Manually given the head region with the pose and feature point locations (e.g. eye corners) in an initial frame, the tracker automatically builds the cylindrical model and recovers the 3D head poses across the rest of the sequence. The initial frame is so selected as to be the most frontal and upright. The tracker tracks the non-rigid motions of the feature points on the stabilized images. Full 3D motion (3 rotations and 3 translations) is recovered for color image sequences in real-time. The performance evaluation on both synthetic and real images demonstrated that it can track as large as 40 degrees and 75 degrees for yaw and pitch, respectively, within 3 degree error range.

3.2 Individualization of eye region model

The system first registers the eye region model to a stabilized face in an initial frame $t = t_0$ by so scaling and rotating the model that both ends of curve1 ($\mathbf{u}_1$) coincide the eye corner points in the image. t_0 is such a frame that contains a neutral eye (an open eye with a center-located iris), which can be different from the initial frame specified in head tracking. Let $\tilde{\mathbf{s}}$ denote the individualized structure parameters. $\tilde{\mathbf{s}}$ is manually given

Table 4: Example results of structure individualization.

	input	normalized	model
(a1) Single-fold eyelid			
(a2) Double-fold eyelid			
(b1) Bright iris			
(b2) Dark iris			
(c1) Bulge			
(c2) Reflection			

in the current implementation through a graphical user interface, and fixed across the entire sequence. Example results of individualization for different appearance factors listed in Fig. 1 are shown in Table 4.

3.3 Tracking of eye motion

The intensity of both the input eye region and the eye region model are so normalized prior to eye motion tracking as to have the same average and standard deviation. Let $\tilde{\mathbf{m}}_t$ (t: time) denote the final estimates of the motion parameters $\mathbf{m}$ across the input sequence. The motion parameter set in the initial frame, $\tilde{\mathbf{m}}_{t=t_0}$, is also manually adjusted simultaneously when the eye region model is individualized.

With the initial motion parameters $\tilde{\mathbf{m}}_{t=t_0}$ and structure parameters $\tilde{\mathbf{s}}$, the system tracks eye motions across the entire sequence starting from $t = t_0$ to finally get $\tilde{\mathbf{m}}_t$ for all t. The system tracks the motion parameter set in the current frame from that in the previous frame based on a gradient descent algorithm. The converged set of parameters in the current frame is used as the initial values in the next frame. We exploit an extended version of the Lucas-Kanade algorithm [9]. The difference from the original Lucas-Kanade algorithm is that our method allows the searched template to be deformable during tracking. The motion parameter set $\tilde{\mathbf{m}}_t$ at a particular frame t is estimated by minimizing the following objective function D :

$$D = \sum \left[T\left(\mathbf{x}; \mathbf{m}_t + \delta\mathbf{m}_t\right) - I\left(W\left(\mathbf{x}; \mathbf{p}_t + \delta\mathbf{p}_t\right)\right) \right]^2 \quad (1)$$

where I is the input eye region image, W is a warp from the coordinate system of the eye region model to that of the eye region image, and $\mathbf{p}_t$ is a vector of the

warp parameters that includes only translation in this implementation. Structure parameters $\mathbf{s}$ don't show up in T because it is fixed across the sequence.

$\delta\mathbf{m}_t$ and $\delta\mathbf{p}_t$ are obtained by solving the simultaneous equations obtained from the first-order Taylor expansion of Eq.(1). $\mathbf{m}_t$ and $\mathbf{p}_t$ are updated by the following Eq.(2):

$$\mathbf{m}_t \leftarrow \mathbf{m}_t + \delta\mathbf{m}_t, \qquad \mathbf{p}_t \leftarrow \mathbf{p}_t + \delta\mathbf{p}_t \qquad (2)$$

The iteration process at a particular frame t converges when the absolute values of $\delta\mathbf{m}_t$ and $\delta\mathbf{p}_t$ become less than preset thresholds. The region surrounded by curve1 ($\mathbf{u}_1$) and curve3 ($\mathbf{l}_1$) is used for the calculation process to put more weight on the structure inside the eye, and to avoid the effect from other facial components including eyebrow that show up in the eye region. When parameter ν_{height} is less than a preset threshold, position of region7, η_x and η_y, are not updated because the *iris* is not visible enough to obtain the reliable position.

4 Experiments

We applied the proposed system to image sequences from two large databases: The Cohn-Kanade AU-coded Face Expression Image Database [10] and the Ekman-Hager Facial Action Exemplars. Facial expressions of 118 subjects from a variety of ethnicities, ages, and both genders are digitized in 576 image sequences (490 in Cohn-Kanade and 86 in Ekman-Hager) with 9530 image frames in total. In-plane and limited out-of-plane motion are included. The initial frames for tracking were chosen to the first frame in each sequence since all the sequences to be analyzed start from neutral expression in these databases.

4.1 Results of motion tracking

Results are evaluated by humans for each factor of the appearance diversity shown in Fig. 1. In terms of the diversity from static structures, the tracking accuracy is evaluated in the last frames of the sequences. In terms of the diversity from dynamic motions, it is evaluated in the frames where the appearance changes due to eye motions reach the maximum intensity, and also in the last frames where the eye comes back to neutral.

4.1.1 Upper eyelids

Most likely failure in tracking upper eyelids was that a furrow on the upper eyelid fold was tracked as the boundary between the upper eyelid and the *palpebral fissure* by mistake. Our system can track the upper eyelid accurately in such a case, as well as single-fold eyelids, thick eyelids (upper eyelids with dark and thick eyelashes), and revealing eyelids (upper eyelids that appear to be single-fold but reveal a furrow unfolded in eye widening) as shown in Table 5.

Table 5: Example results for a variety of upper eyelids

(a) Single-fold eyelid	(b) Double-fold eyelid
(c) Thick eyelids	(d) Revealing eyelids
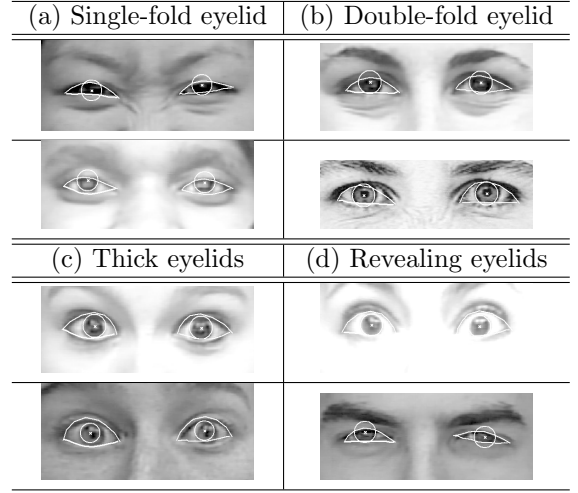
4.1.2 Irises

Most likely failure was that the model for the *iris* matches with other dark portions such as shadow between the inner corner of the eye and the root of nose, especially when iris is bright. Our system can track irises accurately as shown in Table 6.

4.1.3 Bulge with reflection below the eye

An oblique furrow below the bulge tends to be tracked as the boundary between the *palpebral fissure* and the lower eyelid. When the bulge is very bright reflecting environmental illumination, the pattern formed by the oblique furrow and the bright bulge becomes similar to that formed by the boundary and the *sclera*, which makes tracking difficult. Our system can track lower eyelids correctly as shown in Table 7.

4.1.4 Motion

When the upper and lower eyelids get close in eye closure, tracking tended to fail. Besides, parabolic curve models used in past studies didn't match when upper eyelids changed the shape in motion. Of action units defined in FACS [4], Table 8 shows the result for (a) AU5, (b)AU43+7, and (c)AU6+7. Our eye model can track motions of the eye accurately as shown in Table 8 representing a variety of shapes of the upper eyelid.

4.1.5 Performance evaluation

Of 576 image sequences with 9530 frames, only 2 sequences with 40 frames were not tracked well. This happened because the head tracker was not able to stabilize the face well, accordingly eye regions were not obtained in the registration step. One of the causes of the failure in head tracking was that pixels in the face region were almost saturated so that the head tracker was not able to find any texture. The other cause was that the head in the image turned totally to the side so that there

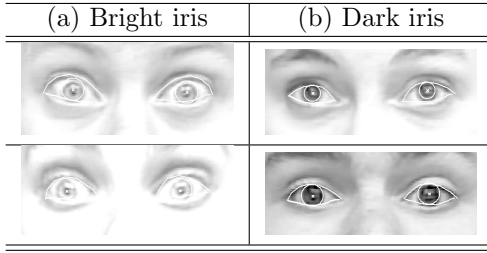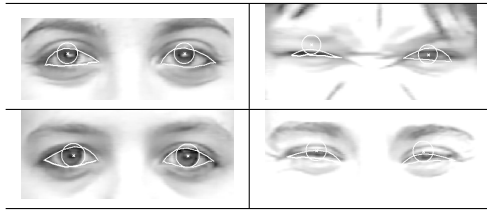Table 6: Example results for different color of iris

(a) Bright iris	(b) Dark iris

Table 7: Example results for different appearance below the eye

Table 8: Example results for motion

	(a) Upper eyelid raising	(b) Blinking and eyelid tightening	(c) Cheek raise and eyelid tightening
#1			
#3			
#5			
#7			
#9			
#11			
#13			

were no face pixels in the stabilized image though head motion was yet correctly tracked.

Of 118 subjects, 5 subjects had very weak edge in the eyelids causing unstable tracking of motion. An algorithm such as enhancing image before calculating Eq. (1) at the geometries of eye parts obtained in the previous frame can improve the robustness for such subjects.

5 Conclusion

A meticulously detailed eye region model and a facial image analysis system exploiting the model were proposed. Having pointed it out that the diversity in the appearance of eye images makes eye tracking problem difficult, we demonstrated that the proposed method had the capability of analyzing the structural individuality and the motion of the eye accurately even in difficult cases.
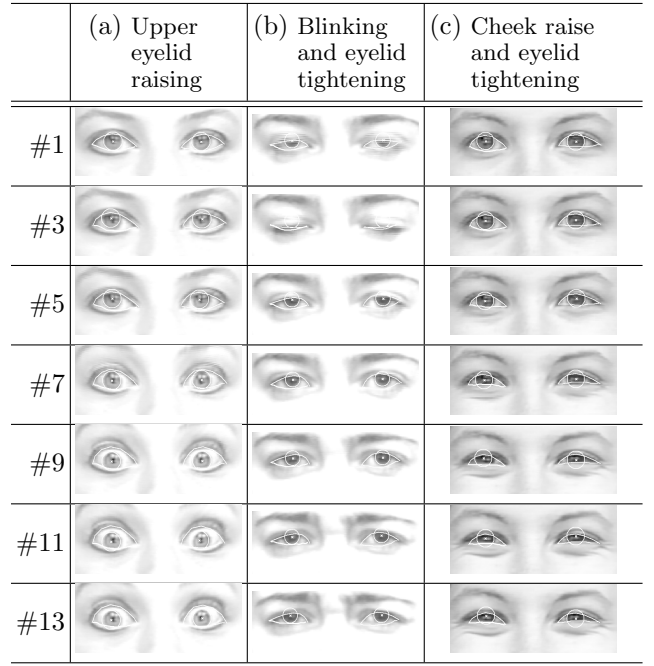
Acknowledgement

References

[1] M. Frank and P. Ekman, "The ability to detect deceit generalizes across different types of high-stake lies," *Journal of Personality & Social Psychology*, vol. 72, pp. 1429–1439, 1997.

[2] C. Padgett *et al.*, "Categorical perception in facial emotion classification," *Cognitive Science*, 1996.

[3] K. Fukuda, "Eye blinks: new indices for the detection of deception," *Psychophysiology*, vol. 40, no. 3, pp. 239–245, 2001.

[4] P. Ekman and W. Friesen, *Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[5] A. Yuille *et al.*, *Active Vision*. MIT Press, 1992, ch. 2, pp. 21–38.

[6] I. Ravyse, H. Sahli, and J. Cornelis, "Eye activity detection and recognition using morphological scale-space decomposition," in *Proc. IEEE International Conference on Pattern Recognition '00*, vol. 1, 2000, pp. 5080–5083.

[7] S. H. Choi, K. S. Park, M. W. Sung, and K. H. Kim, "Dynamic and quantitative evaluation of eyelid motion using image analysis," in *Medical and Biological Engineering and Computing*, vol. 41, no. 2, 2003, pp. 146–150.

[8] J. Xiao, T. Moriyama, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," *International Journal of Imaging Systems and Technology*, vol. 13, pp. 85–94, September 2003.

[9] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1981, pp. 674–679.

[10] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Face and Gesture '00*, 2000, pp. 46–53.