Ear Recognition using Improved Non-Negative Matrix Factorization

Li Yuan, Zhi-chun Mu, Yu Zhang
University of Science and Technology Beijing
E-mail: yuanli64@hotmail.com

Ke Liu
National Natural Science Foundation of China
E-mail: liuke@nsfc.gov.cn

Abstract
An Improved Non-Negative Matrix Factorization with sparseness constraints (INMFSC) is proposed by imposing an additional constraint on the objective function of NMFSC, which can control the sparseness of both the basis vectors and the coefficient matrix simultaneously. The update rules to solve the objective function with constraints are presented. Research of ear recognition and its application is a new subject in the field of biometrics authentication. In practical application, ear is maybe partially occluded by hair etc. So the proposed INMFSC is applied on ear recognition with normal images and partially occluded images. Experiment results show that, compared with the traditional NMFSC, the proposed method not only obtains higher recognition rate, but also improves the sparseness and the orthogonality of coefficient matrix.

1. Introduction
Earlier research has shown that human ear is one of the representative human biometrics with uniqueness and stability [1]. The primary issue of ear recognition is how to extract ear features effectively. At present, some existing ear feature extraction methods for 2D intensity images are as follows: Burge and Burger build adjacency graph out from Voronoi diagram, and use a graph matching based algorithm for authentication [2]. Hurley uses force field transformation to get a set of potential channels and potential wells as ear features [3]. Mu present a long axis based shape and structural feature extraction method [4]. Chang uses PCA algorithm for ear recognition, and gets the conclusion that ear and face does not have much difference on recognition rate [5]. But none of these methods deals with ear occlusion. This problem will arise during passive identification as no assistance on the part of the subject can be assumed.

Non-Negative Matrix Factorization (NMF) [6] is a part-based image representation method. It factors the image database into two matrix factors whose entries are all non-negative and produces a part-based representation of images because it allows only additive, not subtractive, combinations of basis components. A part-based representation can naturally deal with partial occlusion problems, so NMF has received much attention recently.

In this paper, an Improved Non-Negative Matrix Factorization with Sparseness Constraints (INMFSC) is proposed for ear recognition with occluded ear images. The rest of the paper is organized as follows: Section 2 describes NMF and NMFSC. Section 3 presents the improved NMFSC. Section 4 and 5 compare our approach with other NMF methods and conclude the paper.

2. Non-Negative Matrix Factorization with Sparseness Constraints (NMFS)
The problem with NMF is that it does not always get basis vectors for local representation. The reason is that the sparseness level for basis vector and coefficients matrix are not high enough. Local NMF [8] adds three additional constraints on the NMF basis: maximum sparsity in \( H \), maximum expressiveness of \( W \), maximum orthogonality of \( W \). But the convergence of LNMF is time consuming, and LNMF cannot explicitly control the sparseness of the representation.

Through above analysis of NFM and LNMF, we can see how to get the sparsity of basis vector and coefficient matrix is crucial. So Hoyer proposed NMF with sparseness control [9], optional sparseness constraints are added in the objective function to explicitly improve the sparseness of basis vectors and coefficients matrix. NMFSC is defined as following optimization problem:

\[
\min_{W,H} E(W,H) \\
\text{s.t. } W,H \geq 0, \sum W_j = 1 \quad \forall j
\]

(3)

where \( W_j \) is the \( j \) th column of \( W \) and \( h_j \) is the \( j \) th row of \( H \). Here, \( S_u \) and \( S_h \) are the desired sparsenesses of \( W \) and \( H \) respectively. These two parameters are set by the user.

The following projected gradient descent algorithm for NMF with sparseness constraints will get local optimal solution. For convenience, here we assume only applying sparseness constraints on \( W \).

1) Iterative update of \( W \) is projected gradient descent rule:

\[
W := W - \mu_w (WH - V)H^T.
\]

Then project each column of \( W \) to be non-negative, and have unchanged l2 norm, but l1 norm set to achieve desired sparseness \( S_u \) (refer to [10] for projection method).

2) Take standard multiplicative step for \( H \),

\[
H := H \odot (W^TV) \odot (W^TH).
\]

where \( \odot \) and \( \odot \) denote elementwise multiplication and division, respectively. \( \mu_w \) is a small positive constant (stepsizes) set by the user. The choice of which to constrain (\( S_u \) or \( S_h \) or both) depends on the specific application.

The problem with this method is that we cannot apply sparseness constraints on \( W \) and \( H \) at the same time. When \( S_u \) is higher, \( S_h \) is smaller, and when \( S_h \) is higher, \( S_u \) is smaller. We cannot get satisfying sparsity of both simultaneously.

3. Improved NFMSC

The objective of NMFSC is to get coefficients matrix \( H \) with higher sparseness while controlling the sparseness of basis vectors \( W \). So we make some improvement on the objective function. The main idea is to maximize the orthogonality of the vectors comprising the coefficients matrix to get higher sparseness of \( H \). This can not only minimize the redundancy information between feature vectors, but also raise the sparsity of feature vector in \( H \).

The objective function is defined as:

\[
E(W,H) = \| V - WH \|^2
\]

(5)

where \( U = H^TH \), \( \beta \) is a small positive constant.

The improved NMFSC is defined as following optimization problem:

\[
\min_{W,H} E(W,H) \\
\text{s.t. } W,H \geq 0, \sum W_j = 1 \quad \forall j
\]

(6)

\[
\text{sparseness}(w_j) = S_u \quad \forall j
\]

A local solution to the above constrained minimization, as an improved NMFSC factorization, can be found by using the following two step update rules:

\[
H_{a\mu} \leftarrow H_{a\mu} \frac{(W^TV)_{a\mu}}{(W^TH)_{a\mu} + \beta \sum H_{a\mu}}
\]

(7)

\[
W_{ia} \leftarrow W_{ia} - \mu_w [(WH - V)H^T]_{ia}
\]

(8)

then project each column of \( W \) to be non-negative as the traditional NMFSC. Iterative update of coefficient matrix \( H \) is multiplicative rule, it will multiply a non-negative factor for each step. So if \( H \) is initialized to be non-negative, it will always be non-negative updated by (7), and (6) will be minimized. The convergence of the updating rules can be proved in a similar way as [9].

4. Experiments

4.1. Ear recognition without occlusion

We use a subset of USTB ear database [11]. Figure 1 shows four images manually extracted of one subject. First three images of each subject are used for training; the fourth one is used for testing.

Figure 1. Example of USTB Ear Database
Figure 2 shows basis images of the INMFSC, which were all learned from the training ear images. The numbers of bases are 25 and 40 respectively. The images show that the bases are both additive and spatially localized for representing ears.

![Basis images of INMFSC](image)

(a) (b)

Figure 2. Basis images of INMFSC. (a) base number is 25; (b) base number is 40

Nearest neighbor classifier is used for classification. As shown in Table 1, INMFSC gives higher recognition rates than NMF and NMFSC on the USTB ear database. It can be found that when the number of bases increases, the recognition rate does not necessarily increase (Note the INMFSC, when number of bases raise from 60 to 70, the recognition rate falls from 93% to 91%). The probable reason is that the information redundancy between basis vectors increases when the number of bases gets higher.

### Table 1. Comparison of recognition rate

<table>
<thead>
<tr>
<th>Dimension of Feature Vector</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMF</td>
<td>67%</td>
<td>82%</td>
<td>84%</td>
<td>82%</td>
<td>83%</td>
</tr>
<tr>
<td>NMFSC</td>
<td>69%</td>
<td>80%</td>
<td>83%</td>
<td>86%</td>
<td>83%</td>
</tr>
<tr>
<td>INMFSC</td>
<td>74%</td>
<td>88%</td>
<td>87%</td>
<td>93%</td>
<td>91%</td>
</tr>
</tbody>
</table>

#### 4.2. Comparison of Sparseness and Orthogonality of feature vectors

Each column in \( H \) stands for a feature vector. Sparseness measure is based on the relationship between the \( L1 \) norm and the \( L2 \) norm [6] as follows:

\[
\text{sparseness}(X) = \frac{\sqrt{n} - \left( \sum_{i=1}^{n} x_i \right)}{\sqrt{\sum x_i^2}}
\]

where \( n \) is the dimensionality of \( X \). This function evaluates to unity if and only if \( X \) contains only a single non-zero component, and takes a value of zero if and only if all components are equal (up to signs). The sparseness value will interpolate smoothly between the two extremes. Here, we take \( S_w \) as 0.7, \( \beta \) as 0.5. The average of the sparseness of all rows of coefficient matrix is used for comparison. The result is shown in Table 2. It shows that INMFSC gets higher sparseness of feature vectors than NMFSC. But this does not mean that the sparseness is proportional to the recognition accuracy.

### Table 2. Comparison of sparseness

<table>
<thead>
<tr>
<th>Dimension of Feature Vector</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMF</td>
<td>0.495</td>
<td>0.463</td>
<td>0.433</td>
<td>0.409</td>
</tr>
<tr>
<td>INMFSC</td>
<td>0.663</td>
<td>0.621</td>
<td>0.613</td>
<td>0.572</td>
</tr>
</tbody>
</table>

The orthogonality of coefficient matrix is measured by \( \sum_{i,j} H^T H \). The result is shown in Table 3.

We can see that INMFSC gets higher orthogonality than NMFSC, which means less redundancy information between feature vectors.

### Table 3. Comparison of orthogonality \((\times 10^3)\)

<table>
<thead>
<tr>
<th>Dimension of Feature Vector</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMF</td>
<td>1.744</td>
<td>2.034</td>
<td>2.032</td>
<td>2.280</td>
</tr>
<tr>
<td>INMFSC</td>
<td>2.279</td>
<td>2.638</td>
<td>2.927</td>
<td>3.457</td>
</tr>
</tbody>
</table>

#### 4.3. Ear recognition with occlusion

We use the occluded ear image database from USTB ear database [11]. This database has 24 subjects, 4 images per subject (1 without occlusion, 3 with occlusion). Figure 4 shows four images manually extracted of one subject. The ear database without occlusion (mentioned in section 4.1) are used to get the basis space. The gallery consists of images without occlusion (shown as the first image of Figure 3 (a)); the probe consists of occluded images (shown as the last three images of Figure 3 (a)). A sub-region based method is proposed for recognition, each image is split into three sub regions as shown in Figure 3(b).

![Example of occluded ear database](image)

(a) (b)

Figure 3. (a) Example of occluded ear database. (b) Splitting the ear image into three sub-regions

Let \( R \) be the total number of samples, \( N_r \) is the total number of sub-regions. Each training sample is segmented into \( N_r \) parts noted as \( \{ y_{r,i} \} \), \( r = 1, \ldots, R, i = 1, \ldots, N_r \) . We use improved NMFSC to get \( N_r \) sub space \( \{ y'_{r,i} \} \). Suppose \( v \) is an
arbitrary feature in subspace, we can use a Gaussian model to get its probability of belonging to which sub-basis region.

\[ p(v) = \frac{1}{(2\pi)^{d/2} \mid \Sigma \mid^{1/2}} \exp \left\{ -\frac{1}{2} (v - \hat{u})^T \Sigma^{-1} (v - \hat{u}) \right\} \]  \hspace{1cm} (10)

\[ \hat{u} = \sum_{r=1}^{R} \frac{y_{r,i}}{\sum_{r=1}^{R} N_r} \] \hspace{1cm} (11)

\[ \Sigma = (\sqrt{\sum_{r=1}^{R} N_r}) \sum_{r=1}^{R} \sum_{i=1}^{N_r} \left( y_{r,i} - \hat{u} \right) \left( y_{r,i} - \hat{u} \right)^T \] \hspace{1cm} (12)

The \( i \) th subregion of \( r \)th probe sample is \( \{p_{ro_i}\} \). Project each sub-region onto the basis space to get \( \{p_{ro_i} \} \). With (12), we can get \( p_{ro_i} \)'s probability \( (p_i) \) of belonging to the \( i \) th subregion. If \( p_i \) is smaller than the threshold \( T_i \), then this region is determined to be occluded.

The \( i \) th sub-region of \( r \)th gallery sample is \( \{g_{ri}\} \). Project each sub-region onto the basis space to get \( \{g_{ri} \} \). The similarity of gallery and probe subregion is defined as:

\[ s_r = \frac{-g_{ri}^T p_{ro_i}}{\| g_{ri} \| \| p_{ro_i} \|} \] \hspace{1cm} (13)

Then the total similarity of the probe image and the \( r \)th gallery image is:

\[ S_r = \sum_{i=1}^{N_r} p_i s_{ri} \] \hspace{1cm} (14)

The bigger of \( p_i \), the more contribution of the corresponding \( i \) th subregion to the recognition. The final decision is made by:

\[ f = \arg \max_r S_r \] \hspace{1cm} (15)

In this paper, each ear image is segmented into 3 parts, as shown in Figure 3(b). In Figure 4, the white column stands for the recognition rate using the whole ear image for recognition, the grey column stands for the recognition rate using sub-region image. We can see that the sub-region based method gets higher recognition rate. This advantage gets more obvious when the occlusion area percent gets higher.

5. Conclusion

An improved non-negative matrix factorization (INMFSC) algorithm for ear recognition with occlusion and learning part-based subspace representations. The main idea of this INMFSC is to apply sparseness constraints to both basis vectors and coefficient matrix (feature vectors) simultaneously, and raise orthogonality of the coefficient matrix. Experiment results show that INMFSC get higher sparsity and orthogonality of feature vectors, and higher recognition rate on the USTB ear database. We also applied INMFSC to partially occluded ear image recognition. Experiment result has shown that sub-region based method get better performance than the whole-image based method.

6. Acknowledgement

This work is supported by: (1) the National Natural Science Foundation of China under the Grant No. 60375002, and No. 60573058; (2) the key discipline development program of Beijing Municipal Commission of Education.

References