Synthesis of Large Realistic Iris Databases Using Patch-based Sampling

Zhuoshi Wei, Tieniu Tan and Zhenan Sun
Center for Biometrics and Security Research, National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences. P.O. Box 2728, Beijing, P.R.China, 100190 {zswei, tnt, znsun}@nlpr.ia.ac.cn

Abstract

This paper presents a framework to synthesize large realistic iris databases, providing an alternative to iris database collection. Firstly, iris patch is used as a basic element to characterize visual primitive of iris texture, and patch-based sampling is applied to create an iris prototype. Then a set of pseudo irises with intra-class variations are derived from the prototype. Qualitative and quantitative studies reveal that synthetic databases are well suited for evaluating iris recognition systems by achieving three goals: (1) the synthetic iris images bear a close resemblance to real iris images in terms of visual appearance; (2) the proposed framework is able to generate databases with large capacity; (3) statistical performance shows that the synthetic iris images hold all the major characteristics of real iris images.

1. Introduction

Significant progress has been made in developing new iris recognition algorithms. However, most of them are evaluated on relatively small databases, some of which are publicly available[1, 2, 4]. Experiments on these databases suffer from the following drawbacks: (1) None of the algorithms has gone through extensive testing, making the performance on large databases unpredictable. (2) It makes the algorithms database dependent, lacking generalization ability. This scenario raises a practical demand of large databases for evaluating and comparing various iris recognition algorithms. However, collecting large databases not only costs time and effort, but also introduces controversy issue due to privacy concern. Synthesizing artificial iris database provides an alternative to cope with this problem.

Synthetic biometrics are relatively new and have been witnessed a steady development in the past few years. In previous studies of iris synthesis, Cui et al.[3] used PCA and super-resolution for iris synthesis, and firstly brought synthetic irises into biometric domain. Makthal and Ross[7] used Markov Random Field (MRF) to generate iris image. Zuo et al.[10] proposed a model based and anatomy based method to generate iris images. The weaknesses of these methods lie in the unrealistic visual effects of synthetic iris[3, 7] and high complexity of the algorithms[7, 10].

This paper addresses iris database synthesis with a target of overcoming these problems. We introduce a novel iris synthesis technique, which can generate a large realistic iris database in a short time. Motivated by the work of Liang et al.[5] about texture synthesis, we apply patch-based sampling to iris synthesis and it turns out to work well in producing the distinctive characteristic of iris texture.

The remainder of this paper is organized as follows: Section 2 presents the framework of our approach. Section 3 describes the patch-based sampling for iris prototype synthesis. Section 4 addresses the problem of derivatives generation. Experiments and analysis are provided in Section 5. Section 6 concludes the paper.

2. Iris Image Synthesis Framework

Fig.1 presents the complete framework of our approach, showing that there are four steps involved in one class image synthesis: (1) An input sample image is formed. (2) A prototype image is created. (3) Multiple images with intra-class variations are derived from
3. Patch-Based Sampling for Iris Prototype Synthesis

Iris-Texton[8], usually contained in iris patches, are proven to be effective representation of the visual primitive of iris texture. Therefore, we use iris patch as basic element to characterize iris. We introduce patch-based sampling[5], which is capable of synthesizing a wide variety of textures, to synthesize iris texture.

The basic idea of patch-based sampling for iris image synthesis is illustrated in Fig.3. Assuming the grey area of the synthetic image $I_{syn}$ is already generated, for the next target patch $P_k$ to be synthesized, its boundary zone $B_{syn}$ is already known. Then the input sample texture $I_{in}$ is searched exhaustively for zone $B_{in}$, with the same shape and size as $B_{syn}$. If the distance of these two zones $d(B_{syn}, B_{in}) \leq \delta$, we consider $B_{in}$ as a match zone of $B_{syn}$, and the patch with the boundary $B_{in}$ is selected as a candidate patch that could be pasted in the target area of synthetic image $I_{syn}$.

For preserving the stochastic characteristic of iris pattern, a candidate patch is selected randomly from the patch set $\Phi_P$ that consists of a number of patches whose boundary zones $B_{in}$ met the constraint:

$$\Phi_P = \{B_{in}| d(B_{syn}, B_{in}) \leq \delta\} \quad (1) $$

$$k = 1, 2, \ldots, N; j = 1, 2, \ldots, \beta $$

where $\delta$ is the tolerance of the boundary zone, and $B_{in}$ is boundary of the $j^{th}$ candidate patch selected from

the input sample texture $I_{in}$. For a given tolerance $\delta$, if $\Phi_P = \emptyset$, then we simply select $\beta$ patches with the top $\beta$ smallest distance $d(B_{syn}, B_{in})$ to form $\Phi_P$.

The algorithm can be stated as follows:

1. Initiate the image by filling adequate values to the first $w_E$ rows and columns as the known area.
2. For $k \leq N$, find all the patches in $I_{in}$ whose zones match the boundary $B_{syn}$ of the $k^{th}$ patch, and obtain the set $\Phi_P$ using Eq.1; if $\Phi_P = \emptyset$, find the top $\beta$ patches whose boundary zones closest to $B_{syn}$ to form $\Phi_P$.
3. Randomly select a patch from $\Phi_P$ and paste it on the target area of $I_{syn}$. Update the $k^{th}$ boundary zone using: $B_{syn} = (B_{syn} + B_{in})/2$. Set $k = k + 1$.
4. Repeat step 2-3 until $k > N$ ($N$ is total number of patches required).

Besides the above procedure, there are two tips specific for iris synthesis: (1) Qiu et al.[8] had shown that iris textural appearance of Westerners differ from that of Asians. Rich textures equally distribute on the overall iris region for Westerners, whereas they often conglomorate on the inner half iris region for Asians. When employing CASIA[2] samples, since they mainly come from Asian, we use the inner half textures to synthesize the inner region and outer half for the outer region. This strategy is not necessary for UPOL[4] irises as they mainly come from Westerner. (2) The first and last column of rectangular iris texture are actually adjacent, because human iris is circular. So the edge is treated toroidally to make the transition smooth to meet this requirement.

4. Generating Multiple Derivatives

In reality, images of the same iris differ in appearance attributed to a number of factors: distortion, defocus, noises, perturbation, rotation, etc. Several measures are employed to achieve these effects. Nonlinear
deformation described in our previous work[9] is applied to derive irises with different pupil sizes, therefore the distorted effects of texture in pupil dilation and contraction can be simulated. Gaussian filter is used to smooth images to achieve defocus effects. Noisy image is obtained by adding several kinds of noises on them (White Gaussian noise, Salt&Pepper noise, etc.). Perturbation effect is accomplished by shifting image pixel positions randomly within an acceptable threshold, and then fill the blank gaps using bilinear interpolation to smooth the image. Rotation on circular iris can be easily achieved by horizontal translation on its polar coordinates. Fig.4 shows an image prototype and some of its intra-class derivatives.

![Figure 4. Multiple derivatives of an image prototype.](image)

5. Experiments and Analysis

We adopt three indicators to analyze the synthetic irises to find out whether they are valid in real instance: (1) visual appearance; (2) database capacity; (3) statistical performance.

(1) Visual appearance: A set of real and synthetic iris images are displayed in Fig.5. We embedded some synthetic irises into real images by replacing the original iris region (shown in Fig.5(j)) to make them more realistic. We conducted a survey on three groups of people, who are asked to find real irises from a mixture image set (10 real vs. 10 synthetic). Group1 has 15 people without iris background; Group2 consist of 15 people trained with 20 real irises; Group3 is composed by 6 researchers in iris domain. Fig.6(a) gives the test results, which manifests that it is very hard to distinguish real irises from synthetic irises. Therefore it can be concluded that the proposed method can generate highly realistic irises, with textural details well preserved.

(2) Database capacity: By changing input sample size, we conducted extensive experiment to see whether prototypes created by the same iris would be duplicate with each other. Results show that one image can create at least 500 prototypes without duplicates. Therefore the capacity of synthetic database is huge given the existing real iris samples. For instance, the 819 classes data in CASIA-IrisV3-Lamp[2] can produce at least 409,500 classes synthetic irises.

(3) Statistical characteristic: Two synthetic datasets denoted as Syn1 and Syn2, are employed in our experiment. Each contains 800 classes with 20 images per class. Syn1 and Syn2 are derived from the same prototype set, but with different parameters used when generating derivatives. By controlling parameters, the global statistics of genuine Hamming distances (HDs) distribution is predictable. Heavier intra-class variation is allowed in Syn1, triggering higher error rates; whereas only minor variations are allowed in Syn2. Fig.6(c) shows how Syn1 and Syn2 vary in genuine HDs, while sharing the same imposter HDs which is made by prototype matching only. Fig.6(b) shows HDs of CASIA[2] irises for comparison. Fig.6(d) shows ROC (Receiver Operating Characteristic) curves of four databases: CASIA[2], Bath[1], Syn1 and Syn2. Table 1 shows a comparison of their recognition results.

Discussion: Preliminary studies show that the synthetic databases are valid by achieving three goals: (1) synthetic irises are highly realistic; (2) the expected capacity of synthetic databases is large; (3) statistical characteristic of synthetic databases is globally controllable, indicating they are well suited for algorithm evaluation.
6. Conclusion

In this paper, we have proposed a novel iris synthesis method and established an effective paradigm to synthesize large iris databases, with the purpose to overcome the problems of data collection. Patch-based sampling is firstly employed to create prototypes, then a number of intra-class images are derived from each prototype. Our approach produces highly realistic irises. Extensive experiments show that the synthetic irises preserve the major properties of real irises and bear controllable statistics, thus promising for algorithm evaluation.

7. Acknowledgement

This work is supported by the National Basic Research Program of China (Grant No. 2004CB318100), the National Natural Science Foundation of China (Grant No. 60736018, 60335010, 60702024, 60723005), the National Hi-Tech Research and Development Program of China (Grant No. 2006AA01Z193, 2007AA01Z162), and the Chinese Academy of Sciences.

References