Noisy iris image matching by using multiple cues

Tieniu Tan*, Xiaobo Zhang, Zhenan Sun, Hui Zhang

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100190, PR China

Abstract

Noisy iris recognition under visible lighting has recently drawn much attention. This paper proposes an effective method for visible light iris image matching by using multiple characteristics of iris and eye images. The method consists of image preprocessing, iris data matching, eye data matching, and multimodal fusion. Ordinal measures and color analysis are adopted for iris data matching, and texton representation and semantic information are used for eye data matching. After we obtain the four matching scores, a robust score level fusion strategy is applied to generate the dissimilarity measure of the two images under consideration. Extensive experiments on the UBIRIS.v2 database and the NICE.II training dataset demonstrate that the proposed method is effective. Our method significantly outperforms all other algorithms submitted to the Noisy Iris Challenge Evaluation-Part II (NICE.II), an open contest in noisy iris image matching.

1. Introduction

Iris recognition (Daugman, 1993; Ma et al., 2003) is widely regarded as one of the most reliable biometric modalities and significant progress has been made during the past two decades (Bowyer et al., 2008). However, one drawback of most current iris recognition systems is that they require a very high level of user cooperation as well as carefully configured near-infrared (NIR) lighting, which limits the applicability of iris recognition. Recently, iris recognition under visible lighting (Proenca et al., 2010) has attracted much attention. However, due to the physical characteristics of iris, iris texture under visible lighting appears not as stable and clear as that under near-infrared illumination. Moreover, without attentive user cooperation, image quality degrades dramatically due to noise factors such as light reflection, illumination change, eyelashes, eyelids, eye glasses and so on. All these open problems make noisy iris recognition extremely challenging.

We have worked extensively to address the various issues in iris recognition, including those mentioned above. Our work ranges from cooperative systems to non-cooperative ones (Dong et al., 2008a,b, 2009), and from system integration to algorithm development (Ma et al., 2003; Sun and Tan, 2009; He et al., 2009; Dong et al., 2011). Our ongoing long-range iris recognition system has shown promising results (Dong et al., 2009), which can recognize people at a distance. In the past few years, we have also made efforts on visible light iris recognition (Tan et al., 2010), and achieved some encouraging results. The iris segmentation algorithm we developed in that context was ranked as the top algorithm in the Noisy Iris Challenge Evaluation-Part I (NICE.I),1 which aims to evaluate the robustness of iris segmentation algorithms. However, feature representation and matching remain challenging due to noise factors in the images. Therefore, an effective noisy iris matching method is highly desirable.

In this paper, we propose an integrated scheme to match visible light iris images in uncontrolled situations. As shown in Fig. 1, it consists of image preprocessing, feature extraction, matching, and multimodal fusion. First, the normalized eye data and iris data are obtained during image preprocessing. Multiple cues involving texture, color, skin pattern and semantic information are then adopted to characterize the noisy iris image. The dissimilarity score is subsequently obtained via the specific matching strategy. The matching score of two images is finally generated by means of a robust score level fusion strategy.

The remainder of this paper is organized as follows. Section 2 presents a brief analysis and overview of our method as well as an introduction to the experimental datasets. Section 3 describes the preprocessing method. Sections 4 and 5 detail the dissimilarity measures using iris data and eye data respectively. Section 6 describes the fusion strategy for iris and eye data. Section 7 gives the experimental results and discussions. Finally, Section 8 concludes the paper.

* Corresponding author. Tel.: +86 10 82614515; fax: +86 10 62551993.
E-mail addresses: tnt@nlpr.ia.ac.cn (T. Tan), xbzhang@nlpr.ia.ac.cn (X. Zhang), znsun@nlpr.ia.ac.cn (Z. Sun), zhghui@nlpr.ia.ac.cn (H. Zhang).

2. Overview of method and datasets

2.1. Our approach

In visible light iris images (i.e., iris images captured under visible lighting), we observe that iris texture appears not as clear as that in NIR iris images, and changes with the environmental lighting conditions. Moreover, sometimes noises like reflections, light spots, eyelashes, eyelids and glasses frames may occlude the iris region heavily. Therefore, iris texture seen under visible lighting may not be as stable as that in NIR iris images. Fortunately, in addition to iris texture, the iris images contain other distinctive information, including skin pattern, color feature and semantic information. All these traits can be exploited to improve noisy iris image matching.

Usually, iris texture (Ma et al., 2003) is seen as the most important trait of iris images, and most current iris recognition algorithms are based on texture analysis. They have all achieved satisfactory performance. In addition to iris texture, color is another source of information in visible light iris images (Fu et al., 2005), which can be used as a complementary attribute. Moreover, Park et al. (Park et al., 2009) reported that the periocular data, namely the skin pattern around the eye, can also be used for personal identification. Here, we take the whole eye region as a biometric pattern including the iris data and the periocular data. In addition, the structure of eye region provides some important semantic information, such as double/single eyelid, and left/right eye, which can be taken as auxiliary traits for noisy iris image matching.

Based on the above analysis, we propose an integrated framework for noisy iris image matching involving image preprocessing, feature extraction, matching, and multimodal fusion. Fig. 1 shows the flowchart of our proposed method. First, the original input data are preprocessed to obtain the normalized iris data and eye data.

Fig. 2 shows some examples from these two datasets.

2.2. Experimental datasets

The datasets we used are the NICE.II training dataset and the UBIRIS.v2 database (Proenca et al., 2010). The first one is the training dataset used for NICE.II contest,2 and the other one is publicly available.

Images in these two datasets are captured with a high resolution camera under visible lighting during people walking from 8 meters to 4 meters away. UBIRIS.v2 consists of 11,102 images from 522 irises of 261 subjects. The NICE.II training dataset includes 1,000 images from about 170 irises. The image resolution is 300x400. Image quality of the NICE.II training dataset is better than that of UBIRIS.v2.

In addition, in the NICE.II training dataset, a binary image is also provided for each iris image, in which the pixels with intensity 0 denote the valid iris regions, whereas the remaining ones with intensity 255 denote the non iris regions (we call this type of images “mask images” in the rest of this paper). To test our algorithm on UBIRIS.v2 with the same experimental configuration, we also generate the mask images by the algorithm described in (Tan et al., 2010). Fig. 2 shows some examples from these two datasets.

2 Noisy Iris Challenge Evaluation-Part II (NICE.II), http://nice2.di.ubi.pt/.
3. Image preprocessing

As we mentioned in Section 2, both iris data and eye data should be exploited for noisy iris image matching. Therefore in our approach, image preprocessing consists of iris image preprocessing and eye image preprocessing, which are described in the following subsections.

3.1. Iris image preprocessing

Iris image preprocessing usually includes iris detection, iris localization (limbic and pupillary boundary localization), iris segmentation and iris normalization. In the NICE.II contest, the mask images provide enough information for iris detection and iris segmentation, therefore only iris localization and iris normalization are considered here. Fig. 3 illustrates the procedure of iris image preprocessing in our work.

3.1.1. Iris localization

We take iris boundaries as two circles. As we mentioned above, both the original color images and the mask images are provided. Therefore, we can perform iris localization on these two types of images.

The original iris images provide useful information for iris localization as well as noises. Here, we adopt the ItgDiff constellation algorithm presented in (Tan et al., 2010) to localize the limbic and pupillary boundaries. ItgDiff constellation is a modified integrodifferential operator (Daugman, 1993). The ItgDiff constellation method greatly accelerates the computation and also guarantees a global optimum. More details can be found in (Tan et al., 2010).

Pixels in the mask images label the ground truth of the noise-free and noise regions in the corresponding color iris image. Therefore they can provide the edge information for boundary localization. We perform edge detection and curve fitting on the mask images to localize the limbic and pupillary boundaries.

Given an iris image and the corresponding mask image, for either limbic boundary or pupillary boundary, after obtaining two localization results on the two images respectively, we simply choose the one with more edge points around it as the final result. Canny operator is adopted to detect the edge points.

3.1.2. Iris normalization

Since we have obtained the localization results, the rubber sheet model (Daugman, 1993) and linear normalization are adopted to obtain normalized iris images on the R, G, B channels respectively. Meanwhile, normalization is also performed on the mask images.

3.2. Eye image preprocessing

The eye region mainly includes iris, sclera, eyelids, eyelashes and periocular skin. Since we have already obtained the iris location, the remaining work is to normalize eye regions of different images. It is difficult to define the scale of eye region, and usually the iris radius is taken as a measure. Given the predefined normalized iris radius \( R_n \) and the iris radius \( R_i \) in the input image, we resize the image at scale factor \( s = R_n / R_i \). Then we crop a \( 6R_n \times 4R_n \) region centered with iris as the normalized eye image.

4. Image matching based on iris data

4.1. Iris image matching based on ordinal measures

Sun et al. (Sun and Tan, 2009) proposed ordinal measures (OM) for iris texture representation. The ordinal relationship between neighboring image regions is stable and robust. Ordinal measures can be computed by ordinal filters. There are four types of tuning parameters in ordinal filters: the number of lobes (or Gaussian kernels), the scale of each Gaussian lobe, the inter-lobe distance between the centers of two lobes and the orientation (i.e. the angle between the line joining the centers of two lobes and the horizontal line). Ordinal encoding is to convolve the image with an ordinal filter and encode the resultant image with binary bits, as illustrated in Fig. 4.

Recently, He et al. (He et al., 2008) found that, if we divide the iris image into several sub-regions, different sub-regions have different discriminability, and display a certain degree of consistency/correlation. What is more, some of them remain as discriminative as the whole iris image. These properties suggest that it might be possible to achieve similar recognition accuracy while only using parts of the iris regions. In practical applications, many more overlapping sub-regions are obtained in order to completely represent the whole iris image. We adopt the same strategy (SOBoost) of feature selection and iris representation described in (He et al., 2008) to select the most distinctive ordinal filters for accurate and fast iris image matching. More details can be found in (He et al., 2008).

4.2. Iris image matching based on color analysis

It is reported that iris color on its own is discriminative to some extent (Fu et al., 2005), so it is worth combining it with iris texture to improve the performance of iris recognition. However, due to specular reflection and refraction, iris color is sensitive to environment
lighting. Large speckles are likely to occlude iris region which will change its color in the image. In addition, it is insufficient to extract the information by using only one color space, therefore, we use three different color spaces to represent iris color information, namely RGB, HSI, and Ixy. These three color spaces can represent different aspects of color information.

RGB and HSI are two widely used color models, and suitable for different applications. The RGB color model is an additive color model. The HSI color model is close to human color perception, in terms of hue and saturation. The Ixy color space (Ruderman et al., 1998) is proposed to minimize correlation between channels for many natural scenes. It is based on the assumption that human visual system is ideally suited for processing natural scenes. The Ixy space has been successfully used for color transfer between generic natural images and thus constitute the basic elements in early (pre-attentive) visual perception (Zhu et al., 2006).

Two main stages are included in this method: training and testing. In the training stage (see Fig. 5(a)), for each normalized eye image in the training set, the dense SIFT feature extraction (Lowe, 2004) is executed on the R, G, B channels to generate a set of local descriptors. Then the k-means algorithm is adopted to cluster all local descriptors of the images in the training set into K clusters, whose centers are taken as textons or codes. In the testing stage (see Fig. 5(b)), given two input normalized eye images, we use the same method to extract the local descriptors, then two texTons histograms $h_1$, $h_2$ are constructed based on the local descriptors and the learned codes. At last, the dissimilarity score $S_t$ of the input normalized eye images based on texTOn representation is calculated as the chi-square distance between the two histograms, which is defined as follows:

$$ S_t = \frac{1}{N} \sum_{i=1}^{N} \frac{(h_1(i) - h_2(i))^2}{h_1(i) + h_2(i)}, $$

where $N$ is the number of histogram bins.

### 5.2. Eye image matching based on semantic information

Eye region mainly consists of eyelashes, eyelids, pupil and iris in the image. Although eyes are similar to each other in overall structure, some semantic information like left/right eye, and single/double-fold eyelid can also be used for classification. Here, we use the geometrical asymmetry of eyelash distribution as a cue for eye image matching. As we can see in Fig. 6, the eyelashes on the upper eyelid distribute non-uniformly. We find that eyelashes on the side close to tear duct are sparser than those on the other side, and this property can be used to distinguish left eye from right eye.

Based on this semantic trait, we evaluate the density distribution of eyelashes on each side of the eye using two difference filters by assuming that upper eyelid is lighter than eyelashes in images as shown in Fig. 6. The two difference filters have the same scale and two different orientations. And the responses $D_i$, $i=1,2$ are
used to assess the distributions of eyelashes on two sides. At last, the eye image is encoded into one bit $B$ (if $D_1 > D_2$, then $B = 1$, else $B = 0$). Given two eye images, their dissimilarity can be measured by $\text{XOR}(B_1, B_2)$. By passing, the two difference filters are performed on the R, G, B channel respectively, and the final response is the sum of the ones on the three channels.

6. Multimodal fusion

In the preceding sections, we have described four types of feature extraction and matching strategies respectively. These features can be seen as four biometric patterns. They play different roles in classification and have different performances. Hence, we employ an improved score level fusion strategy to combine the four matching scores.

Score normalization is necessary before score level fusion, because the individual matching scores may not be homogeneous. The first step of the score level fusion is to normalize different matching scores into the same order of magnitude. Here, we normalize the four matching scores achieved by different features into $[0, 1]$, where “1” means totally different and “0” exactly the same. We adopt the min–max method as the normalization rule due to its easy implementation and satisfactory performance (Jain et al., 2005). In our work, the maximum and minimum values used for each modality are determined experimentally. We take the values which occur rarely as the outliers, and ignore these values when calculating the maximum and minimum values. The matching score can indicate the degree of confidence about two irises belonging to the same class. These four matching scores generated from ordinal measures, color analysis, texton representation, and
semantic information are denoted by \( S_o, S_c, S_t, S_s \) respectively, where \( S_o, S_c, S_t \) are normalized into \([0,1]\) and \( S_s \) is 0 or 1.

Without changing the distribution of matching scores, we select the weighted sum rule to fuse the four matching scores. The weights are learned from the training dataset via a simple exhaustive search method. The fused matching score is defined as follows:

\[
S = \omega_o S_o + \omega_c S_c + \omega_t S_t,
\]

where \( \omega_i, i = 1, 2, 3 \) are the weights of the three matching scores. To make the fused matching scores distribute more sharply, we adopt a threshold-based decision method to adjust the matching scores. The rules used here are as follows:

1. If \( S_s = 1 \) and \( S < M_1 \), then \( S = M_1 \)
2. If \( S_s = 0 \) and \( S > M_2 \), then \( S = M_2 \)

where \( M_1 \) and \( M_2 \) are learned on the training dataset. After the threshold-based score modification, the matching scores are mapped to a compressed distribution space.

7. Experimental results

7.1. Experiment settings

For iris localization, all images in the UBIRIS.v2 database and the NICE.II training dataset (DB0) are used in our experiments. For image matching, we make the following experiment settings. We use the NICE.II training dataset for SOBoost training and set parameters experimentally. To evaluate the robustness of the proposed method, we conduct test on the UBIRIS.v2 database (Proenca et al., 2010). Firstly, we select 7000 images from the UBIRIS.v2 database, by abandoning images without iris or extremely small irises and seriously blurred images. To test the robustness of our algorithm, we divide the selected 7000 images (All) into seven groups (DB1-DB7) randomly. During experiments, we find that some classes in these 7 groups just include one image, resulting in zero intra-class matching in these classes, which leads to severe imbalance between intra-class pairs and inter-class pairs. In order to make sure that the testing database has similar image quality, inter-class and intra-class distributions as the NICE.II training dataset, we manually select another group (DB8) including 1000 images with enough intra-class pairs. Experiments are also conducted on the NICE.II training dataset. Table 1 shows statistical information of these datasets, in which “OIC” denotes the class with only one image.

7.2. Iris localization

As mentioned above, iris localization is performed not only on the iris images but also on the mask images. Therefore, we make comparison of these two methods and the proposed decision level fusion method on the NICE.II training dataset and the UBIRIS.v2 database. Fig. 7 shows some localization results.

To illustrate the effectiveness of the proposed method (Final), we compare it with iris localization on the original image (Org) and on the mask image (Mask). Both the localization performances of pupillary boundary and limbic boundary are analyzed. Because there is no golden standard to evaluate the performance of iris localization, the normalized center deviation \( D_c \) and normalized radius deviation \( D_r \) are adopted as measures, which are defined as follows:

\[
D_c = |O_0 - O_t| / R_0,
\]

\[
D_r = (R_t - R_0) / R_0,
\]

where \( O_0 \) and \( R_0 \) are the center and radius of manual localization result, and \( O_t \) and \( R_t \) are the center and radius as determined by the algorithm automatically. These two factors are suitable for both pupillary

---

**Table 1**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>No. of images</th>
<th>No. of classes</th>
<th>No. of OICs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB0</td>
<td>1000</td>
<td>171</td>
<td>17</td>
</tr>
<tr>
<td>DB1</td>
<td>1000</td>
<td>396</td>
<td>122</td>
</tr>
<tr>
<td>DB2</td>
<td>1000</td>
<td>385</td>
<td>121</td>
</tr>
<tr>
<td>DB3</td>
<td>1000</td>
<td>399</td>
<td>134</td>
</tr>
<tr>
<td>DB4</td>
<td>1000</td>
<td>406</td>
<td>121</td>
</tr>
<tr>
<td>DB5</td>
<td>1000</td>
<td>405</td>
<td>134</td>
</tr>
<tr>
<td>DB6</td>
<td>1000</td>
<td>397</td>
<td>129</td>
</tr>
<tr>
<td>DB7</td>
<td>1000</td>
<td>408</td>
<td>133</td>
</tr>
<tr>
<td>DB8</td>
<td>1000</td>
<td>131</td>
<td>2</td>
</tr>
<tr>
<td>All</td>
<td>7000</td>
<td>494</td>
<td>8</td>
</tr>
</tbody>
</table>

---

**Fig. 7.** Examples of localization results. The first row presents the original images, the second row and third row show the localization results on the original images and the mask images respectively, and the forth row gives the final results. The left four columns present examples of correct localization whereas the right four columns show examples of mis-localization.
boundary and limbic boundary. Fig. 8 shows the distributions of these two factors on the NICE.II training dataset, and the ones on the UBI-RIS.v2 database are similar.

From this figure, we can see that the performance is not satisfactory by using only the original images. Mask images, however, lead to better results, which indicates that the mask images contain more robust information than the original images, especially for pupillary boundary localization. This is because the mask images are provided as the ground truth. The fusion of these two localization results achieves the highest performance, and this shows that the original images and the mask images provide complementary information for iris localization. Moreover, the $D_c$ and $D_r$ of limbic boundary distribute more sharply than those of pupillary boundary. This suggests that pupillary boundary localization is more challenging than limbic boundary localization.

7.3. Image matching

As we described above, four types of cues for noisy iris image matching are adopted in our algorithm: ordinal measure and SOBoost based iris data matching (OM), color analysis based iris data matching (Color), texton representation based eye data matching (Texton) and semantic information based eye data matching (Semantic). Therefore, we evaluate these measures and the multimodal fusion (Fusion) on our experimental datasets as described in Section 7.1. In addition, the fusion of ordinal measures and color feature of iris data (Iris) is also evaluated.

In the NICE.II contest, a “one-against-all” comparison scheme is conducted for each image, and this scheme gives a set of intra-class dissimilarity values $D^i$ and a set of inter-class dissimilarity values $D^o$. The decidability value $d^*$ is calculated as follows:
\[ d' = \frac{|m_i - m_E|}{\sqrt{(\sigma_i^2 + \sigma_E^2)/2}} \]  

where \( m_i \) and \( m_E \) denote the average values of \( D' \) and \( D'' \), and \( \sigma_i \) and \( \sigma_E \) are the corresponding standard deviations. Fig. 9 shows the matching results in terms of the decidability value \( d' \) on different datasets.

From Fig. 9, we can see that the color analysis based method achieves the lowest performance, and this indicates that the color information on its own is not distinctive enough for iris recognition. The semantic information based eye matching is just a coarse classification, but it also gives high decidability values on different datasets, which means that the semantic information is effective to some extent. The ordinal measures and SOBoost based texture analysis method can not achieve satisfactory performance for visible light iris images matching, and this shows that iris texture varies significantly under visible lighting. The texton representation based eye matching outperforms the ordinal measure based iris data matching and color feature based iris matching. This is because that it considers both iris and periocular patterns. Moreover, compared with the fusion of ordinal measures and color feature of iris data, the texton representation based eye data matching still achieves comparable results on some datasets, which indicates that the eye pattern is more robust than iris pattern in noisy iris images. Our proposed fusion of iris data and eye data method achieves the highest performance, which verifies the effectiveness of multimodal biometrics, and suggests the complementarity of different modalities and features used in this work.

8. Conclusions

In this paper, we have described an effective scheme for matching noisy iris images under visible lighting. It consists of image preprocessing, feature extraction and matching, and multi-modal fusion. In image preprocessing, a decision level fusion method is proposed to localize limbic and pupillary boundaries using the original iris images and the corresponding mask images. For feature representation and matching, multiple cues, including ordinal measures, color histogram, texton representation, and semantic information, are adopted for noisy iris image matching. Regarding multimodal fusion, a robust score level fusion strategy is used to combine the four matching scores into the final dissimilarity measure. Extensive experiments on the UBIRIS.v2 database and the NICE.II training dataset have demonstrated the effectiveness and robustness of the proposed method.

Acknowledgements

This work is funded by National Natural Science Foundation of China (Grant Nos. 60736018, 61075024) and International S&T Cooperation Program of China (Grant No. 2010DFB14110). We also thank Zhaofeng He, Xingguang Li and Man Zhang for their contributions to this work.

References


