Iris Image Classification Based on Hierarchical Visual Codebook

Zhenan Sun, *Member, IEEE,* Hui Zhang, *Member, IEEE,* Tieniu Tan, *Fellow, IEEE,* and Jianyu Wang

Abstract—Iris recognition as a reliable method for personal identification has been well-studied with the objective to assign the class label of each iris image to a unique subject. In contrast, iris image classification aims to classify an iris image to an application specific category, e.g. iris liveness detection (classification of genuine and fake iris images), race classification (e.g. classification of iris images of Asian and non-Asian subjects), coarse-to-fine iris identification (classification of all iris images in the central database into multiple categories). This paper proposes a general framework for iris image classification based on texture analysis. A novel texture pattern representation method called Hierarchical Visual Codebook (HVC) is proposed to encode the texture primitives of iris images. The proposed HVC method is an integration of two existing Bag-of-Words models, namely Vocabulary Tree (VT), and Locality-constrained Linear Coding (LLC). The HVC adopts a coarse-to-fine visual coding strategy and takes advantages of both VT and LLC for accurate and sparse representation of iris texture. Extensive experimental results demonstrate that the proposed iris image classification method achieves state-of-the-art performance for iris liveness detection, race classification, and coarse-to-fine iris identification. A comprehensive fake iris image database simulating four types of iris spoof attacks is developed as the benchmark for research of iris liveness detection.

Index Terms—Iris image classification, Hierarchical Visual Codebook (HVC), iris liveness detection, race classification, coarse-to-fine iris identification

1 Introduction

Iris recognition has become a hot research topic driven by its wide applications in national ID card, border control, banking, etc. Iris is a ring-shaped region of human eye with rich texture information under near infrared illumination. Iris texture is regarded as an epigenetic biometric pattern and stable during life so that iris recognition provides an extremely reliable method for individual authentication [1]. Iris recognition aims to assign a unique identity label to each iris image based on automatic preprocessing, feature analysis and feature matching. State-of-the-art iris recognition methods include Gabor phase demodulation [1], ordinal measures [2], etc.

1.1 Motivation of Iris Image Classification

In traditional iris recognition applications, iris images taken from a human eye are defined as the same class so that the dissimilarity between iris images of different subjects should be identified. However, some applications in iris biometrics need to find the similarity between different subjects, which classify iris images into several specific categories. For example, in iris liveness detection, one needs to classify all iris images into two categories, genuine or fake iris images; in some forensic or commercial applications, the racial information of iris images may be required, e.g. race classification of iris images into Asian and non-Asian subjects. Moreover, the classification of all iris images in the central database into multiple categories may help speed up large-scale iris identification. To meet the requirements of these important applications towards a secure, efficient and convenient society, iris image classification methods are necessary to assign an application specific class label (genuine vs. fake, Asian vs. non-Asian, etc.) to each iris image. Iris liveness detection, race classification, and coarse-to-fine iris identification (or iris indexing) are typical applications of iris image classification, so they can be unified into a general framework, as shown in Fig. 1.

1.2 Problem Statement

Both iris image classification and iris recognition can be globally regarded as the same problem of pattern recognition, i.e. classification of iris images into some pre-defined categories. The only difference is the definition of class labels at macro or micro scale. For recognition, the class label is the identity of a person (individual identity). In classification, the class label may correspond to a group of subjects with similar properties of iris images (group identity). So the solution of iris image classification is significantly different to iris recognition. Iris texture naturally has unique pattern for each subject so we can extract the individually specific features to distinguish different subjects. However,

- *Z. Sun and T. Tan are with the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing 100190, China. E-mail: {znsun, tnt}@nlpr.ia.ac.cn.*
- *H. Zhang is with the Institute of Software, Chinese Academy of Sciences, Beijing 100190, China. E-mail: zhanghui@iscas.ac.cn.*
- *J. Wang is with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China. E-mail: jywang@mail.sitp.ac.cn.*

Manuscript received 10 Sep. 2012; revised 16 Sep. 2013; accepted 11 Oct. 2013. Date of publication 4 Dec. 2013; date of current version 12 May 2014. Recommended for acceptance by D. Maltoni.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier 10.1109/TPAMI.2013.234

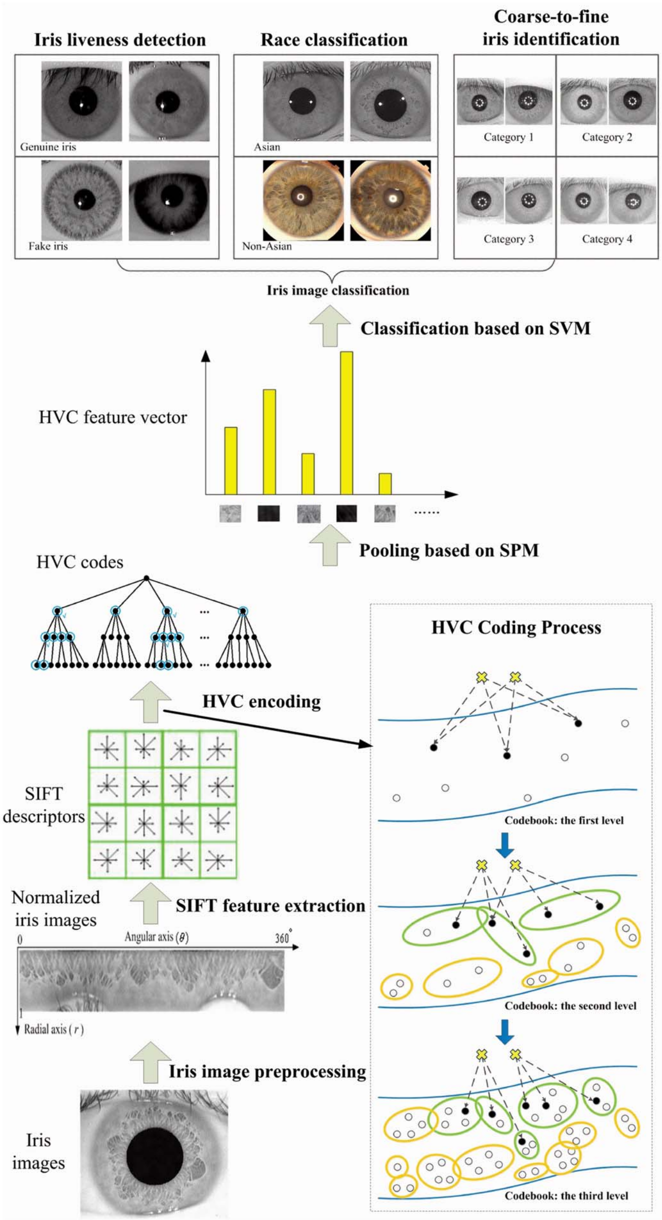


Fig. 1. Flowchart of the proposed iris image classification method.

iris image classification needs to find the stable relationship of similar iris texture features between different subjects. Such an inter-person relationship may be defined manually without support of solid physiological evidences. Therefore, iris image classification is actually a more challenging problem compared with iris recognition. The main challenges of iris image classification are summarized as follows.

1) How to define the common characteristics for each iris image category and the decision boundaries between different categories in the feature space? The visual appearance of human iris is a complex texture pattern. We still cannot achieve a unified definition of texture and texture classes. The definition of iris image categories can be based on ground truth, e.g. genuine vs. fake iris images, Asian vs. non-Asian subjects. However, it is difficult to define the precise decision boundaries between different categories in feature space. Moreover, some applications such as coarse-to-fine iris identification even do not have an explicit definition of iris image categories and the iris images in central database are usually clustered into multiple categories in an automatic and unsupervised manner. So it is inevitable to introduce some ambiguity into the definition of iris image categories.

2) How to represent iris image features suitable for the classification task? Iris images naturally exhibit random texture patterns and we usually use the detailed image features for iris recognition. However, the existing feature representation methods in iris recognition cannot be directly used for iris image classification. On one hand, the iris features for classification should tolerate the within category difference of detailed texture information. On the other hand, the iris features for classification should be distinctive for inter-category iris images. Therefore, the iris features for classification should exhibit both similarity and dissimilarity for cross-identity iris images. For iris images from a same category, the iris features should be clustered closely even these iris images are captured from different subjects. For iris images from different categories, the iris features should be separated at a distance. So it is a grand challenge to develop image feature representation method for iris image classification due to such a contradiction.

1.3 Related Work

Work on iris image classification is driven by specific applications of iris biometrics. So the related works on iris image classification for three kinds of typical applications (e.g. iris liveness detection, race classification, coarse-to-fine iris identification) are summarized as follows.

1) Iris liveness detection. An iris recognition system may suffer from illegal attacks since it is usually used to protect valuable resources [3]. Attack at iris sensor level with forged iris patterns printed on contact lens or paper is a straightforward risk to iris biometrics. So iris liveness detection is necessary to protect an iris recognition system from attacks of fake iris images. There are mainly two approaches to check the liveness of input iris images in the literature. The first approach uses the special features of iris images introduced by optics and illuminators of iris cameras, e.g. specular spots of collimated IR-LED [4] or relationship of reflectance ratio between iris and sclera as a function of wavelength [5]. This approach needs special design of iris sensors. An alternative approach is to check the liveness characteristics of iris images based on texture analysis, which has the advantage of independence of iris sensors. So texture classification of genuine and fake iris images is more flexible for iris liveness detection in practice. Fake iris patterns printed on contact lens, paper, plastic plates and glasses usually generate special high frequency information, so Daugman [6], Ma *et al.* [3] and He *et al.* [7] proposed to detect printed iris images via frequency analysis. Quality based image measures were also used for iris liveness detection [8]. Genuine and fake iris images have distinctive texture patterns, therefore, well developed texture analysis and pattern classification methods can be used for iris liveness detection. The texture features useful for iris liveness detection include gray level co-occurrence matrix [9], statistical distribution of iris texture primitives [10], local binary patterns (LBP) [11] and weighted-LBP [12].

2) Race classification. Automatic identification of the racial attribute of a subject is desirable for many real world applications, such as intelligent marketing, services and forensics, etc. Some attempts have been made to perform race classification based on biometric images. Most of the early work is based on facial images, such as the work in [13], [14] and [15]. It is intuitive that visual appearance of human iris is related to racial information. For example, eyes of Western subjects usually exhibit bright and colorful iris appearance with clear texture patterns. In contrast, eyes of most Asian subjects show brown or dark appearance and it is impossible to obtain detailed texture information in visible lighting. Therefore, we argue that human iris is a biometric trait with both phenotypic and genotypic features, which is a fundamental assumption enabling iris texture based race classification. The phenotypic features of iris biometrics are mainly illustrated in local details of iris texture patterns which are unique to each subject. The genotypic features of iris biometrics are comparatively less well defined and advanced pattern recognition techniques are necessary to find the intrinsic global similarity between iris patterns from the same race. Although we can observe the similarity in both color and texture of iris images between gene-related human eyes (the same race or family, a pair of twins, left/right eyes of the same individual), this paper mainly addresses the problem of automatic race classification based on iris texture features.

We proposed the first texture analysis based racial iris image classification method [16]. Firstly a bank of multichannel 2-D Gabor filters is used to extract the global texture information from iris images and then AdaBoost is adopted to learn a discriminating classifier for race classification. Gabor filters may be not the best operators to describe the detailed texture primitives in iris images. So we developed a specific visual vocabulary namely Iris-Texton to classify Asian and non-Asian subjects based on iris images [17]. Recently we borrowed the ideas of object recognition such as Bag-of-Words model with well optimized codebook for racial iris image classification [18]. And Lyle *et al*. [19] used Local Binary Patterns of periocular biometrics for gender and race classification.

3) Coarse-to-fine iris identification. As a unique and stable biometric modality, iris pattern is used to identify large population in many government and commercial applications. However, large scale iris identification can be much less accurate and less efficient than is commonly believed. A possible solution to speed up large scale iris identification is to classify the large database into a number of categories based on iris texture. Then, a query iris image is the most promising to be identified using the templates from its corresponding category. Such a two stage iris recognition strategy (coarse iris image classification and fine iris matching) is called coarse-to-fine iris identification. The benefit of coarse-to-fine iris identification is multi-folds. 1) It can improve matching efficiency of iris recognition systems. 2) It can reduce the risk of FAR by decreasing the number of inter-class comparisons during iris identification. 3) It can improve the accuracy of iris recognition by integrating global texture information in classification stage and local texture features in recognition stage.

There are a few iris image classification methods proposed for coarse iris image classification. Yu *et al*. described iris surface using fractal dimensions features and classified iris images into four classes based on manually set thresholds [20]. Fu *et al*. used artificial color filter to detect the color information of iris images, and employed margin setting as classifier for coarse iris image classification [21]. In our previous work [22], iris images were grouped into five categories based on statistical description of learned Iris-Textons. Mehrotra *et al*. [23] used energy based histogram of multi-resolution DCT transformation to group iris images. Sunder *et al*. [24] investigated iris macro-features (structures such as moles, freckles, nevi, and melanoma) for iris retrieval and matching.

1.4 Contributions

Compared with iris recognition, iris image classification has not been well defined and addressed in the literature. Existing research work related to iris image classification is scattered in pieces for specific problems. In this paper, we attempt to provide a systematic study of the iris image classification. The main contributions of our work are summarized as follows.

1) This paper systematically addresses the problem of iris image classification, including its definition, core problems, challenges and applications. We propose a general framework for iris image classification based on statistical texture analysis. Three typical applications of iris image classification (iris liveness detection, race classification, coarse-to-fine iris identification) are unified into the general framework. So that research efforts in separated problems can be unified together and advanced solutions to iris image classification will be motivated with the guidance of such a framework. In addition, a common module of iris image classification can be developed for iris recognition systems for various applications inspired by the general framework.

2) A novel texture pattern representation method called Hierarchical Visual Codebook (HVC) is proposed to encode the distinctive and robust texture primitives of iris images. The HVC combines Vocabulary Tree and Locality-constrained Linear Coding to achieve a hierarchical and sparse representation of visual vocabulary. It performs well in typical applications of iris image classification and the results prove that Bag-of-Words model is well suited for iris image classification. In addition, HVC provides a promising approach to learn class-specific visual codebook for general image classification.

3) Our study on HVC based race classification demonstrates the genotypic nature of iris texture. The success of race classification based on iris images indicates that an iris image is not only a phenotypic biological signature but also a genotypic biometric pattern. And coarse-to-fine iris identification also achieves promising results due to the intrinsic similarity of iris patterns between different subjects. However, the success of iris image classification does not contradict the fact of accurate individual authentication. Iris image classification and iris recognition are two different concepts and use texture features at different scales to represent iris pattern. Iris texture at a macro scale can reveal the similarity between iris images but the detailed minute iris features can successfully discriminate individuals. The proposed iris image classification method has wide applications in iris biometrics, e.g. improving security (liveness detection) and efficiency (iris indexing) of iris recognition, analyzing the intrinsic genotypic relationship between iris images from the same race or family, a pair of twins, left/right eyes of the same individual, etc.

4) An iris image database with four types of fake iris patterns (iris texture printed on paper, plastic eyeballs, contact lens and synthetic fake iris images) namely CASIA-Iris-Fake is developed in this paper. The existing research of iris liveness detection are limited on detection of specific fake iris images printed on paper [6], [3], [7], [8], or contact lens [25] due to the lack of a universal benchmark containing a variety of fake iris patterns. The publication of CASIA-Iris-Fake will definitely advance the development of unified countermeasures against iris spoof attacks.

The remainder of this paper is organized as follows. Section 2 proposes a general framework for iris image classification, and the proposed novel iris pattern representation method (Hierarchical Visual Codebook) is described in details. Section 3, 4 and 5 apply the proposed iris image classification method to iris liveness detection, race classification and coarse-to-fine iris identification respectively. Section 6 concludes the paper.

2 Iris Image Classification Based on Hierarchical Visual Codebook

A general framework of iris image classification based on a novel texture pattern representation method called Hierarchical Visual Codebook (HVC) is proposed in this section, as shown in Fig. 1. The framework mainly includes four modules: iris image preprocessing; low level visual feature extraction; statistical iris image representation based on HVC model; iris image classification.

2.1 Iris Image Preprocessing

The first step of the iris image classification framework is iris image preprocessing, i.e. segmentation of the valid iris texture regions from the original iris images and normalization of the ring-shape iris regions into a unified coordinate system. Since the focus of this paper is iris image classification, the iris image preprocessing method in [26] is adopted. Traditionally iris images are normalized into polar coordinate system. In our implementation the size of normalized iris image is 512×80 pixels.

2.2 Low Level Feature Extraction

The objective of low level feature extraction in iris image classification is to obtain the common components of texture primitives across different iris images. So that it is possible to build a statistical representation on the basis of low level visual features for iris image classification. In contrast, low level feature extraction in iris recognition aims to find the unique local features specific to each subject. For example, the Gabor filters [1] and ordinal filters [2] generated iris features in state-of-the-art iris recognition methods are distinctive across different subjects. So low level features used in iris image classification have significantly different purposes to that in iris recognition. And it is not suggested to directly use the well established local features in iris recognition such as Gabor phase information [1] and ordinal measures [2] for iris image classification.

In this paper, we propose to use dense SIFT descriptors [27] as the low level features for iris image classification. Firstly, the gradient information encoded in SIFT provides a generic description of local regions for all iris images. Secondly, the histogram derived from SIFT is distinctive for iris image classification. Thirdly, SIFT is proven as one of the most robust descriptors in image analysis. Therefore we argue that SIFT is well-suited for low feature extraction in iris image classification. In this paper, SIFT descriptors are computed at local regions using a 16×16 window, and each window is divided into 4×4 cells. Image gradients within each cell are quantized into a 8-bin histogram, which results in a 128-D SIFT feature. The SIFT descriptors are densely extracted from the normalized iris images. Finally, 913 SIFT descriptors can be obtained from each normalized iris image as the statistical texture features for iris image classification.

2.3 The Proposed Hierarchical Visual Codebook

Given the low level visual features, it is suggested to obtain the statistical texture representation for iris image classification. So that the individual difference of detailed iris texture can be tolerated and the global texture representation is discriminative enough to distinguish iris images of different categories. Bag-of-Words model (BoW) is demonstrated as the most popular statistical feature representation in object recognition. So BoW is a good starting point to design iris feature representation for classification purpose. The most important issue in BoW is visual codebook learning and coding. The basic idea of our feature representation method for iris image classification is a reliable coarse-to-fine BoW model minimizing coding errors. Therefore our focus is to develop a novel visual codebook adaptive to the iris texture characteristics. Compared with the visual signals in general object recognition or scene classification tasks, iris images do not contain abundant structural information and the main visual features in iris patterns are texture information. In addition, the global texture features between various iris images are much more similar than the visual features between different objects or natural scene images. It indicates that the variations of iris textures are distributed in a relatively small part of the feature space. Therefore it is better to find a visual representation of iris texture to illustrate the detailed differences between iris images from genuine and fake samples (iris liveness detection), Asian and non-Asian subjects (race classification), or Category A and Category B in central database (coarse-to-fine iris identification). Considering these characteristics of iris images, a novel visual feature representation method called Hierarchical Visual Codebook (HVC) is specially developed for iris image classification. The proposed HVC is inspired by two successful algorithms for visual pattern classification, namely Vocabulary Tree [28] and Locality-constrained Linear Coding (LLC) [29].

Vocabulary Tree (VT) was originally proposed to improve quality and efficiency of image retrieval [28]. The basic idea of VT is to hierarchically represent a large set of representative visual words through recursive applications of K-means clustering. Therefore a larger vocabulary can be used to better model the visual contents of images with a much more efficient lookup of visual words. Since iris images are rich of various texture primitives, it is a good idea to extend the idea of VT to iris image classification. The success of a Bag-of-Words model is mainly contributed by a good visual codebook and a good visual coding strategy. Although VT is a good approach to build visual codebook, the hard voting based Vector Quantization (VQ) coding in VT is not a good visual coding strategy. Some coding errors may be introduced into visual representation because of the similarity of some visual words. So we prefer to a soft voting based visual coding strategy such as LLC.

LLC [29] is an effective visual coding scheme which utilizes the locality constraints to project each descriptor into its local-coordinate system with low computational complexity. It aims to reconstruct visual features with locality constraint instead of sparsity constraint based on the following optimization criteria [29]:

$$\min_C \sum_{i=1}^{N} \|x_i - Bc_i\|^2 + \lambda \|d_i \cdot c_i\|^2, \quad s.t. 1^T c_i = 1, \forall i \tag{1}$$

where B is a single level codebook, $\cdot$ is the element-wise multiplication, d_i measures the distance between the visual signal x_i and each vocabulary in the codebook, $C = [c_1, c_2, \ldots, c_N]$ denotes the coding coefficients for X, and λ is a constant to adjust the importance between reconstruction errors and locality constraint.

Vocabulary Tree is an efficient solution for large-scale visual vocabulary and LLC is an effective visual coding method. Obviously it is a good idea to combine the advantages of these two approaches. Therefore a novel visual texture representation called Hierarchical Visual Codebook (HVC) is proposed to inherit the advantages of both Vocabulary Tree and Locality-constrained Linear Coding. In addition the relationship between visual texture primitives in iris images and the overlapping of their feature space are taken into consideration. In general, the HVC method includes codebook learning phase and feature coding phase. These two phases are designed according to the characteristics of iris texture.

In the codebook learning phase, if we learn a normal vocabulary tree following the hierarchical K-means approach introduced in [28], it may be unable to achieve a convergence at the root level of the tree and there may exist a large number of empty clusters in some leaves. Therefore, we adopt a flexible strategy to implement the vocabulary tree. The novel approach is named as flexible vocabulary tree or hierarchical visual codebook. The codebook is denoted as B:$\{B^1, B^2, \ldots, B^{L_{max}}\}$. The maximum number of levels in B is L_{max}, and K_i is the number of clusters partitioned from a parent node in the i-th level. K_1 and K_2 are fixed parameters which are determined experimentally. From the third level, the quantity of child nodes are decided by the number of non-empty clusters during the vocabulary learning process, with an upper limit $K_i (i \geq 3)$. Once the total number of empty clusters in a level is larger than a threshold (e.g. 20% of the maximum number of clusters, this threshold is defined based on experiments and stable high performance can be achieved in a neighboring range of this parameter), the learning process is stopped. The codebook optimization is based on the same method used in [29], which is optional in codebook learning. During the vocabulary tree learning phase in our experiments, the vocabulary learning stops at the third level, and results in a tree with $L_{max} = 3, K_1 = 40, K_2 = 30, K_3 \leq 4$. The parameters K_1, K_2 and K_3 are determined based on the optimal performance achieved in multiple attempts in experiments. Actually there does not exist significant differences in experimental results with variations of these parameters in a certain range. So the parameter settings are consistently used in all experiments of iris image classification.

In our approach, the empty clusters obtained during hierarchical codebook learning are also represented with the same method as the real codes, but these codes have value 0 and are not used for coding. In this section, we use the three levels of hierarchical visual codebook as an example to illustrate the coding algorithm. It is easy to extend the example to the cases with more coding levels. The codebook tree of HVC is denoted as

$$\begin{array}{l} \textit{First level } (B^1 \in R^{D\times N})\text{:}B_a^1,\ a = 1, 2, \ldots, K_1, \\ \textit{Second level } (B^2 \in R^{D\times N})\text{:}B_{ab}^2,\ b = 1, 2, \ldots, K_2, \\ \textit{Third level } (B^3 \in R^{D\times N})\text{:}B_{abc}^3,\ c = 1, 2, \ldots, K_3. \end{array} \tag{2}$$

To decease the quantization errors of VQ coding for VT, the LLC algorithm [29] is used in visual coding. So the visual coding process of HVC is formulated as a constrained optimization problem as follows:

$$\begin{aligned} &\min_C \sum_{n=1}^{N} \{(\|x_n - B^1 c_n^1\|^2 + \lambda \|d_n^1 \cdot c_n^1\|^2) \sum_{n=1}^{N} [(\|x_n - B^2 c_n^2\|^2 \\ &\quad +\lambda \|d_n^2 \cdot c_n^2\|^2) \sum_{n=1}^{N} (\|x_n - B^3 c_n^3\|^2 + \lambda \|d_n^3 \cdot c_n^3\|^2)]\}, \\ &s.t.\ 1^T c_n^1 = 1, 1^T c_n^2 = 1, 1^T c_n^3 = 1, \forall n, \\ &\quad \textit{if } c_{n,abc}^3 \neq 0, \textit{then } c_{n,ab}^2 \neq 0, \forall n, a, b, \\ &\quad \textit{if } c_{n,ab}^2 \neq 0, \textit{then } c_{n,a}^1 \neq 0, \forall n, a \end{aligned} \tag{3}$$

where X is a set of D-dimensional local descriptors extracted from an image, $X = [x_1, x_2, \ldots, x_N] \in R^{D\times N}$, $C = [c_1^1, c_2^1, \ldots, c_N^1, c_1^2, c_2^2, \ldots, c_N^2, c_1^3, c_2^3, \ldots, c_N^3]$ is the set of codes for X. B:$\{B^1, B^2, B^3\}$ is the hierarchical codebook, where the $B^i (i = 1, 2, 3)$ stands for the code collection in the i-th level, where "$\cdot$" denotes the element-wise multiplication, $d_n^t \in R^M$, $t = 1, 2, 3$ is the locality adaptor that gives different freedom for each basis vector proportional to its dissimilarity to the input descriptor x_n, and the parameter λ is a constant to adjust the importance between reconstruction errors and locality constraint for each level visual coding as in LLC [29]. The relational constraints $c_{n,abc}^3$, $c_{n,ab}^2$ and $c_{n,a}^1$ mean that the coding process at lower level just uses a subset of codebook according to the upper level coding results. The coding path down the vocabulary tree is concatenated. Eq. 3 is not a standard optimization problem and it is difficult to find the optimal solution. Instead, an approximate solution is more practical by finding the optimal solution at each level down the vocabulary tree hierarchically.

The reconstruction strategy of LLC feature coding is used to represent an image patch in HVC representation rather than naive voting following a single path down the vocabulary tree. That is to say, the feature vector of an image patch is based on code allocation from selected candidate codes at each level when it is propagated down the tree. We initialize candidate nodes with all the nodes in the first level of B. At the later levels, coding is performed on $K_i \times k, i \geq 2$ candidate codes, where k is the number of the nearest neighbors used at the upper level coding. The solution of the visual coding at the i-th level is derived analytically by

$$\begin{aligned} C^i &= (B^i_{candidate} - 1x^T)(B^i_{candidate} - 1x^T)^T \\ d^i &= exp\left(\frac{\left[\sqrt{\|x-B^i_{candidate_1}\|^2}, \cdots, \sqrt{\|x-B^i_{candidate_{m^i}}\|^2}\right]}{\sigma}\right) \\ \tilde{c^i} &= (C^i + \lambda diag(d^i)))\backslash 1, c^i = \tilde{c^i}/1^T\tilde{c^i} \end{aligned} \tag{4}$$

where $B^i_{candidate,j}, i = 1, 2, 3; j = 1, 2, \cdots, m^i$ is the set of candidate codes for the i-th level, m^i is the number of candidate codes, x is a descriptor, and σ is used to adjust the speed of weight decay for the locality adaptor as LLC [29]. The k codes with the largest value of $\|c^i\|$ are the ones that decide candidate codes for the next level. The HVC coding process follows the down path through the vocabulary tree. Coding results of each level and the number of paths through each node (the number of a visual code has been used for coding) are concatenated as the feature vector of an image patch.

In the coding process of HVC, the LLC coding is performed on several candidate branches of the hierarchical visual codebook for L_{max} times. When there is only one level in HVC, it becomes LLC. So LLC can be regarded as a special case of HVC. In terms of computational cost, the HVC takes more computational time than VT. But it may be more efficient than LLC because only a much smaller size of codebook is used for visual coding at each level.

The proposed HVC method achieves small quantization error owning to the dependency between codes in a down path through the vocabulary tree and sparse coding strategy. The tree structure represents some relationships between the codes and creates overlapping partitions of feature spaces. It can capture salient pattern of local descriptors by local-constrained and parents-constrained coding in each level. The HVC method avoids error accumulating from root level and provides possibility to correct the quantization errors at lower levels by adopting the feature reconstruction strategy for coding.

Max pooling of HVC coding results for all image patches can achieve a powerful statistical feature representation of an iris image. However, histogram representation of HVC features is a description of orderless patch-based visual features so it loses spatial information. A better solution is to combine HVC with the spatial pyramid matching model (SPM) [30] to achieve higher recognition accuracy. But the combination of HVC and SPM will increase the dimensionality of feature vector.

The proposed HVC model has both efficiency and robustness advantages during the codebook learning phase. Firstly, the hierarchical K-means classify all features into a small number of classes, and then classify the subset of features belonging to each class into a small number of classes in the next level clustering. It is clear that learning a small number of codes from a subset of feature pool is more efficient than learning a large number of codes from feature pool. The hierarchical strategy can learn codes from multi-scales. Secondly, the hierarchical K-means clustering shows robustness in clustering tasks. It performs better than K-means when the intra-class variations are large and the inter-class differences are small. The hierarchical K-means is suitable for iris texture codebook learning, since iris texture is highly random pattern but with relatively smaller variations than natural scene images. Some texture primitives are seldom shown in the codebook learning, but they play significant roles in the iris image classification task. These important visual codes with low probability of occurrence may be ignored by the traditional K-means clustering. In contrast, the hierarchical clustering strategy used in HVC is possible to reserve these visual codes in local feature space.

During the coding phase, HVC uses the feature reconstruction strategy for coding, which results in more accurate image representation compared to the VT method [28]. Except the first level coding, the coding process just uses the children nodes of k codes with the largest projections in the upper level as the candidate codebook. This strategy, on one hand, reduces the computational complexity compared with the conventional solutions using all the codes in the vocabulary tree. On the other hand, the HVC replaces the hard voting with feature reconstruction strategy for visual coding, which avoids accumulating errors from root level and provides possibility to correct the quantization errors at lower levels. VT uses hard voting [28], which may cause projection error lasting from the root of the vocabulary tree to leaves. A small quantization error at the root may accumulate into a large quantization error at the leave nodes. HVC coding can solve this problem, because it has more than one path through down the vocabulary tree, which uses codes with different parent nodes to reconstruct a descriptor. It reduces the dependence on upper level coding, and quantization errors can be corrected at the later level coding process.

Given the statistical representation of HVC features, iris image classification becomes a standard pattern recognition problem and well established classifier such as SVM can be used to predict the class labels.

To demonstrate the effectiveness of the proposed iris image classification framework, three typical applications, i.e. iris liveness detection, race classification, coarse-to-fine iris identification, are introduced in Section 3, 4 and 5 respectively.

3 Iris Liveness Detection

There are many ways to make counterfeit iris patterns and Fig. 2 shows some examples. It is a challenging task to discriminate all these kinds of fake iris images and genuine iris images. Here a promising iris liveness detection method is developed based on the proposed iris image classification method.

3.1 Databases

Because of the importance of iris liveness detection, some iris image databases containing iris images with cosmetic contact lenses have been published in public domain in

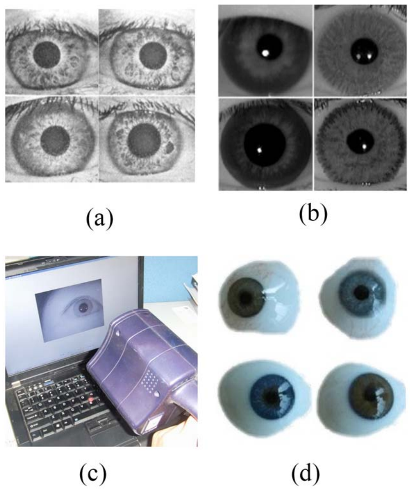


Fig. 2. Some spoof attacks to iris recognition systems. (a) Fake iris patterns on paper. (b) Fake iris patterns on contact lens. (c) Fake iris patterns of re-played video. (d) Fake iris patterns on artificial eyeballs.

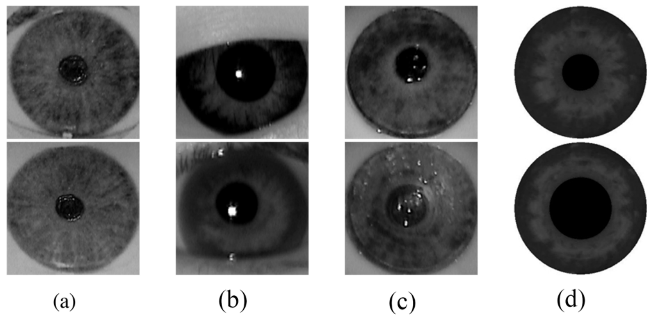


Fig. 3. Some examples of fake iris images. (a) Print. (b) Contact. (c) Plastic. (d) Synth.

recent years. To the best of our knowledge, the Notre Dame Cosmetic Contact Lenses 2013 (or ND-Contact for short in this paper) [31] is the largest one. This dataset contains iris images of subjects without contact lenses, with soft contact lenses, and with cosmetic contact lenses, acquired using an LG4000 iris sensor. In our research, both iris images without contact lenses and with soft contact lenses are regarded as genuine iris images because iris texture patterns are still visible through soft contact lenses to achieve correct identification. And the iris images with cosmetic contact lenses are treated as fake samples.

Although ND-Contact is a good database for research of iris liveness detection, it only has one type of fake iris images, i.e. iris patterns printed on contact lens. Therefore, we developed a more comprehensive database namely CASIA-Iris-Fake for iris liveness detection [32]. This database includes four subsets, namely Print, Contact, Plastic and Synth. In the first three subsets, iris patterns are printed on paper, contact lens, and plastic eyeball model respectively. An iris device IG-H100 [33] is used to capture a large number of fake iris images. The Synth subset contains fake iris images artificially synthesized from iris images with cosmetic contact lens. These four kinds of typical fake iris images have seemingly realistic iris texture and are useful for testing the performance of iris liveness detection methods. Approaches to generate these fake iris patterns are described as follows.

Print: The UPOL iris database [34] contains high-quality iris images with abundant and clear iris texture. So one image of each class is randomly chosen and printed on paper using the Fuji Xerox C1110 printer as the counterfeit input of iris recognition systems. Five fake iris images are captured from each printed iris pattern to construct the Print dataset. There are totally 640 images in this dataset. Fig. 3(a) shows some examples.

Contact: Cosmetic contact lens are popular currently so we collected 57 kinds of cosmetic contact lens with different texture patterns. Some volunteers are asked to wear these contact lens and then an iris device is used to capture iris images of these subjects. There are totally 74 left and right eyes wearing these contact lens. Five fake iris images are captured from each eye to construct the Contact dataset (including 740 images in total). Fig. 3(b) shows some examples.

Plastic: There are 40 inter-class iris images from UPOL database [34] printed on the cover of plastic eyeball models. Ten fake iris image are captured per sample. Therefore there are totally 400 fake iris images in the Plastic dataset. Fig. 3(c) shows some examples.

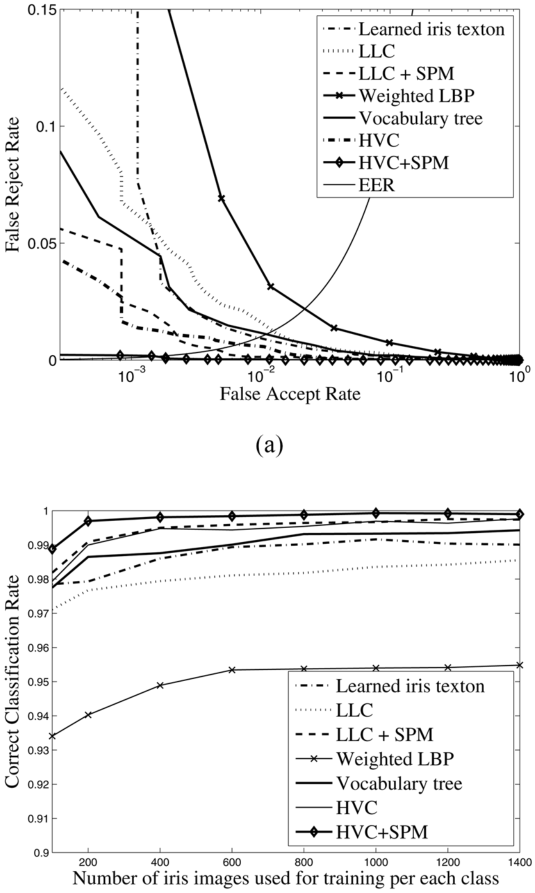
Synth: The idea of iris image synthesis is used to generate fake iris images based on the Contact dataset. We adopt the patch-based sampling method [35] for synthesis, and both intensity and texture features are considered to select sampling patches with smooth transition boundaries. Various intra-class variations (distortion, defocus, noise, perturbation, and rotation) are introduced to generate derivatives for each prototypes. The Synth dataset includes 590 synthesized iris image prototype and their $2,360$ intra-class derivatives. Fig. 3(d) shows some examples.

We use the same device to capture $6,000$ genuine iris images from $1,000$ subjects as the positive samples in the experiments of iris liveness detection.

3.2 Experiments

Three experiments are carried out to test the performance of iris liveness detection methods under different conditions. Firstly, seven genuine/fake iris image classification methods are tested on the overall database of CASIA-Iris-Fake with a mixture of all four fake iris subsets. Secondly, the detailed performance of five representative methods is illustrated in the single dataset of CASIA-Iris-Fake and ND-Contact. Thirdly, the robustness and generalization capability of machine learning based iris liveness detection methods are evaluated on the CASIA-Iris-Fake when the training and testing datasets contain different types of fake iris images.

1) Experiments on the combined CASIA-Iris-Fake database: To evaluate the overall performance of iris liveness detection methods, the four fake iris image datasets of CASIA-Iris-Fake are combined together. Seven state-of-the-art genuine/fake iris image classification methods, including learned iris texton [10], weighted LBP [12], LLC [29] with and without SPM (three levels of SPM$[1 \times 1, 2 \times 2, 4 \times 4]$), Vocabulary Tree [28], and the proposed HVC with and without SPM (two levels of SPM$[1 \times 1, 2 \times 2]$), are tested on the combined fake iris image database. There are 400 fake and 400 genuine iris images randomly selected for normal codebook learning. The normal codebook has $1,024$



(a)

(b)

Fig. 4. Comparison of iris liveness detection methods on the combined database. (a) ROC curves. (b) CCR curves as a function of the number of training samples.

codes; the size of flexible vocabulary tree is mentioned in Section 2.3; the vocabulary tree has the same size as flexible vocabulary tree. These codebook and vocabulary trees are also used in the later experiments on CASIA-Iris-Fake. We use 600 fake and 600 genuine iris images as positive and negative training samples, and others for testing. The ROC curves and the Correct Classification Rate (CCR) curves as a function of the number of training samples are shown in Fig. 4. The performance metrics such as Equal Error Rate (EER), CCR, feature dimensionality and computational cost are shown in Table 1. The computational cost of all methods listed in this paper is tested on the same computer (Intel Core i5-2410M 2.3 GHZ CPU, 2GB RAM, Win7 32bits System) with Matlab 2011 as the programming software.

A number of conclusions can be drawn from the experimental results. Firstly, all iris image classification methods based on BoW can achieve high accuracy in detection of various fake iris patterns. Secondly, the iris liveness detection methods based on learned BoW model (e.g., learned iris texton, VT, LLC, HVC) generally perform better than the BoW model without learning such as LBP. Thirdly, the proposed HVC is the best performing BoW model for iris liveness detection. Fourthly, SPM can slightly improve the accuracy of both LLC and HVC with the cost of much higher dimensionality of feature vector. HVC with only two levels of SPM can already achieve better performance than LLC with three levels of SPM. The experimental results demonstrate the spatial distribution information of iris texture primitives is not critical to iris image classification. But spatial operators such as SPM can provide complementary information to BoW. So it is suggested to combine HVC and SPM for detection of fake iris patterns in highly secure applications.

2) Experiments on the single dataset: To investigate the performance of iris liveness detection on different types of fake iris images, all BoW learning based methods are tested on the five datasets of CASIA-Iris-Fake and ND-Contact individually. The weighted LBP [12] is not used for comparison since it is based on predefined BoW model and it does not perform as well as the BoW learning methods. All the methods are performed on the whole normalized iris image without dividing it into small blocks. For the four datasets of CASIA-Iris-Fake, we use 100 fake and 100 genuine iris images as training samples, and the other images are used as the testing samples. For ND-Contact, we use the setting of training and testing datasets defined by the database provider, i.e. a training set of 3,000 images including 2,000 genuine samples and 1,000 fake samples and a testing set including 800 genuine samples and 400 fake samples.

Experimental results are shown in Table 2. We can see that the best performing iris image classification method for detection of single type of fake iris patterns is the proposed HVC method (with or without SPM). And VT, LLC and HVC can achieve perfect results in ND-contact. To the best of our knowledge, state-of-the-art iris image classification methods can only achieve a CCR around 98% on the ND-contact [25]. So the visual codebook ideas introduced in this paper such as VT, LLC and HVC significantly improve state-of-the-art iris liveness detection performance. To demonstrate the advantage of the proposed HVC over LLC on the ND-contact, only 30% training images in ND-contact are used to train HVC and LLC and the remained 70% training images are also used for testing. The results

TABLE 1
Performance Metrics of Iris Liveness Detection Methods on the Combined Database

Method	CCR (%)	EER (%)	Dimensionality	Computational cost			
				Codebook learning	Feature extraction	Classifier training	Classification
Learned iris texton [10]	98.93	0.96	1,024	Fast	0.69s	0.57s	0.43ms
Weighted LBP [12]	95.34	2.22	576	N/A	1.02s	0.43s	0.41ms
LLC [29]	98.11	1.29	1,024	Fast	0.73s	0.57s	0.41ms
LLC with SPM [29]	99.59	0.50	21,504	Fast	0.77s	0.60s	0.46ms
Vocabulary Tree [28]	99.01	1.10	12,080	Slow	1.31s	0.73s	0.48ms
HVC	99.51	0.69	12,080	Slow	1.22s	0.74s	0.50ms
HVC with SPM	99.79	0.18	60,400	Slow	1.30s	1.42s	0.51ms

TABLE 2
Performance Metrics of Iris Liveness Detection Methods on the Single Dataset

Database	Learned iris texton [10]		LLC with SPM [29]		Vocabulary Tree [28]		HVC		HVC with SPM	
	CCR(%)	EER(%)	CCR(%)	EER(%)	CCR(%)	EER(%)	CCR(%)	EER(%)	CCR(%)	EER(%)
Print	99.54	0.70	99.74	0.15	99.48	0.98	99.61	0.03	99.96	0
Contact	98.18	5.44	98.03	1.55	98.50	2.83	99.64	0.55	99.32	0.64
Synth	98.97	1.05	99.15	0.58	99.40	0.94	99.76	0.19	100	0
Plastic	97.58	1.38	99.52	0.18	99.55	1.13	99.79	0.13	99.97	0
ND-Contact	94.69	4.31	100	0	100	0	100	0	100	0

(CCR and EER) of LLC, LLC+SPM, HVC, HVC+SPM are (94.07% and 3.40%), (98.29% and 0.84%), (98.29% and 0.84%), (99.00% and 0.53%) respectively. So HVC performs better than LLC in a more challenging test on the ND-contact with less training samples.

The experimental results also show that CASIA-Iris-Fake is a more challenging database than ND-contact due to three reasons. Firstly, CASIA-Iris-Fake has four types of fake iris images but there is only one type of fake iris images with cosmetic contact lens in ND-contact. Secondly, the iris sensor used for collecting CASIA-Iris-Fake is IG-H100, which is an earlier version of commercial iris sensors compared with the latest iris sensor LG4000 used for collecting ND-contact. The results show that iris image quality is also important to achieve high accuracy of iris liveness detection. Thirdly, each dataset of CASIA-Iris-Fake has much smaller number of training images (100 positive and 100 negative samples) than that of ND-Contact (2,000 positive and 1,000 negative samples).

The experimental results of iris liveness detection on the cross dataset can be found in the online Appendix, which is available in the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/234 due to page limitation.

3.3 Discussions

We can learn from the experimental results in this section what are the most important issues in development of an effective iris liveness detection method.

(1) A good strategy: The problem of iris liveness detection can be solved in hardware level, software level or a combination of hardware and software. This paper demonstrates that iris image classification provides a good strategy to detect possible spoof attacks to an iris recognition system at software level. Such a solution to iris liveness detection does not need special design of iris devices and it is applicable to all kinds of iris sensors.

(2) A good representation: There are many possible ways to develop an iris image classification method for iris liveness detection. Our experimental results demonstrate that learning visual codebook specific to genuine and fake iris patterns can generate a good feature representation for this purpose. State-of-the-art BoW models in visual recognition such as Vocabulary Tree [28] and LLC [29] are good approaches for training the visual representation of iris liveness detector. This paper proposes a novel visual codebook namely HVC, a combination of VT and LLC, which performs the best in terms of accuracy and generalization for iris liveness detection. The advantage of HVC over VT and LLC is much more significant in real world applications with limited training samples.

(3) A good training dataset: The experimental results on cross datasets (see Appendix, Available online) show that the performance of iris liveness detection methods may degrade greatly when the special types of fake iris patterns in testing have not been used for training. This observation reminds us to include all kinds of fake iris patterns for training. But it is impossible to know all possible fake iris patterns used by the attackers since iris recognition systems are usually installed in a open environment. There are two possible solutions to this problem. One is to develop a robust iris image classification method such as HVC. The other is to update the training data of fake iris patterns online, just like the update of virus database in anti-virus software.

4 Racial Iris Image Classification

Race is a classification system used to categorize humans into genetically differentiated populations or groups defined by phenotype [36]. There are a large number of different races in the world. And the parents of some subjects may come from different races so it is difficult to precisely determine the race category of these subjects, which may generate ambiguity in research of racial classification. Nevertheless, it is meaningful to develop an automatic race classification based on biometric patterns for promising commercial and forensic applications. And a good start point is the research of classifying subjects with significant racial distinction, which is a well defined pattern classification problem. For example, this paper mainly discusses iris biometrics based race classification of typical Asian and non-Asian subjects. In our experiments, almost all Asian subjects are Chinese and all non-Asian subjects are white people living in Europe or USA. So the ground truth of class label (Asian or non-Asian) is clearly defined in our research.

4.1 Databases

Three multi-race iris image databases are used in this paper to evaluate the effectiveness of racial iris image classification methods (Fig. 5).

CASIA: The CASIA multi-race iris image database [37] was collected with an handheld iris device OKI irispass-h [38]. It contains 2,400 iris images of 30 Chinese subjects and 30 European subjects, i.e., 20 images/eye. Randomly selected 500 Asian iris images and 500 non-Asian iris images are used as the training set and all the remained iris images are used as the testing set.

ND: ND-CrossSensor-Iris-2013 Dataset is a large iris image database in the literature and it has race label information for each image [31]. So it is a good database for research of racial iris image classification. However, most subjects in this database are white people and each subject

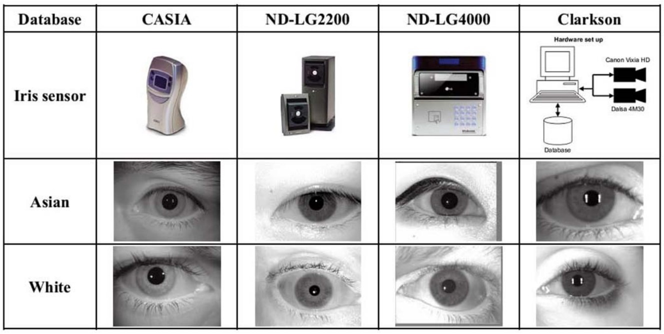


Fig. 5. Illustration of the multi-race iris image databases used for racial iris image classification.

has a large number of iris images across multiple sessions. So we need to select a subset from it namely ND multi-race iris image database for our research of racial iris image classification. There are two iris devices (LG2200 and LG4000) used to collect ND-CrossSensor-Iris-2013 Dataset. In the dataset of LG2200, 1,194 iris images of 60 Asian subjects and 10,660 iris images of 534 white people are used for race classification experiments. In the dataset of LG4000, 1,124 iris images of 60 Asian subjects and 9,641 iris images of 534 white people are used for race classification experiments. To keep the balance between the iris images of Asian and non-Asian subjects, 8,188 iris images of 411 Asian subjects from CASIA-Iris-Lamp [39] are added into the LG2200 and LG4000 dataset. In each dataset, randomly selected 2,000 Asian iris images and 2,000 non-Asian iris images are used as the training set and all the remained iris images are used as the testing set.

Clarkson: The Q-FIRE database collected by Clarkson [40] has iris images captured at a distance and race label information is provided for each subject. So it is interesting to study the possibility of racial iris image classification using a long range iris device. We select 1,251 iris images of 63 white people in this database. However, this database does not have enough iris images of Asian subjects so we add some iris images in the CASIA-Iris-Distance [39] to constitute a subset of 63 Asian subjects with 1,244 iris images. Randomly selected 500 Asian iris images and 500 non-Asian iris images are used as the training set and all the remained iris images are used as the testing set.

To establish a benchmark for research and comparison of race classification methods, the list of all ND, Clarkson and CASIA iris images selected for the experiment can be found in the document [41].

4.2 Experiments

CASIA: We have mentioned in Section 2.2 that SIFT is a suitable low level feature descriptor for iris image classification. So CASIA multi-race iris image database is used to compare the effectiveness of various low level visual features including Gabor, LBP and SIFT for racial iris image classification. The experimental results are shown in Table 3 and a number of conclusions can be drawn as follows.

Firstly, SIFT can achieve much higher accuracy than Gabor and LBP when these low level feature descriptors are combined with HVC for racial iris image classification. Although Gabor filters [1] can achieve state-of-the-art performance in iris recognition, they are not the best choice for iris image classification. This observation demonstrates the difference between the feature representation in iris recognition and iris image classification.

Secondly, HVC is a significantly better feature representation method than LLC in the task of racial iris image classification.

Thirdly, the proposed racial iris image classification method (SIFT and HVC with SPM) can successfully discriminate iris images of Asian and non-Asian subjects with an extremely high accuracy. The experimental results demonstrate that iris texture pattern is an useful indicator of racial category of a subject. Therefore the iris biometrics must be a genotypic biometric trait.

ND: The racial iris image classification results (Table 4) on the ND multi-race iris image database consistently demonstrate the advantages of the proposed HVC model. The results also show that racial iris image classification methods can achieve higher accuracy in LG4000 dataset because LG4000 is a more advanced iris device than LG2200 and it can capture higher quality iris images. This conclusion remind us to use high-quality iris devices to capture clear and detailed iris texture for race classification.

Clarkson: The high accuracy of racial iris image classification achieved in long-range iris recognition applications (Table 5) demonstrates the possibility to predict the racial category of a subject at a distance. So that the proposed racial iris image classification method has wide applications in commercial or forensic areas. For example, if a long-range iris device can recognize that the coming customer is an Asian subject from 3 meters away in a less cooperative manner, the automatic vending machine can recommend the most favorite products of Asian users to him. The results on this database again prove that HVC is the best performing visual representation method for iris image classification.

TABLE 3
Comparison of Race Classification Methods on the CASIA Multi-Race Iris Image Database

Method	EER(%)	CCR(%)
Learned iris texton	18.14	82.07
Gabor and HVC	13.57	86.50
LBP and HVC	25.50	74.86
SIFT and LLC	1.57	98.14
SIFT and LLC with SPM	0.14	99.86
SIFT and HVC	0.14	99.86
SIFT and HVC with SPM	0	99.93

TABLE 4
Comparison of Race Classification Methods on the ND Multi-Race Iris Image Database

Method	LG2200		LG4000	
	EER(%)	CCR(%)	EER(%)	CCR(%)
Learned iris texton	11.95	88.08	10.54	89.79
SIFT and LLC	3.90	96.47	2.90	97.01
SIFT and LLC with SPM	2.09	98.12	1.36	98.87
SIFT and HVC	2.77	97.28	1.35	98.69
SIFT and HVC with SPM	1.71	98.30	0.90	99.15

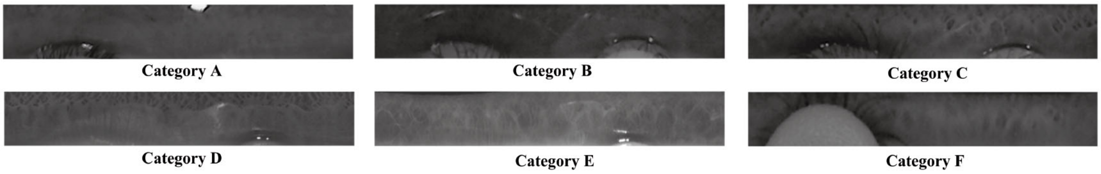


Fig. 6. Some typical iris images of different categories.

5 Coarse-to-Fine Iris Identification

Coarse-to-fine iris identification is a smart strategy for practical large-scale applications. It is a good idea using the proposed HVC method to classify iris images into multiple categories. Some typical iris images from different categories are shown in Fig. 6.

In this paper, a two stage coarse-to-fine iris identification strategy is implemented. Firstly, iris indexing is performed based on iris image classification. Secondly, the query iris image is matched with templates in the retrieved candidate dataset one by one. As we know, iris indexing is a classification task without ground truth category labels. Categorization is performed by manual labeling or automatic clustering. We adopt an automatic clustering strategy which uses K-means clustering to partition iris images into several categories. Each category is represented by the cluster center vector. A training iris image is labeled according to its nearest clustering center.

The process of iris indexing includes training and testing phases. The training phase learns the hierarchical visual codebook, clusters iris images and trains classifiers. The testing phase uses the learned classifier to classify a query iris into one category. The linear SVM classifier [42] is adopted. Since SVM cannot be directly used for multi-class classification, a strategy of one against the rest is used. The iris image classification result cannot guarantee 100% accuracy, so that in practice we can search the high confidence matching template in the most similar category firstly.

5.1 Coarse Iris Image Classification Experiments

The coarse iris image classification experiments are conducted on the large iris image database CASIA-Iris-Thousand [39], which includes $20,000$ iris images of $2,000$ eyes. One image from each eye in the database is randomly selected to construct training set, and remained images are used to construct testing set. The K-means is adopted to cluster iris images into several categories ($2-10$ categories). The iris indexing method proposed in [22] (denote as "Qiu") are adopted as the benchmark method. To evaluate the performance of HVC coding with different low-level features, Gabor features are tested in the HVC framework for iris image representation (denoted as "HVC Gabor"). The iris image classification method using HVC with SIFT features is denoted as "HVC SIFT". The nearest neighbor (NN) and SVM classification methods are implemented for comparison.

1) In the first experiment, the NN classifier is used. Euclidian and Chi-square distance are adopted as the dissimilarity functions. A query image is compared to all the images used for training and classified into the category of its nearest neighbor. Fig. 7(a) shows CCR as a function of the number of categories. It is more challenging to achieve high classification accuracy with the increasing number of categories. The "HVC SIFT" method achieves the best and the "HVC Gabor" method is better than "Qiu" [22]. Results show that SIFT descriptor is more suitable for iris image classification than Gabor features and the HVC coding method can significantly improve the accuracy of BoW model. The difference between Euclidian distance and Chi-square distance is not significant for "HVC SIFT" method in this experiment. But it is much better to use the Chi-square distance as the dissimilarity function for "HVC Gabor" method.

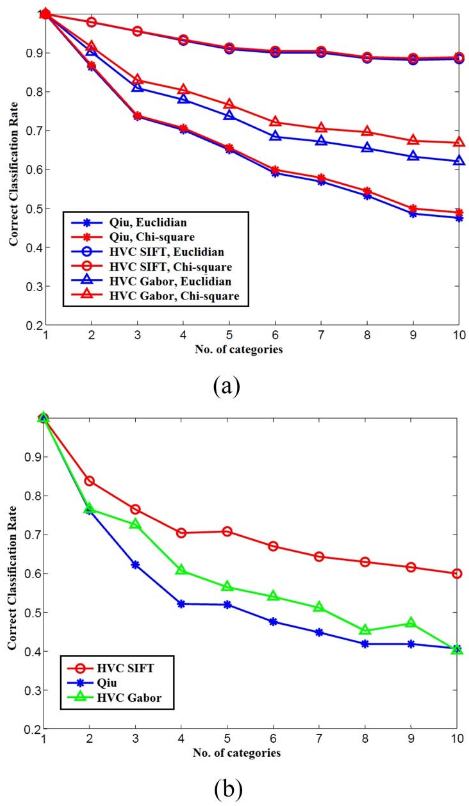


Fig. 7. Performance of iris image classification for coarse-to-fine identification applications. (a) Results based on the nearest neighbor classifier. (b) Results based on the SVM classifier.

TABLE 5
Comparison of Race Classification Methods on the Clarkson Multi-Race Iris Image Database

Method	EER(%)	CCR(%)
Learned iris texton	8.29	91.51
SIFT and LLC	1.87	98.40
SIFT and LLC with SPM	1.74	99.23
SIFT and HVC	1.20	99.13
SIFT and HVC with SPM	0.54	99.26

2) In the second experiment, SVM is used as the classifier. Fig. 7(b) shows CCR performance as a function of the number of categories. Experimental results demonstrate that SVM is worse than the NN classifier in iris image classification, but much faster than the later one. The main reason is that it is difficult to establish the absolute separation plane between iris image categories using SVM. Compared with the iris texton method [22] and "HVC Gabor" method, "HVC SIFT" method also achieves the best performance in this experiment.

The proposed HVC method is effective for coarse-to-fine iris identification. Nearly 90% CCR can be achieved for ten-category iris image classification, which means it is a high probability event to authenticate the identity of query iris image by searching less than 10% templates in the central database. Moreover, the proposed method can be used in iris recognition systems with proper designed classification confidence threshold. Section 5.2 gives an example experiment.

5.2 Coarse-to-Fine Iris Identification Experiments

Iris recognition has been well studied in the literature and a number of iris feature descriptors [1], [2], [43] have been proposed for characterizing the most discriminative and robust features in iris texture. It has been demonstrated that the feature representation models in state-of-the-art iris recognition methods can be unified into a general framework of ordinal measures [2]. So the coarse-to-fine iris identification experiments mainly aim to verify performance improvements of ordinal measures based iris recognition methods after introducing the proposed HVC iris image classification method. The baseline iris recognition method in the experiments uses a tri-lobe ordinal filter (5*5 Gaussian kernel as the basic lobe, horizontal orientation and inter-lobe distance is 10 pixels) to extract 128 bytes of iris feature codes. The CASIA-Iris-Thousand [39], one of the largest iris image database in the public domain, is used as the testing database since it contains iris images of 1,000 subjects. To model iris recognition systems in the field, randomly selected nine iris images of each eye are used to construct the registration database, and the left ones are query images. The ordinal measures [2] and HVC features are extracted, and iris image clusters and SVM classifier for coarse classification are learned from the registration database. During identification, for a query image, if the coarse classification result has high reliability, the image is just compared to images in corresponding category based on the ordinal measure features for identification; otherwise, the query image is compared to the whole registration database. In our experiment, the reliability (note as R) of the coarse classification is calculated by comparing the largest and second largest decision values (note as d_1 and d_2) of SVM results, i.e. $R = (d_1 - d_2)/d_1$. If $R > 0.5$, the classification is regarded as a successful one, where the threshold 0.5 is decided according to experiments and it is adjustable for different applications.

Iris images are classified into 10 categories to show our strategy. The iris identification with and without coarse classification are running with Matlab 2011 on a same computer. The ROC curves of coarse-to-fine iris identification and iris identification without classification are compared in Fig. 8. Results of the coarse-to-fine iris identification show higher accuracy than the identification without coarse classification. One iris image identification without coarse classification needs about 30 seconds, while the coarse-to-fine iris identification only needs about 3 seconds in speed testing. In our experiments, about 80% iris images are classified meeting the condition $R > 0.5$, and the other 20% ones need to be matched with all the registered images. Therefore, the coarse-to-fine iris identification (10 categories) may save over 70% computational cost. With increasing of the size of the iris image database, the number of categories should be increased accordingly, and the efficiency benefit is more significant. In summary, the experimental results demonstrate that the proposed coarse-to-fine iris identification based on HVC significantly improves both effectiveness and efficiency of iris recognition systems.

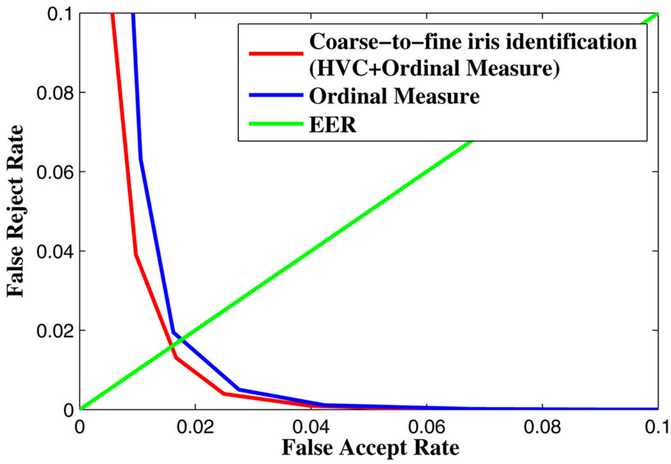


Fig. 8. ROC (DET) curves of iris identification with or without coarse classification.

6 Conclusion

Iris image classification aims to group iris images into multiple categories according to their application related attributes (e.g. liveness, ethnicity, texture category) rather than their identity information. Because iris image classification is a significantly different problem to iris image recognition in terms of definition, challenges, core problems, and applications, specific iris image analysis and pattern classification approaches are needed for iris image classification. Iris image classification is an important research topic but it is not well addressed in the literature. Moreover, the existing research efforts are separated in specific application problems of iris image classification. In this paper, iris image classification is firstly formulated as a generic problem in iris biometrics. Such a formulation is beneficial to unify the research efforts in iris liveness detection, race classification, coarse iris image classification for efficient identification, etc. Moreover, it is possible to develop a generic iris image classification module in an iris recognition system for a variety of applications. So that the computational cost of feature extraction and matching for multiple iris image classification tasks can be greatly reduced.

Our previous work [10], [11], [12], [16], [17] demonstrated the effectiveness of statistical texture analysis for iris image classification. The core idea of iris image classification is to find common texture primitives of different subjects in the same category. This paper aims to learn and encode the most effective texture primitives of iris images for classification. To integrate the advantages of both Vocabulary Tree and Locality-constrained Linear Coding, a novel iris feature representation method called Hierarchical Visual Codebook (HVC) is proposed to encode the distinctive and robust texture primitives of iris images. Each code characterizes a kind of frequently appearing local patches in iris images. The codebook of HVC is organized in a tree structure and the relationship between different visual codes is preserved. The global texture of an iris image is well characterized by the statistical distribution of codes in the feature space with overlapping relationship. HVC is an accurate texture analysis approach which minimizes visual coding errors using a hierarchical quantization strategy. It is an efficient solution suited to represent large-scale and discriminant visual vocabulary of iris texture. Extensive experiments illustrate the effectiveness of this generic HVC method for typical applications of iris image classification. The success of race classification demonstrates the genotypic relationship between iris images of different subjects. Such an evidence may be used for social and forensic applications.

We have attempted to provide a systematic study on iris image classification. More efforts are clearly required for this topic.

1) It is interesting to establish a large-scale multi-race iris image database in public domain. And all researchers around the world are invited to update the iris image samples of different races. So that we can investigate the performance of racial iris image classification with the increasing of the number of races in database. And it is interesting to investigate the genotypic relationship between iris images from different races through the proposed hierarchical visual codebook.

2) Iris recognition is being deployed in many important applications such as national ID card, banking, social benefit, border control, etc. The risk of security attacks to iris recognition systems increases accordingly driven by the great benefit of fraudulent identity authentication. It is predictable that attackers will pay more efforts to develop advanced methods to spoof iris biometrics. So it will become more challenging to develop a reliable security solution to iris recognition with the advancement of iris attack approaches. In this sense, the research of iris liveness detection will never stop since the attack approaches are dynamically updated. It is also a good idea to establish a large database of fake iris patterns in public domain just like the computer/Internet virus sample database in anti-virus software industry.

3) Advanced solutions for coarse-to-fine iris identification are expected. Moreover, HVC is a generic texture analysis method so it is applicable to object recognition, texture classification and other texture like biometric traits.

ACKNOWLEDGMENTS

This work is funded by the National Basic Research Program of China (Grant No. 2012CB316300), the National Natural Science Foundation of China (Grant No. 61273272, 61075024, 61135002) and the International S&T Cooperation Program of China (Grant No. 2010DFB14110). Hui Zhang is the corresponding author of this paper.

REFERENCES

[1] J. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1160, Nov. 1993.
[2] Z. Sun and T. Tan, “Ordinal measures for iris recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2211–2226, Dec. 2009.
[3] L. Ma, T. Tan, Y. Wang, and D. Zhang, “Personal identification based on iris texture analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1519–1533, Dec. 2003.
[4] E. Lee, K. Park, and J. Kim, “Fake iris detection by using purkinje image,” in *Proc. ICB*, Hong Kong, China, 2006, pp. 397–403.
[5] S. Lee, K. Park, and J. Kim, “Robust fake iris detection based on variation of the reflectance ratio between the iris and the sclera,” in *Proc. Biometric Consortium Conf.*, Baltimore, MD, USA, 2006, pp. 1–6.
[6] J. Daugman, “How iris recognition works,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.
[7] X. He, Y. Lu, and P. Shi, “A fake iris detection method based on FFT and quality assessment,” in *Proc. Chinese Conf. Pattern Recognition*, Beijing, China, 2008, pp. 316–319.
[8] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia, “Iris liveness detection based on quality related features,” in *Proc. ICB*, New Delhi, India, 2012, pp. 271–276.
[9] X. He, S. An, and P. Shi, “Statistical texture analysis based approach for fake iris detection using support vector machine,” in *Proc. ICB*, Seoul, Korea, 2007, pp. 540–546.
[10] Z. Wei, X. Qiu, Z. Sun, and T. Tan, “Counterfeit iris detection based on texture analysis,” in *Proc. ICPR*, Tampa, FL, USA, 2008, pp. 1–4.
[11] Z. He, Z. Sun, T. Tan, and Z. Wei, “Efficient iris spoof detection via boosted local binary patterns,” in *Proc. ICB*, Alghero, Italy, 2009, pp. 1080–1090.
[12] H. Zhang, Z. Sun, and T. Tan, “Contact lens detection based on weighted LBP,” in *Proc. ICPR*, Istanbul, Turkey, 2010, pp. 4279–4282.
[13] S. Gutta, H. Wechsler, and P. Phillips, “Gender and ethnic classification of face images,” in *Proc. FG*, Nara, Japan, 1998, pp. 194–199.
[14] G. Shakhnarovich, P. Viola, and B. Moghaddam, “A unified learning framework for real time face detection and classification,” in *Proc. FG*, Washington, DC, USA, 2002, pp. 14–21.
[15] X. Lu and A. Jain, “Ethnicity identification from face images,” in *Proc. SPIE Defense and Security Symp.*, vol. 5404. 2004, pp. 114–123.
[16] X. Qiu, Z. Sun, and T. Tan, “Global texture analysis of iris images for ethnic classification,” in *Proc. ICB*, Hong Kong, China, 2006, pp. 411–418.
[17] X. Qiu, Z. Sun, and T. Tan, “Learning appearance primitives of iris images for ethnic classification,” in *Proc. ICIP*, vol. 2. San Antonio, TX, USA, 2007, pp. 405–408.
[18] H. Zhang, Z. Sun, T. Tan, and J. Wang, “Ethnic classification based on iris images,” in *Proc. Chinese Conf. Biometric Recognition*, Beijing, China, 2011, pp. 82–90.
[19] J. Lyle, P. Miller, S. Pundlik, and D. Woodard, “Soft biometric classification using periocular region features,” in *Proc. BTAS*, Washington, DC, USA, 2010, pp. 1–7.
[20] L. Yu, D. Zhang, K. Wang, and W. Yang, “Coarse iris classification using box-counting to estimate fractal dimensions,” *Pattern Recognit.*, vol. 38, no. 11, pp. 1791–1798, 2005.
[21] J. Fu, J. Caulfield, S. Yoo, and V. Atluri, “Use of artificial color filtering to improve iris recognition and searching,” *Pattern Recognit. Lett.*, vol. 26, no. 14, pp. 2244–2251, 2005.
[22] X. Qiu, Z. Sun, and T. Tan, “Coarse iris classification by learned visual dictionary,” in *Proc. ICB*, Seoul, Korea, 2007, pp. 770–779.
[23] H. Mehrotra, B. Srinivas, B. Majhi, and P. Gupta, “Indexing iris biometric database using energy histogram of DCT subbands,” in *Contemporary Computing*. Berlin, Germany: Springer, pp. 194–204, 2009.

[24] M. Sunder and A. Ross, "Iris image retrieval based on macro-features," in *Proc. ICPR*, Istanbul, Turkey, 2010, pp. 1318–1321.

[25] J. Doyle, P. Flynn, and K. Bowyer, "Automated classification of contact lens type in iris images," in *Proc. ICB*, Madrid, Spain, 2013, pp. 1–6.

[26] Z. He, T. Tan, Z. Sun, and X. Qiu, "Towards accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, Sept. 2009.

[27] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. Freeman, "SIFT flow: Dense correspondence across different scenes," in *Proc. ECCV*, Marseille, France, 2008, pp. 28–42.

[28] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. CVPR*, vol. 2. Washington, DC, USA, 2006, pp. 2161–2168.

[29] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 3360–3367.

[30] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. CVPR*, vol. 2. 2006, pp. 2169–2178.

[31] University of Notre Dame Computer Vision Research Lab. *ND Iris Image Databases* [Online]. Available: http://www3.nd.edu/ cvrl/CVRL/Data_Sets.html

[32] Institute of Automation, Chinese Academy of Sciences. *CASIA-Iris-Fake* [Online]. Available: http://www.cripac.ia.ac.cn/people/znsun/irisclassification/CASIA-Iris-Fake.rar

[33] *IrisGuard H-100* [Online]. Available: http://www.irisguard.com

[34] M. Dobes and L. Machala, *UPOL Iris Database* [Online]. Available: http://www.inf.upol.cz/iris/

[35] Z. Wei, T. Tan, and Z. Sun, "Synthesis of large realistic iris databases using patch-based sampling," in *Proc. ICPR*, Tampa, FL, USA, 2008, pp. 1–4.

[36] Wikipedia, *Race Human Classification* [Online]. Available: http://en.wikipedia.org/wiki/Race_(human_classification)

[37] Institute of Automation, Chinese Academy of Sciences. *CASIA-Iris-Race* [Online]. Available: http://www.cripac.ia.ac.cn/people/znsun/irisclassification/CASIA-Iris-Race.rar

[38] *OKI Irispass-h Handheld Device* [Online]. Available: http://www.oki.com/iris

[39] Institute of Automation, Chinese Academy of Sciences. *CASIA Iris Image Database v4.0* [Online]. Available: http://biometrics.idealtest.org/

[40] S. Schuckers *et al.*, "Quality - Face / Iris Research Ensemble (Q-FIRE) dataset overview," Clarkson Univ., Dept. Electr. Comput. Eng., Tech. Rep. SOP-2009-1001-11, 2010.

[41] *The Benchmark of Racial Iris Classification* [Online]. Available: http://www.cripac.ia.ac.cn/people/znsun/irisclassification/imageset.pdf

[42] R. Fan, K. Chang, C. Hsieh, X. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Aug. 2008, [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/liblinear

[43] C. Tan and A. Kumar, "Towards online iris and periocular recognition under relaxed imaging constraints," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3751–3765, Oct. 2013.

Zhenan Sun is currently a Professor with the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He received the B.E. degree in industrial automation from Dalian University of Technology, Dalian, China, the M.S. degree in system engineering from Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree in pattern recognition and intelligent systems from CASIA in 1999, 2002, and 2006, respectively. Since March 2006, he has joined the NLPR of CASIA as a Faculty Member. He is a member of the IEEE Computer Society, and the IEEE Signal Processing Society. His current research interests include biometrics, pattern recognition, and computer vision, and he has authored/co-authored over 100 technical papers. He is an Associated Editor of *IEEE Transactions on Information Forensics and Security* and IEEE BIOMETRICS COMPENDIUM.

Hui Zhang is currently an Assistant Professor with the Institute of Software, Chinese Academy of Sciences, Beijing, China. She received the B.E. degree from the University of Science and Technology of China, in 2007, and the Ph.D. degree from the Shanghai Institute of Technical Physics of the Chinese Academy of Sciences, in 2012. Her current research interests include biometrics, remote sensing image processing, and pattern recognition.

Tieniu Tan is currently a Professor with the Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), China. He received the B.Sc. degree in electronic engineering from Xi'an Jiaotong University, China, in 1984, and the M.Sc. and Ph.D. degrees in electronic engineering from Imperial College of Science, Technology and Medicine, London, U.K., in 1986 and 1989, respectively. His current research interests include biometrics, image and video understanding, information hiding, and information forensics. He is a Fellow of the International Association of Pattern Recognition.

Jianyu Wang is currently a Research Professor at the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China. He received the B.S. degree in physics from Hangzhou University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, in 1987 and 1990. He is the Vice Chairman of Chinese Society of Space Research, and the Vice Chairman of Remote Sensing of the Environment Branch of the Geographical Society of China. His current research interests include passive and active optoelectronic remote sensing system, Hyperspectral imaging technology, laser remote sensing imaging technology, free-space quantum communication technology.