EYE DETECTION UNDER UNCONSTRAINED BACKGROUND BY THE TERRAIN FEATURE

Jun Wang and Lijun Yin

State University of New York at Binghamton
Department of Computer Science
Binghamton, NY 13902, USA

ABSTRACT

Locating eyes in face images is an important step for automatic face analysis and recognition. In this paper, we present a novel approach for eye detection without finding the face region using topographic features. First, we argue that the eyes show certain topographic pattern if the gray-level face image is treated as a 3D terrain surface. Then a terrain map, which denotes the terrain type of each pixel, is derived from the original image by applying topographic classification approach. From the terrain map, eyes usually locate in the region around the *pit*-labeled pixels because of their intrinsic reflectance characteristic. At last, we construct Gaussian Mixture Model based probabilistic model to describe the distribution of *pit*-labeled candidates. The eye pair detection problem is transformed to maximize the probability that the selected candidate pair belongs to the eye space. Experiments show that our method has certain robustness to uncontrolled background.

1. INTRODUCTION

Facial feature detection is an important step in many face related applications, such as face recognition, face image or video indexing for multimedia material management. The facial features, such as eyebrow, eyes, nose, mouth, and chin, are the prominent features of a face, among which eye is the most frequently used feature for helping identify human face. Intensive research on eye feature detection has been done in the past decades. Some work requires to detect face as an initial processing for detecting eye in order to provide a good estimation of initial eye location, for example, the deformable template based approaches [1, 2, 3, 5] and projection function based method [8]. These approaches of eye detection are computationally sensitive to the initial location of face region.

There are some methods developed for eye detection without the pre-knowledge of face location, including infra-red based approache [9], flow field characteristics based approach [7], and adaboost based method [4]. Among all of them, infra-red light based approach shows the most reliable and accurate detection capability, which is therefore very popular to be used in the commercial and research systems for eye detection.

As we know, because of the specific reflection characteristic of pupil, the eye center exhibits fuscous while the eye white appears bright. If a gray-scale image is treated as a 3D topographic surface with the height of each location being denoted by the intensity of the corresponding pixel, the eye region will show a certain terrain pattern. In particular, the center consists of *pit*-pixels, surrounded by *hillside*-like regions. This give us a hint that the eye can be detected based on its terrain appearance features.

Motivated by the topographic analysis technique [11, 10], in this paper, we propose a novel eye detection method. First, we derive a terrain map from the original gray scale image by applying the topographic primal sketch analysis. Second, we mark the *pit*-pixels in the terrain map as the candidates for eye pair selection. Third, a Gaussian Mixture Model(GMM) based possibility model is learned as a classifier to choose the eye pair from the candidates.

The remainder of the paper is organized as follows. Section 2 gives the background of topographic classification of pixels for gray scale image. Section 3 provides the GMM based classification model for eye pair selection. Section 4 shows the experiments results on real images, followed by the discussion and conclusion in section 5.

2. TOPOGRAPHIC CLASSIFICATION

In topographic primal sketch theory, the gray level images are treated as 3D terrain surfaces. The intensity $I(x, y)$ is represented by the height of the terrain at pixel (x, y). Figure. 1 shows an example of a face image and its terrain surface in the eye region. As we know, the intensity variations on a 2D face image is caused by the face surface orientation and its reflectance. The resulted texture appearance provides an important visual cue in order to classify a variety of facial regions and features [14]. If viewed as a 3D terrain surface, the face image shows some 'waves' in the face region because of the surface and reflectance charac-

0-7803-9332-5/05/$20.00 ©2005 IEEE

teristics. For eyeball, it generally composed of two parts: the black part, mostly is pupil and the white part. The reflectance characteristics of eyeball exhibits in 3D surface as the similar terrain pattern. The center is consists of some *pit*-pixels, surround by some *hillside*-like regions. In other words, there are always appears some pit points in the center of eyes.

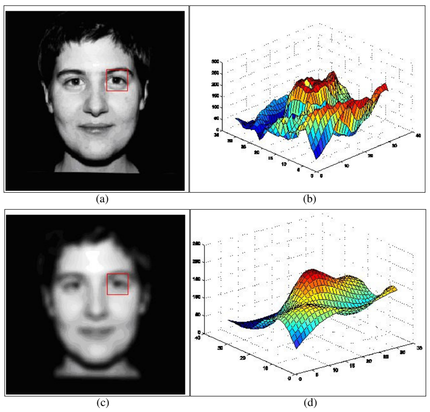

Fig. 1. Face image and the 3D terrain surface of the eye region. The surface is reversed, so the peak denotes the pit in real surface. (a) original face image; (b) terrain surface of the eye region of the original image; (c) face image after Gaussian filter smoothing (filter size 15×15, $\sigma = 3.0$); (d) terrain surface of the eye region of the smoothed image.

Mathematically, we can give the precise definition for *pit* in the continuous surface. Assume that the surface is represented by the equation $z = f(x, y)$. Thus the gradient magnitude $\|\nabla f(x, y)\|$ can be computed as:

$$\|\nabla f(x,y)\| = \sqrt{\left[\frac{\partial f(x,y)}{\partial x}\right]^2 + \left[\frac{\partial f(x,y)}{\partial y}\right]^2} \quad (1)$$

And the Hessian matrix $\mathbf{H}$ is given as:

$$\mathbf{H}(x,y) = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial x \partial y} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix} \quad (2)$$

After applying eigenvalue decomposition to the Hessian matrix, we can get:

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{U}^T = [\mathbf{u_1} \;\; \mathbf{u_2}] \cdot diag(\lambda_1, \lambda_2) \cdot [\mathbf{u_1} \;\; \mathbf{u_2}]^T \quad (3)$$

where λ_1 and λ_2 are the eigenvalues and $\mathbf{u}_1$,$\mathbf{u}_2$ are the orthogonal eigenvectors.

A *pit* occurs where there is a local minim gradient, which means that the gradient is zero and the second directional derivative is positive in all directions: $\|\nabla f(x, y)\| = 0$,$\lambda_1 > 0$ and $\lambda_2 > 0$. Similarly, there are also some other kinds of terrain labels are defined, *peak, ridge, saddle, hill, flat, ravine,* or *pit* [13]. Hill-labelled pixels can be further specified as one of the labels *convex hill, concave hill, saddle hill* or *slope hill*, and saddle hills can be further distinguished as *concave saddle hill* or *convex saddle hill*, saddle as *ridge saddle* or *ravine saddle*[10, 11].

The topographic labeling technique is developed for the continue surface. By using some smoothed differentiation filters, such as the filter based on discrete Chebyshev polynomials, to fit the discrete surface, the topographic labeling techniques is easily to extend to discrete case[11, 12]. The detailed topographic labeling rules for gray scale images can be found in [10].

Generally speaking, the smoothing preprocessing on the gray scale image before calculating the derivatives is necessary in order to reduce the noise influence. In order to achieve the balance between noise elimination and maintain enough image details, it is vital to choose proper parameters of the smoothing filter. An example of a face image and the 3D surface of the eye region after Gaussian smoothing processing is shown in Figure 1.

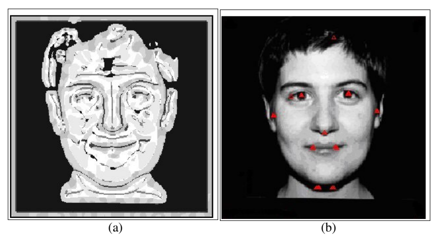

Fig. 2. (a) The topographic classification of the pixels in face region (b) Only the pit-labeled pixels are shown.

By applying topographic classification technique, the pixels in the original gray scale image are labeled by one of the the twelve terrain classes based on its topographic feature. In Figure. 2, an example of the result of topographic classification are shown. Most face region are classified as hill labels. Although the pits-labeled pixel distributes very sparsely, it always occurs in the eyeball regions.

Based on the terrain characteristics of eye region, the *pit*-pixels always occur in the eye region if the image is smoothed by a proper filter without losing too much details. Accordingly, it is conceivable that the *pit*-pixels are set as the candidates for eye detection. In generally, we first merge some pits distributing very closely. This processing will dramatically decrease the number of candidates and reduce the computation cost of classification.

3. LEARNING GMM BASED CLASSIFIER

In order to classify the candidates pit-labeled pixels, we take into account the terrain information around the pixels. Assume that the two points a and b are the centers of two eyeballs and the distance of the two eyes is denoted as d. Two rectangular eye patches with size $0.6d \times 0.3d$, centered at

a and b are cut out along the direction of line ab. The topographic information in the rectangle regions are applied to evaluate the probability of the two points being eye pairs. Theoretically, each pair of *pit*-labeled candidates should be considered, but some pairs of candidates with unreasonable distance are ignored. The terrain type of a reference pixel is discretely represented to the range $1, 2...M$, which is corresponding to the number of types of terrain labels. Thus, after vectorization, the terrain feature can be represented as $\mathbf{t} = \{t_1, t_2, ...t_N\}$, where $1 \leq t_i \leq M$ is the terrain label of the pixel and N is the number of pixels. The terrain maps, consist of the terrain labels of each pixel can be visualized by gray image. Figure.3 c-d shows some visualized terrain maps in our experiments.

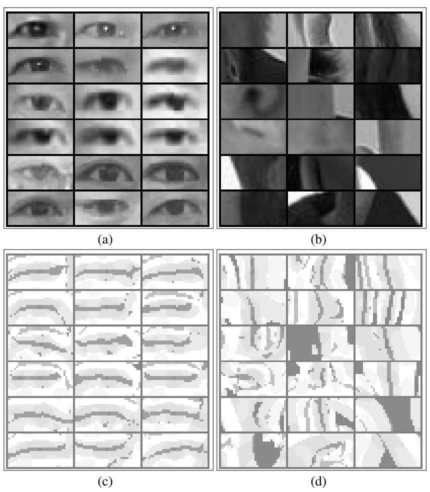

Fig. 3. (a) The samples used as a positive training set, whose corresponding terrain patches are shown in (c); (b) The non-eye samples used as a negative training set, whose corresponding terrain patches are shown in (d).

A training set with 293 face images are constructed for the learning of classifier. First by topographic classification, each pixels of the image are labelled by some certain terrain label. Second, all the pit-labeled pixels are marked as the candidates for training. Third, random pairs of candidates are selected as the training samples, which are roughly consists of two group: eye pairs and non-eye pairs. Figure. 3 shows some examples of these two class of samples. Although the two eyes are symmetrical, there are a little difference in the terrain representation between left eye and right eye. In our method, we treat left eye and right eye respectively.

By GMM, all the samples, eye candidates and non-eye candidates are supposed to distribute in high-dimensionality space obeying several Gaussian distribution. We assume totally three subspaces, each can be described by a Gaussian distributions. The subspaces $\mathcal{E}_l$ and $\mathcal{E}_r$ denote left eye and right eye, represented by the possibility distribution as $\mathcal{N}_l(\mu_l, \mathbf{\Sigma}_l)$ and $\mathcal{N}_r(\mu_r, \mathbf{\Sigma}_r)$. All the non-eye candidates distribute in another subspace $\mathcal{U}$, obeying the distribution $\mathcal{N}_u(\mu_u, \mathbf{\Sigma}_u)$. All these three subspaces constitute the samples space $\mathcal{O}$. In other word, the probabilistic multi-space is characterized as:

$$\mathcal{O} = \{\mu_l, \mathbf{\Sigma}_l, p_l; \mu_r, \mathbf{\Sigma}_r, p_r; \mu_u, \mathbf{\Sigma}_u, p_u\} \quad (4)$$

where p_l,p_r and p_n are the prior possibilities.

If assume the terrain feature of the two candidates a and b are $\mathbf{t}_a$, $\mathbf{t}_b$, the possibility of this candidate pair belongs to eye space $\mathcal{E}$ can be computed as:

$$\begin{aligned} p(\mathcal{E}|\mathbf{t}_a, \mathbf{t}_b) &= \frac{p(\mathbf{t}_a|\mathcal{E}_l) \cdot p_l}{p(\mathbf{t}_a|\mathcal{O})} \cdot \frac{p(\mathbf{t}_b|\mathcal{E}_r) \cdot p_r}{p(\mathbf{t}_b|\mathcal{O})} \\ &+ \frac{p(\mathbf{t}_a|\mathcal{E}_r) \cdot p_r}{p(\mathbf{t}_a|\mathcal{O})} \cdot \frac{p(\mathbf{t}_b|\mathcal{E}_l) \cdot p_l}{p(\mathbf{t}_b|\mathcal{O})} \end{aligned} \quad (5)$$

and the possibility $p(\mathbf{t}_a|\mathcal{O})$ and $p(\mathbf{t}_b|\mathcal{O})$ are similarly calculated as:

$$p(\mathbf{t}|\mathcal{O}) = p(\mathbf{t}|\mathcal{E}_l) \cdot p_l + p(\mathbf{t}|\mathcal{E}_r) \cdot p_r + p(\mathbf{t}|\mathcal{U}) \cdot p_u \quad (6)$$

where $\mathbf{t} = \{\mathbf{t}_a, \mathbf{t}_b\}$ denotes the features of candidate regions.

From the training set, all the parameters shown in Eq.4 can be estimated. The eyes selection is to choose a proper pair of candidates to maximum the possibility calculated by Eq.5.

4. EXPERIMENTS

In this section, we test the proposed eye detection approach on images with both simple and uncontrolled background. In the experiments, the parameters are set as: the Gaussian filter with size 15×15 and standard deviation $\delta = 2.5$ for smoothing procedure, the smoothed differentiation filter with the form of Chebyshev polynomial and convolution kernel size 5 for fitting the discrete surface.

In the experiments, the detection approach is first tested on Japanese Female Facial Expression (JAFFE) database [16]. The JAFFE images of each subject show seven universal facial expressions with the similar background. Among 213 facial images, 204 of them are correctly detected. The algorithm achieves 95.8% correct detection rate. Then we select some face images with complicated background collected from CMU face detection database [17], BioID database [15] and internet for our test. Figure.4 demonstrates some image examples and the corresponding detection results. The variation of images covers different background, face rotation and occlusion.

Moreover, through iteratively locating the eye pairs, the proposed approach is easily to be extended to detect eyes in images with multiple faces. In order to avoid the ambiguity

Fig. 4. Some eye detection results for single-face images with unconstrained background.

eye pairs, such as regarding the eyes from different face as a pair, the appearance features between the eyes should be used to verify the detected eye pairs. Figure.5 gives several eye detection results for multi-face cases.

Fig. 5. Some eye detection results for multiple-face images. The detected eyes in the same rectangle belong to a pair.

In these examples, without locating the face, the eyes can be precisely detected even the face does not occupy the main part of the image or the face is partially occluded. In most cases, the detected eye locations are the centers of pupils. The experimental results show that the proposed method has certain robustness to uncontrolled background. However, it can not locate the eyes when there exist large degree facial rotations.

5. CONCLUSION

In this paper, we present a novel eye detection technique based on topographic primary sketch theory. Because of the specific reflectance characteristic of eyeball, the terrain surface of eyes always exhibits some certain pattern. By topographic labeling, the pixels of gray level image are classified based on their terrain features. The *pit*-labeled pixels are selected out as the candidates. Furthermore, a probabilistic multi-space model is constructed and trained as classifier to detect eye pairs in the candidates. The approach is verified on the face images with simple and unstrained background.

The topographic labeling results are affected by the size of the smoothing filter and convolution kernel when fitting the surface, which means that it can not achieve scale invariant or insensitive. It is a challenging task for us to develop a new strategy to alleviate this problem. Our ongoing work is extending the method to be more robust for real application. Factors in considering various imaging conditions and head motions will be investigated in our future work.

Acknowledgments

This paper is based upon the work supported by the National Science Foundation under grant No. IIS-0414029. We would like to thank Micheal Lyons for providing JAFFE data set for our test.

6. REFERENCES

[1] K. Sobottka, I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking", Signal processing: Image communication, 12, pp. 263-281, 1998.

[2] K.M.Lam and H. Yan, "Locating and Extracting the Eye in Human Face Images", Pattern Recognition, 29(5), pp. 771-9, 1996

[3] N. Oliver, A.P. Pentland, and F. Berard, "Lafter: Lips and face real time tracker", in CVPR, pp 16-21, 1998

[4] D. Cristinacce and T. F. Cootes, "Facial feature detection using AdaBoost with shape constraints", in BMVC, 2003

[5] J. Rurainsky and P. Eisert, "Eye Center Localization Using Adaptive Templates", First IEEE Workshop on Face Processing in Video, Washington DC, June, 2004

[6] Kwan-Ho Lin, Kin-Man Lam and Wan-Chi Siu, "Locating the Eye in Human Face Images Using Fractal Dimensions," IEE Proceedings - Vision, Image and Signal Processing, 148(6), pp. 413-421, December 2001, UK

[7] R. Kothari and J. Mitchell. "Detection of eye location in unconstrained visual images", in ICIP, 1996

[8] Z.-H. Zhou and X. Geng. "Projection functions for eye detection", Pattern Recognition, 2004, 37(5): 1049-1056.

[9] A. Haro, M. Flickner and I. Essa, "Detecting and Tracking Eyes by Using Their Physiological Properties Dynamics and Appearcance", in CVPR, pp. 163-168, 2002

[10] O.D. Trier, Torfinn Taxt and A. K. Jain, "Data Capture from Maps Based on Gray Scale Topographic Analysis", Third Int. Conf. on Document Analysis and Recognition, August 14-16, 1995, Montreal, Canada

[11] L. Wang and T. Pavlidis, "Direct Gray-Scale Extraction of Features for Character Recognition", IEEE Trans. on PAMI, 15(10), October 1993

[12] P. Meer and I. Weiss, "Smoothed Differentiation Filters for Images", J. of Visual Communication and Image Representation, Vol, 3, No. 1, March, pp 58-72, 1992

[13] R.M. Haralick, Layne T. Wasson, and Thomas J. Laffey, "The topographic primal sketch", J. Robotics Res., vol. 2, No. 2, pp 91-118, 1988

[14] L. Yin and S. Royt, Recognizing facial expressions using active textures with wrinkles, in ICME, Baltimore, MD, July 2003. p177-180.

[15] http://www.bioid.com/downloads/facedb/

[16] M. J. Lyons, J. Budynek and S. Akamatsu, Automatic Classification of Single Facial Images, IEEE Trans. on PAMI 21 (12): 1357-1362 (1999).

[17] http://vasc.ri.cmu.edu/idb/html/face/index.html