IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 25, 2018, accepted January 16, 2019, date of publication February 4, 2019, date of current version March 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897213

Hierarchical Clustering Based Band Selection Algorithm for Hyperspectral Face Recognition

QIDONG CHEN[1], JUN SUN[1], VASILE PALADE[2], (Senior Member, IEEE), XIAOQIAN SHI[3], AND LI LIU[4]

[1]School of IoT, Jiangnan University, Wuxi 214122, China
[2]School of Computing, Electronics and Mathematics, Coventry University, Coventry CV1 5FB, U.K.
[3]School of IoT, Wuxi Taihu University, Wuxi 214122, China
[4]Affiliated Hospital, Jiangnan University, Wuxi 214122, China

Corresponding author: Jun Sun (sunjun_wx@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Project 61673194, Project 61672263, Project 61672265, and Project 61876072, in part by the National First-Class Discipline Program of Light Industry Technology and Engineering under Project LITE2018-25, and in part by the Natural Science Foundation for College and Universities in Jiangsu Province under Project 16KJB520051.

ABSTRACT Hyperspectral face recognition is a small sample size problem, where usually less than four hyperspectral cubes are available as training data. At the same time, hyperspectral face image acquires grayscale images over a series of continuous spectra which usually contain large redundant information or noise, especially in the near infrared spectrum bands. Therefore, dimensionality reduction and feature extraction are important tasks on this problem. This paper proposes a hierarchical clustering-based spectrum band selection method, which mitigates the influence of noise and extracts features from each spectra band by using the Gabor filter and the histograms of oriented gradients algorithm, In addition, the fusion of Hog and Gabor features was embedded into the nearest neighborhood-based classifier for performance comparison. The experimental results show that the proposed algorithm is time effective and provides robust performance.

INDEX TERMS Hyperspectral face recognition, band selection, Gabor filter, HOG features, image fusion.

I. INTRODUCTION

Hyperspectral imaging (HSI) technology collects information from different spectral bands including object's morphological, chemical and internal structural features, by recognizing object's spatial information and spectral information [1]. It can be used in several authentication areas, such as evaluating food quality [2], food safety detection [3] and medical treatment like cancer detection [4].

Research on hyperspectral face recognition started in 2003; Pan *et al*. [5] first used a NIR spectral human face database for recognition, where the database has the 31D band from 700 nm to 1000 nm. In the near-infrared range, the face positions on the 31 bands are manually marked, and the average spectral characteristics of the forehead, cheeks, and mouth are used for identification. They also fused the spatial information of the hyperspectral image in order to get a more representation vector for human face recognition, where each pixel in the fusion image was selected from a specific band in the same position. Through this method, a 3D hyperspectral image cube is converted into a 2D image.

Di *et al*. [6] established in 2010 a hyperspectral face database, where each band in this database was selected from every 10nm within the 400-720nm range. During the process, based on the activity of human skin and absorption and reflection characteristics of carotene, hemoglobin, and melanin, six spectral bands were selected as the following experimental bands, namely 530nm, 540nm, 550nm, 570nm, 580nm, and 590nm. Features were extracted using the 2DPCA algorithm. Finally, the K-nearest neighbor classifier was used for classification.

Uzair *et al*. [7] proposed a band fusion algorithm to merge hyperspectral images into one, and used the Partial Least Squares (PLS) regression algorithm to achieve face recognition and classification. In this paper, the hyperspectral image was firstly divided into $m \times n$ blocks by a sliding window. Then, each time the area in the image covered by the sliding window was calculated to a value through a series of computations, and those values merge into a 2D matrix, or the final fused image in the end. Then,

The associate editor coordinating the review of this manuscript and approving it for publication was Dong Wang.

the PLS was used for image classification. The algorithm was tested on available hyperspectral face image databases and compared with 18 existing algorithms to verify its performance. The experimental results show that the proposed band fusion algorithm performed the best. Moreover, 3D-DCT and Partial Least Squares [8], 3D Gabor wavelets [9] and a CRC-based classifier [10] are also used for hyperspectral face classification.

However, a Hughes phenomenon is encountered in hyperspectral image classification problems [11]. That is, the classification accuracy does not increase with bands increase; instead, it decreases with bands increase. A small sample problem ('curse of dimensionality') [12] also exists in the hyperspectral face recognition, which means that the dimensionality reduction is one of the critical issues in a hyperspectral image classification problem. The ''small sample problem'' is also the reason that makes deep neural networks not performing well on this problem. If we consider the hyperspectral face database carefully, we would find that the spectral interval between two adjacent bands is only 10 nm, which results in the correlation between the images of adjacent bands being high. So, the band selection methods are widely used in the hyperspectral face image classification, which can preserve the spectral characteristics of the human face without losing too much information [6], [6]. There is no doubt that the band fusion based algorithm obtained a good performance, but it is time consuming, which is a significant problem when applying this algorithm to the real world.

Martínez-Usómartinez-Uso [13] proposed a band selection algorithm based on hierarchical clustering and applied it to the hyperspectral remote sensing images classification. They firstly calculated the mutual information values, or KL divergence values, between two bands in the hyperspectral remote sensing images, and finally obtained an irrelevance matrix. Then, they used the linkage clustering algorithm to divide all the bands into K clusters based on the calculated irrelevance matrix, and select feature bands according to the weight of each band in each class. It performed well on hyperspectral sensing images classification, but there was only one hyperspectral cube in the experiment, where all the categories were present in the same cube. In this paper, we modified this hierarchical clustering method and applied it to the hyperspectral face image band selection. On the one hand, the hierarchical clustering based band selection method can mitigate the influence of noise, which further enhances the quality of features. Moreover, traditional feature extraction approaches, such as the Gabor filter and the histograms of oriented gradients (HOG), are applied in this paper in order to represent the hyperspectral image cube better, and fusion of Hog and Gabor features are used for the final recognition.

The rest of the paper is organized as follows: We first introduce the basic information theory including mutual information and Kullback-Leibler (KL) divergence in Section 2. Then, the process of the new algorithm is described in Section 3, including hierarchical clustering and the weighted band selection based on the information theory. In Section 4, we described the experimental database we used and the parameters we set. The proposed algorithm is compared to several other algorithms, and the results are analyzed in Section 5. Finally, conclusions and future work are described in Section 6.

II. INFORMATION THEORY

A. MUTUAL INFORMATION

Mutual information is a common method of computational linguistic model analysis [14], which measures the mutuality between two objects. Mutual information is originally a concept from information theory. It is used to represent the relationship between pieces of information and is a measure of the statistical correlation between two random variables [6]. In addition to being applied to computational linguistic model analysis, mutual information is also widely used in neural network learning [15], Wireless Biomedical Implant Systems [16] and printing images registration [17].

For hyperspectral image data $x_1, x_2 \cdots, x_n$, n being the total number of bands, the mutual information value $I(X_i, X_j)$ is:

$$I\left(X_i, X_j\right) = \sum_{x_i \epsilon \Omega} \sum_{x_j \epsilon \Omega} log \frac{p\left(X_i, X_j\right)}{p\left(X_i\right) p\left(X_j\right)} \tag{1}$$

where mutual information I usually is a non-negative value, and it equals to 0 when random variables X and Y are statistically independent. It is determined by $p\left(X_i\right)$ and $p\left(X_j\right)$, which is the probability density of the *ith* band image and the *jth* band image. A bigger value I means the correlation between random variables X and Y is stronger. Ω is the spectra band cluster set. Mutual information value I satisfies the relation:

$$0 \leq I\left(X_i, X_j\right) \leq min\left\{H\left(X_i\right), H\left(X_j\right)\right\} \tag{2}$$

Mutual information value I can also be calculated as follows:

$$I\left(X_i, X_j\right) = H\left(X_i\right) + H\left(X_j\right) - H\left(X_i\right), \quad H(X_i, X_j) \tag{3}$$

where $H(X_i, X_j)$ means the joint entropy, determined by the joint probability density $p(x_j, x_j)$.

Due to the reason that variables X_i and X_j may possibly have a weak correlation or a small entropy, the standard mutual information value NI is defined as:

$$NI\left(X_i, X_j\right) = \frac{2 \cdot I\left(X_i, X_j\right)}{H\left(X_i\right) + H\left(X_j\right)} \tag{4}$$

According to the standard definition, the dissimilarity matrix [5] of two band images can be defined as:

$$D_{NI}\left(X_i, X_j\right) = (1 - \sqrt{NI\left(X_i, X_j\right)})^2 \tag{5}$$

B. KULLBACK-LEIBLER DIVERGENCE

KL divergence is an asymmetric metric, however, the symmetric version of the KL divergence is often used [18], [19]. The most important property of the KL divergence is non-negativity, and that it equals to 0 when probability distributions $p_i(x)$ and $p_j\left(x\right)$ are the same.

TABLE 1. The selected high weight bands.

Cluster Number K	The selected bands (nm)
5	510,600,610,1030,1040
10	490,500,510,590,600,610,620,1020,1030,1040
15	490,500,510,520,530,590,600,610,620,640,660,970,1020,1030, 1040

KL can be used as a distance to determine the two probability distributions and can be interpreted as the cost of using one of the distributions instead of the other. In hyperspectral face band selection, the corresponding probability distribution can be used as a measure of dissimilarity between two images with different spectra band. For hyperspectral image $x_1, x_2 \cdots, x_n$, the KL divergence value is calculated as:

$$D_{KL}\left(X_i, X_j\right) = \sum_{x\in\Omega} p_i(x) log \frac{p_i(x)}{p_j(x)} + \sum_{x\in\Omega} p_j(x)\, log \frac{p_j(x)}{p_i(x)} \quad (6)$$

where Ω is the spectra band cluster, $p_i(x)$ and $p_j(x)$ is the distribution of image i and image j, respectively.

III. HYPERSPECTRAL IMAGE RECOGNITION

A. HIERARCHICAL CLUSTERING

Hierarchical clustering is to decompose or merge datasets layer by layer until a specific condition is met, different connection strategies resulting in different tree structures. According to the various forms of hierarchical decomposition, we can have bottom-up cohesive hierarchical clustering methods and top-down splitting clustering methods. Cohesive hierarchical clustering method starts with each object as a separate cluster. In each iteration, the two nearest clusters are combined according to the given connection method to form a new group until the number of clusters equals to the set. On the contrast, the splitting hierarchical clustering method classifies all the objects into one category initially, then split it into two smaller clusters in each iteration until the termination of the condition is met. In this paper, a bottom-up cohesive hierarchical clustering method is used.

One of the well-known methods in the cohesive hierarchical clustering method is Ward's linkage [20], known as the minimum variance clustering criterion because the Ward's linkage attempts to minimize intra-class variance within each cluster. Some studies have shown that this method is superior to other hierarchical clustering methods [21]. It was first used in the field of hyperspectral remote sensing classification here.

Assume that C_r and C_s are merged into a new cluster, this correlation value or the distance between the cluster (C_r, C_s) and other cluster C_k is

$$\begin{aligned} &D\left[\left(C_k\right), \left(C_r, C_s\right)\right] \\ &\quad = \alpha \cdot D\left(C_r, C_k\right) + \beta \cdot D\left(C_k, C_s\right) + \gamma \cdot D\left(C_r, C_s\right) \\ &\qquad + \delta \cdot \left|D\left(C_k, C_r\right) - D\left(C_k, C_s\right)\right| \end{aligned} \quad (7)$$

where α, β, γ and δ are the coefficients, $\alpha = \frac{n_r+n_k}{n_r+n_s+n_k}, \beta = \frac{n_s+n_k}{n_r+n_s+n_k}, \gamma = \frac{-n_k}{n_r+n_s+n_k}, \delta = \emptyset$, n_r presents the total number bands of r_{th} cluster, and the relevant matrices were initialized through mutual information criteria or KL divergence. Each iteration merges two clusters into one and relevant matrices $D_{N\times N}$ are updated according to Equation (7). Finally, the hyperspectral face image with N bands are divided into K clusters finally.

B. WEIGHTED BAND

After K clusters are determined, the band for presenting each cluster should be selected. There are two strategies: one is to select the largest weighted band to present the cluster; the other one is to get a fusion image based on the weighted band. The weight of each band in the cluster can be calculated as:

$$W_i = \frac{1}{R} \sum\nolimits_{j\in\Omega, j\neq i} \frac{1}{\varepsilon + D_{kl}\left(X_i, X_j\right)} \quad (8)$$

where ε a small non-negative value,R is the total number of band in spectra band sub-cluster Ω. $D_{KL}\left(X_i, X_j\right)$ described the distance or the correlation value between band image i and j. The bigger W_i the band has, the more the band can be able to present the cluster.

It's worth noting that we computed the correlation matrix of all the hyperspectral image cubes and got 15 largest weight bands of each cube to form a matrix, rather than chose K bands using the hierarchical grouping method from this matrix. We did this because K bands formed one cube randomly and it is inappropriate to represent the whole database.

It can be seen that the selected bands are all the middle spectral bands that are corresponding to Wei Di's experimental results, which proved that the hierarchical clustering method is effective. The selection bands are shown in Table 1. Apart from selecting the biggest weight band, we can also use a simple fusion method based on the weight to fuse a new image to present the cluster. In our experiment, we use the mean fusion algorithm to obtain the final image Y, which can be defined as:

$$Y = \frac{1}{R} \sum\nolimits_{i=1}^{R} X_i W_i \quad (9)$$

where R is the total bands of the sub-cluster; X_i is i_{th} band image; and W_i is the weight of each band image.

C. FEATURE EXTRACTION

Feature extraction is one of the most important steps in any machine learning algorithms. In human face classification,

the histogram of oriented gradient and Gabor wavelet are the most widely used feature extraction methods.

Due to the biological relevance and outstanding computational properties [21], the Gabor wavelet is a powerful tool for feature extraction [21], [22] that can extract optimal localization face features in both spatial and frequency domains [23], which are corresponding to the information from different positions in space, different frequencies and different directions. The Gabor filter can be divided into a real part and an imaginary part. The image is smoothed after using the real part; the imaginary part is used to detect the edge. The kernel function of a two-dimensional Gabor filter in the space and spatial frequency domain is given by:

$$g(x,y) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma_{xy}^2}\right)\exp\left(i\left(-2\pi\frac{x'}{\lambda}+\psi\right)\right) \tag{10}$$

$$x' = x\cos\theta + y\sin\theta, \quad y' = y\cos\theta - x\sin\theta \tag{11}$$

where σ_{xy} presents the standard deviations of the Gaussian envelope which characterizes the spatial extent and the bandwidth of the filter, which is determined by the semi-responsive spatial frequency bandwidth. λ presents the wavelength, normally set to 2, but it cannot be bigger than one-fifth of the input image size. Direction is θ, the parameter that specifies the direction of the parallel stripe of the Gabor function, which takes values from 0 to 360 degrees. ψ is the phase offset, which ranges from -180 degrees to 180 degrees. Aspect ratio γ determines the shape of the Gabor function. In our experiments, we set five different wavelengths $\lambda \in [2, 2\sqrt{2}, 4, 4\sqrt{2}, 8]$, and eight directions between 0 to 360 degrees, with a step of 45. Therefore, we use 40 Gabor filter kernels for feature extraction in the end.

The histogram of oriented gradients (HOG) is a feature descriptor with successful application in image recognition. It was proposed by Dalal and Triggs [23] in 2005. Because the image gradients mainly exist at the edges of the image, it needs to grayscale the image firstly, and then use Gamma correction method to normalize the input image to adjust the contrast of the image, reduce the impact of light changes on the image, and suppress noise interference:

$$\Gamma(x,y) = \Gamma(x,y)^{gamma} \tag{12}$$

where Γ is the image after grayscale, the value of *gamma* is always set to 0.5.

The next step is competing for the gradient, which includes size and directions of each pixel to obtain profile and texture information, reducing the impact of light changes. It can be defined as:

$$G_X(x,y) = H(x+1,y) - H(x-1,y) \tag{13}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1) \tag{14}$$

where $G_X(x,y)$, $G_y(x,y)$, $H(x,y)$ is horizon gradient, vertical gradient and the pixel value of pixel $P(x,y)$. The gradient value $G_{x,y}$ and direction $\alpha_{x,y}$ of each pixel is described as:

$$G_{x,y} = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{15}$$

$$\alpha_{x,y} = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{16}$$

The image is divided into multiple cell units, descriptors of these cell units are obtained by Equations (15) and (16) and then combined to form the block HOG descriptor, and the HOG features of the image is a combination of all block feature descriptors.

In order to present the hyperspectral face cube better, we fused the localization face features extracted from Gabor wavelet and the edge HOG features. Finally, the fusion feature $\overrightarrow{FF}$ can be described as:

$$\overrightarrow{FF} = \left[\overrightarrow{HOG}, \overrightarrow{Gabor}\right] \tag{17}$$

D. MAJORITY VOTING METHOD

The majority voting method is a simple and efficient decision-making method, which can maintain a low rate of refusal recognition with a higher correct recognition rate, and was successfully applied in many areas [24], [25]. Mu *et al.* [26] analyzed the superior performance of the majority voting method through experiments. Besides this, images can also be divided into several overlapping small blocks, and then using the majority voting method based classification algorithm to classify each small block of images.

Finally, the classification of the test image was decided according to the recognition results of each small block. This paper selected the nearest neighbor (NN) as the basic classifier to classify the test hyperspectral human face image. All the feature extracted from different band images in the K_i were put into the classifier, and, finally, we got K results for the test image, and the final classification result is the most encountered category. The flowchart of the classification algorithm is shown in Table 2.

IV. EXPERIMENTAL SETTING

A. EXPERIMENTAL DATABASE

In this paper, experiments are performed on the CMU-HSFD [27] and Poly-U [6] databases. Table 3 gives a summary of the database. The CMU database is a hyperspectral face database (CMU-HSFD) obtained with a prototype spectral polarization camera.

Each hyperspectral image contains 65 different bands, ranging from 450nm to1090 nm, with a step size of 10nm. Figure 1 shows a sample subject with all the bands' image. The database contains 147 hyperspectral image cubes of 48 face categories, each of which has 1 to 5 hyperspectral images cubes taken in different time periods and different lighting combinations. Each band has a low signal-to-noise ratio, and most subjects have a blinking or small movements during image acquisition. Some persons in the samples wear glasses or hats.

TABLE 2. Proposed algorithm flowchart.

Algorithm: Hierarchical grouping based band selection method for HSI classification
Input: Hyperspectral human face image database
1. Compute the correlated matrix $D_{N\times N}$ of all cubes
2. Use the linkage clustering algorithm based on $D_{N\times N}$ to merge the bands into 15 clusters of each cube and top 15 weight bands of each cube to form a statistical matrix.
3. Compute the weight of each band in every cluster by Equation (8).
4. Select K largest weight bands from the statistical matrix to present the cluster or get a fusion image based on the weight.
5. Extract HOG feature, Gabor feature and fusion feature of HOG and Gabor, and reduce the feature dimensions by PCA, then put it into the NN classifier to get K results.
6. Get the final classification result using the majority voting method.
Output: Hyperspectral human face image classification accuracy

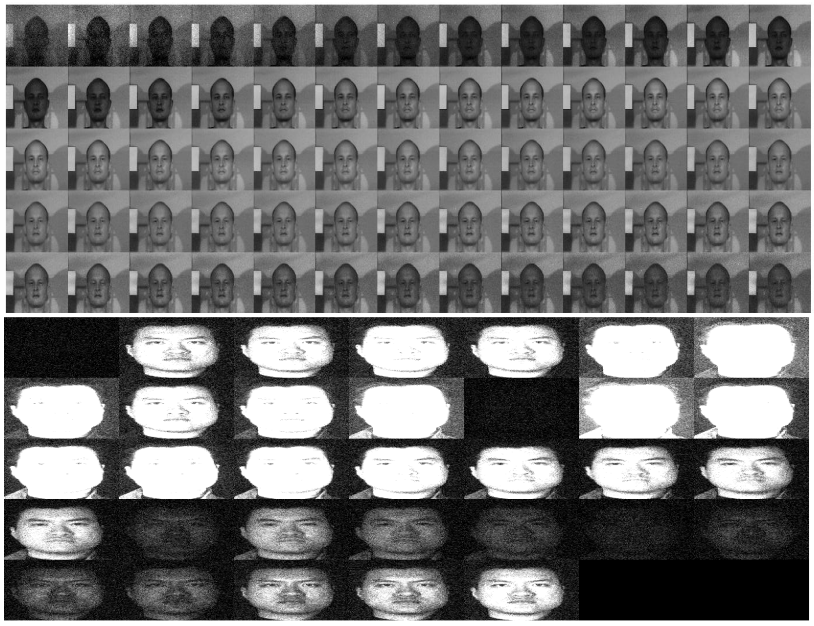


FIGURE 1. CMU (top) and Poly-U (bottom) hyperspectral image example.

TABLE 3. Summary of the CMU and ploy-U databases.

Database name	Categories number	Total images	Total Bands	Spectral range
CMU	48	147	65	450-1090nm
Ploy-U	25	113	33	400-720nm

The Poly-U hyperspectral face database was acquired by the indoor hyperspectral face acquisition system, which has 113 image cubes from 25 volunteers, with age range from 21 to 33. Each volunteer has 4 to 7 hyperspectral image cubes, and each image cube contains 33 spectra bands, ranging from 400 to 720 with a step size of 10nm. Compared to the CMU database, it is clearly that the Poly-U database has more noise, as Figure 1 shows.

B. EXPERIMENTAL PARAMETER

In this paper, we chopped each hyperspectral image cube at first in order to get the face profile, and then we computed the correlation matrix based on the KL divergence or mutual information. Then, to reduce the image dimension and redundant information, we chopped the face area and resized all the selected band image sizes to 40×30 on the CMU database and 45×45 on the Ploy-U database.

For the Gabor wavelet, we set the filter kernel to 40×30, with five different wavelengths and eight different directions, and we got 40 filter kernels to finally extract image features. The grayscale parameter Gamma was set to 0.5, when we obtained the HOG features. And the energy preserving ratio was set to 0.7.

For the number of clusters, according to Adolfo's experiment, the accuracy increases when the clusters' number

TABLE 4. Average recognition accuracy based on 10-fold cross-validation on CMU and ploy-U databases.

Bands number K	CMU Database			Ploy-U Database		
Algorithms	5	10	15	5	10	15
MI-WHCBS-GHF	88.06%	90.94%	82.53%	87.74%	88.45%	81.73%
MI-WHCBS-HOG	**98.06%**	93.43%	87.16%	**96.20%**	94.13%	88.12%
MI-WHCBS-Gabor	90.15%	87.76%	82.54%	91.21%	88.3%	81.57%
KL-WHCBS-GHF	85.97%	82.84%	82.99%	86.37%	83.35%	83.09%
KL-WHCBS-HOG	91.79%	91.36%	89.70%	92.79%	91.86%	88.76%
KL-WHCBS-Gabor	86.12%	84.33%	84.03%	87.12%	85.77%	84.33%
MI-HC-GHF	97.46%	98.06%	96.72%	92.31%	94.86%	**97.83%**
MI-HC-HOG	97.31%	97.31%	97.91%	94.14%	95.28%	96.95%
MI-HC-Gabor	96.42%	97.46%	95.52%	91.26%	94.14%	97.52%
KL-HC-GHF	97.61%	**98.21%**	**98.66%**	93.56%	95.54%	**98.77%**
KL-HC-HOG	96.42%	96.12%	95.22%	94.62%	96.92%	98.46%
KL-HC-Gabor	96.87%	97.61%	97.91%	92.52%	95.44%	98.45%

increase and becomes stable eventually. Based on this theory, the test have compared the accuracy with three different numbers, namely is 5, 10, 15.

For all the experiments on the CMU dataset, we select 58 training samples that contain 2 samples of 29 categories randomly, and the remaining 67 samples are used to test our algorithms firstly. Then, we select 48 training samples that contain one sample of 48 classes, and the remaining 99 samples for comparing the algorithm we proposed with the band fusion based algorithm. However, it is difficult to compare the accuracy on the Poly-U dataset because it is updated frequently. Therefore, we choose 50 image cubes randomly from each volunteer for training, and the remaining 63 image cubes for testing. All the experiments results were obtained under a 10-fold validation method.

It should be noted that all the experiments in this paper were run on MATLAB 2015a platform, and the computer's operating system was Windows 7 with 1.6GHz dual-core Intel Core i5 processor and 8GB 2133MHz LPDDR3 memory.

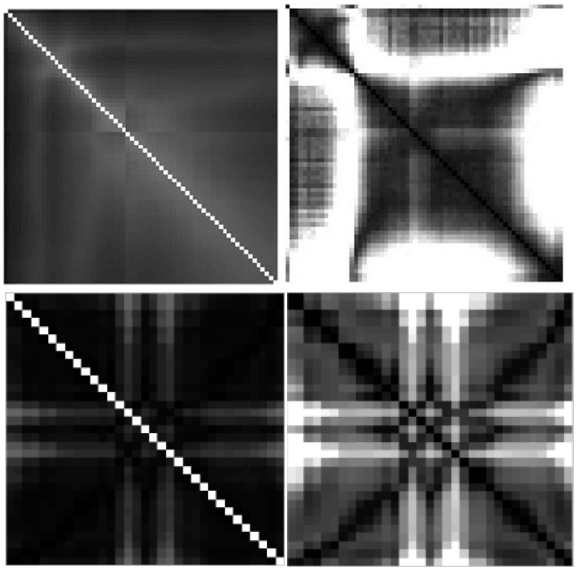


FIGURE 2. (1) The top figures are the visualization of the mutual information (left) based correlation matrix and KL divergence based correlation matrix (right) of the CMU cube. (2) The bottom figures are the visualization of the mutual information (left) based correlation matrix and KL divergence based correlation matrix (right) of the Ploy-U cube.

V. RESULTS AND ANALYSIS

A. ORRELATION MATRIX VISUALIZATION

The visualization of correlation matrix based on KL divergence and mutual information is shown in Figure 2. These two figures show that there exists a strong correlation between different bands. In the left figure, the white pixel presents the high relationship, while in the right picture is on the contrast. The correlation matrix size is determined by the bands number of hyperspectral image cube, which is 65*65 on the CMU database, and 33*33 on the Ploy-U database.

The time cost of computing the mutual information based matrix is $O(\lambda^3)$, while for the KL divergence based is $O(\lambda^2 MN)$. Compared to the mutual information, the KL divergence based computing method needs less RAM and time cost. In our experiment, we chose the KL divergence measure to compute the correlation matrix.

B. ACCURACY RESULTS AND ANALYSIS

In the experiment, we proposed four methods based on the group clustering method with different feature extraction methods, where these four methods are described one-by-one below. MI and KL mean mutual information and Kullback-Leibler divergence, respectively. WHCBS means these algorithms use the fusion image to classify as the Equation (9) shows, while HC means to select the largest weight band spectral image in each cluster to classify. Besides it, GHF means the fusion of the features obtained by Gabor wavelet and HOG is used in the algorithm.

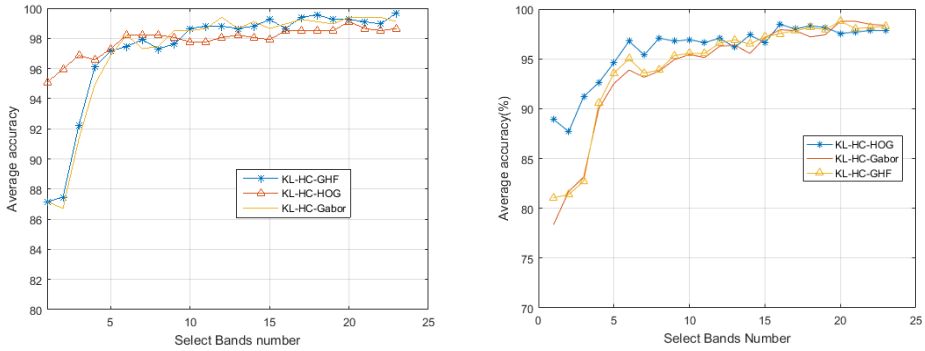


FIGURE 3. The average accuracy of different select bands number and extraction feature based on the largest weight selection method. The selected cluster number ranges from 1 to 24 on the CMU database (left) and the Poly-U database (right).

TABLE 5. Average recognition accuracy based on 10-fold cross-validation.

Bands number K	5	10	15
CMU			
KL-HC-HOG	97.46%	97.21%	99.4%
KL-HC-GHF	97.31%	98.36%	99.7%
Ploy-U			
KL-HC-HOG	94.62%	96.92%	98.46%
KL-HC-GHF	93.56%	95.54%	98.77%

TABLE 6. Time consumption based on 10-fold validation.

Bands number K	5	10	15
CMU			
KL-HC-HOG	12.3s	27.4s	45.5s
KL-HC-Gabor	1137.1s	2753.6s	5625.4s
KL-HC-GHF	1273.s	2860.3s	6042.7s
Ploy-U			
KL-HC-HOG	10.7s	24.3s	39.5s
KL-HC-Gabor	847.1s	2654.1s	4627.4s
KL-HC-GHF	983. 7s	2690.5s	4746.5s

Table 4 shows that the accuracy of KL divergence based

We also tested the change of the average accuracy with a different number of select bands. Figure 3 shows that the average accuracy increased with increasing the selection number. The fluctuation phenomenon occurs because the training samples were selected randomly each iteration, but the overall trend is growing by the selection bands number.

There is no doubt that the accuracy of feature fusion is higher because it can extract more information than only by extracting HOG features. Also, as shown in Table 5, the highest average accuracy was 99.7% and 99.4% for the KL-HC-GHF and KL-HC-HOG, respectively. We can also see that the accuracy of the KL-HC-Gabor and KL-HC-GHF are both increased when the number of clusters increased, and reach to the peak when there were 20 groups. Moreover, the accuracy declined along with the number of selected bands when we used fusion image for training and recognition.

However, despite the little higher accuracy, the time consumption fro the KL-HC-Gabor and KL-HC-GHF are approximately one hundred times higher than for the KL-HC-HOG, as shown in Table 6. This proved that KL-HC-HOG is an efficient algorithm with low computation cost. It should be noted that the total number of testing samples is 67 for the CMU database and 63 for the Ploy-U database.

Deep learning models are successfully implemented in the field of face recognition these days due to their strong feature extraction capabilities, including recurrent neural networks [28] and convolution neural networks [29]. However, in this paper, we emphasized on reducing redundant information by selecting parts of bands based on a hierarchal clustering method. Deep learning models may not be suitable for hyperspectral face recognition. Firstly, we usually have only less than four hyperspectral cubes available as training data, and deep learning networks are often powerless for such small sample problems. In addition, face image cube acquires grayscale images over a series of continuous spectra, which contains large redundant information or noise, which may make deep learning models obtain worse features for identification due to the inadequate training.

C. COMPARING WITH THE BAND FUSION BASED ALGORITHM

Uzair *et al*. [7] proposed in 2015 a band fusion algorithm by extracting the spectral-space features to merge hyperspectral images into one, and used the Partial Least Squares (PLS) regression algorithm to achieve face recognition and classification.

As shown in Figure 4, the bright lines are the edges of the fusion image, and it is obviously compared to the original one. Due to this phenomenon, we observed that the many distinguishing features exist in the edges, which proves that

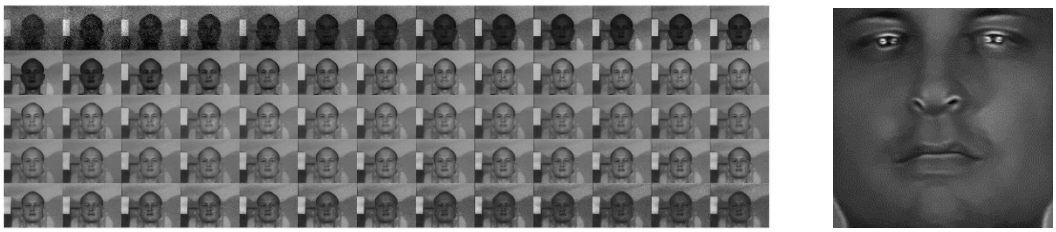


FIGURE 4. (left) An overall visualization of the example; (right) Chopped fusion image based on the band fusion algorithm.

TABLE 7. Accuracy and time-consumption comparison.

Algorithms	Training/Testing num.	Bands selected num.	Time cost (s)	Accuracy(%)
Band fusion(CMU)	48/99	4	2379.5	98.8
KL-HC-HOG(CMU)	48/99	14	21.2	96.9
KL-HC-HOG(CMU)	58/67	20	23.6	99.4
Band fusion(Poly-U)	50/63	5	1768.3	94.92
KL-HC-HOG(Ploy-U)	50/63	6	13.2	97.45

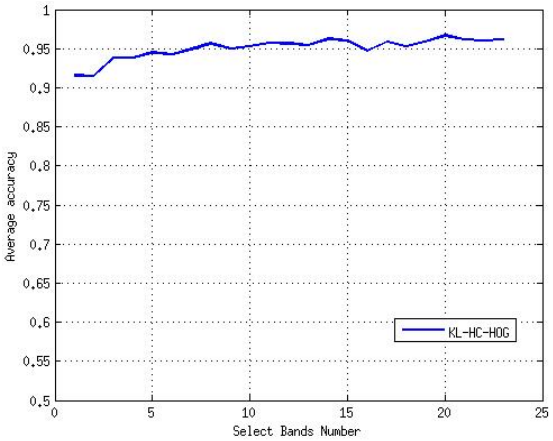


FIGURE 5. The average accuracy change of KL_HC_HOG based on 10-fold validation with the selected bands' number increasing. It should be noted that there are 48 hyperspectral image cubes used for training, the remaining 99 image cubes are used for testing.

the directional gradient histogram (HOG) can extract more significant features from the side.

The database we split in this experiment is the same as Uzair's [3] set that only one cube of each category was used as the training example, the remaining 99 examples were used for testing on CMU database. We found that the KL-HC-HOG algorithm was stable even when we only selected two bands from each cube for training in Figure 5.

Table 7 shows the accuracy and time consumption of two algorithms in different situations. For the band fusion based algorithm, it should be noted that we selected the 570nm, 640nm, 720nm, 1000nm spectral bands for fusion for the CMU database, and 530nm, 540nm, 550nm, 630nm, 670nm for the Ploy-U database, as in the Uzair's set [3]. We found that despite the fact that the band fusion based algorithm outperforms KL-HC-HOG by 1.9%, and its time cost is a hundred times better than KL-HC-HOG. If we choose 2 examples of each category for training, the accuracy reaches to 99.4% on the CMU database with 20 selected bands and 97.45% on the Ploy-U database with 6 selected bands, while the time consumption is still meager compared to the band fusion based algorithm.

VI. CONCLUSIONS

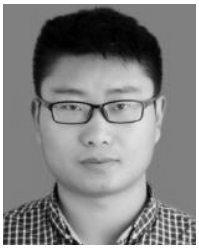
This paper firstly compared four algorithms based on group clustering and information theory, including KL divergence and mutual information, and finally selected the most effective algorithm, which is KL-HC-HOG. Then, we compared this algorithm with the band fusion based algorithm. After fusing all the images into one, we found that all the distinguishing features are the edges feature, which proved that selecting the HOG as the feature extraction method was correct. Besides it, we also found that despite a little decrease in the accuracy, our algorithm is more time-efficient. If we select more training spectral bands, the accuracy exceeds the band fusion based algorithm, reaching to 99.4% and it still costs little time for training and testing. It proved that the algorithm we proposed is faster and easy to be implemented.

In the paper, we emphasized on designing an efficient band selection method to reduce redundant information. In the future, we will first use high performance computing systems to speed up the computations, such as CUDA programming and parallel computing method like message passing interface (MPI). In addition, we will use more advanced machine learning algorithms based on hierarchical clustering methods, and possibly deep learning models for hyperspectral face recognition.

REFERENCES

[1] T. Tian and J. Zhu, "Max-margin majority voting for learning from crowds," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2015, pp. 1621–1629.

[2] Y. Liu, H. Pu, and D.-W. Sun, "Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications," *Trends Food Sci. Technol.*, vol. 69, pp. 25–35, Nov. 2017.
[3] J.-H. Qu, Q. Wei, and D.-W. Sun, "Carbon dots: Principles and their applications in food quality and safety detection," *Crit. Rev. Food Sci. Nutrition*, vol. 58, no. 14, pp. 2466–2475, 2018.
[4] G. Mooradian, M. Weiderhold, A. E. Dabiri, and C. Coyle, "Hyperspectral imaging methods and apparatus for non-invasive diagnosis of tissue for cancer," U.S. Patent 5 782 770 A, Jul. 21, 1998.
[5] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.
[6] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 6, pp. 1354–1361, Nov. 2010.
[7] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.
[8] M. Uzair, A. Mahmood, and A. S. Mian "Hyperspectral face recognition using 3D-DCT and partial least squares," in *Proc. BMVC*, Sep. 2013, pp. 51–57.
[9] L. Shen and S. Zheng, "Hyperspectral face recognition using 3D Gabor wavelets," in *Proc. 21st Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 1574–1577.
[10] G. Chen, C. Li, and W. Sun, "Hyperspectral face recognition via feature extraction and CRC-based classifier," *IET Image Process.*, vol. 11, no. 4, pp. 266–272, Apr. 2017.
[11] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994.
[12] E. Keogh and A. Mueen, "Curse of dimensionality," *Encyclopedia of Machine Learning*, 2010, pp. 257–258.
[13] A. Martìnez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[14] J. Li and D. Jurafsky. (2016). "Mutual information and diverse decoding improve neural machine translation." [Online]. Available: https://arxiv.org/abs/1601.00372
[15] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.
[16] Y. Liao, M. S. Leeson, Q. Cai, Q. Ai, and Q. Liu, "Mutual-information-based incremental relaying communications for wireless biomedical implant systems," *Sensors*, vol. 18, no. 2, p. 515, 2018.
[17] J. Hong-Qi, J. Chuan-Xia, and Z. Rong-Li, "Printing images registration based on mutual information and hybrid algorithm," *Packag. Eng.*, 2018.
[18] L. Karacan, A. Erdem, and E. Erdem, "Image matting with KL-divergence based sparse sampling," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 424–432.
[19] J. Chen, H. Matzinger, H. Zhai, and M. Zhou. (2018). "Centroid estimation based on symmetric KL divergence for Multinomial text classification problem." [Online]. Available: https://arxiv.org/abs/1808.10261
[20] J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.
[21] A. Vinay, V. S. Shekhar, K. N. B. Murthy, and S. Natarajan, "Face recognition using Gabor wavelet features with PCA and KPCA—A comparative study," *Procedia Comput. Sci.*, vol. 57, pp. 650–659, 2015.
[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
[23] B. Ameur, S. Masmoudi, A. G. Derbel, and A. B. Hamida, "Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition," in *Proc. 2nd Int. Conf. Adv. Technol. Signal Image Process.*, Mar. 2016, pp. 453–458.
[24] J. Zhu and T. Tian "Max-margin majority voting for learning from crowds," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2015.
[25] S. S. More and P. P. Gaikwad, "Trust-based Voting Method for Efficient Malware Detection," *Procedia Comput. Sci.*, vol. 79, pp. 657–667, 2016.
[26] X. Mu, P. Watta, and M. H. Hassoun, "Analysis of a plurality voting-based combination of classifiers," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Feb. 2009, pp. 304–309.
[27] L. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," Robotics Inst., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-02-25, Oct. 2002.
[28] J. Schmidhuber, "Reducing the ratio between learning complexity and number of time varying variables in fully recurrent nets," in *Proc. ICANN*, 1993.
[29] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. (Nov. 2014). "Show and tell: A neural image caption generator." [Online]. Available: https://arxiv.org/abs/1411.4555. doi: 10.1109/CVPR.2015.7298935.



QIDONG CHEN received the B.S. degree from the School of IoT, Hainan University, Haikou, China, in 2015. He is currently pursuing the Ph.D. degree in computer science with Jiangnan University, Wuxi, China. He had special insights in the field of large-scale global optimization problems. His current research interests include large-scale optimization problems with surrogate-assisted models, hyperspectral image recognition, and natural language processing. He has recieved the Second Prize of the artificial intelligence application, which is held by Intel and Paratera in 2017.

JUN SUN received the Ph.D. degree in control theory and engineering and the M.Sc. degree in computer science and technology from Jiangnan University, China, in 2009 and 2003, respectively, where he is currently a Full Professor with the Department of Computer Science and Technology. He is also a Vice Director of the Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence. His major research areas and work are related to computational intelligence, machine learning, and bioinformatics. He has published more than 150 papers in journals, conference proceedings, and several books in the above-mentioned areas.

VASILE PALADE (M'02–SM'04) received the Ph.D. degree from the University of Galati, Galaţi, Romania, in 1999. He was a Lecturer with the Department of Computer Science, University of Oxford, Oxford, U.K., between 2004 and 2014. He is currently a Reader of pervasive computing with the Faculty of Engineering, Environment and Computing, Coventry University. He has published more than 100 papers in journals and conference proceedings, and several books. His research interests include computational intelligence with application to bioinformatics, fault diagnosis, and Web usage mining.

XIAOQIAN SHI received the B.S. degree from the School of Information and Computer Science, Anqing Normal College, Anqing, China, in 2015, and the M.S. degree from the School of IoT, Jiangnan University, Wuxi, China, in 2018. She is currently a Lecturer with the Taihu College, Wuxi. She had special insights in the field of hyperspectral image recognition. Her current research interests include large-scale optimization problems with surrogate-assisted models and hyperspectral image recognition.

LI LIU received the B.Sc. degree in computer technology and its applications from Suzhou University, in 1997, the M.Sc. degree in computer science and applications from Jiangnan University, in 2002, and the Ph.D. degree in the area of intelligent computing with applications in pharmacokinetics, in 2008. She is currently a Chief of the Affiliated Hospital, Jiangnan University. Her research interests include image processing and computational modeling for medical applications. She has published more than 50 published journal and conference papers in the above-mentioned areas.