논문 2010-47SP-5-20

서프 및 하프변환 기반 운전자 동공 검출기법

(Face and Iris Detection Algorithm based on SURF and circular Hough Transform)

아텀 렌스키*, 이 종 수**

(Artem Lenskiy and Jong Soo Lee)

요 약

본 논문에서는 얼굴과 동공을 검색하는 새로운 기법을 제시하며, 안전운행을 위한 운전자의 동공 감시에 적용한 실험결과를 포함하고 있다. 제시된 기법은 세 단계 주요 과정을 거치는데, 먼저 스킨칼라 세그먼테이션 기법으로 얼굴을 찾는 과정으로 이는 지금까지 사용된 휴리스틱모델이 아닌 학습과정 모델에 기반을 두고 있다., 다음에 얼굴 특징 세그먼테이션으로 눈, 입, 눈썹 등의 부분을 검출 하는데, 이를 위해 얼굴 각 부분에서 추출한 고유 특징들에 대한 PDF 추정을 사용하고 있다. 마지막으로 서큘러 하프 변환기법으로 눈 안의 동공을 찾아낸다. 제시된 기법을 조명이 다른 웹 얼굴 영상과 운전자의 CCD 얼굴 영상에 적용하여 동공을 찾아내는 실험을 하여, 높은 동공 검출율을 확인하였다.

Abstract

The paper presents a novel algorithm for face and iris detection with the application for driver iris monitoring. The proposed algorithm consists of the following major steps: Skin-color segmentation, facial features segmentation, and iris positioning. For the skin-segmentation we applied a multi-layer perceptron to approximate the statistical probability of certain skin-colors, and filter out those with low probabilities. The next step segments the face region into the following categories: eye, mouth, eye brow, and remaining facial regions. For this purpose we propose a novel segmentation technique based on estimation of facial class probability density functions (PDF). Each facial class PDF is estimated on the basis of salient features extracted from a corresponding facial image region. Then pixels are classified according to the highest probability selected from four estimated PDFs. The final step applies the circular Hough transform to the detected eye regions to extract the position and radius of the iris. We tested our system on two data sets. The first one is obtained from the Web and contains faces under different illuminations. The second dataset was collected by us. It contains images obtained from video sequences recorded by a CCD camera while a driver was driving a car. The experimental results are presented, showing high detection rates.

Keywords : skin-color segmentation, facial features segmentation, circular Hough transform, iris detection

I. Introduction

Facial feature localization can be considered as a preliminary step in the development of many computer vision applications, such as security systems, car driver vigilance tracking, emotion recognition, and so on.

From a hardware perspective iris detection can be divided into two groups. Some authors base their algorithms on infra-red (IR) cameras[1] or cameras with IR illumination[2~4]. Pupils in the reflected IR spectrum appear as bright spots making them easy to detect. The other group uses color cameras. In this case there is the possibility to take color information

* 학생회원, 울산대학교 전기전자정보시스템공학부
(Student Member, School of Electrical, Electronics and IT, University of Ulsan)
** 정회원, 울산대학교 컴퓨터정보통신공학부
(Regular Member, School of Computer Engineering and Information Technology, University of Ulsan)
접수일자: 2010년4월21일, 수정완료일: 2010년8월16일

into account for skin-color segmentation purposes[5]. Using skin-color information the facial region can be detected relatively easily.

For eye localization a number of approaches have been recently proposed. Some of them feed raw eye images directly into a classifier, such as support vector machines[6] or neural networks[5, 7~8]. Others, prior to classification, extract robust and compact features. The most popular features are wavelet coefficients or their derivatives[8~9], Gabor transform [8~9] or secondary derivatives of a Gaussian kernel are also popular . There are also methods utilizing the geometrical properties of eyes by calculating generalized projection functions[11]. One more approach is to consider average intensity. Usually, the intensity of the eye region is lower compared to the rest of the face. This concept is successfully applied in [12] and [13]. In the latter work Zernike Moments served as features as they are invariant to rotations.

As soon as eye regions are segmented, they are analyzed according to iris positioning. For the purpose of finding iris centers a few methods are commonly known. One of them[14] is based on the Circular Hough transform, while the other[15] applies a set of geometrical rules to localize the iris.

In this paper we propose a novel skin-color segmentation algorithm which allows us to robustly and quickly analyze the driver's face. The detected face is segmented into four facial classes using a proposed facial segmentation algorithm. The detected eye regions are analyzed using Hough transform to find iris centers.

The paper is organized as follows. Section II describes the skin-color model along with a segmentation algorithm and morphological and geometrical analysis. Section III describes the facial feature segmentation algorithm and the iris localization technique. Experimental results and discussions are presented in Section IV. Section V contains conclusion remarks.

II. Face detection

1. Skin-color model

Instead of heuristic approaches previously applied to segment skin-colors, we propose an adjustable model obtained through the learning process. Our adaptive model is based on a two layer perceptron. The number of input neurons corresponds to the number of color components in RGB space. The number of neurons in the hidden layer is not restricted and can be customized accordingly to the complexity of the training set. Using a trial and error method we set up a number of hidden neurons equal to 10.

For training purposes, we constructed two subsets of samples, The first subset is for positive, i.e. skin-color samples, and the second for negative non-skin color samples. We built a 3-dimensional histogram, where the intensity of each color component corresponds to a value on an orthogonal axis and the number of pixels with a particular color *(r,g,b)* appeared in all training images represents a histogram value (fig.1). By normalizing the histogram, we obtain a skin-color membership function. Dividing each histogram value by the total number of pixels in the histogram a statistical PDF is obtained. In our case, the maximum value of the normalized RGB histogram is equal to 1. This is the maximum value that a sigmoid function in a hidden layer neuron can take. Therefore, a better MLP performance is achieved. For the training set, the original 256x256x256 RGB space is reduced to the space of 64x64x64 for two primary reasons. Firstly, a decrease in dimension leads to a fewer number of neurons needed for skin-color domain approximation..

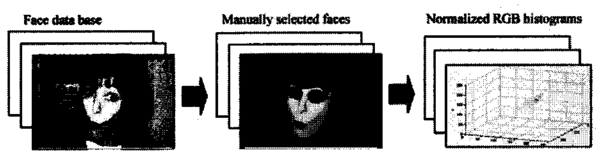

그림 1. 훈련샘플 DB구축
Fig. 1. Training samples preparation.

Secondly, more skin-colors might be outside the training set and the quality of skin segmentation could be also reduced, but, our experimental results show an acceptable segmentation quality.

The subset with negative samples represents colors from outside the skin-color domain. They are defined by uniformly positioning negative values (equals -1) at every second value along each axis, but avoiding positions with skin-colors.

2. Skin-color segmentation

The training process results into hidden W^H and output W^O matrices, and correspondent bias vectors b^H and b^O. The whole procedure of filtering non-skin color pixels is presented as follows:

$$d = W^O f(W^H v + b^H) + b^O \qquad (1)$$

where $v = \{r\ g\ b\}$ is an input vector representing a particular image pixel and f is a hyperbolic tangent sigmoid function

$$f(x) = \frac{2}{(1+e^{-2x})} - 1$$

3. Skin-color model

The result of skin-color segmentation contains a number of components. Many of them such as arms and clothes are leftover from the segmentation process and are subject to elimination. To identify whether the remaining components are faces, prior knowledge of geometrical facial proportions are exploited. First, an opening operation is executed to fill-in empty spaces caused by skin-color segmentation, and then connected components are extracted. Each of the target components is analyzed to conform to facial proportions and size. The size check removes components with small areas. It is expected that the number of pixels in a face candidate exceeds 20% of the total number of pixels in the image. In the next step all remaining components are processed to subtract those which do not conform to facial proportions. For this matter a two dimensional covariance matrix is calculated based on component pixels' coordinates. For each component a ratio among the greatest and the smallest eigenvalues is calculated. Components with a ratio of less than four are eliminated. Eigenvalues and vectors are also used to extract the elliptical area with the face. The extracted facial candidates are then analyzed to find iris centers.

III. Facial features detection

1. SURF based eye regions segmentation

In our approach, eyes are treated as textures. Each eye does not completely replicate the appearance of the other eye, although it looks very similar, therefor it can be considered as a texture patch. We propose a supervised facial segmentation method based on a salient features classification. Firstly, a data set of training pairs is selected. Each pair contains an image with a face and hand segmented image, where corresponding facial regions are marked with different colors. We mark eye regions with green, eye brows as yellow, lips as red and the rest of the face with blue. Although, we are mainly concern about eye groups, we added other categories for a better separability in the feature space.

Each of the images is processed to extract speeded-up robust features (SURF). We chose SURF features due to their comparably short descriptor vectors (64 dimensions) and robustness to changes in illumination.

Extracted features are arranged into four different groups depending on their locations. Those features with locations that fell within the green segment are placed into the eye group. Those features which fell into the red segment are arranged into the lip group, while features from yellow regions are placed in the eye-brow group. The fourth class corresponds to the blue region and contains features which didn't fall in any of the above three classes. These are features from other parts of the face. Due to the excessive number of features obtained from a vast amount of images we ran two algorithms to reduce the size of the vocabulary. The first one eliminates outlier descriptors. We consider them either as not-informative or as features which by accident fell into an image segment of a different type during hand segmentation. The second algorithm is a clustering algorithm, which generalizes the data by finding new feature representatives. These new features serve as a vocabulary in the segmentation process. The segmentation process can be summarized in the following list of steps:

1) Extract SURF descriptors from an input image

2) Find minimal Euclidian distances between each extracted descriptors and descriptors from each class in the vocabulary

$$v_{j=1..m}=\begin{bmatrix}\min_{i=1..N1}(\|d_j^{/}-d_i^{I}\|)\\ \min_{i=1..N2}(\|d_j^{/}-d_i^{II}\|)\\ \min_{i=1..N3}(\|d_j^{/}-d_i^{III}\|)\\ \min_{i=1..N4}(\|d_j^{/}-d_i^{IV}\|)\end{bmatrix} \tag{3}$$

where $d^{/}$ - descriptors from an input image, d^{I}, d^{II}, d^{III},d^{IV} - descriptors from corresponding vocabulary classes. M is a number of extracted descriptors N1, N2, N3, N4 -number of descriptors in correspond classes.

3) Apply the Parzen Window method to estimate the probability density functions(PDF) that pixel belongs to each of the classes:

$$p_k(x,y)=\sum_{j=1}^{m} e^{-\frac{v_{j,k}}{2\sigma_1^2}} e^{-\frac{\|coord(d_j^{/})-[x,y]^T\|}{2\sigma_1^2}} \tag{4}$$

where k = 1..4 represents class index, $coord(d_j^{/})$ returns descriptor coordinates in the image.

4) Each pixel is classified into one of the classes by applying the following rule:

$$t(x,y)=\max_{k=1..4}p_k(x,y) \tag{5}$$

If the greatest probability is less than a threshold, then a pixel's class is considered as unknown (marked black in the segmented image).

To decrease segmentation time in step 3 we work with a proportionally smaller field and at the end of the algorithm the resultant field is interpolated to the original size. Intermediate results of the eye segmentation process are shown on figure 2.

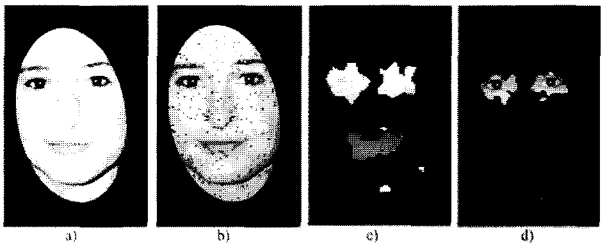

그림 2. a)얼굴검출 결과, b)기호 '+'는 디스크립터 위치, c)얼굴특징 세그먼테이션 결과, d)검출된 눈영역

Fig. 2. a)Face detection result, b)Symbol '+' depicts descriptor's locations, c) Result of facial features segmentation, d) Detected eye regions.

The result of eye region segmentation usually contains a number of eye candidates. In most of cases, images contain two large regions which indeed contain eyes and a few smaller regions. We proceed with a morphological analysis to fill-in spaces within each of the regions and then extracted connected components. The two biggest components are selected as potential eye regions for the further analysis, while others are omitted. Each of the selected components is analyzed separately.

2. Iris detection

In order to localize an iris we apply circular Hough transform[16]. Hough transform operates on a binary image which we obtain through applying a Canny edge detector. To avoid artificial edges appearing due to the intensity differences between a black background and an eye (fig. 3 (a)), we apply the Canny edge detector to the extended rectangle eye region (fig. 3 (b)). Before the Hough transform is applied, an elliptical area containing an eye region is estimated (fig. 3 (c)) and then, edges corresponding to the elliptical region are extracted (fig. 3 (d)) followed by Circular Hough transform analysis (fig. 3

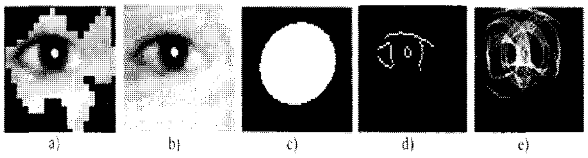

그림 3. a) 연결된 성분, b) a)성분 내의 직사각형, c) a)의 코배리언스행렬로부터 추출된 타원, d) c) 타원내에서 검출된 에지, e) d)에서 계산된 지름들에 대한 누적공간

Fig. 3. a) One of the selected connected components, b) Rectangular region containing connected component (a), c) Ellipse estimated through covariance matrix based on pixels' coordinates from (a), d) Detected edge within ellipse (c), e) Accumulator space for one of the radii calculated from (d).

(e)). Circular Hough transform requires defining the radius of searching circles. We run Circular Hough for a range of radii which is defined proportionally to the area of the elliptical face candidate.

Briefly Circular Hough transform can be summarized in the following list of steps:

1) For each edge point draw a circle in the accumulator space (a, b, r).

$$x = a + r\cos(\theta)$$
$$y = a + r\sin(\theta)$$

2)At the coordinates which lay on the perimeter of the drawn circle of radius r the accumulator value is incremented (fig. 3(e)).

3) The highest values in accumulator space correspond to centers of circles.

To find the best matching radii we observed that with an increasing radius, the potential number of pixels falling on the curve of a circle increases as well (fig. 4 (a)), for ideal circles it follows $l(r) = 2\pi$. Therefore, we are looking for local maxima of the estimated circle's perimeter$l(\widehat{r})$. We simply find local maxima by calculating the difference $\Delta l(\widehat{r})$ of perimeters for neighboring radii (fig.4 (b)) and selecting those r_i which lead to a positive difference:

$$\Delta \hat{l}_i = \hat{l}(r_{i+1}) - \hat{l}(r_i)$$
$$r = \{r_i, 0 \le \Delta \hat{l}_i\}$$

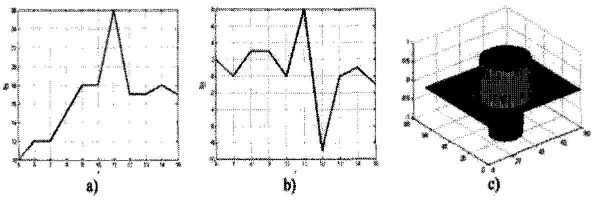

그림 4. a) 증가지름과 원주위 화소수와의 관계, b)근접 원주위 화소수 차이, c) 눈 형태 예

Fig. 4. a) Relationship between the increasing radius and the number of pixels falling on the circle perimeter, b) Difference between numbers of pixels falling on the circle perimeter for neighbor radii, c) An example of an eye template.

In the next step, we apply the iris template to determine the final radius and coordinates of the circle. For each of the radii obtained in the previous step a template in the form of a ring is generated. The radius of an inner circle is equal to a radius r_j from the list of selected radii and filled with negative ones. The radius of the outer circle is equal to $1.5r_j$ and filled with positive ones. Therefore, the iris template is a ring with negative values inside, positive values on the ring and zeros outside. An example of an eye template is shown in (fig. 4(c)). The final match is defined by selecting the circle which leads to the maximum of the following criteria:

$$\max_k \sum_j \sum_i T_{j,i}(x_k, y_k, r_k) I_{j,i} \tag{6}$$

where $T_{j,i}$ and $I_{j,i}$ is the template and eye images correspondingly, k is an index in the list of detected circles, and r_k, (x_k, y_k, r_k) are radius, and coordinates of the center of the circle.

그림 5. 동공 국지화 결과

Fig. 5. The result of iris localization.

Combining face candidate detection, eye region detection and detection of iris positions in a sequence of steps we are able to localize irises on the image (fig. 5)

Ⅳ. Experiments and performance analysis

To test our iris detection system we carried out two set of experiments. One on the facial image database[17] commonly used to benchmark face detection algorithms and the other on the images we captured from a CCD camera.

The image database contains 450 images of 27 unique people with different background and lightning. Among images from the database we exclude images 400,402,403 and 328-336. The first group contains images drawn in black and white and in the second people are far from the camera, which does not resemble distance between the camera and the driver. Therefore, we analyzed 439 images. Among them 45 images were selected for the training process. In the 413(94%) images faces and irises were correctly detected, The face detection rate is quite similar to those of the techniques in [18~19], however our technique detects irises precisely requiring the less time interval. An iris is considered successfully detected if the distance between the real iris center and the center of the detected iris is less than half of the radius of iris circumference.

In the second set of experiments we captured frames from a CCD camera with the resolution of 768x576. The camera was attached to the car dashboard. The video sequences were recorded while a driver was driving a car. Five people of different ethnicities participated in the experiment. We selected 60 images for the training purpose, among video frames recorded from three drivers. Images were taken of various head orientations, such as when a driver was looking into the left and right side view mirrors or the dashboard. Each processed video was visually analyzed. The detection rates for each person participating in the experiment are shown in table 1.

표 1. 검출 성공률
Table 1. Detection rates.

Person	Iris	total #	true	false	Rate
#1	Left	1007	972	15	96.5%
	Right	1007	968	39	96.1%
#2	Left	901	868	33	96.3%
	Right	901	868	33	96.3%
#3	Left	916	889	27	97.0%
	Right	916	887	29	96.8%
#4	Left	892	835	57	93.6%
	Right	892	860	32	96.4%
#5	Left	667	593	74	88.9%
	Right	667	617	50	92.5%

그림 6. CCD획득 영상의 처리결과
Fig. 6. Video processing results for CCD images captured.

Left and right irises were analyzed separately. The third column in the table indicates the total number of frames in a corresponding video sequence. The fourth and fifth columns show numbers of combined true positive and negative, and false positive and negative detected iries correspondingly.

Some examples of detected irises are shown in figure 6.

To reach the best possible results using the proposed system, the system should be calibrated for a particular camera and the acquisition device. Depending on the camera, color perceptions may change. Furthermore, depending on the frame grabber there is a possibility of interlacing effects appearing during the sampling process, and therefore it may affect the quality of the feature space. To make the algorithm less dependent on the camera choice, it is suggested that the loosen color mode should be used, meaning the skin-color domain covers slightly more colors around the skin-color domain.

V. Conclusion

In this paper we presented a face and iris detection algorithm. It employs skin-color information to exclude non-skin-color regions, texture information to detect eyes and geometrical information to detect irises.

The proposed novel skin segmentation algorithm based on multi-layer perceptron, allows us to robustly segment skin-colors of people with various ethnicities. We also proposed a new facial segmentation schema using SURF salient features. To detect the iris within eye regions we applied the circular Hough transform for a range of radii. Final position and radius were selected using two criteria: roundness and dark/light ratio. Our experiments show a high detection rate of 94% for the facial image database.

In contrast to other rigid algorithms [5, 14] taking heuristic approaches, we used the learning process in both skin and eye segmentation steps. This increased the detection rate. Our technique can be easily recalibrated for a particular camera or even an individual person.

Reference

[1] Nakayama, M., Katahara, S., Aoki, M., Detection of driver's hazardous situation using coarse far-infrared image, Intelligent Vehicles Symposium, 2008 IEEE , vol., no., pp.325-328, 4-6 June 2008.

[2] Riad I. Hammoud, G. Witt, R. Dufour, A. Wilhelm, T. Newman, On Driver Eye Closure Recognition for Commercial Vehicles, Proceedings of SAE Commercial Vehicles Engineering Congress and Exhibition, Chicago, IL, USA, Oct 8, 2008.

[3] Batista Jorge P., A Real-Time Driver Visual Attention Monitoring System IbPRIA 2005, LNCS 3522, pp. 200-208, 2005.

[4] Zhao Shuyan and Grigat Rolf-Rainer. 2006. Robust Eye Detection under Active Infrared Illumination, Pattern Recognition. ICPR 2006. Vol. 4 pp. 481 - 484.

[5] Wang Rong-ben, Guo Ke-you, Shi Shu-ming, Chu Jiang-wei, A monitoring method of driver fatigue behavior based on machine vision, Intelligent Vehicles Symposium, 2003. Proceedings. IEEE , vol., no., pp. 110-113.

[6] Jee Hyungkeun, Lee Kyunghee, Pan Sungbum, Eye and face detection using SVM, Conference on Intelligent Sensors, Sensor Networks and Information, 2004, pp. 577 - 580.

[7] Rowley H.A., Baluja S., Kanade T., Neural Network- Based Face Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, Vol. 20 , Issue 1, pp. 23 - 38.

[8] R. Motwani, M. Motwani, F. Harris, Eye Detection using Wavelets and ANN, in Proceedings of GSPx, Santa Clara, September 2004.

[9] Kun He, Jiliu Zhou, Yu Song, Qiang Qiao, Multiresolution eye location from image, Proceedings of Signal Processing, ICSP '04. 2004, vol. 2, pp. 901-905.

[10] King-Hong Cheung, Jane You , Wai-Kin Kong, Zhang, D., A study of aggregated 2D Gabor features on appearance-based face recognition, Proceedings of Image and Graphics, 2004. pp. 310-313.

[11] Zhi-Hua Zhou, Xin Geng, Projection Functions for Eye Detection, Pattern recognition, 2004, vol. 37, n5, pp. 1049-1056

[12] Lin Daw-Tung, Yang Chen-Ming, Real-time eye detection using face circle fitting and dark-pixel filtering, Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, Vol. 2 (2004), pp. 1167-1170 Vol.2.

[13] Kim Hyoung-Joon, and Kim Whoi-Yul, Eye Detection in Facial Images Using Zernike Moments with SVM, ETRI Journal, vol.30, no.2, Apr. 2008, pp.335-337 25

[14] D'Orazio T., Leo, M., Cicirelli, G., Distante, A. 2004. An algorithm for real time eye detection in face images. Inter. Conf. on Pattern Recognition. Vol 3 pp.278 - 281.

[15] Perez C.A., Lazcano V.A., Estvez P.A., Real-Time Iris Detection on Coronal-Axis-Rotated Faces, IEEE Transactions on Systems, Man and Cybernetics - Part C-Applications and Reviews, Vol 37,No.5, pp.971-978. Sept. 2007,

[16] Duda R. O., and Hart P. E. 1972. Use of the Hough Transformation to Detect Lines and Curves in Pictures. Communications of Association for Computing Machinery, 15(1),

pp.11-15.

[17] Markus Weber, "Frontal face dataset", California Institute of Technology, 2003, http://www.vision.caltech.edu/htmlfiles/archive.html.

[18] Tran Le Hong Du, Duong Anh Duc, Duong Nguyen Vu, Ridge and Valley based Face Detection, International Conference on Research, Innovation and Vision for the Future, 2006.

[19] Dazhi Zhang, Boying Wu, Jiebao Sun, Qinglei Liao A Face Detection Method Based on Skin Color Model, Proceedings of the 11th Joint Conference on Information Sciences (2008).

저 자 소 개

Artem Lenskiy(학생회원)

2002 BS CS Novosibirsk State Technical University, Russia.

2004 MS CS Novosibirsk State Technical University, Russia.

2010 Ph.D EE University of Ulsan, Korea.

<Research interests : image processing, pattern recognition, intelligent autonomous vehicle, self-similar processes>

Jong Soo Lee(정회원)

1973 BS EE Seoul National University Seoul, Korea

1981 MS EE VPI&SU USA

1985 Ph.D EE VPI&SU USA

1985-1993 ETRI, KAITECH

1994-presnet Univ. of Ulsan

<Research interests : Multimedia and English, 3D video/user interface, AR/VR applications>