Robust classification for occluded ear via Gabor scale feature-based non-negative sparse representation

Baoqing Zhang
Zhichun Mu
Chen Li
Hui Zeng
Robust classification for occluded ear via Gabor scale feature-based non-negative sparse representation

Baoqing Zhang
Zhichun Mu
University of Science and Technology Beijing
School of Automation and Electrical Engineering
Beijing 100083, China
E-mail: mu@ies.ustb.edu.cn

Chen Li
North China University of Technology
College of Information Engineering
Beijing 100144, China

Hui Zeng
University of Science and Technology Beijing
School of Automation and Electrical Engineering
Beijing 100083, China

Abstract. The Gabor wavelets have been experimentally verified to be a good approximation to the response of cortical neurons. A new feature extraction approach is investigated for ear recognition by using scale information of Gabor wavelets. The proposed Gabor scale feature conforms to human visual perception of objects from far to near. It can not only avoid too much redundancy in Gabor features but also tends to extract more precise structural information that is robust to image variations. Then, Gabor scale feature-based non-negative sparse representation classification (G-NSRC) is proposed for ear recognition under occlusion. Compared with SRC in which the sparse coding coefficients can be negative, the non-negativity of G-NSRC conforms to the intuitive notion of combing parts to form a whole and therefore is more consistent with the biological modeling of visual data. Additionally, the use of Gabor scale features increases the discriminative power of G-NSRC. Finally, the proposed classification paradigm is applied to occluded ear recognition. Experimental results demonstrate the effectiveness of our proposed algorithm. Especially when the ear is occluded, the proposed algorithm exhibits great robustness and achieves state-of-the-art recognition performance. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.OE.53.6.061702]

Subject terms: Gabor scale feature; non-negative sparse representation; ear recognition; partial occlusion.

Paper 130665SS received May 7, 2013; revised manuscript received Jul. 23, 2013; accepted for publication Aug. 7, 2013; published online Dec. 2, 2013.

1 Introduction
As the technology to identify an individual based on his or her physical or behavioral attributes, biometrics has been a hotspot of pattern recognition and modern security technique development in the past few years. Highly accepted in commercial, civilian, and forensic applications, it uses biometric features such as fingerprints, iris, face, and voice for recognition.

Face, fingerprint, and iris have been recognized as the three most popular biometrics modalities for personal identification. Nevertheless, we are convinced that each biometric has its strengths and weaknesses and no single biometric can be expected to meet all the requirements imposed by all applications. 1 Therefore, great research efforts are further needed to explore other potential biometric modalities that can be conveniently acquired at a low cost. The ear is one such promising modality since the ear has a rich and stable structure that does not change significantly as an individual grows. 2 The geometry and shape of the ear have been observed to have significant variations among individuals. 3 Additionally, the acquisition of ear images does not necessarily require a person’s co-operation and is also considered to be nonintrusive by most people in comparison with other biometrics modalities such as fingerprints or iris. Because of these qualities, the interest in ear recognition systems has grown significantly in recent years. 4,5

Even though current ear recognition systems have reached a certain level of maturity, their success is limited to controlled conditions, 6 in which additional restrictions are imposed on the subjects and on the imaging conditions, e.g., the subject needs to be with small pose changes, the ear images to be without occlusion, and the lighting to be constant. However, one challenging problem inevitable in real application scenarios is ear occlusion because ears are often occluded by some objects such as hair or hat. 7 Actually, occlusion poses a significant obstacle to robust real-world ear recognition. In most ear recognition-related research, 1,8 ear recognition under partial occlusion is addressed as an open challenging problem. Hence, an effective and robust algorithm is essentially needed for ear recognition with various levels of occlusion at any location.

2 Related Prior Work
Because a three-dimensional (3-D) representation of a subject can be adapted to any rotation, scale, and translation, using 3-D models for recognition is a promising solution to enhance the accuracy of ear recognition systems under pose or illumination variations. 9 However, most of the current 3-D ear recognition systems tend to be computationally expensive. And to acquire accurate scan results, more restrictions are imposed on the subject, which are definitely not suitable for nonintrusive human identification in real-world applications. Therefore, the focus of our work has been to exploit the two-dimensional (2-D) ear images that can be conveniently acquired from a low-cost digital camera.

A variety of approaches have been explored to extract discriminant features from the 2-D ear images for personal identification. According to the feature extraction approaches, ear recognition methods can be divided into holistic feature-based and local feature-based approaches. The holistic methods extract statistical features from the 2-D image.
Chang et al. used standard principal component analysis (PCA) for face and ear recognition and concluded that ear and face did not have much difference on the recognition performance. Hurley et al. applied force field transform for ear recognition, which assumes that pixels have a mutual attraction proportional to their intensities and inversely to the square of the distance between them. The local feature-based methods use local information rather than the holistic information for recognition, and therefore are more robust to small pose changes and illumination variations. Kumar and Wu proposed an automatic ear recognition approach using the phase information of log-Gabor filters to encode the local structure of the ear. The rank-1 performance of this method ranges from 92.60% to 95.93% on the public database containing 221 subjects. Other local feature-based methods include local binary pattern (LBP), scale-invariant feature transform (SIFT), Gabor wavelets, and non-negative matrix factorization.

### 2.1 Latest Research on Occluded Ear

Even though a large amount of work has been devoted to ear recognition for application in real scenarios, only limited work has focused on ear recognition under occlusion.

Most researchers proposed to deal with ear recognition under partial occlusion using local feature- or model-based methods. Yuan and Mu proposed a 2-D ear recognition approach based on local information fusion to deal with ear recognition under partial occlusion. In that work, neighborhood preserving embedding was used for feature extraction. Experimental results on the USTB ear dataset and UND dataset illustrated that the most meaningful region of the ear could be represented by using only a few subwindows and the multiclassifier model got higher recognition rate than using the whole image for recognition. Nevertheless, methods of this kind involve partition of the ear and classifier fusion. Consequently, partition strategy and fusion strategy will have a great impact on the recognition performance.

Bustard and Nixon proposed to treat the ear as a planar surface and then create a homography transform using SIFT feature matches for ear registration and recognition. On XM2VTS ear dataset containing 63 subjects, they got rank-1 recognition rate of 92% for 30% occlusion from above and rank-1 recognition rate of 92% for 30% occlusion from left side. Arbab-Zavar and Nixon proposed a model-based approach for ear recognition. The model was a part-wise description of the ear derived by a stochastic clustering on a set of scale-invariant features of a training set. The ear’s boundary structures were further described by log-Gabor filters. On the same set of 63 subjects from XM2VTS ear dataset, they got a recognition rate of 89.4% for 30% occlusion from above or the left side; other occlusion locations should be further discussed because the ears may be occluded at any locations in real application scenarios. Hence, great efforts are essentially needed to improve ear recognition technologies for wide deployment in surveillance and other commercial applications.

### 2.2 Our Work

The proposed G-NSRC paradigm treats the Gabor scale features as the dictionary of our G-NSRC model for ear recognition, and then, the test ear is represented as a linear additive (without subtraction) combination of the Gabor scale features extracted from all the training ear samples. Compared with SRC in which the sparse coding coefficients can be negative, allowing the data to “cancel each other out” by subtraction, our G-NSRC conforms to the intuitive notion of combing parts to form a whole and therefore is more consistent with the biological modeling of visual data. Various levels of occlusion at any locations are evaluated more thoroughly to demonstrate the effectiveness of the proposed algorithm. Extensive experimental results show that the proposed classification paradigm achieves better recognition performance and is found to be more robust to large-scale ear occlusion.
3 Gabor Scale Feature Extraction

3.1 Gabor Filters

The characteristics of the Gabor wavelets, especially for frequency and orientation representations, are similar to those of the human visual system. The 2-D Gabor wavelets can be defined as follows:

\[ \psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma} e^{-\|k_{u,v}\|^2/\sigma} e^{i k_{u,v} \cdot z}, \]

where \( z = (x, y) \) denotes the pixels of an image, \( u \) defines the orientations of Gabor filters, \( v \) defines the scales of Gabor filters, and \( \| \cdot \| \) denotes the norm operation. \( k_{u,v} \) is defined as \( k_{u,v} = k_v e^{i \phi_v} \), where \( k_v = k_{\text{max}}/f^v \) and \( \phi_v = \pi u/4 \). \( k_{\text{max}} \) is the maximum frequency and \( f \) is the spacing factor between kernels in the frequency domain.

The Gabor representation of an image \( I(z) = I(x, y) \), called the Gabor image, is the convolution of the image with the Gabor kernels

\[ G(z, u, v) = I(z) \ast \psi_{u,v}(z). \]  

For each image pixel, we have two Gabor parts: the real part and the imaginary part. So, filtering coefficient \( G(z, u, v) \) is a complex number that can be rewritten as

\[ G(z, u, v) = M(z, u, v) \cdot \exp[i\theta(z, u, v)]. \]

\( M(z, u, v) \) is the magnitude and \( \theta(z, u, v) \) is the phase. In most Gabor-based feature extraction research, the Gabor feature is defined via concatenation of the magnitude coefficients. This feature extraction technique generates redundant features resulting in extremely high dimensionality of the Gabor features since Gabor filters with multiple scales and orientations are adopted. Consequently, extracting Gabor features is computationally intensive, making the features impractical for real-time applications. So, dimensionality reduction approaches such as PCA and LDA (Refs. 27 and 28) are commonly used to reduce the Gabor feature dimensionality. However, this procedure tends to destroy the visual perception characteristic hidden in the Gabor feature.

3.2 Gabor Scale Feature

The magnitude information of Gabor transform reflects the variations of local energy in the image, so it can tolerate image deformation to some extent. In this article, we propose to extract Gabor scale feature contained in the magnitude information of Gabor transform. First, the input image is convoluted with the Gabor kernel functions, and thus magnitude information across different orientations and different scales is obtained. Then, for a certain scale \( v \), magnitude information of all the orientations at this scale is cumulated to formulate the scale feature. In comparison with conventional Gabor features, the Gabor scale feature can not only reduce the feature dimensionality by a factor of the Gabor orientation parameter \( v \), but also reduce the redundancy between the orientation information across different scales. What is more, as the Gabor transform strongly responds to edges, Gabor scale feature tends to extract structural information of the ear in different scales, which is more important for ear identification as ear key feature points and textural information is not obvious in comparison with face. In addition, the Gabor scale feature conforms to human visual perception of objects from different scales, i.e., from far to near. The whole process of the Gabor scale feature extraction of ear images is presented in Algorithm 1 below.

In this article, we use Gabor kernels with three different scales, \( v \in \{0, 1, 2\} \) and four different orientations, \( u \in \{0, 1, 2, 3\} \). We set parameters \( f = \sqrt{2} \), \( k_{\text{max}} = \pi/2 \), \( \sigma = 2\pi \). Figure 1 illustrates the process of Algorithm 1 to obtain the Gabor scale features of an input ear image. From Fig. 1, we see that the Gabor scale feature tends to extract structural information compared with conventional Gabor features and conforms to human visual perception of objects from different scales, i.e., from far to near.

4 Classification Based on Non-negative Sparse Representation of Gabor Scale Feature

This section describes the framework of our proposed G-NSRC for ear recognition.

4.1 Non-negative Sparse Representation

As a data adaptive representation, sparse representation has been applied in different research areas. And it has been strongly supported by studies of human vision. Olshausen and Field research has indicated that neural networks in the visual system could be performing sparse coding of learned features that are qualitatively similar to the receptive fields of simple cells in V1. Sparse representation has exhibited great promise in pattern recognition areas.

Using an over-complete dictionary matrix \( D \in \mathbb{R}^{m \times n} \) that contains \( n \) prototype signal atoms for columns \( \{d_i\}_{i=1}^n \), a signal \( y \) can be represented as a sparse linear combination of these atoms:

\[ y = Dx, \quad y \in \mathbb{R}^m. \]

Obviously, if \( D \) is a full rank matrix and \( m < n \), the above formulation is an underdetermined linear system. Thus, this system has infinitely many solutions. The solution with the fewest number of nonzero coefficients is certainly an appealing one. Then, the sparse representation can be formulated as the following optimization problem:

\[ x_0 = \arg \min \|x\|_0 \quad \text{s.t.} \quad y = Dx. \]

The \( \ell^0 \) norm is defined as \( \|x\|_0 = \{i: x_i \neq 0\} \), i.e., the number of nonzero entries in vector \( x \).

**Algorithm 1** Gabor scale feature extraction of ear images.

**Input:** an ear image \( I(z) = I(x, y) \), Gabor kernel functions \( \psi_{u,v}(z) \), for \( v \) \( u \).

**Output:** Gabor scale features at the different scales, \( \text{Gaborscale}(z, v) = \sum_v M(z, u, v). \)

---

**Notes:**

1. Zhang et al.: Robust classification for occluded ear via Gabor scale feature.
2. Optical Engineering 061702-3 June 2014/Vol. 53(6)
3. Zhang et al.: Robust classification for occluded ear via Gabor scale feature...
In standard sparse representation described above, the data are represented as a combination of elementary features involving both additive and subtractive interactions. As the coefficients can be negative, the data are allowed to “cancel each other out” by subtraction, which is contrary to the intuitive notion of combing parts to form a whole. In fact, non-negativity is more consistent with the biological modeling of visual data and often leads to better performance for data representation. In the research work of Lee and Seung, the authors forcefully argued for non-negative representation. Other arguments for non-negative representation come from the biological modeling, where such constraints are related to the non-negativity of neural firing rates.

Non-negative sparse representation distinguishes itself in that it enforces non-negative constraints on the dictionary and sparse coding coefficients, i.e., all the elements must be equal to or greater than 0. This non-negativity constraint leads sparse representation to a parts based representation of the object in the sense that it allows only additive, not subtractive, combination of the original data. Therefore, it is an ideal representation of data in the pattern recognition, where it is natural to consider the data as a combination of parts to form a whole. The non-negative sparse representation model can be formulated as

\[ x_0 = \arg \min \|x\|_0 \quad \text{s.t.} \quad y = Dx \]  

(6)

with additional constraints \( \forall i, j: D_{ij} \geq 0, x_i \geq 0 \) which is different from the sparse representation model.

4.2 Gabor Scale Feature-Based Non-negative Sparse Representation Classification

The design of appropriate dictionaries plays a key role in the sparse representation-based classification algorithm. As a local feature description method, our proposed Gabor scale feature can not only enhance the ear structural features but also tolerate image variations and local deformation to some extent. Additionally, the proposed feature conforms to human visual perception of objects from different scales. So, in this article, we propose to use Gabor scale features extracted from all the training ear images on each image dataset as the dictionary of our G-NSRC model.

Assuming that there exist \( n \) training ear images from \( k \) distinct classes with \( n_i \) images belonging to the \( i \)’th class, we first extracted Gabor scale feature from every training ear sample identified with the vector \( g_{ij} \in \mathbb{R}^m \), called atoms of our G-NSRC. Here, \( i \) denotes the index of the class, \( i = 1, 2, \ldots, k \), and \( j \) denotes the index of the training sample, \( j = 1, 2, \ldots, n_i \). All the atoms from the \( i \)’th class are arranged as columns of a matrix \( G_i = [g_{i1}, g_{i2}, \ldots, g_{in_i}] \).

Then, the dictionary of our G-NSRC can be defined as the concatenation of the \( n \) atoms from all \( k \) classes:

\[ G = [G_1, G_2, \ldots, G_k] = [g_{11}, g_{12}, \ldots, g_{kn_k}] \]  

(7)

Thus the linear representation of \( y_G \), denoting the Gabor scale feature of the test ear \( y \) from the \( i \)’th class, can be modeled in terms of all the atoms in the dictionary as: \( y_G = Gx \), where \( x = [0, \ldots, 0, a_{i1}, a_{i2}, \ldots, a_{iin_i}, 0, \ldots, 0]^T \in \mathbb{R}^m \) is a coefficient vector whose entries are 0 except for those from the same class as the test ear signal \( y \). Vector \( x \) is called the sparse coding coefficient.
In SRC, the test data is represented as a combination of elementary features involving both additive and subtractive interactions. Here, we propose to express the test ear sample as a linear additive (with nonsubtraction) combination of all the atoms. Our proposed G-NSRC treats Gabor scale feature as the atoms of the dictionary $G$. According to the feature extraction process presented in Sec. 3.2, the elements in the dictionary $G$ are all non-negative, so additional constraints only need to be enforced on the sparse coefficient $x_i$, i.e., $x_i \geq 0$, for $i = 1, 2, \ldots, m$. So, the proposed G-NSRC model for ear recognition is given below:

$$x_0 = \arg \min \|x\|_0 \quad \text{s.t.} \quad y_G = Gx_0 \geq 0. \tag{8}$$

Similar to SRC, after the non-negative sparse coding coefficient $x$ is computed by the algorithm given in Sec. 4.3 below, classification is then made by

$$\text{identity}(y) = \arg \min_r r_i(y_G) = \|y_G - e_0 - G\delta_i(x_0)\|_2, \tag{9}$$

where $\delta_i(x) \in \Re^p$ is a new vector whose only non-negative nonzero entries are those that are associated with class $i$.

When the test ear image is occluded, the occluded test ear sample $y_1$ is rewritten as

$$y_G = y_1 + e_0 = Gx_1 + e_0 = [G; I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = \hat{G}w_0w_0^T \geq 0,$$

for $i = 1, 2, \ldots, m, \ldots, m + \text{size}(I). \tag{10}$

Then, classification strategy in Eq. (9) should be modified as

$$\text{identity}(y_1) = \arg \min_r r_i(y_1) = \|y_G - e_0 - G\delta_i(x_0)\|_2. \tag{11}$$

The solution of our G-NSRC model-based ear recognition algorithm is presented in Sec. 4.3:

**Algorithm 2  Gabor scale feature-based non-negative sparse representation classification.**

1. **Input:** a dictionary matrix consisting of Gabor scale features extracted from all the training samples $G = [G_1, G_2, \ldots, G_k] = [g_{i1}, g_{i2}, \ldots, g_{ik}]$ for $k$ classes; the Gabor scale feature of the test ear sample $y_0$.

2. Normalize the columns of $G$ to have unit $l^p$ norm.

3. Solve the quadratic programming problem

$$Q_0(x) = \frac{1}{2}\|y_G - Gx\|_2^2 + \lambda\sum_i x_i,$$

using update rule described in Sec. 4.3, $\hat{x} = \arg \min_x Q_0(x)$.

4. Compute the residuals $r_i(y) = \|y_G - G\delta_i(\hat{x})\|_2$. And then, for a test ear sample with occlusion $y_1$, identity($y_1$) = $\arg \min_r r_i(y_1) = \|y_1 + e_0 - G\delta_i(x_0)\|_2$, where $\delta_i(x_0)$ is a vector whose only non-negative nonzero entries are those in $x_0$ that are associated with the $i$th class.

5. **output:** identity($y$).

### 4.3 Non-negative Sparse Solution of G-NSRC

Although in many applications, greedy strategies such as orthogonal matching pursuit (OMP) show better performance than convex optimization, it is suggested in Ref. 35 that $l^1$-minimization leads to more stable active sets, and it is preferred for the classification tasks. Here, we propose to approximate the sparse solution of Eq. (6) using $l^1$-minimization. By replacing $P^0$ with the $l^1$ norm, the G-NSRC model can be modeled as

$$x_1 = \arg \min_x \|x\|_1 \quad \text{s.t.} \quad y = Dx \tag{12}$$

with constraints $x_i \geq 0$, for $i = 1, 2, \ldots, m$. $D$ denotes the dictionary $G$ of the proposed G-NSRC, and $y$ denotes the Gabor scale feature of the test ear sample $y_G$. The transformation of $P^0$ norm to $l^1$ norm makes the optimization problem a convex optimization problem. This can be rewritten as a standard problem in linear optimization under quadratic and linear inequality constraints.

Thus, for an appropriate Lagrange multiplier $\lambda$ which controls the tradeoff between sparseness and accurate reconstruction, the solution to the problem (12) is precisely the solution to the unconstrained optimization problem

$$O_F(D, x) = \frac{1}{2}\|y - Dx\|_2^2 + \lambda\sum_i x_i. \tag{13}$$

As the atoms of $D$ are defined as the Gabor scale feature, therefore $D$ is known. And then, Eq. (13) is quadratic with respect to $x$. The global minimum can be found using optimization methods such as quadratic programming and gradient descent. In this article, an efficient algorithm that is extremely simple to implement is proposed below.

Let $M = [1, 1, \ldots, 1]$, the objective function $O_F$ can be rewritten as

$$O_F(x) = \frac{1}{2}\text{tr}[(y - Dx)(y - Dx)^T] + \lambda\sum_i x_i$$

$$= \frac{1}{2}\text{tr}(yy^T) - \text{tr}(yx^TD^T) + \frac{1}{2}\text{tr}(Dxx^TD) + \lambda\text{tr}(xM). \tag{14}$$

Let $\alpha_i$ be the Lagrange multiplier for constraints $x_i \geq 0$, and $\alpha = [\alpha_i]$. The Lagrange function $l_F$ is

$$l_F = \frac{1}{2}\text{tr}(yy^T) - \text{tr}(yx^TD^T) + \frac{1}{2}\text{tr}(Dxx^TD) + \lambda\text{tr}(xM) + \text{tr}(\alpha x). \tag{15}$$

The partial derivations of $l_F$ with respect to $x$ is

$$\frac{\partial l_F}{\partial x} = -D^Ty + D^TDx + \lambda M + \alpha. \tag{15}$$

Using the Kuhn–Tucker condition $\alpha_i x_i = 0$, the following equation can be gotten:

$$(-D^Ty)x_i + (D^TDx)x_i + (\lambda M)x_i = 0. \tag{16}$$

So the following update rule can be gotten:

$$x_{i+1} = x_i(D^Ty)/(D^TDx_i + \lambda M). \tag{17}$$
The above update rule is the same as the one given in the work of Hoyer.34 As proved in Hoyer’s work, iteration of this update rule was guaranteed to reach the global minimum. What should be pointed out here is that although the proposed algorithm achieves a global solution to Eq. (13), this still may be a local solution to Eq. (6).

4.4 G-NSRC for Ear Recognition

The complete G-NSRC for ear recognition is summarized in Algorithm 2.

5 Experimental Results

In this section, we will investigate the use of our proposed G-NSRC for ear recognition. Extensive experiments are carried out on subset of UND database Collection J2 and USTB ear database to validate the claims of the previous sections.

5.1 Experiments to Validate the Effectiveness of Gabor Scale Feature

A subset of 100 subjects are selected from UND database Collection J2,37 which consists of 415 subjects. An improved Adaboost algorithm is used to detect and locate ear area automatically.38 Figure 2 shows the detection result. The red rectangle in Fig. 2(b) is the detected ear area.

There are a total of 603 ear images from 100 subjects used in this work. Some of the ear images undergo obvious pose changes or illumination variations and some are with occlusion. Typical cropped ear images from one subject are shown in Fig. 3.

We randomly select one test ear image per subject and the remaining ear images are used as training samples. What should be pointed out is that we do not carry out any pre-processing such as de-noising, illumination normalization, or pose normalization of the ear image. The LBP is a powerful illumination invariant texture descriptor39 and has been applied in different research areas. We compare our method with LBP and Gabor features via concatenation of the Gabor filtering coefficients. To reduce the dimensionality of the Gabor features, we choose PCA for dimensionality reduction. The recognition performance of the three methods from rank 1 to rank 5 is presented in Table 1.

The second experiment to validate the effectiveness of the proposed Gabor scale feature uses regular ear images from our USTB ear database III,40 which contains 79 subjects and is publicly available for academic research. In this ear database, the subject rotated his head from 0 to 45 deg toward the left side. Images of the head were acquired at the following angles: 0, 5, 10, 15, 20, 25, 30, 35, 40, and 45 deg. Two images were recorded at each angle, resulting in a total of 20 images per subject. Figure 4 presents a typical subject from this ear database.

The detection results using the improved Adaboost algorithm are presented in Fig. 5.

In the ear recognition experiment, for each subject, we randomly select five images for each person under different pose variations for testing and the remaining images for training. The cumulative match characteristic (CMC) curves of rank 50 from each of the recognition experiments are presented in Fig. 6.

From the experimental results illustrated in Table 1 and the CMC curves in Fig. 6, we can see that our proposed Gabor scale feature is more effective than the conventional Gabor feature and LBP on both UND and USTB ear database III. Hence, the proposed Gabor scale feature is effective for ear recognition.

5.2 The Proposed G-NSRC for Ear Recognition under Occlusion

Most of the ear databases available for academic research suffer from illumination variations, pose changes, and occlusion simultaneously. In our USTB ear database III described in Sec. 5.1, all images are photographed with color CCD camera under the white background and constant lighting. What is more, a total of 20 ear images are acquired for each subject, sufficient for the sparse representation-based classification recognition approaches. Because of these properties, this database is suitable for specialized research on ear...
recognition under occlusion, excluding other influencing factors such as large pose changes and illumination variations.

We first carry out our experiment on partially occluded ear image database from the USTB ear database III. This dataset consists of 144 ear images from 24 subjects. There are three different levels of occlusion, i.e., part occlusion (disturbance from some hair), trivial occlusion (little hair), and regular occlusion (natural occlusion) to simulate different levels of occlusions by hair in real application scenarios as shown in Fig. 7. From left to right, the percent occluded by hair is 0%, 15%, 25%, and 35%. In the recognition experiment, the five normal images without occlusion are used for training and the occluded images are used for testing. In this experiment, Gabor scale feature extracted from all the training ear images belonging to this ear database are treated as the dictionary of the proposed G-NSRC.

Our G-NSRC approach uses the whole image for ear recognition under occlusion, without parting the whole image into subblock and fusing the results of multiclassifiers. Recognition performance comparison of the algorithms proposed in Refs. 7 and 41 with our G-NSRC is presented in Table 2.

However, partially occluded ear image database from the USTB ear database III comprises a small number of subjects, only 24 subjects, and it only addressed ear occlusion from above. Other occlusion locations should be further discussed since the ears may be occluded at any locations in real scenarios. So, the second experiment is carried out on regular ear images from our USTB ear database III as described in Sec. 5.1.

Inspired by Ref. 22, we randomly occlude the test ear image with 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%. Recognition performance comparison of the algorithms proposed in Refs. 7 and 41 with our G-NSRC is presented in Table 2.
and 50\% by replacing a block of each test ear image with an unrelated image as shown in Fig. 8 to evaluate our proposed G-NSRC for ear recognition under various levels of occlusion at any location. The location of occlusion is randomly chosen for each image and is unknown to the computer, which is rational in real ear recognition applications. In our work, various levels of occlusion at any locations are evaluated more thoroughly to demonstrate the effectiveness of the proposed algorithm for ear recognition under occlusion compared with other occluded ear recognition research.

The whole process of our G-NSRC for an occluded test ear image is illustrated in Figs. 9 and 10. The feature extraction process of the proposed G-NSRC is illustrated in Fig. 9. Figure 9(a) shows a test ear image with 25\% randomly located occlusion from the first class of USTB ear database III. Figure 9(b) shows the Gabor features of the occluded test ear image. Figures 9(c), 9(d), and 9(e) present the Gabor scale features at different scales of the test ear image. Figure 10 illustrates the classification process. Figures 10(a) and 10(b) plot the non-negative sparse representation coefficients and representation residual using G-NSRC for recognition. Figures 10(c) and 10(d) plot the sparse representation coefficients and representation residual using G-SRC for recognition. We see that the G-SNRC correctly classify the 25\% occluded test ear to the first class of the database.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>0%</th>
<th>15%</th>
<th>25%</th>
<th>35%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPE (Ref. 8)</td>
<td>94</td>
<td>91</td>
<td>87</td>
<td>80</td>
</tr>
<tr>
<td>INMF (Ref. 41)</td>
<td>88</td>
<td>83</td>
<td>67</td>
<td>53</td>
</tr>
<tr>
<td>G-NSRC</td>
<td>98</td>
<td>96</td>
<td>91</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 2  Recognition rate for occluded ear.

![Fig. 8](image8.png) Occluded test ear images from the USTB ear database III.

![Fig. 9](image9.png) An example of Gabor scale feature extraction for a test ear with random block occlusion: (a) a 25\% occluded test ear image from the first class of USTB ear database III. (b) Gabor features of the occluded test ear. (c) Gabor scale features at the first scale. (d) Gabor scale features at the second scale. (e) Gabor scale features at the third scale.

![Fig. 10](image10.png) Representation coefficient and residual of the 25\% occluded test ear image from class 1: (a) and (c) plot the coefficients of G-NSRC and G-SRC. (b) and (d) illustrate the representation residual associated with each class by G-NSRC and G-SRC, respectively.
while the occluded ear sample is wrongly classified using G-SRC. Although the representation coefficients are both sparse for G-SRC and G-NSRC, the main difference lies in that the representation coefficients of our proposed G-NSRC are all non-negative. It demonstrates non-negativity is more consistent with the biological modeling of visual data and our G-NSRC is more robust to image occlusion compared with G-SRC.

Random projection has been previously studied as a general dimensionality-reduction method for numerous clustering problems as well as for learning nonlinear manifolds. Similar to random-faces defined by Ref. 22, we define the linear projection generated by Gaussian random matrix as random-ears.

The recognition rates of rank 1 using G-NSRC, G-SRC, random-ears + SRC and PCA + NN (used as a baseline) are given in Fig. 11. It should be clarified here that for G-NSRC and G-SRC, we use the proposed Gabor scale features extracted from all the training ear images in this database as the dictionary.

From the results described in Fig. 11, we can see that our G-NSRC and G-SRC outperform the other two methods. With the increase of occlusion percent, the advantage of G-NSRC over other three approaches is getting higher. When the occlusion percent is equal to 25%, G-NSRC can still achieve a recognition rate of 92.7%. None of the three other methods can achieve a recognition rate of 35%. That is because, as a local feature, our proposed Gabor scale feature is more effective to encode ear structural information. Moreover, the non-negativity of the proposed G-NSRC is more consistent with the biological modeling of visual data. Therefore, when the ear suffers from occlusion, our proposed G-NSRC achieves much better recognition performance than the other three approaches.

Table 3 illustrates the recognition performance comparisons between our proposed G-NSRC and conventional Gabor feature-based sparse representation classification algorithm (Gabor + SRC and Gabor + NSRC). For Gabor + SRC and Gabor + NSRC, we use Gabor feature extracted from the training ear images in this database as the dictionary of SRC and NSRC.

As can be seen in Table 3, for the same Gabor feature, Gabor + NSRC outperforms Gabor + SRC greatly. It demonstrates that non-negativity leads to better performance for data representation. Obviously, the proposed G-NSRC achieves the best performance and shows great robustness to ear occlusion, especially when the occlusion percent surpasses 10%. That is because as a local feature, our proposed Gabor scale feature can not only avoid too much redundancy in conventional Gabor features but also tend to extract more precise structural information such as geometry and shape of the ear. The experimental results in Table 3 demonstrates that combining visual perception characteristic of Gabor wavelets and non-negative sparse representation, the proposed G-NSRC is more robust to ear occlusion, especially for large-scale occlusion.

6 Conclusions
In this article, a new feature extraction approach is proposed by using scale information of Gabor wavelets. The new Gabor scale feature extracts structural information of the ear in different scales and effectively provides a measure of the image’s local properties. Then, combining visual perception characteristic of Gabor features and non-negative sparse representation, we propose G-NSRC for ear recognition under occlusion. Compared with SRC in which the sparse coding coefficients can be negative, allowing the data to “cancel each other out” by subtraction, our G-NSRC conforms to the intuitive notion of combing parts to form a whole and therefore is more consistent with the biological modeling of visual data. Various levels of occlusion at any locations are evaluated more thoroughly to demonstrate the effectiveness of the proposed algorithm for ear recognition under occlusion in this article. Extensive experimental results show that the proposed classification paradigm achieves better recognition performance and is found to be more robust to large-scale ear occlusion. In the future work, we will consider studying greedy strategies to acquire the sparse representation coefficients of the underdetermined systems.

<table>
<thead>
<tr>
<th>Occlusion ratio</th>
<th>0%</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>30%</th>
<th>35%</th>
<th>40%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabor + SRC</td>
<td>99.5</td>
<td>94.3</td>
<td>64.6</td>
<td>35.2</td>
<td>18.7</td>
<td>10.1</td>
<td>6.3</td>
<td>8.6</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Gabor + NSRC</td>
<td>99.8</td>
<td>95.2</td>
<td>83.9</td>
<td>75.9</td>
<td>74.7</td>
<td>58.5</td>
<td>47.6</td>
<td>37.7</td>
<td>29.4</td>
<td>17.7</td>
</tr>
<tr>
<td>G-NSRC</td>
<td>99.8</td>
<td>99.5</td>
<td>98.7</td>
<td>97.7</td>
<td>96.2</td>
<td>92.7</td>
<td>87.6</td>
<td>85.3</td>
<td>75.2</td>
<td>51.9</td>
</tr>
</tbody>
</table>

Fig. 11 The recognition rate (%) of different methods under various levels of random occlusion.
Acknowledgments

This article is supported by the National Natural Science Foundation of China (Grant No. 61170116), the National Natural Science Foundation of China (Grant No. 61005009), Beijing Municipal Natural Science Foundation (Grant No. 4102039), and the Doctoral Fund of Ministry of Education of China (Grant No. 20100006110014). The authors would like to thank the anonymous reviewers for their insightful comments that helped to significantly strengthen this article.

References


Zhang et al.: Robust classification for occluded ear via Gabor scale feature...

Baoqing Zhang is a PhD candidate from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. His current research interests include computer vision, pattern recognition, and biometrics.

Zhichun Mu is a professor in the School of Automation and Electrical Engineering at University of Science and Technology Beijing, China. His research interests include pattern recognition, computer vision, artificial intelligence, and its applications.

Chen Li received his BS degree in 2005 and PhD degree in 2013 both from University of Science and Technology Beijing, China. She is currently working in the College of Information Engineering, North China University of Technology. Her research interests include biometrics, computer vision, 3-D reconstruction, and pattern recognition.
Hui Zeng is an associate professor in the School of Automation and Electrical Engineering at University of Science and Technology Beijing, China. Her research interests cover 3-D computer vision, image processing, and pattern recognition.