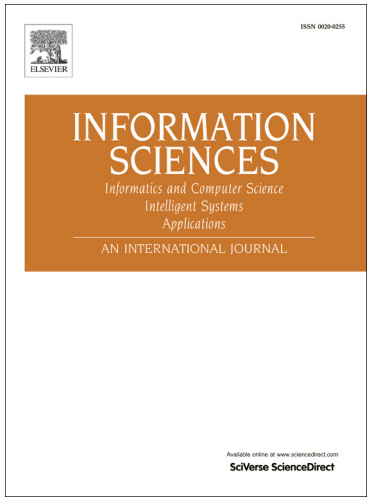
Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework

Zhu, Zexuan; Jia, Sen; He, Shan; Sun, Yiwen; Ji, Zhen; Shen, Linlin

DOI:
10.1016/j.ins.2014.11.045

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhu, Z, Jia, S, He, S, Sun, Y, Ji, Z & Shen, L 2015, 'Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework', *Information Sciences*, vol. 298, pp. 274-287. https://doi.org/10.1016/j.ins.2014.11.045



Three-Dimensional Gabor Feature Extraction for Hyperspectral Imagery Classification Using a Memetic Framework

Zexuan Zhu, Sen Jia, Shan He, Yiwen Sun, Zhen Ji, Linlin Shen

PII: S0020-0255(14)01135-9
DOI: http://dx.doi.org/10.1016/j.ins.2014.11.045
Reference: INS 11282

To appear in: *Information Sciences*

Received Date: 28 May 2014
Revised Date: 29 September 2014
Accepted Date: 30 November 2014

Please cite this article as: Z. Zhu, S. Jia, S. He, Y. Sun, Z. Ji, L. Shen, Three-Dimensional Gabor Feature Extraction for Hyperspectral Imagery Classification Using a Memetic Framework, *Information Sciences* (2014), doi: http://dx.doi.org/10.1016/j.ins.2014.11.045

Three-Dimensional Gabor Feature Extraction for Hyperspectral Imagery Classification Using a Memetic Framework

Zexuan Zhu[a], Sen Jia[a], Shan He[b], Yiwen Sun[c], Zhen Ji[a], Linlin Shen[a,*]

[a]*College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China*
[b]*School of Computer Science, University of Birmingham, B15 2TT, UK*
[c]*Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China*

Abstract

Feature extraction based on three-dimensional (3D) wavelet transform is capable of improving the classification accuracy of hyperspectral imagery data by simultaneously capturing the geometrical and statistical spectral-spatial structure of the data. Nevertheless, the design of wavelets are always proceeded with empirical parameters, which tends to involve a large number of irrelevant and redundant spectral-spatial features and results in suboptimal configuration. This paper proposes a 3D Gabor wavelet feature extraction in a memetic framework, named M3DGFE, for hyperspectral imagery classification. Particularly, the parameter setting of 3D Gabor wavelet feature extraction is optimized using memetic algorithm so that discriminative and parsimonious feature set is acquired for accurate classification. M3DGFE is characterized by an efficient fitness evaluation function and a pruning local search. In the fitness evaluation function, a new concept of redundancy-free relevance based on conditional mutual information is proposed to measure the goodness of the extracted candidate features. The pruning local search is specially designed to eliminate both irrelevant and redundant features without sacrificing the discriminability of the obtained feature subset. M3DGFE is tested on both pixel-level and image-level classification using real-world hyperspectral remote sensing data and hyperspectral face data, respectively. The experimental results show that M3DGFE achieves promising classification accuracy with parsimonious feature subset.

Keywords: Memetic Algorithm, Gabor Wavelet Transform, Feature Extraction, Feature Selection, Hyperspectral Imagery Classification.

1. Introduction

Hyperspectral imaging captures an image of objects with wavelengths ranging from the visible spectrum to the infrared region. The technology has allowed more accurate image classification, object discrimination, and material identification thanks to the availability of rich information on both spectral and spatial distributions of the analyzed targets. However, hyperspectral imagery data usually contain tens and thousands of images simultaneously collected

*Corresponding author. E-mail addresses: llshen@szu.edu.cn

from various spaced spectral bands. When only a limited number of labeled samples are available, it is a great challenge for classification of such data [9]. In addition, noise imposed by sensors and the environment also deteriorates the performance of learning algorithms.

Feature selection and extraction methods have been widely used to address the aforementioned problems. Feature selection [13, 19] selects relevant features and removes noisy/redundant ones in the original feature space, so that the classification accuracy could be improved or not substantially deteriorated. Feature extraction methods [20] transform the given features into other space to generate a new set of features possessing high information packing properties. The most discriminative information is concentrated to relative small number of selected new features with which superior classification accuracy is permitted.

Principle component analysis (PCA) [36], linear discriminant analysis (LDA) [15], and wavelet transform [17] are among the most commonly used feature extraction methods for hyperspectral imagery classification. PCA identifies a subspace where data variances are maximized by orthogonally transforming possibly correlated features into a smaller set of linearly uncorrelated principal components [1, 29]. LDA is related to PCA in that it also tries to find a linear combination of features but explicitly attempts to model the difference between the classes of data [5, 3]. Wavelet transform, in a solid and formal mathematical framework, has attracted increasing attention and served as another powerful solution for feature extraction of hyperspectral imagery classification [26, 51, 41, 12, 24].

Many feature extraction methods mentioned above have been shown to be effective in improving the hyperspectral imagery classification accuracy. However, most of them consider only the spectral signature of each pixel or an individual spectral band, whereas the important spatial information is ignored. Since hyperspectral imagery is naturally a three-dimensional (3D) data cube containing both spatial and spectral dimensions, it is believed that spectral and spatial structures of hyperspectral data should be considered simultaneously to further improve the classification accuracy. In this regard, feature extraction of hyperspectral imagery data should treat the 3D cube as a whole and a few 3D feature extraction methods have been proposed. For instances, Qian et al. [38, 37] proposed a 3D discrete wavelet transform (3D-DWT) to extract spectral-spatial features from hyperspectral remote sensing data. Bau et al. [4] introduced a 3D Gabor filterbank as a tool for extracting spectral-spatial features to represent image regions in hyperspectral region classification. Two authors of this work, i.e., Shen and Jia [45], proposed a 3D Gabor wavelet transform based feature extraction method for hyperspectral imagery classification. Particularly, a set of well-designed Gabor wavelets with different frequencies and orientations was applied to extract signal variances in joint spatial-spectrum domains. We further extended [45] in [46] by introducing a filter-ranking feature selection method based on symmetrical uncertainty and approximate Markov blanket to select discriminative 3D Gabor features. As a result, comparable or better classification accuracy was achieved with much more parsimonious feature set.

On one hand, 3D feature extraction methods have been shown to obtain better classification accuracy than many other state-of-the-art feature selection/extraction methods [38, 37, 4, 45]. On the other hand, 3D feature extraction would generate a larger number of spectral-spatial features, therefore dimension reduction like feature selection is necessarily needed after feature extraction. For example, in [37], a structured sparse logistic regression was applied

after the 3D-DWT extraction to select discriminative spectral-spatial features. A stepwise greedy feature selection was adopted in [4] to identify Gabor filters that optimize the discriminability among different classes. A sequential feature selection and fusion process was developed in [45] to identify the most discriminative Gabor features after 3D Gabor wavelet transform. In [46], a filter-ranking feature selection method was imposed to rank the Gabor features based on their symmetrical uncertainty to the class labels and then the redundant ones were eliminated based on approximate Markov blanket. It is easy to implement the two-phrase feature extraction applied in [37, 4, 45, 46], but it could introduce bias and retain a large number of irrelevant and redundant features in the first phrase where wavelet transform is empirically configured. The greedy or filter-ranking feature selection used in the second phase could also get trapped in local optimal feature sets. Moreover, most of the existing 3D feature extraction methods [38, 37, 4, 45, 46] were targeted at pixel-level classification, but very few are applicable to image-level classification (the differences between pixel-level and image-level classification are described in Section 2.2).

In this study, a novel memetic 3D Gabor wavelet feature extraction namely M3DGFE is proposed for both pixel-level and image-level hyperspectral imagery classification. Particularly, M3DGFE conducts 3D Gabor feature generation and selection simultaneously in a memetic algorithm (MA) framework [32]. The evolutionary search mechanism of MA optimizes both the parameter configurations of 3D Gabor wavelet transform and the selection of feature subset. In this way, Gabor wavelets transform no longer relies on empirical parameter setting and discriminative features can be picked out as desirable signatures for final classification. The performance of M3DGFE is evaluated on both pixel-level and image-level classification using two real-world hyperspectral remote sensing data and one hyperspectral face data, respectively. Comparison studies between M3DGFE and other state-of-the-art feature selection/extraction methods show that M3DGFE obtains superior classification accuracy with compact feature sets.

This work is an extension of our conference paper [54], where the prototype of the framework merely targeted at pixel-level classification was first proposed. Significant improvements have been made in both theory and experiments in this work. Particularly, the memetic framework is extended to handle both pixel-level and image-level classification problems, and a novel fitness function evaluating feature relevance is introduced. Much more extensive experimental results are also provided to demonstrate the efficiency of the proposed method. The main contributions of this study are three-fold:

1) a general memetic framework is proposed for 3D Gabor feature extraction of hyperspectral imagery classification;

2) a novel redundancy-free relevance (RFR) measure is put forward to efficiently evaluate the fitness of candidate feature subsets. RFR enables feature selection methods to identify relevant features and at the same time eliminate irrelevant and redundant features;

3) both pixel-level and image-level classification problems especially with small sample size are studied using various state-of-the-art feature selection/extraction methods, which could provide insights for other researchers facing similar issues.

The remainder of this paper is organized as follows. Section 2 describes the fundamentals of 3D Gabor wavelet feature extraction and Section 3 introduces the proposed memetic 3D Gabor feature extraction framework. Section 4 presents the experimental results of M3DGFE and other compared methods on three hyperspectral imagery datasets. Finally the conclusion is given in Section 5.

2. Three-Dimensional Gabor Wavelet Feature Extraction

Gabor wavelet is closely related to the human visual system and it has been used as a powerful tool to maximize joint time/frequency and space/frequency resolutions for signal analysis [16]. Gabor wavelets have been successfully used to extract features for texture classification [50], face recognition [43], medical image registration [44], etc.

2.1. Three-Dimensional Gabor Wavelet Transform

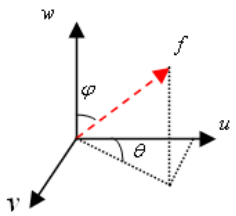


Figure 1: Three-dimensional frequency domain.

In this study, 3D Gabor wavelet transform [44] is applied to hyperspectral image cube to reveal the signal variances in joint spatial-spectrum domains. A circular 3D Gabor wavelet in spatial-spectrum domains (x, y, b) is defined as follows:

$$\Psi_{f,\theta,\varphi}(x,y,b) = \frac{1}{S} \times \exp\left(j2\pi(xu+yv+bw)\right) \times \exp\left(\frac{(x-x_c)^2+(y-y_c)^2+(b-b_c)^2}{-2\sigma^2}\right) \quad (1)$$

where $u = f\sin\varphi\cos\theta$, $v = f\sin\varphi\sin\theta$, and $w = f\cos\varphi$. Variable S is a normalization scale, f is the central frequency of the sinusoidal plane wave, φ and θ are the angles of the wave vector with w-axis and $u-v$ plane in frequency domain (u, v, w) (as shown in Figure 1), and σ is the width of Gaussian envelop in (x, y, b) domain. (x_c, y_c, b_c) is the position for signal analysis.

Let V be a 3D hyperspectral image cube of an $X \times Y$ region captured in B spectral bands. $V(x, y, b)$ is the signal information of a sampled spatial location (x, y) captured in spectral band b. The response of signal to wavelet $\Psi_{f,\theta,\varphi}(x, y, b)$ represents the strength of variance with frequency amplitude f and orientation (φ, θ). The response of $V(x, y, b)$ to $\Psi_{f,\theta,\varphi}(x, y, b)$ is defined as:

$$\Theta_{f,\theta,\varphi}(x,y,b) = \left|(V \otimes \Psi_{f,\theta,\varphi})(x,y,b)\right| \quad (2)$$

where $\otimes$ denotes the convolution operation and $|\cdot|$ calculates the magnitude of the response. $\Theta_{f,\theta,\varphi}(x, y, b)$ reveals the information of signal variances around location (x, y, b) with center frequency f and orientation (θ, φ) at joint spatial-spectrum domains.

2.2. Pixel-Level Classification vs. Image-Level Classification

This work studies 3D Gabor wavelet feature extraction for hyperspectral imagery classification on two levels, i.e., pixel-level and image-level. In pixel-level classification such as material recognition in hyperspectral remote sensing data [37, 45], each pixel in a location (x, y) across all B spectral bands, i.e.,$V(x, y, *)$, is treated as a learning target, and the task of classification is to assign a class label for each pixel. In such case, after 3D Gabor wavelet transform, a pixel at $V(x, y, *)$ yields a set of responses $\{\Theta_{f,\theta,\varphi}(x, y, 1), \Theta_{f,\theta,\varphi}(x, y, 2), \ldots, \Theta_{f,\theta,\varphi}(x, y, B)\}$ to a single 3D Gabor wavelet. By considering together all 3D Gabor wavelets, each pixel sample therefore can be featured by a numeric vector that concatenates the pixel's response sets to all wavelets. A classifier needs to be trained to classify each pixel to different categories, e.g., road, grass and water, based on the numeric feature vector.

In image-level classification like hyperspectral face recognition [14], each 3D hyperspectral cube is considered as a sample of a category and the target is to label a class for the whole image cube. In this case, each pixel on an image can serve as a feature point for classifying the image. After transformed with a single 3D Gabor wavelet, a 3D hyperspectral cube $V(*, *, *)$ of size $X \times Y \times B$ can be represented by a response set containing the responses of all pixels across all bands to the wavelet, i.e., $\{\Theta_{f,\theta,\varphi}(1, 1, 1), \Theta_{f,\theta,\varphi}(1, 1, 2), \ldots, \Theta_{f,\theta,\varphi}(X, Y, B)\}$. Accordingly, when transformed with all 3D Gabor wavelets, a 3D hyperspectral cube sample is represented as the concatenation of all response sets, each of which is subject to a unique wavelet. Pixel-level classification normally studies pixels within a single 3D hyperspectral cube V, whereas image-level classification usually considers a set of 3D hyperspectral cubes say $\{V^1, V^2, \ldots, V^M\}$.

2.3. 3D Gabor Feature

In this study, we define a 3D Gabor feature of all data samples, i.e., pixels or 3D hyperspectral cubes, as a combination of the samples' spectral-spatial property and responses to a specific 3D Gabor wavelet. In the following text, a 3D Gabor feature is denoted as G:

$$G = \begin{cases} \left[\Theta_{f,\theta,\varphi}(1, 1, b), \ldots, \Theta_{f,\theta,\varphi}(x, y, b), \ldots, \Theta_{f,\theta,\varphi}(X, Y, b)\right] & \text{Pixel} - \text{level classification} \\ \left[\Theta^1_{f,\theta,\varphi}(x, y, b), \ldots, \Theta^m_{f,\theta,\varphi}(x, y, b), \ldots, \Theta^M_{f,\theta,\varphi}(x, y, b)\right] & \text{Image} - \text{level classification} \end{cases} \tag{3}$$

where G in pixel-level classification is of length $X \times Y$ and characterized by parameters $\{f, \theta, \varphi, b\}$, whereas in image-level classification G is of length M, i.e., equal to the number of 3D hyperspectral cubes considered, and characterized by parameters $\{f, \theta, \varphi, x, y, b\}$. $\Theta^m_{f,\theta,\varphi}(x, y, b)$ denotes the response of location $V(x, y, b)$ in the m-th cube to the 3D Gabor wavelet.

With appropriately selected parameters, a 3D Gabor feature is capable of capturing the desirable signatures of imagery classification from a specific aspect. The key issue is how to identify the optimal parameters to generate the best 3D Gabor feature set in terms of both classification accuracy and compactness. By searching the space of f, θ, φ, and b with or without (x, y), one can find the solution with satisfactory classification accuracy. The following section will introduce the proposed MA framework for searching the optimal 3D Gabor feature set. Take pixel-level classification for example, the procedure of 3D Gabor feature extraction is illustrated in Figure 2.

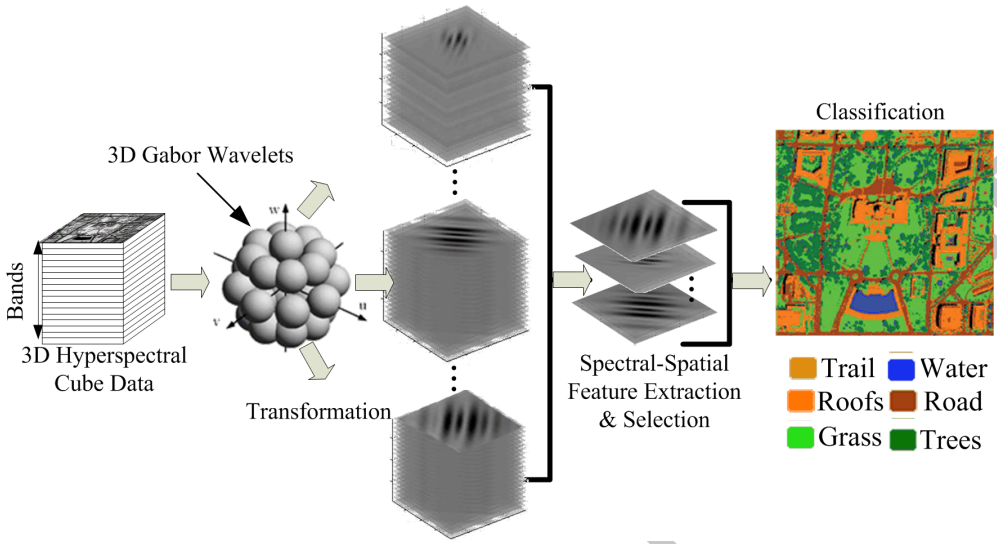


Figure 2: Three-dimensional Gabor wavelet feature extraction for pixel-level hyperspectral imagery classification

3. Memetic Algorithm for 3D Gabor Feature Extraction

Memetic algorithm (MA) [32, 27], the most well-known paradigm of memetic computing [10, 35, 33, 7], is widely recognized as a synergy of population-based global evolutionary algorithm and individual learning or local search heuristic. Taking advantage of both global and local search, MAs are capable of obtaining better performance than their conventional counterparts in various complex real-world search problems such as capacitated arc routing [47, 30], robot control [34], digital IIR filter design [49], protein structure prediction [23], image segmentation [25], and feature selection [52, 53].

In this study, a genetic algorithm (GA) [22] based MA framework is proposed to optimize the 3D Gabor feature extraction for hyperspectral imagery classification. In this framework, the parameters used to generate 3D Gabor features are optimized with GA based global search and a pruning local search is introduced to fine-tune the GA solutions, especially by eliminating both irrelevant and redundant features. The proposed framework named M3DGFE is outlined in Algorithm 1 and more details of it are provided in the following subsections.

3.1. Chromosome Encoding

At the beginning of M3DGFE, a population of chromosomes is randomly generated with each chromosome encoding a set of candidate 3D Gabor features. A chromosome (as shown in Figure 3) is designed as a string of n 4-tuple (i.e., $\{f_i, \theta_i, \varphi_i, b_i\}$ for pixel-level classification) or 6-tuple (i.e., $\{f_i, \theta_i, \varphi_i, x_i, y_i, b_i\}$ for image-level classification) genes, each of which can be used to generate a corresponding 3D Gabor feature based on Eqs. (1), (2), and (3). The length of a chromosome is variable when it undergoes local search and crossover operation.

In each gene, f_i is in $[0, 0.5]$; θ_i and φ_i take real values in $[0, \pi]$; x_i, y_i, and b_i indicate the spectral position of the pixel in the 3D hyperspectral cube. The search space of 3D Gabor features is intrinsically continuous. Nevertheless,

Algorithm 1 The procedure of M3GDFE

BEGIN

1: Randomly initialize a population of chromosomes encoding parameters for generating 3D Gabor features;
2: **While** stopping criteria are not satisfied **do**
3: Generate 3D Gabor features based on the parameters encoded in each chromosome using Eqs. (1), (2), and (3);
4: Evaluate the fitness of each chromosome in the population based on Eq. (9);
5: Perform **Pruning Local Search** on each chromosome to eliminate both irrelevant and redundant features;
6: Evolve the population based on selection, crossover and mutation operators;
7: **End While**

END

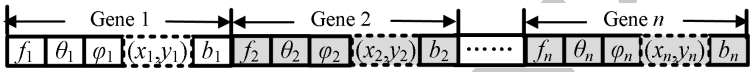


Figure 3: A 3D Gabor feature chromosome.

features characterized by similar wavelet parameters are redundant for capturing similar signatures. Therefore, one can sample f, θ, and φ in certain intervals to reduce the redundancy of the extracted features and meanwhile significantly reduce the search space. Particularly, the frequency f takes the values of $[0.5, 0.25, 0.125, 0.0625]$ with orientations θ and φ set as the values of $[0, \pi/4, \pi/2, 3\pi/4]$ in this study. Since the frequency vector points to the same direction with different θ when $\varphi = 0$, there are in total 52 wavelets available for feature extraction. Let B denotes the total number of bands. Each pixel in a 3D hyperspectral cube is represented with $52B$ 3D Gabor features and the complexities of searching the optimal feature set for pixel-level and image-level classification are 2^{52B} and 2^{52XYB}, respectively.

The feature space of image-level classification could be much larger than that of pixel-level classification. Nonetheless, since pixels in a local region are likely similar to each other, downsampling of representative pixels in a local region could be used in image-level classification to substantially reduce the feature space while maintaining good classification accuracy. Yet, the feature space of size 2^{52B} or 2^{52XYB} still poses a big challenge for most search algorithms. M3DGFE is designed to handle the problem by capitalizing on the merits of GA based global search and pruning local search.

3.2. Fitness Evaluation

After the initialization, the chromosome population evolves until some predefined stopping criterion is satisfied. The stopping criterion could be a convergence to the global optimal or a maximum computational budget is reached. In each evolution generation, the goodness of the 3D Gabor feature set encoded in each chromosome should be evaluated based on a fitness function.

Since the final objective is to classify pixels/images, classification accuracy should be the first choice for chromosome fitness evaluation. However, the evaluation of classification accuracy based on a classifier could be very time

consuming especially when evolutionary algorithms like GA and MA need thousands of fitness evaluations to reach a satisfying solution. Instead of using classification accuracy, we propose a novel criterion namely redundancy-free relevance (or RFR for short) for chromosome fitness evaluation. Particularly, RFR measures the relevance of the encoded features in a chromosome to the class labels, excluding the redundancy detected among the features. It is a computationally efficient measure to approximate the classification accuracy.

Before introducing the definition of RFR, the preliminary knowledge of feature relevance and redundancy is presented as follows. Let C be a vector of the class labels of the data samples, and G_i be the 3D Gabor feature encoded by the i-th gene, i.e., $\{f_i, \theta_i, \varphi_i, b_i\}$ or $\{f_i, \theta_i, \varphi_i, x_i, y_i, b_i\}$, of a given chromosome $\mathfrak{X}$. The relevance of G_i to C is measured in terms of mutual information $I(G_i; C)$ [42]:

$$I(G_i; C) = H(G_i) - H(G_i|C) \tag{4}$$

where $H(G_i)$ is the entropy of G_i, and $H(G_i|C)$ denotes the conditional entropy of G_i given C. $H(G_i)$ and $H(G_i|C)$ are defined as follows:

$$H(G_i) = -\sum_{g \in G_i} p(g) \log p(g) \tag{5}$$

$$H(G_i|C) = -\sum_{c \in C} \sum_{g \in G_i} p(g, c) \log p(g|c) \tag{6}$$

where $p(g)$ is the probability mass function of g, $p(g, c)$ is the joint probability mass function of g and c, and $p(g|c)$ is the conditional probability mass function of g given c.

The redundancy between two features is measured based on conditional mutual information. Given two features G_i and G_j encoded in $\mathfrak{X}$, the conditional mutual information between G_i and C given G_j is defined as:

$$I(G_i; C|G_j) = H(G_i|G_j) - H(G_i|C, G_j) \tag{7}$$

where the calculations of $H(G_i|G_j)$ and $H(G_i|C, G_j)$ are similar to Eq. (6). $I(G_i; C|G_j)$ measures the conditional information shared by G_i and C given G_j. So, if $I(G_i; C|G_j) < \epsilon$, where ϵ is a small constant, G_i gives little discriminatory information of C in the existence of G_j. If $I(G_j; C) > I(G_i; C)$ and $I(G_i; C|G_j) < \epsilon$ both hold true, G_i is of very low relevance to C or high redundancy to G_j. In either case, G_i's relevance to C could be ignored if G_j has already been included in $\mathfrak{X}$.

Based on the definitions of feature relevance and redundancy, the RFR of G_i is defined as follows:

$$\mathrm{RFR}(G_i) = \begin{cases} 0 & \text{if } \exists G_j,\ I(G_j; C) > I(G_i; C) \text{ and } I(G_i; C|G_j) < \epsilon \\ I(G_i; C) & \text{otherwise} \end{cases} \tag{8}$$

The relevance of a feature is counted only if it is not redundant to any other features. The fitness of the chromosome $\mathfrak{X}$ is thus calculated by summing up the RFR of all encoded 3D Gabor features:

$$\mathrm{Fitness}(\mathfrak{X}) = \sum_{i=1}^{|\mathfrak{X}|} \mathrm{RFR}(G_i) \tag{9}$$

where $|\mathfrak{X}|$ denotes the number of genes contained in $\mathfrak{X}$. The fitness evaluated in Eq. (9) is expected as a good estimation of the classification accuracy. The more RFR is captured in $\mathfrak{X}$, the more accurate classification could be achieved. Not fitting the classification accuracy directly, RFR-based fitness evaluation can also alleviate the overfitting [39] or selection bias [2] problem especially when the training sample size is small.

3.3. Pruning Local Search

Algorithm 2 The procedure of pruning local search

INPUT: a chromosome $\mathfrak{X}$;

BEGIN

1: **For** $i = 1$ to $|\mathfrak{X}| - 1$ **do**
2: **For** $j = i + 1$ to $|\mathfrak{X}|$ **do**
3: **If** $I(G_i; C) > I(G_j; C)$ and $I(G_j; C|G_i) < \epsilon$ **then**
4: Remove the j-th gene from $\mathfrak{X}$;
5: **Else If** $I(G_j; C) > I(G_i; C)$ and $I(G_i; C|G_j) < \epsilon$ **then**
6: Remove the i-th gene from $\mathfrak{X}$;
7: Continue line 1;
8: **End If**
9: **End For**
10: **End For**

END

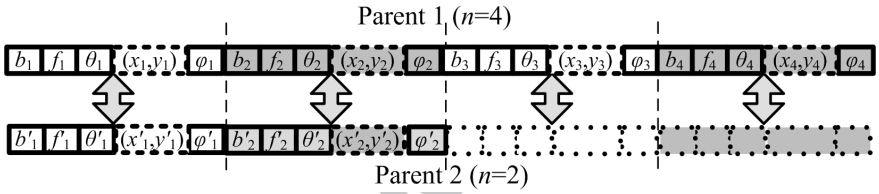
Since only the relevant but not redundant features contribute to the fitness of a chromosome, both irrelevant and redundant 3D Gabor features encoded in each chromosome can be removed for the sake of reducing computational complexity. After fitness evaluation, each chromosome undergoes a local search to prune both irrelevant and redundant features. The procedure of the pruning local search is outlined in Algorithm 2, where features are checked pairwise every time and those of low relevance or high redundancy are removed. In practice, irrelevant and redundant features can be identified when computing RFR, so the pruning local search can be done at the same time as fitness evaluation, i.e., the pruning local search causes very little extra computational cost.

3.3.1. Evolutionary Operators

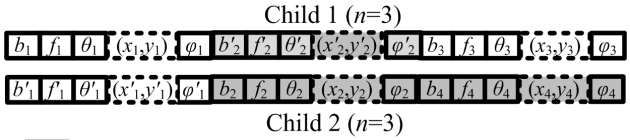
Following the fitness evaluation and local search, the population is evolved using evolutionary operators including linear ranking selection, uniform crossover, and mutation. Here, it is notable that the uniform crossover is performed on each 4-tuple or 6-tuple gene rather than on each component of a gene to ensure the consistency of the encoded 3D Gabor features during the crossover. Moreover, because two parent chromosomes could have different lengths, an appending operation is imposed to make their length equal so that the conventional crossover is applicable. For

example, as shown in Figure 4, given two parent chromosomes of four and two genes respectively, two dummy genes are appended to the shorter one so that the two parent chromosomes have the same length. Afterward, the uniform crossover is applied on each gene and then the dummy genes are removed, resulting in two children chromosomes each of three genes. The crossover also has to guarantee the uniqueness of the genes in every children chromosome. The details of the crossover operator are summarized in Algorithm 3. After crossover, conventional mutation operator is applied to the children chromosomes by randomly changing each position of a chromosome at a predefined mutation rate .

Unlike many other feature selection/extraction methods that require a predefined number of selected features, M3DGFE dynamically varies the number of selected features encoded in each chromosome during the pruning local search and crossover operation. The final number of selected features is determined automatically in the best chromosome in terms of fitness value.



(a) Two parent chromosomes before crossover



(b) Two children chromosomes after crossover

Figure 4: An example of crossover.

4. Experiments

The performance of M3DGFE is evaluated on both pixel-level and image-level classification using three real-world hyperspectral imagery datasets.

4.1. Pixel-Level Classification of Hyperspectral Remote Sensing Imagery Data

4.1.1. Datasets

In many real-world applications of hyperspectral remote sensing imagery data, pixel classification is an important task for terrains/objects identification. In this experiment, the proposed M3DGFE is applied to the pixel-level classification of two most widely used hyperspectral remote sensing imagery datasets namely Indiana pines AVIRIS (Indiana) [6] and Kennedy Space Center (KSC) [31, 45].

Algorithm 3 The procedure of crossover

INPUT: two parent chromosomes $\mathfrak{X}1$ and $\mathfrak{X}2$;

OUTPUT: two children chromosomes $\mathfrak{C}1$ and $\mathfrak{C}2$;

BEGIN

Set $\mathfrak{C}1 = \mathfrak{X}1$, $\mathfrak{C}2 = \mathfrak{X}2$, and $k = \max(|\mathfrak{C}1|, |\mathfrak{C}2|)$;
Append $k - |\mathfrak{C}1|$ and $k - |\mathfrak{C}2|$ dummy genes to $\mathfrak{C}1$ and $\mathfrak{C}2$, respectively;
For $i = 1$ to k **do**
    Generate a random number r in $[0, 1]$;
    If $r < 0.5$ and $\mathfrak{C}1_i \notin \mathfrak{C}2$ and $\mathfrak{C}2_i \notin \mathfrak{C}1$ **then** {//$\mathfrak{C}1(2)_i$ indicates the i-th gene in $\mathfrak{C}1(2)$}
        Exchange $\mathfrak{C}1_i$ and $\mathfrak{C}2_i$;
    End If
End For
Remove all dummy genes from $\mathfrak{C}1$ and $\mathfrak{C}2$;

END

The Indiana data represent a section of a scene taken over northwest Indiana's Indiana Pines by the AVIRIS sensor in 1992. It contains 10366 pixels, 220 bands, and 16 classes. The KSC images were acquired over the KSC, Florida, on March 23, 1996 using NASA's airborne visible infrared imaging spectrometer (AVIRIS). In the original 224 bands, 48 bands (numbered 1-4, 102-116, 151-172, and 218-224) are identified as water absorption and low SNR bands, leaving 176 spectral bands for classification. Following [45], eight classes representing various land cover types are defined. The information of the two datasets is summarized in Tables 1 and 2, respectively.

4.1.2. Experimental Design

Classification using ALL feature bands is involved as the baseline and the other four state-of-the-art feature selection/extraction methods including ReliefF [40], Mutual Information (MI) based filter ranking method [18], Principle Component Analysis (PCA) [20], and the 3D-DWT method proposed in [37] are considered for comparison study. M3DGFE is also compared to the counterpart GA-based 3D Gabor feature extraction (G3DGFE), i.e., M3DGFE without the pruning local search, to test the effect of local search. G3DGFE and M3DGFE use the same parameter configurations with population size 50, crossover probability 0.6, and mutation rate 0.1. The maximum number of genes in each chromosome n is empirically set to 100. Both G3DGFE and M3DGFE are stopped when a maximum iteration number of 100 or a search convergence is reached. As stated in Section 3.3, the pruning local search consumes very little computational cost, therefore the computational budgets of G3DGFE and M3DGFE are nearly equivalent to each other.

Unlike G3DGFE and M3DGFE, whose number of selected features are not deterministic, ReliefF and MI need a predefined number of selected features. To make the comparison fair, ReliefF and MI are set to select the same

Table 1: Information of classes and samples of Indiana data

Class	Land Cover Type	#Samples(pixels)
C1	Stone-Steel-Towers	95
C2	Hay-windrowed	489
C3	Corn-mintill	834
C4	Soybean-notill	968
C5	Alfalfa	54
C6	Soybean-clean	614
C7	Grass-pasture	497
C8	Woods	1294
C9	Buildings-Grass-Trees-Drives	380
C10	Grass-pasture-mowed	26
C11	Corn	234
C12	Oats	20
C13	Corn-notill	1434
C14	Soybean-mintill	2468
C15	Grass-trees	747
C16	Wheat	212

Table 2: Information of classes and samples of KCS data

Class	Land Cover Type	#Samples(pixels)
C1	Willow swamp	108
C2	Cabbage palm/oak hammock	132
C3	Slash pine	163
C4	Oak/broadleaf hammock	74
C5	Hardwood swamp	248
C6	Water	330
C7	Spartina marsh	181
C8	Citrus	142

number of features as M3DGFE. For PCA, the dimensionality reduction is accomplished by empirically choosing enough eigenvectors to account for 95% of the variance in the original data. It is noted that the methods deal with different types of features. Particularly, the feature selection methods ReliefF and MI select spectral bands in the

original feature space. The feature extraction methods PCA and 3D-DWT apply orthogonal linear transform and 3D discrete wavelet transform to the original data, respectively, and then select eigenvectors and 3D-DWT features accordingly. Whereas both G3DGFE and M3DGFE extract and select 3D Gabor features. Despite acquiring different types of features, the performance of all compared methods can be objectively evaluated in terms of classification accuracy.

All methods are challenged with small sample size. In each run, only 5% randomly sampled pixels of a dataset are used for training and the remaining unseen 95% pixels for test. The average classification accuracy and the number of selected features of each method are reported over 30 independent runs. The classification accuracy is evaluated using K-Nearest-Neighbor (KNN) [11] with $K = 1$ and Support Vector Machine (SVM) [48] with linear kernel for each method. Exceptionally, following [37], the sparse group lasso (SGLasso) feature selection and classification method is used for 3D-DWT. Here, SVM is implemented with LIBSVM [8] and the default parameter setting is adopted. ReliefF, MI, PCA, and KNN are all implemented in Weka environment [21].

4.1.3. Experimental Results

The classification accuracy of each data class, the overall accuracy (OA), and the number of selected features of all methods on Indiana and KSC data are summarized in Tables 3 and 4, respectively. Wilcoxon rank-sum test[1] [28] at a 0.05 significant level is performed between M3DGFE and each of ALL, MI, ReliefF, PCA, 3D-DWT, and G3DGFE. The results show that M3DGFE and G3DGFE obtain significantly better overall accuracy than the other methods using both KNN and SVM, which suggests that the 3D Gobor feature extraction framework indeed can capture or even enhance the desirable signatures for pixel classification. The performance of M3DGFE and G3DGFE is competitive, but M3DGFE manages to obtain more compact 3D Gabor feature set. The pruning local search used in M3DGFE plays a key role in eliminating both irrelevant and redundant features while maintaining or improving the classification accuracy. The band selection methods MI and ReliefF fail to improve the classification accuracy with respect to the baseline performance using all bands. PCA is inferior to the other methods on these two datasets.

To show the discriminative ability of the selected features visually, we take Indiana data for example and plot the prediction results of the whole image in Figure 5 based on the features obtained on 5% training pixels. It is consistent with the observation in Tables 3 and 4 that G3DGFE and M3DGFE predict more accurately than the other methods.

The above results have demonstrated that 3D Gabor features are superior to capture the geometrical and statistical spectral-spatial structure of hyperspectral remote sensing data, and thus leading to improved classification accuracy. To evaluate the efficiency of the memetic framework, we also pit M3DGFE against the greedy 3D Gabor features selection and fusion method (Fused-Gabor for short) [45] as well as the filter-ranking Gabor feature selection method (Filtered-Gabor for short) [46]. The Fused-Gabor method first applies a predefined set of 3D Gabor wavelets to the imagery cube to extract 3D Gabor features, and then a greedy feature selection method kicks in, where a group of

[1]The p-values of Wilcoxon rank-sum tests performed in Tables 3 to 6 are provided in http://csse.szu.edu.cn/staff/zhuzx/M3DGFE/p-values.pdf

3D Gabor features is iteratively added to the candidate selected feature set so that the classification accuracy is best improved. The selection procedure is repeated until no improvement can be achieved. Finally, all selected 3D Gabor features are fused for the final classification. The Filtered-Gabor method is different from the Fused-Gabor method in that a filter-ranking feature selection method is introduced to replace the greedy feature selection and fusion method. In particular, the extracted 3D Gabor features are ranked and sorted based on their symmetrical uncertainty to the class labels. Afterward, the irrelevant and redundant features are eliminated using approximate Markov blanket.

The results of M3DGFE, Fused-Gabor, and Filtered-Gabor in terms of number of selected features, overall classification accuracy, and time cost of classification using the corresponding selected features are tabulated in Table 5. M3DGFE is shown to attain better classification accuracy with KNN, whereas Fused-Gabor and Filtered-Gabor win with SVM. Fused-Gabor and Filtered-Gabor tend to select more irrelevant and redundant features than M3DGFE, which deteriorates the performance of KNN since it is based on noise-susceptible Euclidean distance. By contrast, SVM is more robust against noise and thus can benefit from the involvement of more features that could provide more discriminative information in proportion. Accordingly, Fused-Gabor and Filtered-Gabor obtain more accurate classification than M3DGFE with SVM. Overall, the classification accuracy of M3DFE is comparable to the other two methods regarding both KNN and SVM, yet M3DGFE manages to reduce the number of selected features substantially, and thus leading to much less classification time. It is also worth noting that both Fused-Gabor and Filtered-Gabor need to generate a large number of 3D Gabor features before feature selection kicks in, which inevitably boosts the memory or disk space consumption. Using the same parameter values of $\{f, \theta, \varphi, b\}$ as described in Section 3.1, Fused-Gabor and Filtered-Gabor end up with space complexity $\mathbf{O}(52XYB)$, whereas the space complexity of M3DGFE is merely $\mathbf{O}(nPB)$, where n is the maximum number of genes in a chromosome, P is the population size, and normally nP is much smaller than $52XY$.

4.2. Image-Level Classification of Hyperspectral Face Data

4.2.1. Dataset and Experimental Design

The publicly available HK-PolyU Hyperspectral Face Database [14] is used to evaluate the performance of M3GDFE on image-level classification. The multi-spectrum data were obtained by using a CRI's VariSpec LCTF to filter the light with wavelength less than 400nm and greater than 720nm. The spectral range produces 33 bands in all with a step length of 10nm. The face images were captured from 48 young volunteers (13 females and 35 males) at different sessions. In this experiment, we follow [14] to use the frontal hyperspectral images of 25 individuals (each of four 3D hyperspectral cubes). The eye coordinates were manually located and each face was cropped and rotated with reference to the eye locations and resized to size 64×64. Figure 6 shows 32 bands of an example hyperspectral face.

According to [14], the first six and last three bands can be excluded due to the high noise, and two hemoglobin absorption bands around 540 and 580nm should be selected to better describe the skin characteristics. Particularly, the two band subsets, one consists of bands at 530, 540 and 550 nm, the other contains bands at 570, 580 and 590 nm, were

suggested in [14]. In this study, only the three bands at 530, 540 and 550 nm are considered for the sake of reducing computational complexity. As such, each sample face is characterized by 64×64 feature pixels and three feature bands, i.e., totally 12288 features. In line with the study on hyperspectral remote sensing data, the performance of M3GDFE on hyperspectral face data is compared with that of ALL, MI, ReliefF, PCA, and G3GDFE. Since there are four samples for each individual (one 3D hyperspectrial cube captures one sample face), a four-fold cross validation scheme is used to evaluate the performance of the methods. Twenty independent runs of four-fold cross validation are carried out to estimate the average classification accuracy of all methods with both KNN and SVM. The best classification accuracy reported in [14] using BS-WFD method is also included for comparison.

4.2.2. Experimental Results

The results of the compared methods on hyperspectral face data are reported in Table 6. To test the significant differences of the methods' performance, Wilcoxon rank-sum test at a 0.05 significant level is performed between M3DGFE and each of ALL, MI, ReliefF, PCA, and G3DGFE. Similar to the results of hyperspectral remote sensing data, M3DGFE is observed to significantly outperform ALL, MI, ReliefF, and PCA. G3DGFE and M3DGFE obtain similar classification accuracy, but M3DGFE selects significantly fewer features. M3DGFE also shows obviously higher accuracy than BS-WFD. Comparing the results of this study directly with that published in [14] may not be appropriate due to the different preprocessing, classifier, and performance evaluation schema used. Yet the comparison confirms the efficiency of the proposed method to a certain degree.

To see whether M3DGFE can identify meaningful feature pixels, we also plot the 30 most frequently selected locations in spatial domain in Figure 7. It is shown that important facial points like eye corners, eyebrows, and mouth are included. It is interesting to see that most of the selected locations distribute on one side of the face. The reason for this observation could be the use of RFR-based fitness evaluation. As the front face has relatively symmetric patten, the important points on the one side of the face could be redundant to the counterparts on the other side. Accordingly, the points on the side of better discriminative quality, probably thanks to better lighting and/or shooting angle, are preferred for classification.

5. Conclusion

In this paper, a 3D Gabor feature extraction based on memetic algorithm (M3DGFE) is proposed for hyperspectral imagery classification. Particularly, M3DGFE optimizes the 3D Gabor wavelet transform based feature generation and selection such that desirable 3D Gabor features capturing the signal variances in joint spatial-spectrum domains are identified and the classification accuracy is improved. Both pixel-level and image-level classification problems of small sample size are studied and various state-of-the-art feature selection/extracton methods are involved in comparison with M3DGFE. The experimental results on real-world hyperspectral imagery datasets demonstrate that M3DGFE is efficient in identifying discriminative features and eliminating irrelevant and redundant ones. It is expected to serve as a competitive alterative for solving the increasing complicate hyperspectral imagery classification problems.

6. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China, under grants 61471246, 61171125, 61272050, and 61205092, the NSFC-RS joint project under grants IE111069, the Guangdong Foundation of Outstanding Young Teachers in Higher Education Institutions under grant Yq2013141, Shenzhen Scientific Research and Development Funding Program under grants JCYJ20130329115450637, KQC201108300045A, and ZYC201105170243A, Guangdong Natural Science Foundation under grant S2012010009545, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, MOE of China, under grand 20111568.

References

[1] A. Agarwal, T. El-Ghazawi, H. El-Askary, J. Le-Moigne, Efficient hierarchical-pca dimension reduction for hyperspectral imagery, in: 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 353–356.

[2] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proceedings of the National Academy of Sciences 99 (2002) 6562–6566.

[3] T.V. Bandos, L. Bruzzone, G. Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, IEEE Transactions on Geoscience and Remote Sensing 47 (2009) 862–873.

[4] T.C. Bau, S. Sarkar, G. Healey, Hyperspectral region classification using a three-dimensional gabor filterbank, IEEE Transactions on Geoscience and Remote Sensing 48 (2010) 3457–3464.

[5] A. Berge, A.C. Jensen, A.H.S. Solberg, Sparse inverse covariance estimates for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 45 (2007) 1399–1407.

[6] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, J. Calpe-Maravilla, Composite kernels for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 3 (2006) 93C97.

[7] F. Caraffini, F. Neri, L. Picinali, An analysis on separability for memetic computing automatic design, Information Sciences 265 (2014) 1–22.

[8] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[9] C.I. Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification, Kluwer Academic/Plenum Publishers, New York, 2003.

[10] X. Chen, Y.S. Ong, M.H. Lim, K.C. Tan, A multi-facet survey on memetic computation, IEEE Transactions on Evolutionary Computation 15 (2011) 591–607.

[11] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1967) 21–27.

[12] A. Daamouche, F. Melgani, Swarm intelligence approach to wavelet design for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 6 (2009) 825–829.

[13] M. Dash, H. Liu, Feature selection for classification, Intelligent Data Analysis 1 (1997) 131–156.

[14] W. Di, L. Zhang, D. Zhang, Q. Pan, Studies on hyperspectral face recognition in visible spectrum with feature band selection, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40 (2010) 1354–1361.

[15] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of eugenics 7 (1936) 179–188.

[16] D. Gabor, Theory of communication. part 1: The analysis of information, Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering 93 (1946) 429–441.

[17] A. Grossmann, J. Morlet, Decomposition of hardy functions into square integrable wavelets of constant shape, SIAM journal on mathematical analysis 15 (1984) 723–736.

[18] B. Guo, S.R. Gunn, R. Damper, J. Nelson, Band selection for hyperspectral image classification using mutual information, IEEE Geoscience and Remote Sensing Letters 3 (2006) 522–526.

[19] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.

[20] I. Guyon, S. Gunn, M. Nikravesh, L. Zadeh, Feature Extraction, Foundations and Applications, Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, 2006.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (2009) 10–18.

[22] J.H. Holland, Adaptation in Natural Artificial Systems. 2nd ed., MIT Press, Cambridge MA, 1992.

[23] M.K. Islam, M. Chetty, Clustered memetic algorithm with local heuristics for ab initio protein structure prediction, IEEE Transactions on Evolutionary Computation 17 (2013) 558–576.

[24] S. Jia, Y. Qian, J. Li, W. Liu, Z. Ji, Feature extraction and selection hybrid algorithm for hyperspectral imagery classification, in: 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 72–75.

[25] L. Jiao, M. Gong, S. Wang, B. Hou, Natural and remote sensing image segmentation using memetic computing, IEEE Computational Intelligence Magazine 5 (2010) 78–91.

[26] P. Kempeneers, S. De Backer, W. Debruyn, P. Coppin, P. Scheunders, Generic wavelet-based hyperspectral classification applied to vegetation stress detection, IEEE Transactions on Geoscience and Remote Sensing 43 (2005) 610–614.

[27] N. Krasnogor, Studies on the theory and design space of memetic algorithms, Ph.D. thesis, University of the West of England, Bristol, U.K., 2002.

[28] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, The Annals of Mathematical Statistics 18 (1947) 50–60.

[29] F. Mathieu, C. Jocelyn, B. Jón Atli, et al., Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas, EURASIP Journal on Advances in Signal Processing 2009 (2009) 1–14.

[30] Y. Mei, K. Tang, X. Yao, A memetic algorithm for periodic capacitated arc routing problem, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 41 (2011) 1654–1667.

[31] J.T. Morgan, Adaptive Hierarchical Classifier with Limited Training Data, Ph.D thesis, University of Texas at Austin, 2002.

[32] P. Moscato, Memetic Algorithm: A Short Introduction. In New Ideas in Optimization, McGraw-Hill, London, 1999.

[33] F. Neri, C. Cotta, Memetic algorithms and memetic computing optimization: A literature review, Swarm and Evolutionary Computation 2 (2012) 1–14.

[34] F. Neri, E. Mininno, Memetic compact differential evolution for cartesian robot control, IEEE Computational Intelligence Magazine 5 (2010) 54–65.

[35] Y.S. Ong, M.H. Lim, X.S. Chen, Research frontier: Memetic computation - past, present & future, IEEE Computational Intelligence Magazine 5 (2010) 24–31.

[36] K. Pearson, Liii. on lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (1901) 559–572.

[37] Y. Qian, M. Ye, J. Zhou, Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features, IEEE Transactions on Geoscience and Remote Sensing 51 (2012) 2276–2291.

[38] Y. Qian, J. Zhou, M. Ye, Q. Wang, Structured sparse model based feature selection and classification for hyperspectral imagery, in: 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1771–1774.

[39] J. Reunanen, Overfitting in making comparisons between variable selection methods, The Journal of Machine Learning Research 3 (2003) 1371–1382.

[40] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of relieff and rrelieff, Machine learning 53 (2003) 23–69.

[41] F. Schmidt, S. Douté, B. Schmitt, Wavanglet: an efficient supervised classifier for hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 45 (2007) 1374–1385.

[42] C.E. Shannon, W. Weaver, A mathematical theory of communication, American Telephone and Telegraph Company, 1948.

[43] L. Shen, L. Bai, A review on gabor wavelets for face recognition, Pattern Analysis and Applications 9 (2006) 273–292.

[44] L. Shen, L. Bai, 3d gabor wavelets for evaluating spm normalization algorithm, Medical Image Analysis 12 (2008) 375.

[45] L. Shen, S. Jia, Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification, IEEE Transactions on Geoscience and Remote Sensing 49 (2011) 5039–5046.

[46] L. Shen, Z. Zhu, S. Jia, J. Zhu, Y. Sun, Discriminative gabor feature selection for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters 10 (2013) 29–33.

[47] K. Tang, Y. Mei, X. Yao, Memetic algorithm with extended neighborhood search for capacitated arc routing problems, IEEE Transactions on Evolutionary Computation 13 (2009) 1151–1166.

[48] V. Vapnik, The nature of statistical learning theory, springer, 1999.

[49] Y. Wang, B. Li, T. Weise, Two-stage ensemble memetic algorithm: Function optimization and digital iir filter design, Information Sciences 220 (2013) 408–424.

[50] T.P. Weldon, W.E. Higgins, D.F. Dunn, Efficient gabor filter design for texture segmentation, Pattern Recognition 29 (1996) 2005–2015.

[51] X. Zhang, N.H. Younan, C.G. O'Hara, Wavelet domain statistical hyperspectral soil texture classification, IEEE Transactions on Geoscience and Remote Sensing 43 (2005) 615–618.

[52] Z. Zhu, S. Jia, Z. Ji, Towards a memetic feature selection paradigm, IEEE Computational Intelligence Magazine 5 (2010) 41–53.

[53] Z. Zhu, Y.S. Ong, M. Zurada, Simultaneous identification of full class relevant and partial class relevant genes, IEEE-ACM Transactions on Computational Biology Bioinformatics 7 (2010) 263–277.

[54] Z. Zhu, L. Shen, Y. Sun, S. He, Z. Ji, Memetic three-dimensional gabor feature extraction for hyperspectral imagery classification, in: Advances in Swarm Intelligence, Springer, 2012, pp. 479–488.

Table 3: Classification accuracy and selected feature size of the compared methods on Indiana data

		ALL	MI	ReliefF	PCA	3D-DWT*	G3DGFE	M3DGFE
C1	KNN	78.31±15.22$^{\approx}$	76.94±16.87$^{\approx}$	48.31±18.11^{-}	72.03±15.16$^{\approx}$	**86.00±8.32**$^{+}_{+}$	72.76±23.99$^{\approx}$	71.53±25.81
	SVM	81.72±16.77^{+}	78.25±17.78$^{\approx}$	71.47±19.85$^{\approx}$	77.61±15.18$^{\approx}$		71.15±26.92$^{\approx}$	69.36±25.77
C2	KNN	92.59±4.64^{+}	91.75±5.69^{-}	92.63±4.52^{-}	61.38±8.17^{-}	**98.86±0.64**$^{\approx}_{+}$	98.09±2.04$^{\approx}$	98.36±1.93
	SVM	94.90±3.65^{-}	94.62±3.91^{-}	96.25±3.44$^{\approx}$	93.87±6.17^{-}		97.46±3.86$^{\approx}$	97.14±3.55
C3	KNN	40.17±3.45^{-}	54.18±4.30^{-}	50.58±3.98^{-}	17.71±3.13^{-}	63.96±5.25$^{-}_{-}$	**92.00±2.93**$^{\approx}$	91.86±3.20
	SVM	53.50±4.91^{-}	44.16±7.89^{-}	42.24±8.53^{-}	2.10±4.40^{-}		87.91±5.43^{-}	90.73±4.68
C4	KNN	46.76±6.07^{-}	60.23±6.18^{-}	48.69±6.12^{-}	22.92±4.50^{-}	71.99±3.08$^{-}_{-}$	91.12±2.83$^{\approx}$	**91.92±2.50**
	SVM	57.15±6.08^{-}	52.99±7.90^{-}	40.91±7.45^{-}	7.58±10.17^{-}		86.99±4.38$^{\approx}$	86.05±5.12
C5	KNN	15.56±11.26^{-}	26.85±15.59^{-}	24.74±13.37^{-}	5.34±5.97^{-}	45.69±13.65$^{-}_{-}$	**83.61±24.96**$^{\approx}$	83.61±25.13
	SVM	28.32±18.08^{-}	37.86±22.07^{-}	36.88±23.08^{-}	0.00±0.00^{-}		74.26±26.42$^{\approx}$	77.51±24.74
C6	KNN	28.86±6.21^{-}	43.13±6.52^{-}	37.08±6.41^{-}	13.60±2.95^{-}	61.94±3.98$^{-}_{-}$	**88.99±4.62**$^{\approx}$	88.58±5.24
	SVM	46.83±7.01^{-}	57.96±10.28^{-}	61.90±8.52^{-}	0.58±1.73^{-}		80.76±8.38$^{\approx}$	81.76±7.53
C7	KNN	66.51±6.15^{-}	69.55±6.82^{-}	63.23±6.59^{-}	23.60±5.35^{-}	89.81±3.16$^{\approx}_{\approx}$	**91.71±3.38**$^{\approx}$	91.38±3.72
	SVM	84.60±4.74^{-}	70.13±9.99^{-}	50.57±9.43^{-}	4.56±5.97^{-}		85.82±6.05$^{\approx}$	88.29±4.98
C8	KNN	85.15±3.79^{-}	83.63±4.24^{-}	84.09±3.20^{-}	65.55±6.54^{-}	96.92±1.42$^{-}_{-}$	**98.64±1.14**$^{\approx}$	98.46±1.38
	SVM	90.51±3.24^{-}	88.72±4.46^{-}	87.90±3.52^{-}	96.98±1.63^{-}		98.02±1.51$^{\approx}$	98.16±1.64
C9	KNN	25.97±5.80^{-}	28.17±6.25^{-}	25.10±6.18^{-}	19.60±5.65^{-}	64.93±8.53$^{-}_{-}$	95.32±3.44$^{\approx}$	**96.61±2.40**
	SVM	45.58±9.00^{-}	34.43±9.06^{-}	34.68±8.67^{-}	4.33±5.18^{-}		87.74±7.81$^{\approx}$	89.92±6.18
C10	KNN	43.90±28.48^{-}	35.88±27.08^{-}	16.06±14.30^{-}	4.94±5.83^{-}	55.83±18.13$^{-}_{-}$	**73.52±34.22**$^{\approx}$	73.27±33.85
	SVM	54.12±31.56^{-}	44.96±31.76^{-}	14.34±18.70^{-}	0.00±0.00^{-}		70.79±33.98$^{\approx}$	72.78±33.81
C11	KNN	27.13±9.11^{-}	23.50±7.73^{-}	24.54±8.12^{-}	9.61±4.41^{-}	46.71±5.82$^{-}_{-}$	**90.53±5.91**$^{\approx}$	89.67±6.01
	SVM	37.84±10.44^{-}	40.37±12.27^{-}	46.12±13.57^{-}	0.00±0.00^{-}		90.31±7.56$^{\approx}$	89.95±7.91
C12	KNN	18.83±19.46^{-}	17.38±15.40^{-}	25.69±20.42^{-}	2.19±4.00^{-}	42.11±14.89$^{\approx}_{\approx}$	48.04±39.74$^{\approx}$	53.72±41.11
	SVM	18.64±20.49^{-}	15.10±15.79^{-}	31.92±27.52^{-}	0.00±0.00^{-}		54.21±40.77$^{\approx}$	**55.05±41.28**
C13	KNN	44.99±4.25^{-}	57.80±4.99^{-}	54.19±4.64^{-}	32.38±3.08^{-}	80.42±2.03$^{-}_{-}$	92.52±2.45$^{\approx}$	**93.50±2.25**
	SVM	65.55±4.58^{-}	56.90±6.51^{-}	49.46±6.84^{-}	33.45±6.62^{-}		87.96±3.46$^{\approx}$	86.76±3.92
C14	KNN	63.70±4.31^{-}	70.16±3.02^{-}	63.29±3.35^{-}	49.42±3.81^{-}	83.49±2.33$^{-}_{-}$	96.46±1.51$^{\approx}$	**97.06±1.05**
	SVM	71.04±3.25^{-}	60.05±5.20^{-}	55.19±6.55^{-}	92.54±4.20$^{\approx}$		93.82±2.23$^{\approx}$	93.77±2.50
C15	KNN	85.14±4.30^{-}	83.13±5.51^{-}	80.43±7.38^{-}	54.61±7.64^{-}	95.85±2.06$^{\approx}_{\approx}$	95.97±1.79$^{\approx}$	**96.48±1.95**
	SVM	91.09±3.86^{-}	83.21±4.63^{-}	72.52±5.72^{-}	84.36±5.07^{-}		94.58±2.66$^{\approx}$	94.80±2.63
C16	KNN	87.39±5.24^{-}	82.30±8.80^{-}	81.89±8.30^{-}	32.07±10.95^{-}	92.44±5.94$^{\approx}_{\approx}$	93.53±6.02$^{\approx}$	**94.84±5.63**
	SVM	91.90±5.72$^{\approx}$	83.80±10.67^{-}	94.95±6.08$^{\approx}$	3.13±6.95^{-}		89.71±9.24$^{\approx}$	92.75±6.08
OA	KNN	59.32±1.03^{-}	65.54±1.10^{-}	60.88±1.97^{-}	39.06±1.43^{-}	81.09±1.26$^{-}_{-}$	94.03±0.60^{-}	**94.41±0.70**
	SVM	70.26±1.15^{-}	63.94±1.79^{-}	59.23±1.53^{-}	51.34±0.61^{-}		90.66±1.39$^{\approx}$	90.96±1.17
#Features		220	62	62	27†	1031	78.00±1.93^{-}	62.27±2.84

* According to [37], sparse group lasso (SGLasso) feature selection and classification method is used for 3D-DWT. Wilcoxon rank-sum test at a 0.05 significant level is performed between M3DGFE and each of the other methods in terms of classification accuracy using the same classifier. The significance of difference between M3DGFE and G3DGFE is also tested in terms of number of selected features. Superscripts -, +, and ≈ indicate that the performance of the corresponding method is significantly worse than, significantly better than, and similar to that of M3DGFE, respectively. The comparison results between 3D-DWT and M3DGFE using KNN and SVM are indicated with superscript and subscript, respectively. † Number of eigenvectors to account for 95% of the variance in the original data. Bold typefaces emphasize the best accuracy obtained in each class or the overall accuracy.

Table 4: Classification accuracy and selected feature size of the compared methods on KCS data

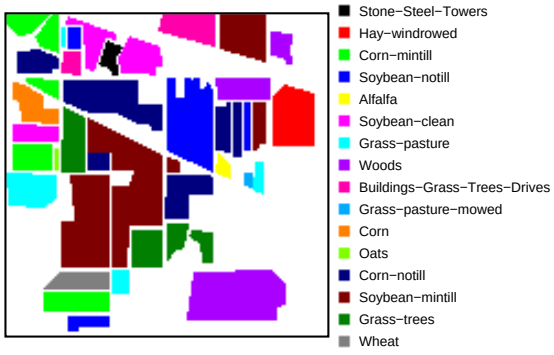
		ALL	MI	ReliefF	PCA	3D-DWT*	G3DGFE	M3DGFE
C1	KNN	85.76±13.98$^{\approx}$	85.97±11.61$^{\approx}$	80.31±15.14^{-}	57.97±16.54^{-}	81.57±14.21$^{\approx}_{-}$	89.59±19.92$^{\approx}$	90.59±20.11
	SVM	85.98±13.37$^{\approx}$	84.02±13.76$^{\approx}$	77.21±16.66^{-}	33.95±35.42^{-}		90.52±20.6$^{\approx}$	**90.91±20.12**
C2	KNN	57.1±16.04^{-}	52.39±15.37^{-}	51.27±12.92^{-}	38.06±14.25^{-}	**87.28±8.14**$^{\approx}_{+}$	86.58±11.36$^{\approx}$	85.71±12.76
	SVM	74.14±15.73$^{\approx}$	65.49±16.54^{-}	63.35±14.13^{-}	37.74±28.5^{-}		81.65±16.73$^{\approx}$	79.87±17.04
C3	KNN	75.95±8.94^{-}	79.44±10.61^{-}	71.37±12.00^{-}	62.94±13.07^{-}	94.48±2.86$^{\approx}_{\approx}$	94.95±10.09$^{\approx}$	**95.79±9.09**
	SVM	83.29±10.82^{-}	85.14±9.77^{-}	81.99±12.11^{-}	59.81±25.44^{-}		91.4±13.56$^{\approx}$	93.74±11.66
C4	KNN	20.71±12.96^{-}	15.38±10.83^{-}	16.25±11.04^{-}	13.60±10.56^{-}	36.86± 13.28$^{-}_{-}$	71.16±33.23$^{\approx}$	**72.04±32.70**
	SVM	27.52±17.79^{-}	22.25±15.4^{-}	22.51±16.30^{-}	3.05±11.37^{-}		63.20±31.95$^{\approx}$	64.78±30.33
C5	KNN	44.64±11.45^{-}	43.18±11.81^{-}	44.14±12.03^{-}	31.47±10.19^{-}	61.91±10.97$^{-}_{-}$	86.75±11.80$^{\approx}$	**89.09±11.23**
	SVM	51.13±15.29^{-}	51.57±12.08^{-}	50.23±13.68^{-}	25.12±35.31^{-}		89.04±9.94$^{\approx}$	87.99±13.24
C6	KNN	64.45±6.68^{-}	60.81±8.62^{-}	58.27±8.13^{-}	39.79±6.91^{-}	88.34±3.17$^{-}_{-}$	99.65±0.82$^{\approx}$	**99.92±0.40**
	SVM	71.51±9.41^{-}	71.38±10.46^{-}	66.74±12.38^{-}	62.03±35.15^{-}		99.05±1.39$^{\approx}$	98.45±3.18
C7	KNN	99.92±0.41$^{\approx}$	99.85±0.51$^{\approx}$	98.83±4.59$^{\approx}$	99.66±0.99$^{\approx}$	**99.77± 0.41**$^{\approx}_{+}$	99.21±1.96$^{\approx}$	99.74±1.04
	SVM	99.94±0.31^{+}	99.53±1.31^{+}	98.97±4.94$^{\approx}$	99.68±0.98^{+}		97.15±4.35$^{\approx}$	97.66±3.68
C8	KNN	36.42±13.3^{-}	34.38±14.03^{-}	36.83±13.8^{-}	21.99±9.32^{-}	45.00±7.23$^{-}_{-}$	95.18±7.21$^{\approx}$	**96.43±5.94**
	SVM	38.87±15.45^{-}	38.68±16.03^{-}	38.29±15.4^{-}	6.86±16.96^{-}		89.47±10.64$^{\approx}$	90.76±10.44
OA	KNN	62.63±2.39^{-}	60.97±3.44^{-}	59.19±3.17^{-}	46.91±2.41^{-}	77.95±2.52$^{-}_{-}$	92.69±2.19$^{\approx}$	**93.51±2.40**
	SVM	68.63±2.94^{-}	67.56±3.82^{-}	65.00±3.63^{-}	46.69±3.68^{-}		90.85±2.88$^{\approx}$	90.93±3.08
#Features		176	52	52	2†	226	76.77±2.88^{-}	51.70±3.75

* According to [37], sparse group lasso (SGLasso) feature selection and classification method is used for 3D-DWT. Wilcoxon rank-sum test at a 0.05 significant level is performed between M3DGFE and each of the other methods in terms of classification accuracy using the same classifier. The significance of difference between M3DGFE and G3DGFE is also tested in terms of number of selected features. Superscripts -, +, and ≈ indicate that the performance of the corresponding method is significantly worse than, significantly better than, and similar to that of M3DGFE, respectively. The comparison results between 3D-DWT and M3DGFE using KNN and SVM are indicated with superscript and subscript, respectively. † Number of eigenvectors to account for 95% of the variance in the original data. Bold typefaces emphasize the best accuracy obtained in each class or the overall accuracy.

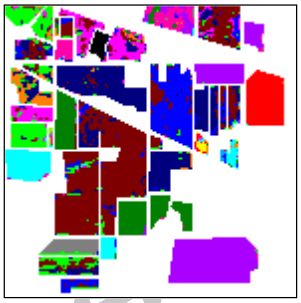
Table 5: Comparison results of M3DGFE, Fused-Gabor, and Filtered-Gabor.

Data	Method	*#Features*	KNN		SVM	
			OA	Time(s)	OA	Time(s)
Indiana	Fused-Gabor	1496	92.44±1.03^{-}	123.71±2.25^{-}	**95.31±0.95**$^{+}$	26.23±0.16^{-}
	Filtered-Gabor	136	92.89±0.89^{-}	10.17±0.23^{-}	94.66±1.05^{+}	2.32±0.03^{-}
	M3DGFE	**62**	**94.41±0.70**	**5.14±0.15**	90.96±1.17	**0.85±0.02**
KSC	Fused-Gabor	809	89.33±3.65^{-}	2.80±0.01^{-}	92.47±3.76$^{\approx}$	0.44±0.05^{-}
	Filtered-Gabor	110	89.53±3.25^{-}	0.05±0.01^{-}	**94.62±3.32**$^{+}$	0.05±0.01^{-}
	M3DGFE	**52**	**93.51±2.40**	**0.03±0.01**	90.93±3.08	**0.03±0.01**

Wilcoxon rank-sum test at a 0.05 significant level is performed between M3DGFE and each of Fused-Gabor and Filtered-Gabor in terms of classification accuracy and time cost. Superscripts -, +, and ≈ indicate that the performance of the corresponding method using the same classifier is significantly worse than, significantly better than, and similar to that of M3DGFE, respectively. Bold typefaces emphasize the best result obtained by the methods in each column.

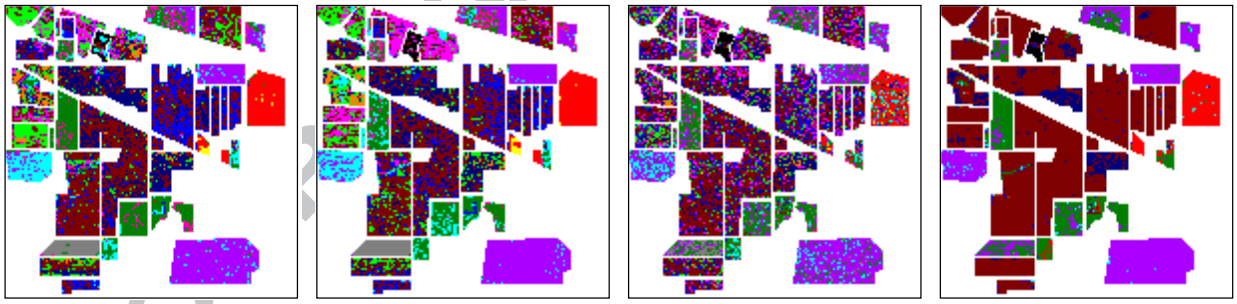


(a) Ground Truth



(b) 3D-DWT+SGLasso.

(c) ALL+KNN (d) ALL+SVM (e) MI+KNN (f) MI+SVM



(g) ReliefF+KNN (h) ReliefF+SVM (i) PCA+KNN (j) PCA+SVM

(k) G3DGFE+KNN (l) G3DGFE+SVM (m) M3DGFE+KNN (n) M3DGFE+SVM

Figure 5: Prediction results of the compared feature selection/extraction methods using different classifiers on Indiana data.

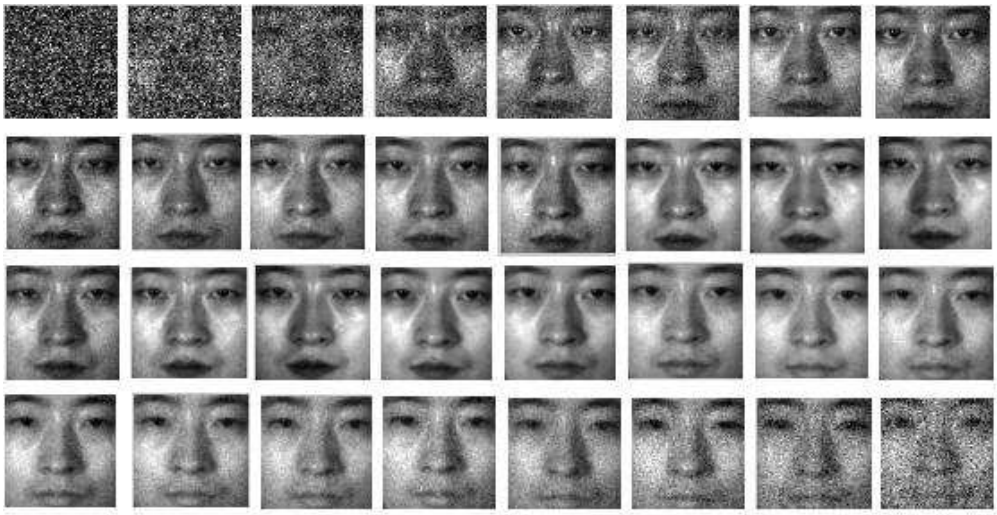


Figure 6: The 32 bands of an example hyperspectral face image.

Table 6: Classification accuracy and selected feature size of the compared methods on the 3D hyperspectral face data.

	ALL	MI	ReliefF	PCA	BS-WFD*	G3DGFE	M3DGFE
KNN	63.40±7.07^{-}	60.80±6.14^{-}	50.60±8.32^{-}	68.00±6.93^{-}	79.33	95.20±3.22$^{\approx}$	95.80±2.40
SVM	80.00±7.48^{-}	63.80±6.51^{-}	55.80±7.53^{-}	77.60±4.08^{-}		95.80±3.46$^{\approx}$	**96.80±2.96**
#Features	12288	65	65	12†	6	82.25±1.83^{-}	65.40±1.04

* BS-WFD [14] denotes the band subset fusion-based (2D)^{2}PCA with weighted averaging fusing method. Wilcoxon rank-sum test at a 0.05 significant level is performed between M3DGFE and each of the other methods in terms of classification accuracy. The significance of difference between M3DGFE and G3DGFE is also tested in terms of number of selected features. Superscripts -, +, and ≈ indicate that the performance of the corresponding method is significantly worse than, significantly better than, and similar to that of M3DGFE, respectively. Because the standard deviation result of BS-WFD is available in [14], statistical test between BS-WFD and M3DGFE is not performed. † Number of eigenvectors to account for 95% of the variance in the original data. Bold typefaces emphasize the best overall accuracy.

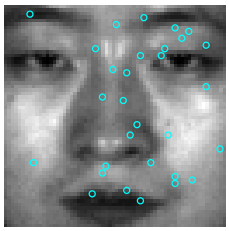


Figure 7: The 30 most frequently selected locations by M3DGFE on the face.