Pattern Recognition 46 (2013) 2613–2621

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Eye localization from thermal infrared images

Shangfei Wang [a,*], Zhilei Liu [a], Peijia Shen [a], Qiang Ji [b]

[a] Key Lab of Computing and Communication Software of Anhui Province, School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, Anhui, PR China

[b] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

ARTICLE INFO

Article history:
Received 28 October 2012
Received in revised form
7 February 2013
Accepted 4 March 2013
Available online 21 March 2013

Keywords:
Eye localization
Eyeglass detection
Long wave thermal infrared images

ABSTRACT

By using the knowledge of facial structure and temperature distribution, this paper proposes an automatic eye localization method from long wave infrared thermal images both with eyeglasses and without eyeglasses. First, with the help of support vector machine classifier, three gray-projection features are defined to determine whether a subject is with eyeglasses. For subjects with eyeglasses, the locations of valleys in the projection curve are used to perform eye localization. For subjects without eyeglasses, a facial structure consisting of 15 sub-regions is proposed to extract Haar-like features. Eight classifiers are learned from the features selected by Adaboost algorithm for left and right eye, respectively. A vote strategy is employed to find the most likely eyes. To evaluate the effectiveness of our approach, experiments are performed on NVIE and Equinox databases. The eyeglass detection results on NVIE database and Equinox database are 99.36% and 95%, respectively, which demonstrate the effectiveness and robustness of our eyeglass detection method. Eye localization results of within-database experiments and cross-database experiments on these two databases are very comparable with the previous results in this field, verifying the effectiveness and the generalization ability of our approach.

1. Introduction

Face and facial expression recognition, which are the primary communication functions of humans, have drawn growing attention in the research areas related to human–computer interaction and psychology. In the recent years, considerable progress has been made in the field of face and facial expression recognition using visible images and videos [1–3]. However, most of the existing methods are not robust enough for employment in uncontrolled environments. Illumination change is the most important factor, because it can significantly influence the appearance of visible images. Nevertheless, long wave thermal infrared (IR) images, which record the temperature distribution, are not sensitive to illumination condition [4,5]. Thus, IR-based face and facial expression recognition algorithms may improve the recognition performance in uncontrolled environments, and can be regarded as a crucial complementarity to visible face and facial expression recognition [5–9]. As eye is one of the most significant features of human face, automatic eye localization is required for face and expression recognition in thermal spectrum. Despite of much progress in eye detection for visible images [10], there has been little progress for eye detection for thermal images. Since the thermal camera acquires absolute temperature values of a face, the geometric and appearance features of thermal images are more blurred than visible images, although, recent work on camera focus can significantly reduce blur effect caused by the temperature accuracy [11,12]. So, it is more difficult to perform eye localization from infrared thermal images [13]. Most existing research either manually marks the eyes [7] or automatically detects them in a co-registered visible image [9,14], making the realistic use of thermal infrared imagery alone impossible.

Even though some geometrical features and appearance features are lost in thermal images, infrared thermal images retain the information of the basic facial structure and the temperature distribution of face. For instance, eye regions are almost symmetrical, the eyebrows and the nose are usually cold [15,16], and the cheek is warm. The distinction between the intensity of background and that of face is significant in most cases, since the emissivity of human skin is different from that of other material. The regions of eyeglasses are dark, because thermal radiation is opaque to eyeglasses.

According to the above knowledge, we propose an automatic eye localization method from long wave infrared thermal images with and without eyeglasses. Since eyeglasses are opaque in the thermal infrared spectrum and therefore show black in thermal

* Corresponding author. Tel.: +86 551 63602824.
E-mail addresses: sfwang@ustc.edu.cn (S. Wang), leivo@mail.ustc.edu.cn (Z. Liu), speijia@mail.ustc.edu.cn (P. Shen), qji@ecse.rpi.edu (Q. Ji).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.03.001

images, it is necessary to perform eye localization on samples with eyeglasses and without eyeglasses separately. First, we perform eyeglass detection. Three features are defined to describe the valleys in the gray-level integration projection curve caused by eyeglasses. Then, a Support Vector Machine (SVM) is used to distinguish the samples with or without eyeglasses. For samples with eyeglasses, projection curve is used to localize the eye centers. For eye localization on samples without eyeglasses, we firstly detect the face region according to the temperature diversity between the facial area and the background. In the training phase, Haar features [17,18] are extracted to represent the intensity feature information of facial regions. We use eight kinds of feature sets to seek the best description of the intensity variation. A facial structure consisting of 15 sub-regions is proposed, and then eight kinds of Haar-like features, including two edge features, four liner features, one center-surrounding feature and one diagonal feature, are extracted from these sub-regions. After that, the AdaBoost algorithm is used to select features from each Haar-like feature sets in each sub-region. The selected features from 15 sub-regions are then combined to form a feature vector, based on which, eight SVM classifiers are trained. The structure parameters of the 15 sub-regions are also calculated from the training samples. In the testing phase, the structure of 15 salient sub-regions is used by sliding on the left/right half part of the face, and eight feature vectors are extracted. A voting strategy is used to determine whether the pixel is an eye or not. The pixel with the largest vote is declared an eye.

The proposed eye localization method is evaluated on the thermal sub-database of Natural Visible and Infrared Expression (NVIE) database [19] and Equinox database [20,21]. Both within database and cross-database experiments are conducted, demonstrating the effectiveness and generalization ability of our approach.

The remainder of this paper is organized as follows. Related works are introduced in Section 2. Section 3 provides the details of our eye localization method, including eyeglass detection, eye localization on samples with eyeglasses and without eyeglasses. Section 4 introduces the experimental conditions and results. Section 5 concludes the paper.

2. Related works

Depending on the wavelength, infrared images can be divided into Short Wave InfRared images (SWIR, 950–1700 nm), Medium Wave InfRared images (MWIR, 3000–5000 nm), and Long Wave InfRared images (LWIR, 8000–14,000 nm). Our work belongs to the category of eye detection from the long wave infrared thermal images. There exist several works of eye location from short wave [22,23] and middle wave [13,22] infrared images. Eye detection from short wave infrared images is relatively easier since most of the eye features are visible. As noted in [13], eye detection from middle wave infrared images is, however, challenging. This is because in the middle wave infrared domain limited features can be extracted from the eye region. Except for eyelashes and eyebrows, other important eye features such as irises, pupils, and superficial blood vessels of the conjunctiva are unclear. For the long wave infrared thermal images used in our paper, eye detection is even more challenging since even eyelashes and eyebrows are difficult to detect. Thus, the eye localization in the long wave infrared images conducted in this paper is more challenging than all these works. To the best of our knowledge, only one work deals with the eyeglass detection in long wave infrared thermal images [24,4], and three works [14,15,25] have been reported to detect facial components from long wave infrared thermal images without eyeglasses.

Due to the imaging theory of infrared thermal images, the eyeglasses are opaque, which block the temperature information of the eyes region [4]. The presence or absence of the eyeglasses may affect the performance of the facial temperature based recognition task [5,26]. For example, Wang and Liu [26] have conducted a series of experiments on Equinox Database to validate the influence of eyeglass on infrared facial recognition. The study shows that recognition results are poor while conducting cross-corpus face recognition experiment on datasets with and without eyeglasses, while good and relative stable recognition results can be achieved on a database, in which all the subjects with eyeglasses on or off. Thus, eyeglass detection is necessary in our eye localization task from infrared thermal images, since eye detection is the first step in most face recognition task. Heo et al. [4,24] proposed a eyeglass detection method using Ellipse Fitting. Since the eyeglasses regions in thermal face images are symmetric and elliptic, the most similar ellipses can be seen as the eyeglass regions. First, a thermal image was binarized. Then, the data points in the binary image were connected using the Freeman chain coding with 8-connectivity. A non-iterative ellipse-fitting algorithm was applied to each set of connected components to produce an ellipse. The two most similar ellipses within the face region were detected as the possible eyeglasses. The authors conducted their eyeglass detection experiment on the thermal images selected from the Equinox database, and achieved the eyeglass detection accuracy of 93.79%.

For eye localization from thermal images without eyeglasses, the first work is performed by Selinger and Socolinsky [14]. They first located face region using a boosted tree-like class-cover catch digraph classier with a maximum rejection bias [27]. Then eye locations were searched in the upper half of the face area using a boosted cascade of Haar-like feature classifiers. The mean absolute error (Mean) and the standard deviation (StDev) of the error in the x- and y-coordinates for the left and right eyes were used to evaluate the performance of their method. They conducted the experiment on the samples without eyeglasses from Equinox database.

The second work is conducted by Trujillo et al. [15], who extracted interest points based on the intensity features to represent facial components. They used Harris features and k-means clustering to detect eyes and mouth regions. They evaluated their approach on a gallery set composed of 30 individuals with 3 expressions (i.e. surprise, happiness, and anger) and 3 poses from OCTBVS database. Since the main purpose of their work was to recognize facial expression from thermal images, the detailed experimental results on eyes and mouth detection were not provided.

The third work is performed by Martinez et al. [25]. They presented an automatic method for detecting eyes, nostrils, and mouth in facial thermal imagery. The detection of eyes and nostrils was performed using SVM and Gentle Boost algorithm based on Haar-like features. Then a region of interest (ROI) for mouth detection was selected based on the detections of the eyes and nostrils. Patch entropy and self-similarity [28] were combined to detect the mouth open/close states and their positions. Their approach was evaluated on their own database including 78 images of 22 subjects, and got a correct detection rate for eyes of 83%.

The studies above illustrate the development of eye detection from thermal images. However, no research has considered eye localization from thermal images with and without eyeglasses simultaneously. Thus the existing study cannot be directly applied to real situation, where we cannot know whether the subjects wear eyeglasses or not in advance. Thus, in this paper, we first propose the method to recognize whether there exist eyeglasses. For samples without eyeglasses, we propose a facial structure consisting of 15 sub-regions to fully capture the temperature distribution around the eyes for robust eye localization.

3. Methods

In this section, we introduce the main process of our eye localization method consisting of eyeglass detection, eye localization from images with eyeglasses, and eye localization from images without eyeglasses. The framework of our method is shown in Fig. 1.

3.1. Eyeglass detection

Glasses are opaque in the thermal infrared spectrum, and therefore show black in thermal images. Thus the intensity of the eyeglass region is much lower than other regions of the images. It causes valleys in the gray-level integration projection curves. Thus, we define three features as shown in Fig. 2.

First, the original thermal image is converted into grayscale image I, and the horizontal integration projection $P_h(I)$ of image I is calculated according to Eq. (1).

$$P_h(I) = \sum_{x=1}^{W} I(x,y) \tag{1}$$

where $I(x,y)$ represent the gray value of point (x,y) in grayscale image, and W is the width of this image. As shown in Fig. 2(b), the horizontal projection curve generates obvious valleys at the eyeglass regions. Thus, we define the first feature R_1 to describe this valley V_1 as shown in Eq. (2),

$$R_1(V_1) = \frac{H_v(V_1)^2}{W_v(V_1)} \tag{2}$$

where H_v and W_v are the height and width of this valley, as shown in Fig. 2(b). If R_1 is larger than a threshold, the position of this valley is likely to be the vertical position of eyeglass (Y_v). We select this region as the region of interest as shown in Fig. 2(c).

Second, the Otsu threshold algorithm [29] is employed to generate the binary image of this region, as shown in Fig. 2(c). Similar to Eq. (1), the vertical projection of the binary image can be represented by

$$P_v(B) = \sum_{y=1}^{H} B(x,y) \tag{3}$$

where $B(x,y)$ represents the binary value of point (x,y) in the binary image, and H is the height of this image.

Third, the x-coordinates ($\hat{x}_l$ and $\hat{x}_r$) of two maximum vertical valleys (VV_l and VV_r) are selected to represent the horizontal central coordinates of left and right eyeglasses, which are shown in Fig. 2(c). In order to represent the horizontal effect of eyeglasses, another parameter R_2, is calculated by adding the R_1 features of these two valleys VV_l and VV_r,

$$R_2(VV_l,VV_r) = R_1(VV_l) + R_1(VV_r) \tag{4}$$

In addition, the x-coordinate (x_c) of the peak of vertical projection curve is also calculated, which is shown in Fig. 2(c). Generally, x_c is the midline position of ROI, because the temperatures of two eyes' centers are lower than that of both sides. Then, we divide the ROI into two sub-regions through x_c. Similar to Eq. (1), the y-coordinates ($\hat{y}_l$ and $\hat{y}_r$) of the maximum horizontal valleys (HV_l and HV_r) of these two sub-regions' horizontal projection curves are obtained separately. Then, similar to R_2, another feature named R_3 is calculated by adding the R_1 features of valleys HV_l and HV_r as following:

$$R_3(HV_l,HV_r) = R_1(HV_l) + R_1(HV_r) \tag{5}$$

Finally, based on these three features, the SVM classifier with polynomial kernel function is adopted to recognize the samples with or without eyeglasses.

3.2. Eye localization

Through eyeglass detection, all the samples will be classified as samples with eyeglasses and samples without eyeglasses. For the samples with eyeglasses, their eye locations can be determined by $(\hat{x}_l, \hat{y}_l)$ and $(\hat{x}_r, \hat{y}_r)$ calculated from Section 3.1. For the samples

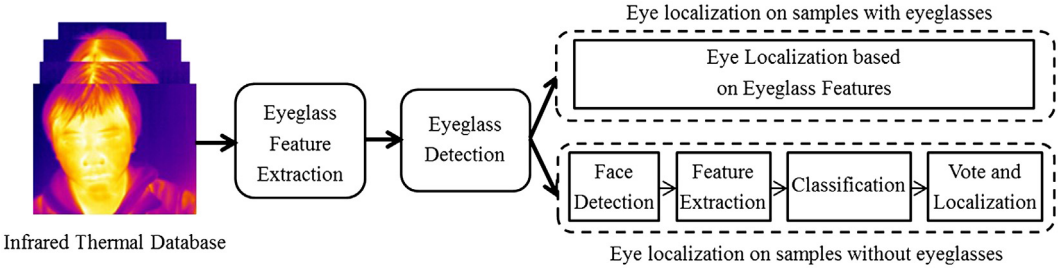

Fig. 1. Framework of eyeglass detection and eye localization.

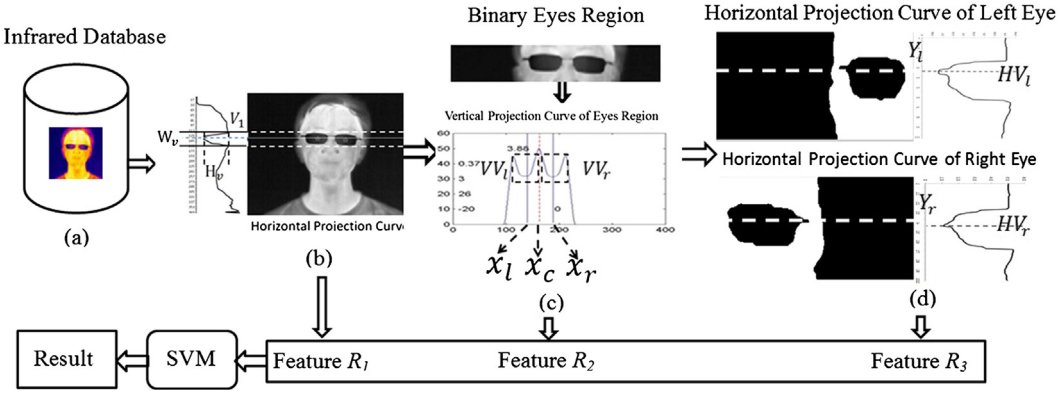

Fig. 2. Glass detection process.

without eyeglasses, the details of eye localization method are introduced as follows.

3.2.1. Face detection

In order to reduce the search area of eyes, we first detect the face automatically. In most cases, the temperatures of human faces are different from those of the environment, so it is feasible to detect a face from thermal images. The Otsu threshold algorithm [29] is adopted to binarize infrared thermal images. Then the horizontal and vertical projection curves are calculated from binarization images, according to Eqs. (1) and (3). After that, the largest gradient of the projection curve is used to detect the face boundary automatically. Finally, face images are normalized to $H_f \times W_f$, in which H_f and W_f are the height and width of face images. In order to enhance the detail of the thermal face, histogram equalization is applied to the normalized face.

3.2.2. Feature extraction and selection

The geometrical and appearance features of eyes in thermal images are so weak that it is even difficult to detect eyes precisely by human beings in some cases. Therefore, it is very important to find useful characteristics in thermal images for eye localization. To do this, an average thermal face is calculated from the training database, as shown in Fig. 3. From the average face, we find the temperature distributions on different facial regions are different, which is further analyzed by an Analysis of Variance (ANOVA) on the mean of sub-regions' temperature in Section 4.3. For example, eyebrows and nose are the coldest region on a human face [16], the cheek is warm, and the left and right eyes are symmetric and are slightly cold. To extract useful features from these areas, we identify a structure of 15 sub-regions around them, as shown in Fig. 3. For the left eye, we assume the center of sub-region 1 is located in the left eye. Then the center of sub-region 6 is the right eye, and the center of sub-region 11 is nose. The centers of other sub-regions are determined by L and S, which are the horizontal distance between two eyes and the vertical distance between eyes and nose, respectively, as shown in Fig. 3. It is similar for the right eye. During the training phase, the centers of sub-regions 1, 6 and 11 are manually located, and the mean and variance of L and S can be computed from all the training samples, which will be used in the testing phase.

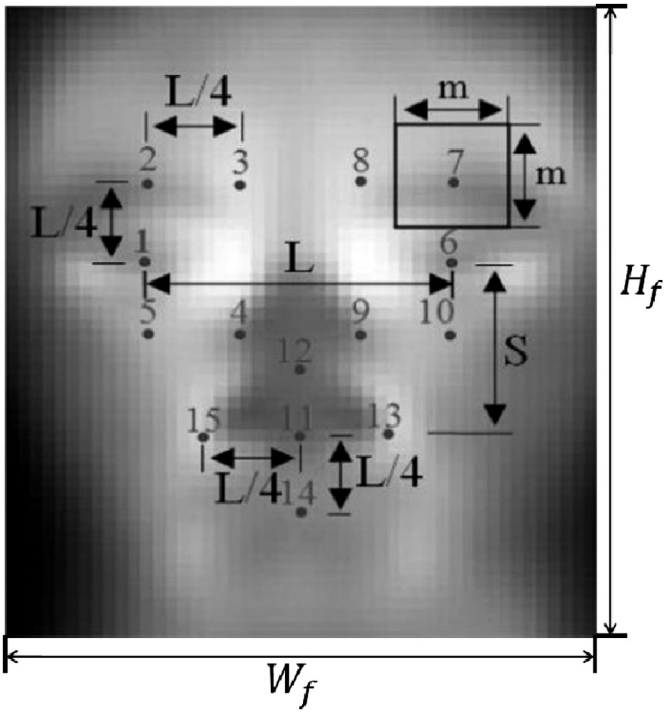

Fig. 3. Fifteen sub-regions on the average face.

Then, eight kinds of Haar-like feature sets, including two edge features (Fig. 4(a)), four line features (Fig. 4(b)), one center-surrounding feature (Fig. 4(c)) and one diagonal feature (Fig. 4(d)) are extracted from each sub-region with size of $m \times m$ [18].

Since the feature dimension of each Haar-like set is very large, feature selection is required. Motivated by the work of Viola and Jones [17], we use AdaBoost algorithm to select the most representative features from each Haar-like feature set. Then the selected features from 15 sub-regions for the same Haar-like set are linearly combined. SVM with linear kernel is used as the eye classifier. Since there are eight Haar-like feature sets, eight left eye classifiers and eight right eye classifiers are trained for left and right eye localization, respectively.

3.2.3. Testing method

In the testing phase, the face is detected and normalized first by the method described in Section 3.2.1. When detecting the left eye, we focus on the upper left part of the face by a sampling step of n pixels. The sampled pixels are regarded as the left eye candidates. Then eight kinds of Haar-like feature vectors are extracted from 15 sub-regions centered around the assumed left eye coordinates, based on the structure obtained from the training phase. After that, the candidate is voted by eight well learned classifiers using the corresponding feature vectors, respectively. The pixel with the most votes is declared as the detected left eye coordinates. A similar process is performed for the right eye localization.

4. Experiments

4.1. Experimental conditions

Two thermal facial image databases, NVIE [19] database and Equinox [20] database, are adopted in this study. The NVIE database contains both spontaneous and posed expressions of 215 subjects (157 males and 58 females), recorded simultaneously by a visible camera and an infrared thermal (wave band 8–14 μm, resolution: 320×240) camera, with illumination provided from three different directions. The posed database also includes expression images with and without eyeglasses. The Equinox database contains four kinds of images, visible, LWIR (8–12 μm, resolution: 240×320), MWIR (3–5 μm), and SWIR (0.9–1.7 μm), with illumination provided from three different directions, front, left, and right, and three expressions on every condition. Since the method we propose is performed on infrared thermal images, the thermal sub-database of NIVE and LWIR samples of Equinox are selected.

We conduct our eye localization experiment in two different ways, within database experiment and cross-database experiment. Within database experiment means training and testing on the same database. In contrast, we conduct our cross-database experiment by training on the NVIE database, but testing on the Equinox database, or training on the Equinox database but testing on the NVIE database, to verify the robustness and effectiveness of our method.

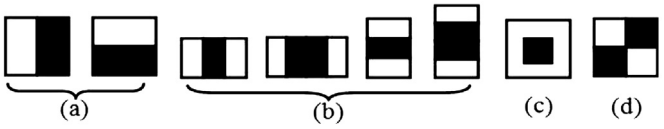

Fig. 4. Haar-like feature sets. (a) Edge features; (b) line features; (c) center-surrounding feature; (d) diagonal feature.

While performing the within experiment of glass detection on the NVIE database, the training set is composed by 1632 images with eyeglasses and 1602 images without eyeglasses, regarding as the positive samples and negative samples, respectively. All of which are selected from the infrared posed expression database of the NVIE database. The infrared spontaneous expression database of NVIE is regarded as the testing set, including 128,557 infrared images (93,027 with eyeglasses, 35,550 without eyeglasses) from 215 subjects. While for Equinox database, 320 images are selected as the training set, composed by 160 with eyeglasses and 160 without eyeglasses. The remaining images are selected as the testing images.

After the samples with eyeglasses are screened out, we performed our eye localization experiment on samples with eyeglasses. For the samples without eyeglasses, we first conduct our eye localization experiment training on NVIE database [19]. A set of 2067 infrared thermal frontal face images from NVIE database are used as training samples, consisting of 1669 posed images and 398 first frame of spontaneous expression image sequences. During training, the centers of sub-regions 1, 6, and 11 of the structure are manually labeled, thus for each image, a set of 15 sub-regions is obtained, which is used as the positive samples for localizing left and right eye in the training phase. The same amount of negative samples are randomly selected from non-eye areas with the same structure of 15 points as positive samples. These samples are used to train eight classifiers for left eye and right eye, respectively. The remaining 35,550 thermal images without eyeglasses from NVIE database are used as the test samples to validate the effectiveness. The eye localization experiment conducted on the Equinox database contains a training set and a testing set both with 420 images. In the training phase, we find that, after face localization and normalization, L is about half of the face width for most samples, while S varies slightly. We suppose S obeys Gaussian distribution, and obtain its mean S_m and variance S_v from training samples. In the testing phase, L is set to $W_f/2$. S is set to S_m, S_m+2S_v, S_m-$2S_v$, respectively. Thus, three kinds of 15 sub-region structures are used during testing. The sampling step, n during testing phase is set to 2 pixels.

In our experiments, the width of face W_f is normalized to be 50, and the height H_f is resized by the same scaling. The resolutions of Haar base detector and sub-region are both 12×12, thus 11,781 features, consisted of eight Haar-like features sets, are extracted from each sub-region.

For images with eyeglasses, we use the eye detection rate to measure the effectiveness of our approach. For images without eyeglasses, motivated by the work of Martinez [25] and Selinger [14], we use two parameters to measure the effectiveness of our eye localization approach. One is error, which is the displacement from automatically located centers of the target eyes to the true (manually annotated) center [25], defined as

$$err_{eye}=\frac{\max(\|P_l-\hat{P}_l\|,\|P_r-\hat{P}_r\|)}{\|P_l-P_r\|} \tag{6}$$

where, P_l and P_r are the true coordinates of the left and right eye, respectively. $\hat{P}_l$ and $\hat{P}_r$ are the automatically detected coordinates of left and right eyes, respectively. $\|.\|$ represents the L_2 norm.

The others are the mean absolute error and the standard deviation of the error [14] in the x- and y-coordinates for the left and right eyes as follows:

$$\begin{cases} AbsMeanError_x=\dfrac{\sum_{i=1}^{N}|\hat{x}_i-x_i|}{N} & AbsMeanError_y=\dfrac{\sum_{i=1}^{N}|\hat{y}_i-y_i|}{N} \\ StDev_x=\sqrt{\dfrac{\sum_{i=1}^{N}(|\hat{x}-x_i|-AbsMeanError_x)^2}{N}} & StDev_y=\sqrt{\dfrac{\sum_{i=1}^{N}(|\hat{y}-y_i|-AbsMeanError_y)^2}{N}} \end{cases} \tag{7}$$

where, N is the number of samples, (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ are the true and automatic localization coordinates, respectively.

4.2. Analyses of the effectiveness of features on glass detection

To validate the effectiveness of our method on capturing the features of wearing eyeglasses, we perform two-sample Kolmogorov–Smirnov (K–S) test [30] on the three features (R_1, R_2, and R_3) extracted from the training set of NVIE database and Equinox database. Results given in Table 1 show that the significance values of all these three features are near or equal to zero on both databases. In other words, these features have significant differences between samples with and without eyeglasses. Thus, they can be used as the criterions to classify samples with or without eyeglasses. The combination of them can achieve good performance.

4.3. Analyses of the effectiveness of our proposed facial structure

We divide 15 sub-regions into 9 groups according to their locations and temperature similarities, as shown in Table 2. Then an ANOVA is performed to analyze the significant difference among the temperature mean of different groups. Based on the analysis results, we can see that the mean temperature of all group pairs except 3 pairs are with significant difference at 0.05 level, as shown in Table 3. It demonstrates that the proposed structure captures the characteristics of the temperature variations around eyes. Thus, it may be helpful for eye localization in thermal images.

4.4. Within database experimental results

4.4.1. Eyeglass detection results and analyses

In our work, the average glass detection rates on NVIE and Equinox database are 99.36% and 95%, respectively. Several successfully localized samples are shown in Fig. 5. Compared with the detection result of Jingu Heo's method [24], who conducted experiment on the thermal images selected from the Equinox database and achieved 93.79% of detection rate, our method is more effective. Since the projection curve technique we use has no

Table 1
K–S test results on samples with and without eyeglasses

Sig.	R_1	R_2	R_3
NVIE database	0.00	7.5371e−190	0.00
Equinox database	6.04e−82	4.452e−89	6.04e−82

Table 2
Groups of facial sub-regions

Group	1	2	3	4	5	6	7	8	9
Sub-region	1, 6	2, 7	3, 8	4, 9	5,10	11	12	13,15	14

Table 3
The significant (Sig.) of mean between different region groups

Sig.	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9
Group 1									
Group 2	0.00								
Group 3	0.00	0.00							
Group 4	0.00	0.07	0.00						
Group 5	0.00	0.11	0.00	0.00					
Group 6	0.00	0.00	0.00	0.00	0.00				
Group 7	0.00	0.00	0.00	0.00	0.00	0.00			
Group 8	0.00	0.00	0.68	0.00	0.00	0.00	0.00		
Group 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

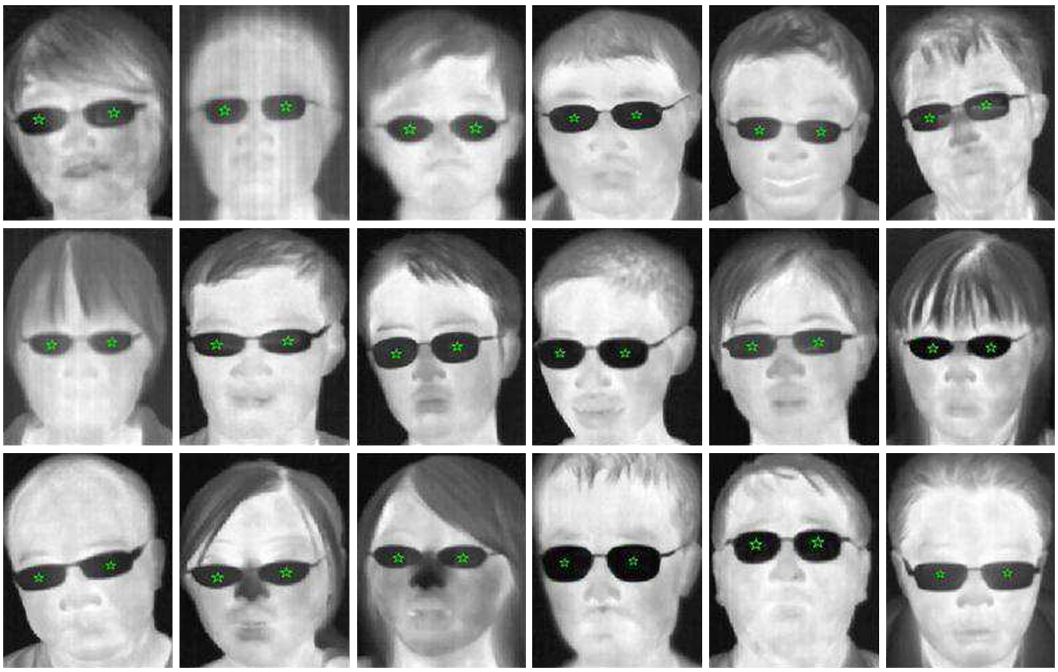

Fig. 5. Some successfully localized samples with eyeglasses.

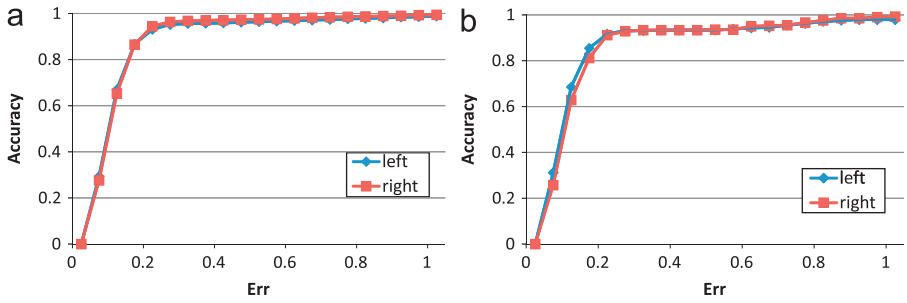

Fig. 6. Eye localization results of within database experiments. (a) Train on NVIE, test on NVIE; (b) train on Equinox, test on Equinox.

complicated calculation, the method we propose is feasible and effective.

4.4.2. Experimental results and analyses of eye localization from images without eyeglasses

Fig. 6 shows the experimental results of eye localization on the NVIE and Equinox databases, respectively. From Fig. 6 and Table 4, we can see that when error < 0.15 is regarded as success, we

Table 4
Eye detection experimental performances on samples without eyeglasses.

References	In this paper		Martinez [25]
Method	Our method		Haar features and GentleBoost algorithm
Database	NVIE	Equinox	78 images of 22 subjects
Accuracy (err < 0.15)	86.00%	83.00%	83.00%

achieve accurate rate of localization around 86% on NVIE database and 83% on the Equinox database. Since the number of training and testing samples from NVIE database vastly exceed that from Equinox database, it is reasonable that the accurate rate training and testing on the Equinox database are slightly a little lower than that on the NVIE database. Some examples of results corresponding to the error accepted (error < 0.15) in both databases are shown in Fig. 7. As shown in Table 4, compared with the results of 83% achieved by Martinez using Haar features and GentleBoost algorithm in their database [25], whose image quality is higher than that of the Equinox database, our results are pretty competitive. The encouraging performances on the NVIE database and Equinox database demonstrate that our method is effective and robust to the changes of facial expressions.

The mean and standard deviation of eye localization errors of training and testing on NVIE database as well as Equinox database experiments are shown in Table 5, respectively. Comparing the results on NVIE and Equinox database, we can find that the mean error testing on Equinox goes up to 2.1286, larger than that of the error training and testing on NVIE database. The reason for that is because the training samples from NVIE database are larger than that from Equinox database. The classifiers trained from former experiment have a better adaptability than that from the latter.

In order to compare with the results of Equinox databases in [14], we recalculated the mean and standard deviation after the outliers, where the detected eye coordinates are at least 10 pixels away from the ground-truth location, are removed from our set as in [14]. From the results shown in Table 6, we can see that both the mean errors and the standard deviations are less than the results obtained by Selinger in [14]. It validates the effectiveness of the facial structure consisting of 15 sub-regions proposed in this paper.

4.4.3. Cross-database experimental results

To verify the robustness of our eye localization method on samples without eyeglasses, we conduct the cross-database experiments on NVIE database and Equinox database separately.

The cross-database experiment results on NVIE database and Equinox database are shown in Fig. 8. We first use the model training from NVIE database, to test on the Equinox database. The results are shown in Fig. 8(a). When error < 0.15 is regarded as success, we achieve accurate rate of localization around 75% testing on Equinox database. Since the classifiers are trained on the NVIE database, it is reasonable and acceptable that the accurate rate of testing on the Equinox database is lower than that of testing on the NVIE database. And for the same reason, the mean and standard deviation, as shown in Table 7, are a little higher than that of testing on the NVIE database. Besides, the model training on the Equinox database is also tested on the NVIE database. The results are shown in Fig. 8(b). When error < 0.15 is regarded as success, the average accurate rate of localization is around 68% testing on the NVIE database. Since the training set from Equinox database is much less than that from NVIE database, it is reasonable that the accurate rates of cross-database experiment testing on the NVIE are lower than that testing on the Equinox database. The mean absolute errors and the standard deviation of errors are also computed, as shown in Table 7. From the table, we can see that the range of the means of errors for cross-database experiment test on the Equinox (from 1.7833 to 2.9381) is narrower than that of the test on NVIE (from 1.4976 to 4.457). In addition, compared with the means and standard deviations of errors for within-database experiment given in Table 5, the values of means and standard deviations of errors are increased in this cross-database experiment, which is also acceptable because that biases are exist between different databases.

5. Conclusions

In this paper, we present an effective method for eye localization on long wave thermal infrared samples both with eyeglasses and without eyeglasses. For eye localization on samples with eyeglasses, we use the valleys and peaks of projection curves to

Table 5
Means and standard deviations of eye detection errors in within database experiment.

Errors	Test on NVIE, train on NVIE		Test on Equinox, train on Equinox	
	Mean	StDev	Mean	StDev
Left x	1.4902	2.3622	1.6405	3.0102
Left y	1.9357	4.6394	2.0524	4.0406
Right x	1.4498	2.3284	1.7881	2.6687
Right y	1.7718	4.0948	2.1286	4.0253

Table 6
Compared with the results of Equinox database in [14].

Errors	Test on Equinox, train on Equinox		Results reported in [14]	
	Mean	StDev	Mean	StDev
Left x	1.1160	1.0478	1.9477	2.0254
Left y	1.1788	1.0327	1.5738	1.6789
Right x	1.3243	1.1581	2.8054	2.0702
Right y	1.2872	1.2646	1.5338	1.6821

Fig. 7. Some successfully localized samples without eyeglasses (err < 0.15).

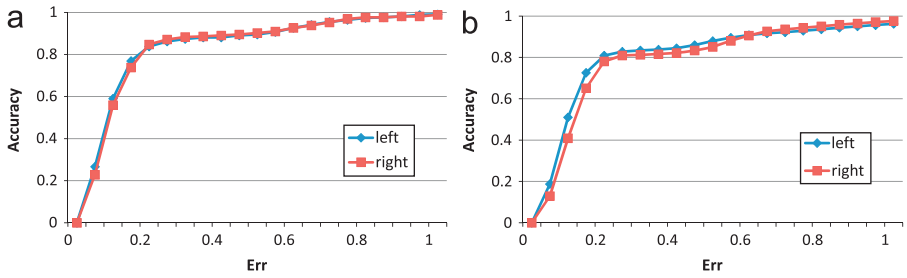

Fig. 8. Eye localization results in cross-database experiments. (a) Train on NVIE, test on Equinox; (b) train on Equinox, test on NVIE.

Table 7
Means and standard deviations of eye detection errors in cross-database experiments.

Errors	Train on NVIE, test on Equinox		Train on Equinox, test on NVIE	
	Mean	StDev	Mean	StDev
Left x	1.7833	2.5986	1.4976	2.4733
Left y	2.9381	4.5205	3.9866	6.5674
Right x	2.0595	3.2315	1.5178	2.5488
Right y	2.7619	4.3417	4.457	6.2856

represent the information of eye regions. For eye localization on samples without eyeglasses, a structure consisting of 15 sub-regions is proposed to extract the Haar-like features to capture the temperature distributions of the eyes and their adjacent facial regions. Eight classifiers are learned from the combination features selected by Adaboost algorithm for left and right eye, respectively. A vote strategy is used to find the most likely eyes. The results of ANOVA demonstrate that our structure captures the useful characteristics of the temperature distributions around eye. The eye localization experiments performed on NVIE and Equinox database verify the effectiveness and generalization ability on multi-expression infrared thermal samples.

Compared with the related four works, our contributions are as follows: (1) we are the first to perform eye localization on samples both with eyeglasses and without eyeglasses, providing the possibility of application. (2) We are the first to define three parameters to capture the temperature character caused by eyeglasses, and obtain good performance of eyeglass detection. (3) Since infrared thermal images reflect the temperature distribution of human faces, we propose a 15 sub-region structure to capture both the temperature distribution of the eyes and that of the adjacent regions for robust eye localization. (4) We evaluate our eye localization approach of samples without eyeglasses on the sub-database of NVIE database, including 35,550 images, which is much larger than the previous three research. The results show that our methods perform better than the previous methods. (5) We are the first to evaluate eye detection from thermal images by a cross-corpus experiment. It demonstrates the generalization ability of our approach.

Conflict of interest statement

None declared.

Acknowledgments

This paper is supported by the NSFC (61175037, 61228304), Special Innovation Project on Speech of Anhui Province (11010202192), Project from Anhui Science and Technology Agency (1106c0805008), and Youth Creative Project of USTC.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (December (4)) (2003) 399–458.
[2] Zhihong Zeng, M. Pantic, G.I. Roisman, T.S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (January (1)) (2009) 39–58.
[3] Vinay Bettadapura, Face expression recognition and analysis: the state of the art, CoRR, abs/1203.6722, 2012.
[4] Seong G. Kong, Jingu Heo, Faysal Boughorbel, Yue Zheng, Besma R. Abidi, Andreas Koschan, Mingzhong Yi, Mongi A. Abidi, Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition, International Journal of Computer Vision 71 (February (2)) (2007) 215–233.
[5] George Bebis, Aglika Gyaourova, Saurabh Singh, Ioannis Pavlidis, Face recognition by fusing thermal infrared and visible imagery, Image and Vision Computing 24 (7) (2006) 727–742.
[6] M. Akhloufi, A. Bendada, J.C. Batsale, State of the art in infrared face recognition, Quantitative InfraRed Thermography Journal 5 (1) (2008) 3–26.
[7] Zhilei Liu, Shangfei Wang, Emotion recognition using hidden Markov models from facial temperature sequence, in: Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction, Volume Part II, ACII'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 240–247.
[8] Masood Mehmood Khan, Robert D. Ward, Michael Ingleby, Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature, ACM Transactions on Applied Perception 6 (February (1)) (2009) 6:1–6:22.
[9] D.A. Socolinsky, A. Selinger, A comparative analysis of face recognition performance with visible and thermal infrared imagery, in: Proceedings of the 16th International Conference on Pattern Recognition, 2002, vol. 4, 2002, pp. 217–222.
[10] Dan Witzner Hansen, Qiang Ji, In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 32(3): 478–500 (2010).
[11] Radek Benes, Pavel Dvorak, Marcos Faundez-Zanuy, Virginia Espinosa-Duro, Jiri Mekyska, Multi-focus thermal image fusion, Pattern Recognition Letters (0) .
[12] Marcos Faundez-Zanuy, Jiri Mekyska, Virginia Espinosa-Duro, On the focusing of thermal images, Pattern Recognition Letters 32 (11) (2011) 1548–1557.
[13] T. Bourlai, Z. Jafri, Eye detection in the middle-wave infrared spectrum: towards recognition in the dark, in: 2011 IEEE International Workshop on Information Forensics and Security (WIFS), 29 December 2011, pp. 1–6.
[14] A. Selinger, D.A. Socolinsky, Face recognition in the dark, in: Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '04), 2004, june 2004, p. 129.
[15] Leonardo Trujillo, Gustavo Olague, Riad Hammoud, Benjamin Hernandez, Automatic feature localization in thermal images for facial expression recognition, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Workshops, vol. 03, CVPR '05, IEEE Computer Society, Washington, DC, USA, 2005, p. 14.
[16] Dvijesh Shastri, Ioannis Pavlidis, Automatic initiation of the periorbital signal extraction in thermal imagery, in: Proceedings of the 2009 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS '09, IEEE Computer Society, Washington, DC, USA, 2009, pp. 182–187.
[17] Paul Viola, Michael J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (May (2)) (2004) 137–154.
[18] Takeshi Mita, Toshimitsu Kaneko, Osamu Hori, Joint Haar-like features for face detection, in: Proceedings of the 10th IEEE International Conference on Computer Vision, vol. 2, ICCV '05, IEEE Computer Society, Washington, DC, USA, 2005, pp. 1619–1626.
[19] Shangfei Wang, Zhilei Liu, Siliang Lv, Yanpeng Lv, Guobing Wu, Peng Peng, Fei Chen, Xufa Wang, A natural visible and infrared facial expression database for expression recognition and emotion inference, IEEE Transactions on Multimedia 12 (November (7)) (2010) 682–691.
[20] Equinox Database. ⟨http://www.equinoxsensors.com/products/hid.html⟩.

[21] D. Socolinsky, A. Selinger, A comparative analysis of face recognition performance with visible and thermal infrared imagery. International Conference on Pattern Recognition (ICPR), 2002, 217–222.
[22] C. Whitelam, Z. Jafri, T. Bourlai, Multispectral eye detection: a preliminary study, in: 20th International Conference on Pattern Recognition (ICPR), August 2010, pp. 209–212.
[23] Tang Jin, Chen Shouming, Xie Xiuzhen, J. Gu, Eyes localization in an infrared image, in: IEEE International Conference on Automation and Logistics (ICAL '09), August 2009, pp. 217–222.
[24] Jingu Heo, Seong G. Kong, Besma R. Abidi, Mongi A. Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition (2004) 122. IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR 2004, July 2004, pp. 94–99.
[25] Brais Martinez, Xavier Binefa, Maja Pantic, Facial component detection in thermal imagery, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2010, pp. 48–54.
[26] Shangfei Wang, Zhilei Liu, Infrared face recognition based on histogram and k-nearest neighbor classification, in: Proceedings of the 7th international conference on Advances in Neural Networks, Volume Part II, ISNN'10, 2010, Springer-Verlag, pp. 104–111, Berlin, Heidelberg.
[27] D.A. Socolinsky, J.D. Neuheisel, C.E. Priebe, J. De Vinney, D. Marchette, Fast Face Detection with a Boosted CCCD Classifier, Technical Report, Johns Hopkins University, 2002.
[28] Eli Shechtman, Michal Irani, Matching local self-similarities across images and videos, in: IEEE Conference on Computer Vision and Pattern Recognition 2007 (CVPR'07), June 2007.
[29] Nobuyuki Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (January (1)) (1979) 62–66.
[30] F.J. Massey, The Kolmogorov–Smirnov test for goodness of fit, Journal of the American Statistical Association 46 (253) (1951) 68–78.

Shangfei Wang, received the M.S. degree in circuits and systems, and the Ph.D. degree in signal and information processing from the University of Science and Technology of China, Hefei, China, in 1999 and 2002. From 2004 to 2005, she was a postdoctoral research fellow in Kyushu University, Japan. She is currently an Associate Professor of School of Computer Science and Technology, USTC. Dr. Wang is an IEEE member. Her research interests cover computation intelligence, affective computing, multimedia computing, information retrieval, and artificial environment design. She has authored or coauthored over 50 publications.

Zhilei Liu, received the Bachelor degree in School of Mathematics and Information Science from the Shandong University of Technology, Zibo, Shandong Province, China, in 2008. He is studying for Ph.D. degree in School of Computer Science and Technology from the University of Science and Technology of China, Hefei, Anhui Province, China. His research interest is Affective Computing.

Peijia Shen, received his B.S. degree in Department of Computer Science and Technology from the HeFei University of Technology, Hefei, Anhui, China. He received his M.S. degree from the School of Computer Science and Technology from the University of Science and Technology of China in 2012, Hefei, Anhui, China.

Qiang Ji received his Ph.D. degree in Electrical Engineering from the University of Washington. He is currently a Professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He recently served as a program director at the National Science Foundation (NSF), where he managed NSF's computer vision and machine learning programs. He also held teaching and research positions with the Beckman Institute at University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Dept. of Computer Science at University of Nevada at Reno, and the US Air Force Research Laboratory. Prof. Ji currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI. Prof. Ji's research interests are in computer vision, probabilistic graphical models, information fusion, and their applications in various fields. He has published over 160 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including NSF, NIH, DARPA, ONR, ARO, and AFOSR as well as by major companies including Honda and Boeing. Prof. Ji is an editor on several related IEEE and international journals and he has served as a general chair, program chair, technical area chair, and program committee member in numerous international conferences/workshops. Prof. Ji is a fellow of IAPR.