CrossMark

Multiple features learning for ship classification in optical imagery

Longhui Huang[1] · Wei Li[1] ⓘ **· Chen Chen[2] ·**
Fan Zhang[1] · Haitao Lang[3]

Abstract The sea surface vessel/ship classification is a challenging problem with enormous implications to the world's global supply chain and militaries. The problem is similar to other well-studied problems in object recognition such as face recognition. However, it is more complex since ships' appearance is easily affected by external factors such as lighting or weather conditions, viewing geometry and sea state. The large within-class variations in some vessels also make ship classification more complicated and challenging. In this paper, we propose an effective multiple features learning (MFL) framework for ship classification, which contains three types of features: Gabor-based multi-scale completed local binary patterns (MS-CLBP), patch-based MS-CLBP and Fisher vector, and combination of Bag of visual words (BOVW) and spatial pyramid matching (SPM). After multiple feature learning, feature-level fusion and decision-level fusion are both investigated for final classification. In the proposed framework, typical support vector machine (SVM) classifier is employed

---

✉ Wei Li
liwei089@ieee.org

Chen Chen
chenchen@crcv.ucf.edu

Fan Zhang
zhangf@mail.buct.edu.cn

Haitao Lang
langht@mail.buct.edu.cn

[1] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

[2] Center for Research in Computer Vision, University of Central Florida, Orlando, FL, USA

[3] Faculty of Science, Beijing University of Chemical Technology, Beijing 100029, China

Springer

to provide posterior-probability estimation. Experimental results on remote sensing ship image datasets demonstrate that the proposed approach shows a consistent improvement on performance when compared to some state-of-the-art methods.

1 Introduction

Ship recognition and classification algorithms in ocean area are important to enhancing maritime safety and security [12, 17]. The goal of vessel/ship classification is to recognize the type of ships in a given image with as much details as possible. Beyond the typical recognition challenges caused by partial occlusions and variations in scale [38], ships are especially difficult to recognize since ships' appearance is usually significantly affected by external factors such as weather conditions (i.e., cloudy, sunny, etc), viewing geometry, and sea state [29]. Moreover, wide variation within-class in some types of vessels also makes vessels classification more complicated and challenging [2].

In recent years, numerous efforts have been investigated in the problem of vessel recognition in optical remote sensing imagery, achieving many diverse solutions to the problem [10, 11, 17]. In [15, 26], several well-studied feature extraction methods in facial recognition were applied to a small set of ship images. The feature extraction methods include global/holistic features: Principal Components Analysis (PCA) [25], Linear Discriminant Analysis (LDA) [4], Independent Component Analysis (ICA) [3], Random Projections (Rand) [42] and Multilinear PCA (MPCA) [25]; and local features: hierarchical multiscale Local Binary Pattern (HMLBP) [14] and Histogram of Oriented Gradients (HOG) [9]. HMLBP operates on multiple gray-scale texture to obtain scale-invariant local spatial structure and texture information of the image. HOG is essentially an image descriptor that represents the image by the distribution of local intensity gradients or edge directions, to capture the local object appearance and shape in the image. On the basis of previous work in [15], Rainey et al. [35] explored the efficacy of several object recognition algorithms at classifying ships and other ocean vessels in commercial panchromatic satellite imagery.

In addition to the above mentioned methods, the Bag of (visual) words (BOVW) [19], as the one of the most popular approaches in image analysis and classification applications, was also applied to ship classification. The BOVW, which is inspired by the bag of words representation used in text classification tasks, represents an image as a histograms of frequencies of a set of local descriptors such as Scale-Invariant Feature Transform(SIFT) [24]. In order to increase robustness of the feature, Rainey et al. [5] replaced Scale-Invariant Feature Transform (SIFT) descriptors in the BOVW mode with Pyramid of Histograms of visual Words (PHOW [34]) descriptors, which are extracted on a dense grid of key points and provide greater accuracy on resized non-degraded data. An algorithm for vessel classification, which combined BOW histograms with sparse representation based classification (SRC), was proposed in [36]. The original BOW and BOW with ℓ_2-normalized term frequency (TF) weighting schemes were studied in [28] to the vessel type classification task in the maritime domain. In [1], the Local Binary Patterns (LBP [21]), which has high distinctiveness and low computational complexity, was developed.

According to the way of extracting features, these approaches can usually be partitioned into two categories: global (holistic) and local. Global features usually describe the image

using a lower-dimensional space or histogram, and form the feature space by utilizing the entire image. Thus the global features are often easily implemented and have low computational cost, but their performances for classification are limited. Local features usually use a set of local descriptors to characterize the ship image. However, when the shapes of two classes are relatively similar, the local features are not discriminative enough to distinguish. Although the aforementioned feature extraction approaches (i.e., local features such as BOVW, and global features such as MPCA) have achieved satisfying performances in vessel classification, each type of feature has its own advantages and limitations. Moreover, it is known that different features have their own characters to capture different information of ships and certain features may be only suitable for one specific pattern. To this end, the global feature, which is complementary to the local features, is utilized to represent more discriminative information of a class.

Inspired by this observation, a multiple features learning (MFL) framework is proposed for ship classification in this work. The proposed method employs three types of features to simultaneously capture global structures and local details information of various ships. We employ Gabor-based multi-scale completed local binary pattern (MS-CLBP [7]), which fuses the benefits of Gabor wavelet and CLBP descriptor, to extract the global feature from images, and we treat it as the first type of features in our work. This global feature captures the scale-invariant and orientation-invariant spatial structure and texture information from the entire image. The local feature extraction method based on patch-based MS-CLBP and Fisher vector (FV [37]) is used to obtain a local feature representation of ship images which is considered as the second type of features. The patch-based MS-CLBP and FV has been demonstrated great success in image classification [16]. The BOVW, as a classical feature extract method in the object recognition domain, is employed in the proposed algorithm. We further combine BOVW and spatial pyramid matching (SPM [20]) as the third type of features, which is able to overcome the orderless of bag-of-features representation and enhance the spatial order structure of local features. The FV and BOVW are both visual code models, but the FV employs a soft assignment strategy while the BOVW employs a hard assignment strategy. In the FV model, we use MS-CLBP operator to extract local texture information from image patches. The BOVW model utilizes the SIFT descriptor to roughly represent edge information in an image patch. Note that both the FV and BOVW are employed in order to obtain more comprehensive local features of ship images.

In this paper, we simultaneously extract global and local features to comprehensively capture the characteristics of ships. After feature extraction, fusion strategies are followed for a final classification. It is well-known that the fusion strategies mainly include feature-level and decision-level fusion. In our scheme, feature-level fusion and decision-level fusion are both investigated in the proposed classification framework. Feature-level fusion combines different feature vectors together into a single feature vector. Decision-level fusion, which performs on probability outputs of each individual classification pipeline and combines the distinct decisions into a final one, can be divided into two types: "hard" fusion at the class-label level and "soft" fusion at the probability level. However, hard-fusion (e.g., majority voting (MV) rule [47]) may lead to a rough result. In this paper, we choose a "soft" fusion scheme, namely logarithm opinion pool (LOGP [22]), and verify its effectiveness by comparing it with the hard-fusion method (i.e., MV) in the experimental analysis.

There are two main contributions of this work. First, a multiple feature learning (MFL) method using three types of features including global and local features is proposed. In

the proposed classification framework, feature-level and decision-level fusion are both investigated. Second, Gabor-based MS-CLBP is employed to extract global feature to compensate for the local feature, therefore taking full advantage of the complementary nature between global and local features. The proposed method is extensively evaluated using two public optical remote sensing ship image datasets. The experimental results verify the effectiveness of our proposed method as compared to some state-of-the-art algorithms.

The remainder of the paper is organized as follows. Section 2 describes the proposed approach, including the three types of features extraction and the feature-level as well as decision-level soft fusion strategy. Section 3 introduces two experimental datasets (i.e., BCCT200−resize [35] and VAIS [48]). Section 4 reports the experimental results and provides some analysis. Finally, Section 5 makes several concluding remarks.

2 Proposed classification framework

The flowchart of the proposed method is illustrated in Fig. 1. There are two fusion strategies for the proposed classification framework, i.e., feature-level fusion and decision-level fusion approaches, which are illustrated in Fig. 1a and b, respectively.

As shown in Fig. 1a, we first extract three kinds of features of the input image respectively. The first one is Gabor-based MS-CLBP. We use multi-orientation (e.g., $\pi/8$, $\pi/4$, $\pi/2$, etc.) Gabor filters to obtain multiple Gabor feature images, where the MS-CLBP is then employed. The second one is a local feature which is extracted by patch-based MS-CLBP and FV. We partition the image and its multi-scale versions into dense patches and using the CLBP descriptor to extracts a set of local patch descriptors. Then, FV encoding is used to encode the local patch descriptors into a discriminative representation. The third one is the combination of BOVW and SPM. We employ the dense SIFT descriptor by partitioning an image into small patches. Then k-means clustering is employed to generate the codebook which presents the visual words of the BOVW. The SPM is further employed to calculate the local features, which enhances the spatial order structure of local features. Finally, we employ the typical support vector machine (SVM) classifier [6] to obtain final classification performance. Details of the features extraction is presented in Section 2.1.

The decision-level fusion framework is shown in Fig. 1b. For each input image, three kinds of features are extracted separately. Then, for each of the three types of features, the typical SVM classifier is employed to calculate the probability estimates, respectively. Finally, the proposed method merges outputs of each individual classification pipeline using decision-level soft-fusion (i.e., LOGP) to obtain final classification result.

2.1 Feature extraction

2.1.1 Gabor-based MS-CLBP

Inspired by the success of Gabor filters and LBP in computer vision applications, we employ Gabor-based MS-CLBP as the first type of features in the proposed framework to extract global features of ship images. The features extraction process is summarized in Algorithm 1.

(a)

(b)
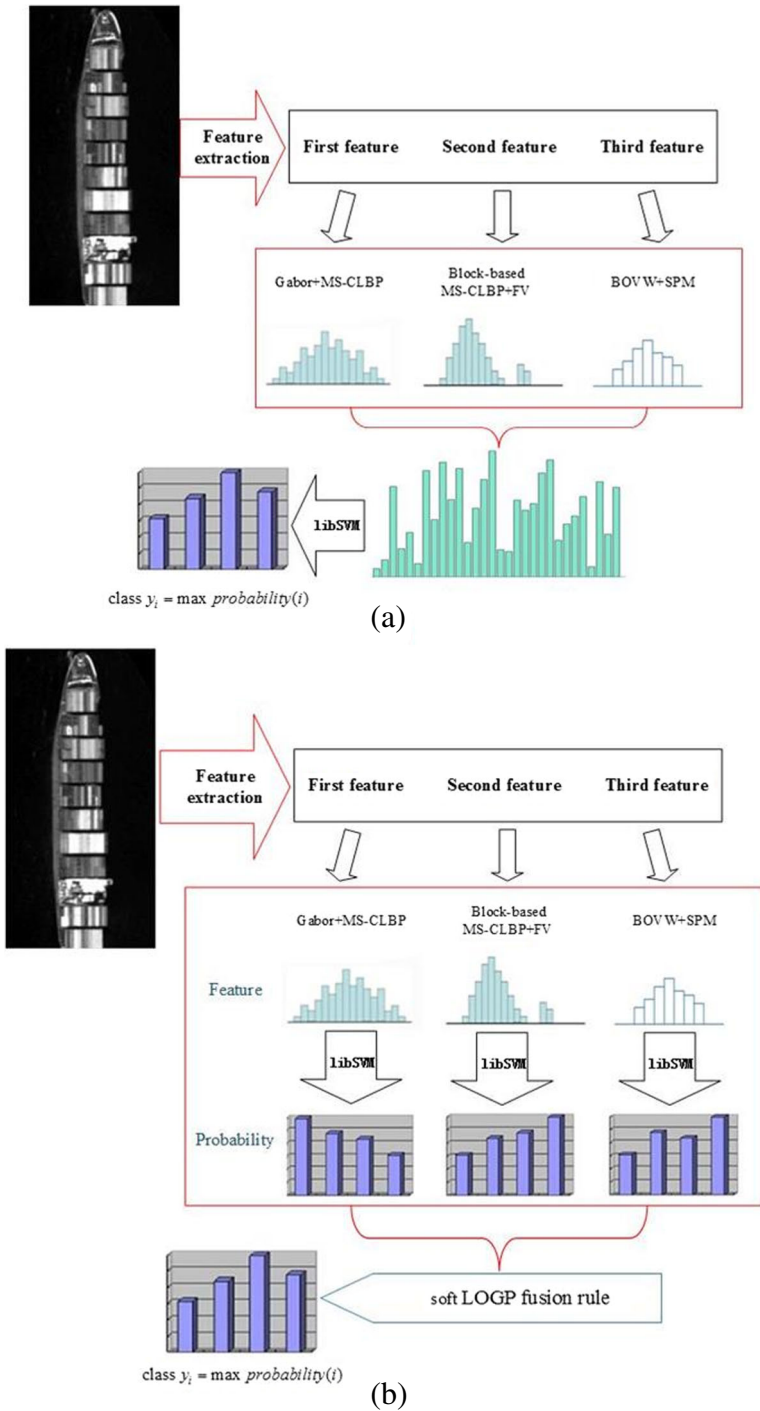
Fig. 1 Flowchart of the proposed multiple features learning framework: **a** feature-level fusion, **b** decision-level fusion

---

Algorithm 1 Gabor-based MS-CLBP

Input: A testing image

Step 1: Obtain the Gabor texture images by convolve the input image with Gabor function (1);

Step 2: Extract the MS-CLBP feature of the Gabor texture images;

Step 3: Concatenate the all MS-CLBP feature vectors to produce the final feature vector **F**[1].

Output: **F**[1].

A Gabor wavelet [8, 30] transform can be viewed as an orientation-dependent bandpass filter. Its impulse response is defined by a sinusoidal wave multiplied by a Gaussian function. In 2-D (x, y) coordinate system, a Gabor filter, which includes a real component and an imaginary term, can be defined as,

$$G_{\varepsilon,\theta,\psi,\sigma,\gamma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \tag{1}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$. Here, ε is the wavelength of the sinusoidal factor, θ is the orientation separation angle (e.g., $\pi/8$, $\pi/4$, $\pi/2$, etc.) of Gabor kernels, ψ represents the phase offset, σ is the standard derivation of the Gaussian envelope, and γ is the spatial aspects ratio specifying the ellipticity of the support of the Gabor function. Note that $\psi = 0$ and $\psi = \pi/2$ return the real and imaginary parts of the Gabor filter, respectively. Parameter σ is determined by ε and spatial frequency bandwidth bw as,

$$\sigma = \frac{\varepsilon}{\pi}\sqrt{\frac{\ln 2}{2}\frac{2^{bw} + 1}{2^{bw} - 1}}. \tag{2}$$

The LBP [27] is a texture operator, which is able to characterize the spatial structure information of local image texture, and it has been widely employed in object recognition (e.g., texture classification, face recognition, etc). Given a pixel in the image, which gray value is denoted as g_c. Its neighboring pixels are equally spaced on a circle of radius. The resulting LBP for g_c in decimal number can be computed by comparing it with its neighbors,

$$LBP_{m,r}(g_c) = \sum_{p=0}^{m-1} s\left(g_p - g_c\right)2^p = \sum_{p=0}^{m-1} s\left(d_p\right)2^p,$$

$$s(d_p) = \begin{cases} 1, & d_p \geq 0 \\ 0, & d_p < 0 \end{cases} \tag{3}$$

where g_p is the gray value of the neighbors, $d_p = (g_c - g_p)$ represents the difference between the center pixel and each neighbor, r represents the radius of a circle, and m is the total number of involved neighbors. If the coordinate of g_c is $(0, 0)$, the coordinates of g_p are denoted as $(-r\sin(2\pi i/m), r\cos(2\pi i/m))$. After the LBP coded image is generated by compute the LBP values of all pixels in the image, a histogram is calculated to represent the texture image. Nevertheless, the LBP only uses the sign information of d_p while ignoring the magnitude information.

In the improved CLBP [13], the sign and magnitude information, i.e., CLBP-Sign (CLBP_S) and CLBP-Magnitude (CLBP_M), is complementary. The CLBP_S is equivalent to the traditional LBP operator and the CLBP_M operator is expressed as,

$$CLBP\_M_{m,r}(g_c) = \sum_{p=0}^{m-1} f\left(\left|g_p - g_c\right|, \eta\right) = \sum_{p=0}^{m-1} f\left(\left|d_p\right|, \eta\right),$$

$$f(x, y) = \begin{cases} 1, & d_p \geq \eta \\ 0, & d_p < \eta \end{cases} \tag{4}$$

where η is a threshold that is usually set to the mean value of $\left|d_p\right|$. The syncretic CLBP can describe both the spatial and depth information by mapping the CLBP_S and CLBP_M into histograms. To make the feature scale invariant, the multi-scale representation of CLBP, termed as MS-CLBP, is considered. In our work, the multi-scale CLBP representation is formed by concatenated the CLBP_S and CLBP_M histogram features extracted at different scales which are obtained by down-sampling the original image using the bicubic interpolation [7]. An example of the implementation of a 3-scale CLBP operator is illustrated in Fig. 2.

2.1.2 Patch-based MS-CLBP and FV

As noted in [6], image representation method based on patch-based MS-CLBP and FV can well extract the local feature of image and has achieved great performance in remote sensing image scene classification. In this work, we first use regular grid to partition an entire image into $B \times B$ overlapped patches. For simplicity, the overlap between two patches is half of the patch size (i.e., $B/2$) in both horizontal and vertical directions. For each patch, we employ MS-CLBP with the second implementation capture the spatial pattern and the contrast of local image texture. If M patches are extracted from the multi-scale images, we can obtain a set of local patch descriptors and form a feature matrix denoted by $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M]$, where $\mathbf{h}_i$ is the CLBP histogram feature vector extracted from patch i.

After local feature extraction, FV encoding is used to encode the local patch descriptors into a discriminative representation. Let $\mathbf{X} = \{\mathbf{x}_i, i = 1, \cdots, N\}$ be the set of local patch
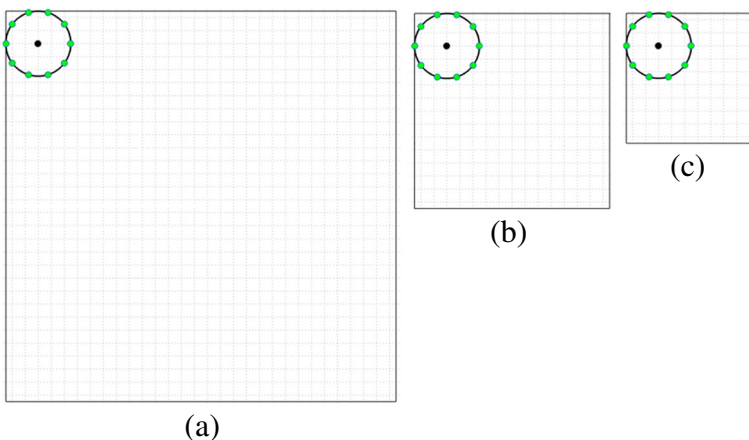

Fig. 2 An example of a 3-scale CLBP operator ($m = 10, r = 1$): **a** scale $= 1$, **b** scale $= 1/2$, **c** scale $= 1/3$

descriptors extracted from an image. A Gaussian mixture model (GMM), which probability density function is denoted as u_λ with parameter λ, is trained on the training images using Maximum Likelihood (ML) estimation [23]. The image can be characterized by gradient vector,

$$\mathbf{G}_\lambda^\mathbf{X} = \frac{1}{N} \nabla_\lambda \log u_\lambda(\mathbf{X}). \tag{5}$$

The gradient of the log-likelihood describes the direction along which parameters are to be adjusted to best fit the data. The GMM u_λ can be represented as

$$u_\lambda = \sum_{t=1}^{T} \omega_t u_t(\mathbf{x}). \tag{6}$$

We denote $\lambda = \{\omega_t, \mu_t, \Sigma_t, t = 1, ..., T\}$, where ω_t, μ_t and Σ_t are the mixture weight, mean vector, and covariance matrix of Gaussian μ_t, and T is the number of Gaussian in the GMM. Under an independence assumption, the covariance matrices are diagonal, i.e., $\Sigma_t = \text{diag}(\sigma_t^2)$. We make use of Fisher kernel (FK) of [18] to measure the similarity between two samples $\mathbf{x}$ and $\mathbf{y}$, and it is defined as,

$$S(\mathbf{x}, \mathbf{y}) = \mathbf{G}_\lambda^{\mathbf{x}'} \mathbf{F}_\lambda^{-1} \mathbf{G}_\lambda^{\mathbf{x}}, \tag{7}$$

where $\mathbf{F}_\lambda$ is the Fisher information matrix of u_λ: $\mathbf{F}_\lambda = \mathbf{E}[\nabla_\lambda \log u_\lambda(\mathbf{x}) \nabla_\lambda \log u_\lambda(\mathbf{x})']$. Following the diagonal closed-form approximation of [31], the FK can be rewritten as a dot-product between normalized vectors $\mathbb{G}_\lambda$ with,

$$\mathbb{G}_\lambda^\mathbf{x} = \mathbf{F}_\lambda^{-1/2} \mathbf{G}_\lambda^\mathbf{x}. \tag{8}$$

Let $\gamma_i(t)$ be the occupancy probability, i.e., the probability of descriptor $\mathbf{x}_i$ generated by the Gaussian u_t,

$$\gamma_i(t) = \frac{\omega_t u_t(\mathbf{x}_i)}{\sum\limits_{j=1}^{T} \omega_j u_j(\mathbf{x}_i)}. \tag{9}$$

Let $\mathbb{G}_{\mu,t}^\mathbf{X}$ (resp. $\mathbb{G}_{\sigma,t}^\mathbf{X}$) be the gradient with respect to the mean μ_t (resp. standard deviation σ_t) of Gaussian t. Mathematical derivations lead to,

$$\mathbb{G}_{\mu,t}^\mathbf{X} = \frac{1}{N\sqrt{\omega_t}} \sum_{i=1}^{N} \gamma_i(t) \left(\frac{\mathbf{x}_i - u_t}{\sigma_t}\right), \tag{10}$$

and

$$\mathbb{G}_{\sigma,t}^\mathbf{X} = \frac{1}{N\sqrt{2\omega_t}} \sum_{i=1}^{N} \gamma_i(t) \left[\frac{(\mathbf{x}_i - u_t)^2}{\sigma_t} - 1\right]. \tag{11}$$

The final gradient vector is just the concatenation of the $\mathbb{G}_{\mu,t}^\mathbf{X}$ and $\mathbb{G}_{\sigma,t}^\mathbf{X}$ vectors for $t = 1, ..., T$. Therefore, the dimensionality of the FV is $(2D+1) \times T$, where D denotes the size of the local descriptors.

Each training image yields a feature matrix representing the local patch descriptors using patch-based MS-CLBP operator. Then, all the feature matrices of the training data are used to estimate the GMM parameters via the Maximum Likelihood (ML) algorithm. After that, the FV is utilized to generate the final feature representation. The features extraction process is summarized in Algorithm 2. Figure 3 illustrates the detailed procedure for generating FV of training images. For the testing image, we use the patch-baesd MS-CLBP feature
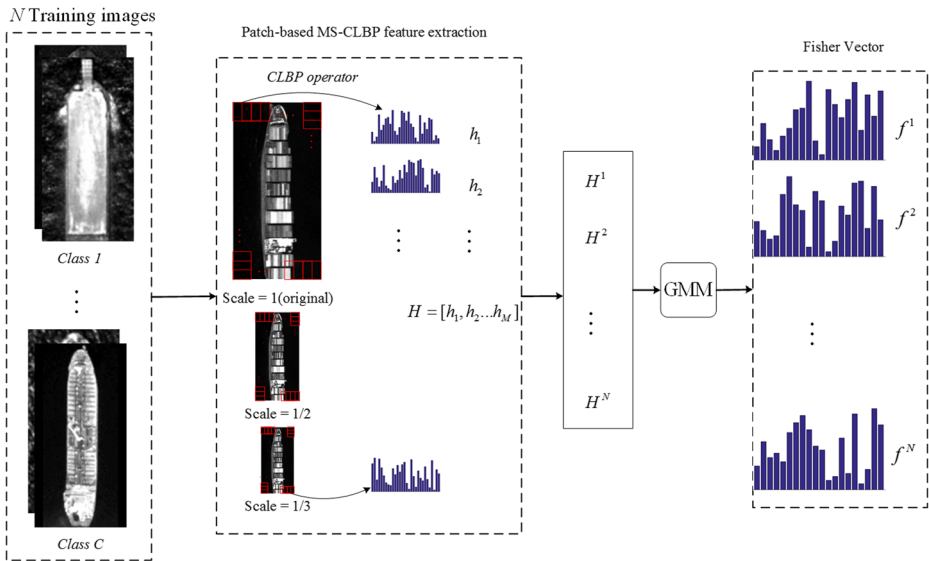
Fig. 3 Fisher vector representation of the training images

extraction method shown in the Fig. 3 to obtain local descriptors. Then the Fisher Kernel representation is utilized based on the GMM which is estimated from the training data to obtain the final FV feature. The detailed process for generating FV is shown in Fig. 4.

Algorithm 2 Patch-based MS-CLBP and FV

Input: Available training images and a testing image.
Step 1: Partition each training images into $B \times B$ overlapped patches using regular grid.
Step 2: Obtain a set descriptors by employing MS-CLBP operator over the patches.
Step 3: Obtain a GMM u_λ using Maximum Likelihood estimation.
Step 4: Partition the testing image. The partition process is the same with step 1. Extract the local descriptors, similar to step 2.
Step 5: Obtain gradient vector G_λ^X according to (5).
Step 6: Obtain final Fisher Vector feature representation $\mathbf{F}^2$ according to (10) and (11).
Output: $\mathbf{F}^2$.

2.1.3 BOVW and SPM

In recent years, the BOVW model, which is a classical local feature representation method, has demonstrated its outstanding performance in several object recognition tasks such as face recognition. Since traditional BOVW model ignores spatial and structural information, we affiliate SPM to the BOVW framework since the SPM can capture the spatial arrangement of images. The BOVW features extraction process is summarized in Algorithm 3, and the overall description of combination feature of BOVW and SPM is summarized in Algorithm 4.

A testing image
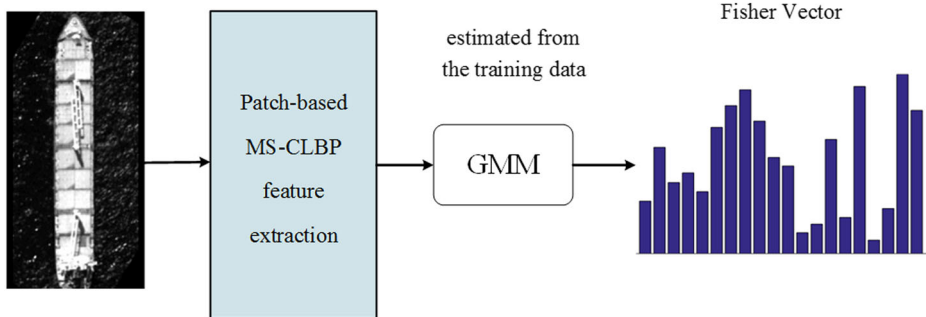
Fig. 4 Fisher vector representation of a testing image

First of all, we partition image into small blocks using regular grid and then we extract local descriptor utilizing the SIFT operator on each block whose center is viewed as a key point. An entire image is partitioned into $\varpi \times \varpi$ overlapped blocks and the overlap between two blocks is half of the block size (i.e., $\varpi/2$) in both horizontal and vertical directions.

In doing this, we can obtain a set of local descriptors. Then, k-means clustering is employed to generate the codebook which presents the visual words of the BOVW. After that, we encode the patches based on the codebook using vector quantization and calculate the frequent histogram. The SPM is further employed to enhance the spatial order structure of local features. In the SPM framework, an image is partitioned into $2^l \times 2^l$ segments in different scales (e.g., $l = 0, 1, 2$), and the BOVW histogram within all the segments is calculated respectively. The final feature representation of the image is the concatenation of all the histograms. An example of the implementations of the SPM is illustrated in Fig. 5.

Algorithm 3 BOVW model

Input: Available training images and a testing image
Step 1: Partition each training images into $\varpi \times \varpi$ overlapped patches.
Step 2: Obtain a set descriptors by employing SIFT operator over the patches.
Step 3: Generate codebook D utilizing k-means clustering over the descriptors.
Step 4: Partition the testing image. The partition process is the same with step 1. Extract the local descriptors $\mathbf{X}\{x_i, i = 1, 2, ...N\}$, similar to step 2.
Step 5: Encode $\mathbf{X}$ based on D, calculate the occurrence histogram vector.
Output: Histogram vector.

Algorithm 4 BOVW and SPM

Input: a testing image
Step 1: Partition input images into $2^l \times 2^l$ segments, l represents the layer of SPM.
Step 2: Calculate BOVW histogram within all the segments according to Algorithm 3.
Step 3: Obtain the final feature representation $\mathbf{F}^3$ by concatenating all the BOVW histograms.
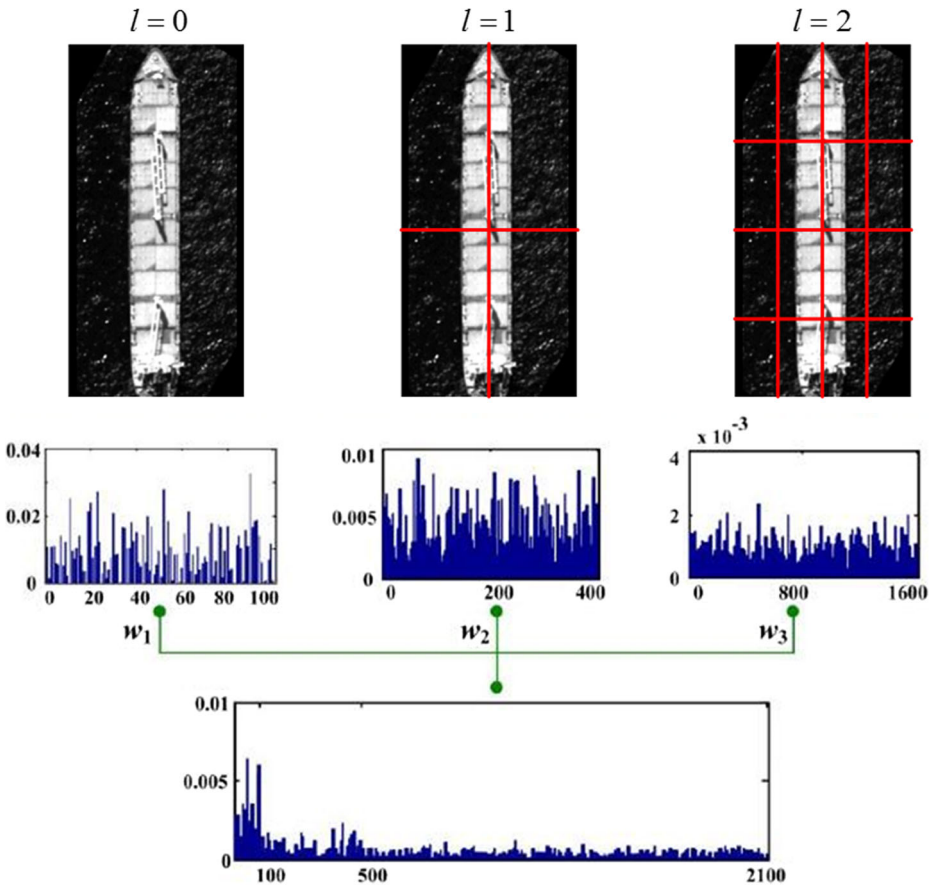**Output**: $\mathbf{F}^3$.

Fig. 5 Example of a three-level spatial pyramid matching ($l = 0, 1, 2$)

2.2 Feature-level fusion

After feature extraction, feature-level fusion and decision-level fusion are both investigated for final classification. Figure 1a illustrates the feature-level fusion [43] employed in this work. For multiple feature learning, the Gabor-based MS-CLBP extracts spatial structure and texture feature; the local feature extract method based on patch-based MS-CLBP and FV is visual code model with soft assignment strategy; and another local feature extracts method combined of BOVW and SPM is visual code model with hard assignment strategy. For various classification tasks, these features have their own advantages and disadvantages, and it is difficult to determine which one is always optimal [21]. Thus, we use feature-level fusion strategy to fusion three types of features by straightforward stacking feature vectors into a composite one. The specific implementation of feature fusion process is represented in Algorithm 5. To modify the scale of feature values, feature normalization before feature stacking is a necessary preprocessing step.

Algorithm 5 Feature-level fusion

Input: Available training images and a testing image

Step 1: Extract first feature vector $\mathbf{F}^1$ ($\mathbf{F}^1 \in {}^{M1}$) according to Algorithm 1. Extract first feature vector $\mathbf{F}^2$ ($\mathbf{F}^1 \in R^{M2}$) according to Algorithm 2. Extract first feature vector $\mathbf{F}^3$ ($\mathbf{F}^1 \in R^{M3}$) according to Algorithm 4.

Step 2: Fusion three feature vector into a composite vector $\mathbf{F} = [\mathbf{F}^1, \mathbf{F}^2, \mathbf{F}^3]$ ($\mathbf{F} \in R^{(M1+M2+M3)}$).

Step 3: Employ SVM classifier to obtain classification result.

Output: classification result.

2.3 Decision-level fusion

Decision-level fusion [32] is to merge results from each individual classification pipeline and combines distinct classification results into a final decision which can improve the accuracy of a single classifier that uses a certain type of features. Score level fusion [39, 45, 46] is a special case of decision-level fusion, and it is equivalent to the soft fusion of decision-level fusion. The goal here is to utilize score level fusion combine the posterior-probability estimations provided from each individual classification pipeline. In this paper, we employ LOGP soft-fusion rule, also known as product rule [44] in score level fusion scheme. The process is illustrated in Fig. 1b.

Assume $p_i(y_k|\mathbf{x})$ be the conditional class probability of the ith classifier in class k ($0 < k \leq C$) (C is the number of classes and y_k indicates the kth class to which a sample $\mathbf{x}$ belongs.). The LOGP rule [33] utilizes individual conditional class probabilities of each classifier to estimate a global membership function $\mathbf{P}(y_k|\mathbf{x})$, which is a weighted product of these output probabilities. The final class label is given according to,

$$y^* = \arg \max_{k=1,\ldots,C} = \mathbf{P}(y_k|\mathbf{x}), \tag{12}$$

where the global membership function is defined as,

$$\mathbf{P}(y_k|\mathbf{x}) = \prod_{i=1}^{3} p_i(y_k|\mathbf{x})^{\alpha_i}, \tag{13}$$

or

$$\log \mathbf{P}(y_k|\mathbf{x}) = \sum_{i=1}^{3} \alpha_i p_i(y_k|\mathbf{x}), \tag{14}$$

where $\{\alpha_i\}_{i=1}^3$ is the classifier weights uniformly distributed over all of the classifiers. The overall description of decision-level fusion is summarized in Algorithm 6.

Algorithm 6 Decision-level fusion

Input: Available training images and a testing image
Step 1: Extract first feature vector $\mathbf{F}^1$ according to Algorithm 1. Employ SVM classifier to obtain posterior-probability estimation $p_1(y_k | x)$.
Step 2: Extract first feature vector $\mathbf{F}^2$ according to Algorithm 2. Employ SVM classifier to obtain posterior-probability estimation $p_2(y_k | x)$.
Step 3: Extract first feature vector $\mathbf{F}^3$ according to Algorithm 4. Employ SVM classifier to obtain posterior-probability estimation $p_3(y_k | x)$.
Step 4: Obtain a global membership function $P(y_k | x)$ according to (13) or (14).
Step 5: Obtain final class label according to (12).
Output: class label.

---

3 Experimental datasets

To evaluate the efficacy of the proposed MFL approach for ship classification, we conduct extensive experiments using optical image datasets. In the experiments, a library for SVM[1] is employed for classification, which is able to provide posterior-probability estimation for each type of features.

The first dataset is an overhead satellite scene referred to as BCCT200-resize which is detailed in [35]. It consists of small grey-scale ship images chipped out of larger electro-optical satellite images by the RAPIER Ship Detection System. The chips have been pre-processed to be rotated and aligned to have uniform dimensions and orientation. The dataset contains 4 different ship categories (i.e., barges, cargo ships, container ships, and tankers). Each class contains 200 images of size 150×300 pixels. Examples of each class in this data set can be seen in Fig. 6.

The second dataset is the original BCCT200 dataset, in which the images are unprocessed and display ships under various orientations and resolutions. Such variations makes the dataset more challenging. The dataset includes the following classes: barges, container ships, cargo ships and tankers. 200 images have been collected from each class which are of non-uniform size. Example images of each class are shown in Fig. 7.

In order to facilitate a fair comparison, we follow the same experimental setup reported in [35] for the above two datasets. Five-fold cross-validation is performed, in which the dataset is randomly partitioned into five equal and disjoint subsets. There are 40 images from each ship class in a subset. For each run, a different subset is held out for testing and the remaining four are used for training. The classification accuracy is average results over the five cross-validation evaluations.

The third dataset used in our experiments, referred to as VAIS [48], is the world's first publicly available dataset of paired visible and infrared ship imagery. The dataset consists of 2865 images (1623 visible and 1242 IR), of which there are 1088 corresponding pairs. The dataset includes 6 coarse-grained categories: merchant ships, sailing ships, medium passenger ships, medium "other" ships, tugboats and small boats. The area of the visible bounding boxes ranged from $644 - 6350890$ pixels, with a mean of 181319 pixels and a median of 13064 pixels. Example bounding box images are shown in Fig. 8.

---

[1]http://www.csie.ntu.edu.tw/~cjlin/libsvm/
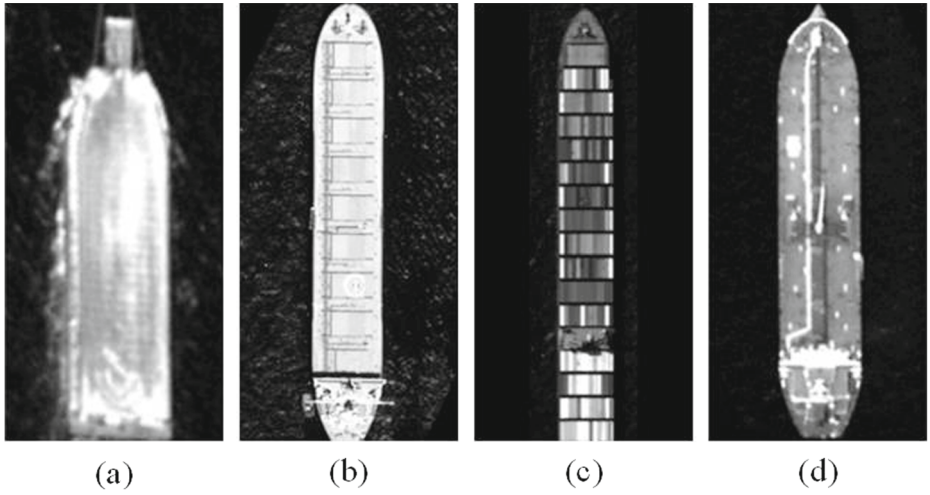
Fig. 6 Example images from each class of the BCCT200-resize data: **a** barge, **b** cargo ship, **c** container ship, and **d** tanker
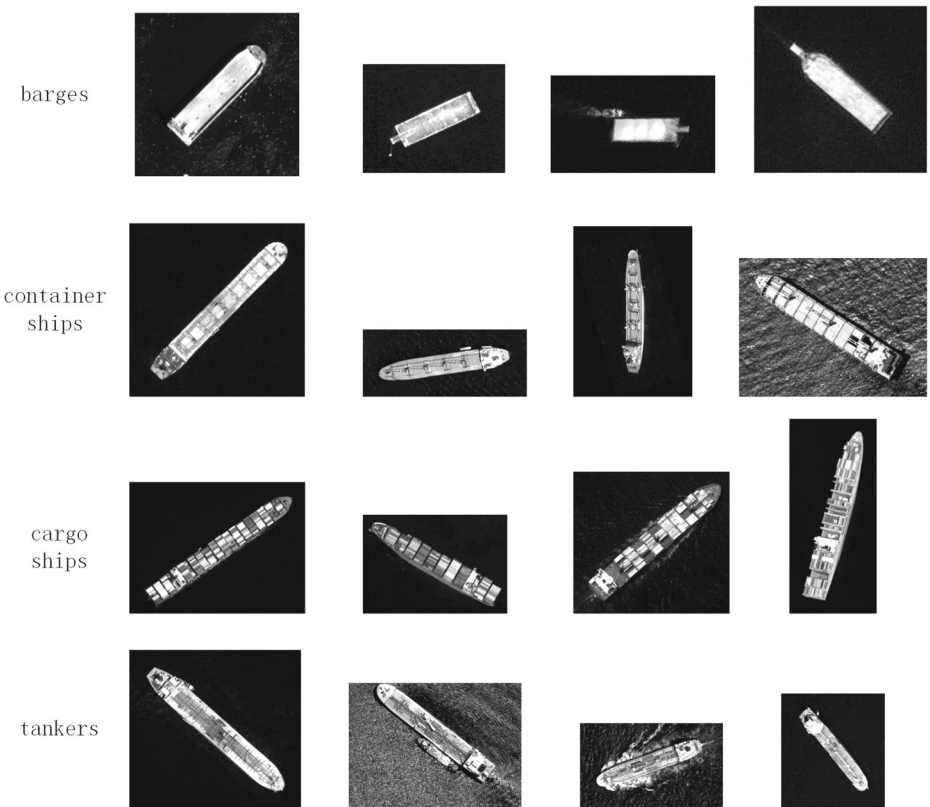
Fig. 7 Example images from each class of the original BCCT200 data
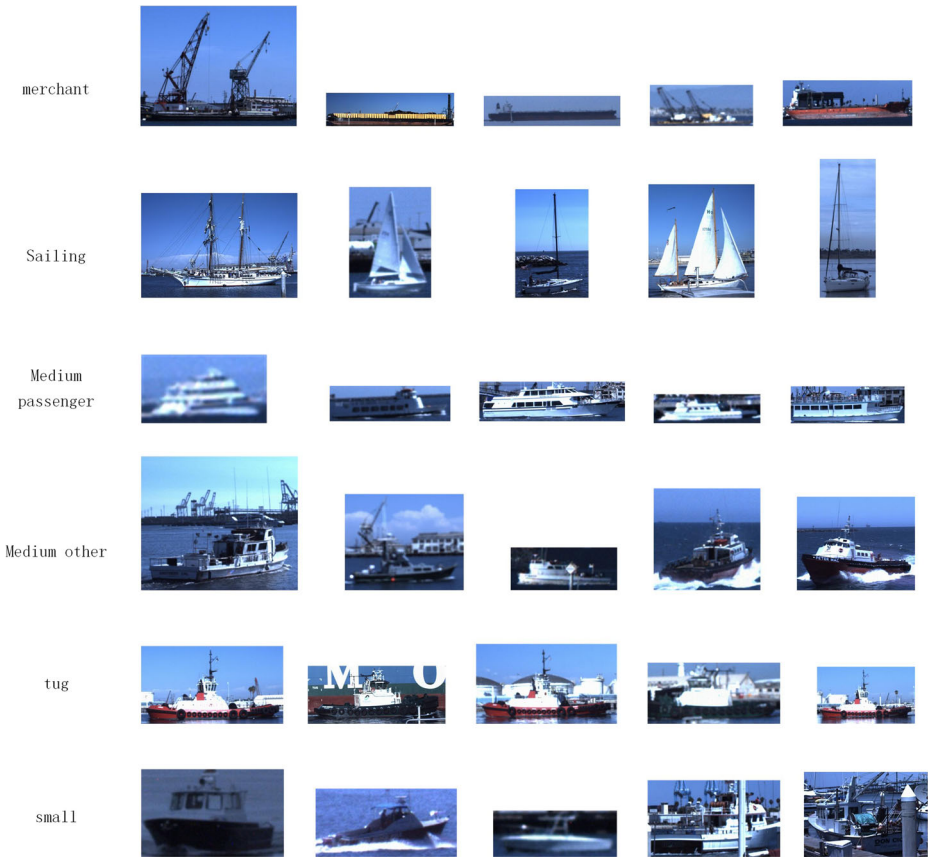
Fig. 8 Five visible image samples from each of the main categories of the VAIS data

The dataset is partitioned into "official" train and test splits. All images in this dataset greedily assigned from each named ship to either partition. This resulted in 539 image pairs and 334 singletons for training, and 549 image pairs and 358 singletons for testing. In our experiments, we only use the visible ship imagery category. Followed the same pre-process of image as deep convolutional neural network (CNN) algorithm in [48], we resize each crop to the size using bicubic interpolation. Note that the original images in these two experimental datasets are color images, the images are converted from the RGB color space to the YCbCr color space, and the Y component (luminance) is used for ship classification.

4 Experimental results and analysis

4.1 Parameters setting

First of all, we investigate optimal parameters for each type of features. For the first one (i.e., the Gabor-based MS-CLBP), according to [30], the orientations θ of Gabor filtering

set as $\left[0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\right]$ and the spatial frequency bandwidth set 5 for both two experimental datasets. The result of Gabor filtering of a sample image is shown in Fig. 9. Then we estimate the optimal parameter pair (m, r) for the CLBP operator. In this experiment, we vary the CLBP parameter sets, and fix parameters for others. The classification results are listed in Tables 1 and 2 for two experimental datasets, respectively.

Since the number of neighbors (i.e., m) and the radius (i.e., r) directly impact the dimensionality of the CLBP histogram features. For example, larger m will increase the feature dimensionality and computational complexity. Based on the results in Tables 1, 2 and 3, we choose $(m, r) = (10, 6)$ for the BCCT200-resize dataset; $(m, r) = (12, 7)$ for the original BCCT200 dataset and $(m, r) = (10, 5)$ for the VAIS dataset in terms of classification accuracy as well as computational complexity. Thus, the dimensionality of the CLBP features (CLBP_S and CLBP_M histograms concatenated) for both the BCCT200-resize dataset and the VAIS dataset is set to 216.

For the second type of features (i.e., patch-based MS-CLBP and FV), since the feature extracts dense local patches, the size of patch $(B \times B)$ is studied. Varying patch sizes B result varying number of local patches, therefore the number of Gaussians (i.e., K) in the GMM is studied simultaneously. In the experiment, B is chosen from a reasonable region, and Fig. 10 illustrates the classification results with varying patch sizes B as well as different numbers
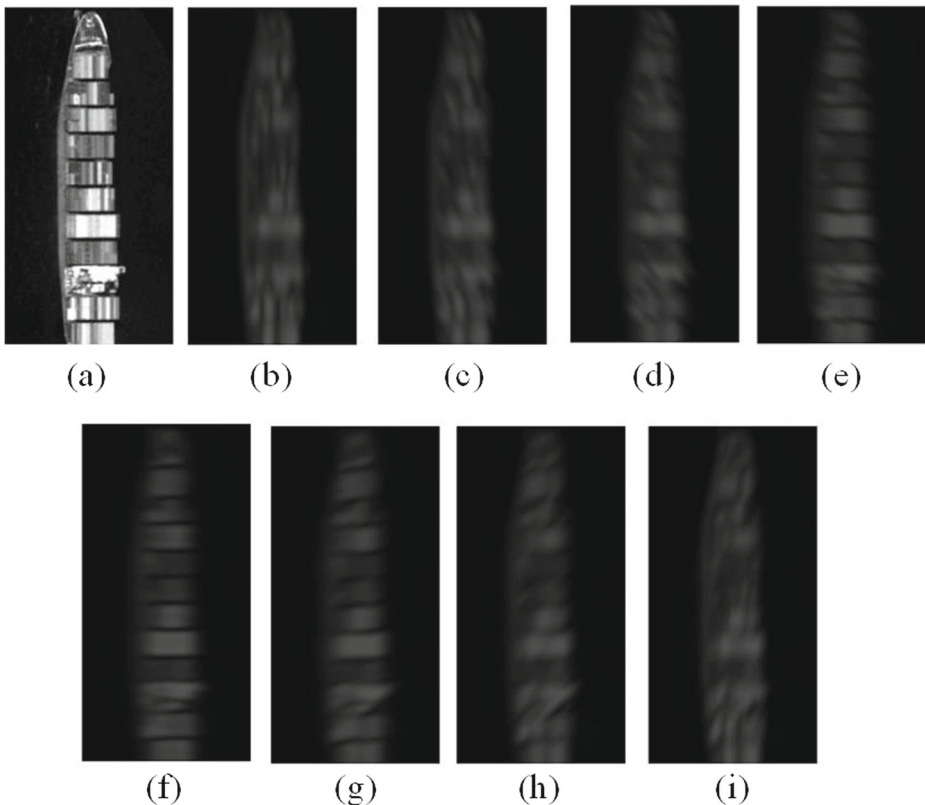
Fig. 9 Example of Gabor filter: **a** input image, **b–i** 8 Gabor filtered images corresponding to 8 different orientations for the input image

Table 1 Classification accuracy (%) of with the first feature different parameters (m, r) for the BCCT200-resize dataset

r	1	2	3	4	5	6	7	8
$m = 8$	86.88	89.38	90.00	91.25	90.63	90.00	90.00	90.63
$m = 10$	88.75	88.75	90.00	90.00	91.25	93.13	91.88	91.88
$m = 12$	86.88	89.38	90.63	90.63	91.88	91.88	90.00	90.00

of Gaussians K in the GMM for the two experimental datasets. Optimal parameters of B and K are further determined from the tuning results in Fig. 10. That is, the optimal patch size B is 28 and the number of Gaussians K is 20 for the BCCT200-resize dataset; as for the original BCCT200 dataset, the optimal patch sizes B is 22 and the number of Gaussians K is 25. For the VAIS dataset, the optimal patch size B and the number of Gaussians K both be set to 20. For the MS-CLBP parameters, the scale is set to $1/[1:3]$ and another parameter pair is chosen $(m, r) = (6, 3)$, empirically.

For the third type of features (i.e., the BOVW and SPM), the size of block (i.e., $\varpi \times \varpi$) is studied for the dense SIFT descriptors. The experiment results with varying patch sizes for the two experimental dataset are illustrated in Fig. 11. It is obvious that optimal block size ϖ is 24 for both BCCT200-resize dataset and original BCCT200 dataset, and $\varpi = 26$ achieves the best classification performance for the VAIS dataset. For k-means clustering, blocks are randomly selected and the codebook size is set to 1024.

4.2 Classification results and analysis

The proposed MFL strategy is compared with some state-of-the-art methods under the same experimental setup to verify the effectiveness of the proposed multiple feature learning approach. That is, for the BCCT200-resize dataset, 80% of the images from each class are used for training and the remaining images are used for testing; for the VAIS dataset, its "official" train and test data is used for training and testing, respectively. For the BCCT200-resize dataset, five-fold cross-validation is performed, in which the dataset is randomly partitioned into five equal and disjoint subsets. For each run, a different subset is held out for testing and the remaining four are used for training. The classification accuracy is the average over the five cross-validation evaluations. Specifically, the proposed method is compared with the feature fusion method that combines Gabor feature and MS-CLBP feature. The improvement of the proposed method over the Gabor + MS-CLBP and other existing methods verifies our multiple feature learning method is more effective. The comparison result for the BCCT200-resize, BCCT200 and VAIS datasets are shown in Tables 4, 5 and 6, respectively. Obviously, the proposed method achieves superior classification performance over other existing methods, which demonstrates the effectiveness of the

Table 2 Classification accuracy (%) of with the first feature different parameters (m, r) for the original BCCT200 dataset

| r	1	2	3	4	5	6	7	8
$m = 8$	65.63	71.25	74.38	73.13	73.13	73.75	72.50	71.25
$m = 10$	63.13	68.13	68.13	69.38	70.63	71.25	72.50	75.50
$m = 12$	65.00	67.50	70.00	75.00	77.50	76.88	77.50	75.63

Table 3 Classification accuracy (%) of with the first feature different parameters (m, r) for the VAIS dataset

r	1	2	3	4	5	6	7	8
$m = 8$	71.20	73.47	73.73	75.33	75.33	73.60	73.07	70.00
$m = 10$	73.20	74.53	75.20	76.27	77.73	77.47	76.27	74.40
$m = 12$	72.27	74.27	75.33	76.53	77.33	77.33	76.13	73.87

proposed MFL for vessel classification. Especially, for BCCT200-resize dataset, the MFL gains about 10% higher overall accuracy than deep learning method and the MFL gains about 4% higher overall accuracy for VAIS dataset than Convolutional Neural Networks (CNN). Therefore, the proposed approach, which combines multiple hand-crafted features, has obvious advantages. Moreover, excepting of typical SVM, we employ another classifier (i.e. extreme learning machine, ELM [21]) under the MFL framework. We find a phenomenon from Tables 4, 5 and 6 that the feature-level strategy is preferable for VAIS dataset and the decision-level strategy is preferable for BCCT200-resize and BCCT200 datasets. For various ship image datasets, the decision-level and feature-level strategies have their own advantages and disadvantages, and it is hard to determine which one is always optimal.
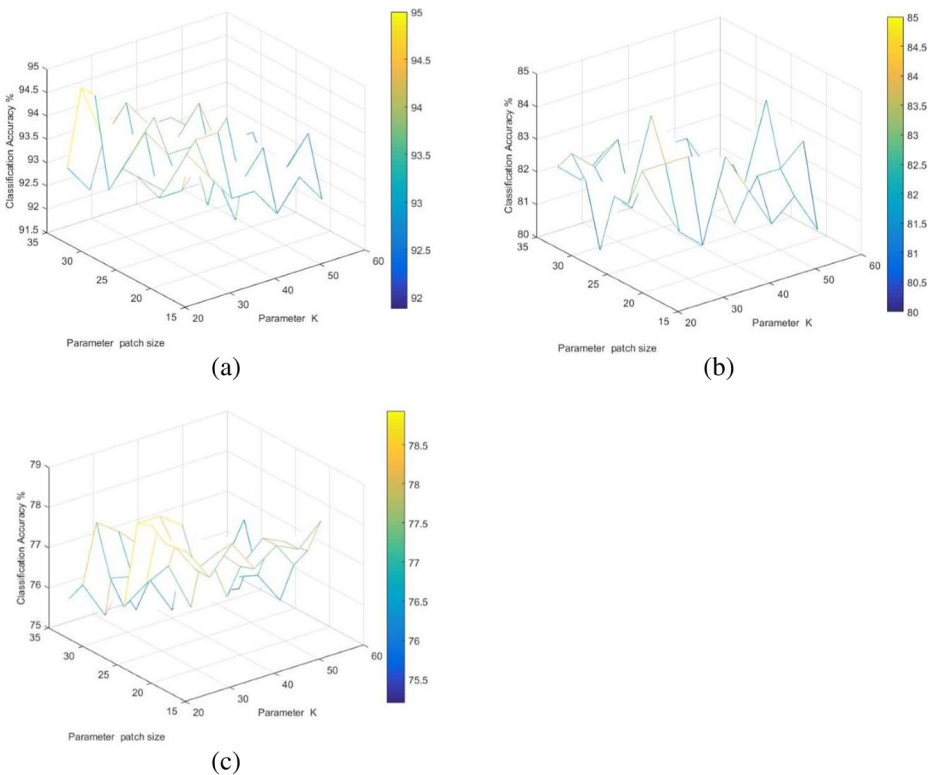
Fig. 10 Parameter tuning of and for the second feature using three experimental data sets: **a** the BCCT200-resize, **b** the original BCCT200, **c** the VAIS
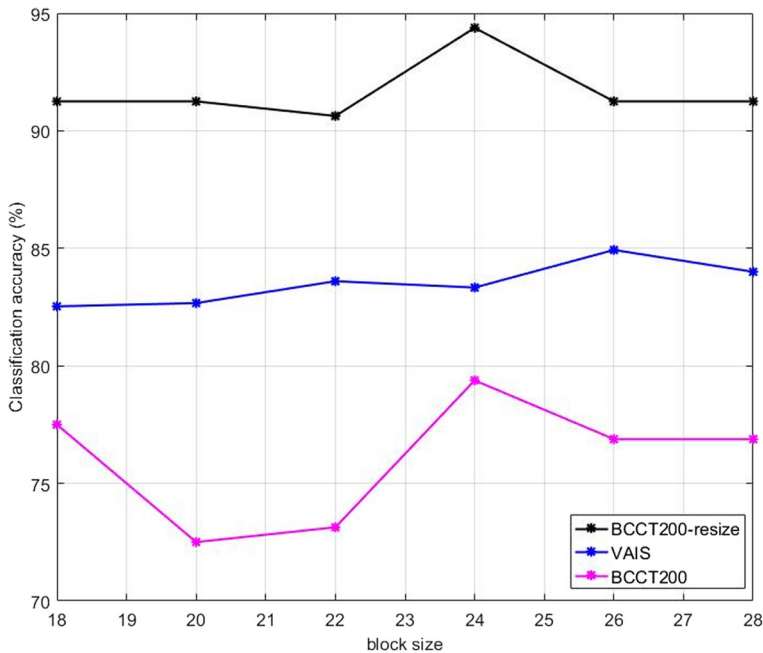
Fig. 11 Classification accuracy (%) versus varying block sizes for the three experimental dataset

In addition, to demonstrate the enhanced discriminative power of multiple features fusion strategy, we compare the classification performance of the proposed multiple feature learning approach with the performance of methods that use each individual features in this framework with LIBSVM classifier. The accuracy per class from the aforementioned methods is reported in Tables 7, 8 and 9. It is apparent that the proposed MFL outperforms all the individual feature based approaches. For the BCCT200-resize dataset, the global feature representation method, i.e., Gabor + MS-CLBP, achieves highest accuracy for cargo category. The first local feature representation method, i.e., MS-CLBP+FV, achieves maximum accuracy for container category in the original BCCT200 dataset. For the VAIS dataset, the second local feature representation method, i.e., BOVW+SPM, achieves maximum accuracy for Medium passenger category. Nevertheless, the proposed MFL obtains superior accuracy for others classes and also the highest overall accuracy for both two experimental datasets.

Confusion matrix of the proposed MFL with feature-level fusion strategy for the BCCT200-resize dataset is listed in Table 10, and that of the proposed MFL with decision-level fusion strategy is listed in Table 11. The confusion matrixes also are the average over the five cross-validation evaluations. The major confusion occurs between class 2 (i.e., cargo) and class 3 (i.e., container), since some cargo images is similar to the container images. Tables 12 and 13 show the confusion matrix of feature-level fusion and decision-level fusion for the original dataset, respectively. It is obvious that the major confusion occurs within class 3 (i.e., container), class 4 (i.e., tanker), and class 2 (i.e., cargo). Tables 14 and 15 display the confusion matrix of feature-level fusion and decision-level fusion for the VAIS dataset, respectively. It is observed that the major confusion occurs within class 2 (i.e., sailing), class 1 (i.e., merchant), class 3 (i.e., Medium passenger), and class 5 (i.e.,

Table 4 Comparison of classification accuracy (%) with some state-of-the-art methods for the BCCT200-resize dataset

Method	Accuracy(%)
PCA + SVM [35]	77.1
LDA + SVM [35]	74.1
ICA + SVM [35]	77.4
MPCA +SVM [35]	79.1
BOW + SVM [35]	76.8
HOG +SVM [35]	81.6
hierarchical multiscale LBP (HMLBP) [35]	90.8
Sparse Representation with Feature Vectors(SRFV) [36]	78.25
HyperNEAT [40]	75.00
Deep Learning HyperNEAT [41]	83.70
Gabor + MS-CLBP	90.63
MFL (feature-level) + ELM	94.38
MFL (decision-level) + ELM	94.63
MFL (feature-level) + SVM	94.00
MFL (decision-level) + SVM	94.63

Table 5 Comparison of classification accuracy (%) with some state-of-the-art methods for the original BCCT200 dataset

Method	Accuracy(%)
PCA + NN [40]	74.5
V-BOW [28]	78.50
BOW + SVM [35]	76.3
BOW + SRC-L_0 [35]	74.8
BOW + SRC-L_1 [35]	75.6
Gabor + MS-CLBP	76.00
MFL (feature-level) + ELM	85.88
MFL (decision-level) + ELM	85.88
MFL (feature-level) + SVM	85.88
MFL (decision-level) + SVM	86.87

Table 6 Comparison of classification accuracy (%) with some state-of-the-art methods for the VAIS dataset

Method	Accuracy(%)
Gnostic Field [48]	82.4
CNN [48]	81.9
Gnostic Field + CNN [48]	81.0
Gabor + MS-CLBP	77.73
MFL (feature-level) + ELM	87.60
MFL (decision-level) + ELM	85.07
MFL (feature-level) + SVM	85.33
MFL (decision-level) + SVM	85.07

Table 7 Classification accuracy (%) per class for the BCCT200-resize dataset

class	Classification methods				
	Gabor + MS-CLBP	MS-CLBP + FV	BOVW + SPM	MFL (feature-level)	MFL (decision-level)
Barge	98.00	99.50	100	100	99.50
Cargo	85.00	81.50	78.50	82.00	83.50
Container	87.50	94.50	96.50	97.00	97.00
Tanker	92.00	97.00	94.50	97.00	98.50
OA (%)	90.63	93.13	92.38	94.00	94.63

Table 8 Classification accuracy (%) per class for the original BCCT200 dataset

class	Classification methods				
	Gabor + MS-CLBP	MS-CLBP + FV	BOVW + SPM	MFL (feature-level)	MFL (decision-level)
Barge	84.00	94.50	89.00	94.50	95.00
Cargo	68.00	72.00	61.00	74.00	76.00
Container	79.50	88.50	84.50	85.50	87.50
Tanker	72.50	87.50	80.00	89.50	89.00
OA (%)	76.00	85.62	78.63	85.88	86.87

Table 9 Classification accuracy (%) per class for the VAIS dataset

class	Classification methods				
	Gabor + MS-CLBP	MS-CLBP + FV	BOVW + SPM	MFL (feature-level)	MFL (decision-level)
Merchant	73.24	70.42	67.61	78.87	77.46
Sailing	41.86	50.00	48.84	51.16	44.19
Medium passenger	47.19	61.29	95.16	77.42	82.26
Medium other	90.40	89.39	94.95	96.97	96.97
Tug	83.07	86.58	90.73	91.69	92.65
Small	50.00	60.00	65.00	65.00	60.00
OA (%)	77.73	78.80	84.53	85.33	85.07

Table 10 Confusion matrix of the proposed MFL (feature-level fusion) for the BCCT200-resize dataset

Class	1	2	3	4	Recall(%)
Barge (1)	40	0	0	0	100
Cargo (2)	1.2	32.8	4.4	1.6	82.00
Container (3)	0.8	0.4	38.8	0	97.00
Tanker (4)	0.8	0	0.4	38.8	97.00
Precision (%)	93.46	98.80	88.99	96.04	

Table 11 Confusion matrix of the proposed MFL (decision-level fusion) for the BCCT200-resize dataset

Class	1	2	3	4	Recall(%)
Barge (1)	39.8	0	0	0.2	99.50
Cargo (2)	0.8	33.4	3.6	2.2	83.50
Container (3)	0.6	0.6	38.8	0	97.00
Tanker (4)	0.2	0.2	0.2	39.4	98.50
Precision (%)	96.14	97.66	91.08	94.26	

Table 12 Confusion matrix of the proposed MFL (feature-level fusion) for the original BCCT200 dataset

Class	1	2	3	4	Recall(%)
Barge (1)	37.8	0	0.6	1.6	94.50
Cargo (2)	1	29.6	4.6	4.8	74.00
Container (3)	1	2	34.2	2.8	85.50
Tanker (4)	1.2	2.4	0.6	35.8	89.50
Precision (%)	92.2	87.06	85.5	79.56	

Table 13 Confusion matrix of the proposed MFL (decision-level fusion) for the original BCCT200 dataset

Class	1	2	3	4	Recall(%)
Barge (1)	38	0	0.4	1.6	95.00
Cargo (2)	1.4	30.4	4	4.2	76.00
Container (3)	0.8	2.4	35	1.8	87.50
Tanker (4)	1.4	2.2	0.8	35.6	89.00
Precision (%)	91.35	86.86	87.06	82.41	

Table 14 Confusion matrix of the proposed MFL (feature-level fusion) for the VAIS dataset

Class	1	2	3	4	5	6	Recall(%)
Merchant(1)	56	2	1	2	3	7	78.87
Sailing(2)	9	44	14	4	14	1	51.16
Medium passenger(3)	3	2	48	0	4	5	77.42
Medium other(4)	0	2	0	192	4	0	96.97
Tug(5)	1	6	16	2	287	1	91.69
Small(6)	0	1	6	0	0	13	65.00
Precision (%)	81.16	77.19	56.47	96.00	91.99	48.15	

Table 15 Confusion matrix of the proposed MFL (decision-level fusion) for the VAIS dataset

Class	1	2	3	4	5	6	Recall(%)
Merchant(1)	55	1	1	3	4	7	77.46
Sailing(2)	11	38	11	5	20	1	44.19
Medium passenger(3)	4	0	51	0	5	2	82.26
Medium other(4)	0	1	0	192	5	0	96.97
Tug(5)	1	4	15	2	290	1	92.65
Small(6)	0	0	6	0	2	12	60.00
Precision (%)	77.46	86.36	60.71	95.05	88.96	52.17	

Table 16 Classification accuracy (%) of the proposed MFL using different fusion rules

Fusion rule	Data		
	BCCT200-resize	BCCT200	VAIS
MV (hard fusion)	94.00	85.75	84.13
LOGP (soft fusion)	94.63	86.87	85.07

tug), or between class 6 (i.e., small) and class 3 (i.e., Medium passenger). Sailing ships contain small sails up, small sails down and large sails down. Some sails down ships are similar to merchant ships. The small ships include speedboat, jet-ski, smaller pleasure and larger pleasure. Thus some small ships and medium passenger ships have relatively high similarity.

To evaluate the effectiveness of the decision-level fusion strategy (i.e., LOGP) of the proposed framework, we compare it with a popular hard fusion rule of majority voting [21]. Table 16 provides the classification results comparison between these two fusion schemes (i.e., LOGP and MV) using two experimental datasets. As evident from the results, the soft fusion strategy provides slightly better performance than the hard fusion rule of MV. The reason is that the fusion at the posterior probability level provides more flexibility than the hard counting, especially for multi-class classification tasks.

5 Conclusion

In this paper, we proposed a multiple features learning classification framework for ship classification in optical remote sensing imagery. The overall classification framework consists of multiple features including Gabor-based MS-CLBP, patch-based MS-CLBP and Fisher Vector, and combination of the BOVW and SPM. These global and local features are complementary to each other and the combination of them is a powerful and comprehensive representation of ship images. It was found that the proposed method became more discriminative than the entire individual feature based approaches. Experimental results on two optical vessel dataset verified that the proposed decision-level soft fusion classification method consistently achieves superior classification performance over other state-of-the-art algorithms.

Given the recent tremendous success of deep learning technique, especially CNN, in image classification, a combine of deep learning features and hand crafted features for ship classification will be investigated in our future work.

References

1. Arguedas VF (2015) Texture-based vessel classifier for electro-optical satellite imagery. In: IEEE international conference on image processing (ICIP), 2015, pp 3866–3870
2. Barnum J (1986) Ship detection with high-resolution hf skywave radar. IEEE J Ocean Eng 11(2):196–209
3. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13(6):1450–1464
4. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
5. Bosch A, Zisserman A, Munoz X (2007) Image classification using random forests and ferns. In: 2007 IEEE 11th international conference on computer vision, pp 1550–5499

6. Chang C, Lin C (2011) Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):389–396

7. Chen C, Zhang B, Su H, Li W, Wang L (2015) Land-use scene classification using multi-scale completed local binary patterns. Signal Image and Video Processing 1–8

8. Clausi DA, Jernigan ME (2000) Designing gabor filters for optimal texture separability. Pattern Recogn 33(11):1835–1849

9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 886–893

10. Guo W, Xia X, Wang X (2014) A remote sensing ship recognition method based on dynamix probablity generative model. Expert Syst Appl 41(14):6446–6458

11. Guo W, Xia X, Wang X (2015) A remote sensing ship recognition method for entropy-based hierarchical discriminant regression. Optik-International Journal for Light and Electron Optics 126(20):2300–2307

12. Guo W, Xia X, Wang X (2015) Variational approximate inferential probability generative model for ship recognition using remote sensing data. Optik-International Journal for Light and Electron Optics 126(23):4004–4013

13. Guo Z, Zhang L, Zhang D (2010) A completed modeling of local binary pattern operator for texture classification. IEEE Trans Image Process 19(6):1657–1663

14. Guo Z, Zhang L, Zhang D, Mou X (2010) Hierarchical multiscale lbp for face and palmprint recognition. In: 2010 IEEE international conference on image processing, Hong Kong, pp 4521–4524

15. Harguess J, Rainey K (2011) Are face recognition methods useful for classifying ships? In: Applied imagery pattern recognition workshop, pp 1–7

16. Huang L, Chen C, Li W, Du Q (2016) Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors. Remote Sens 8(6)

17. Huang S, Xu H, Xia X (2016) Active deep belief networks for ship recognition based on BvSB. Optik-International Journal for Light and Electron Optics 24:11688–11697

18. Jaakkola TS, Haussler D (1998) Exploiting generative models in discriminative classifiers. Adv Neural Inf Proces Syst 11(11):487–493

19. Jun Y, YuGang J, Hauptmann AG, Chong Wah N (2007) Evaluating bag-of-visual-words representations in scene classification. In: ACM Sigmm international workshop on multimedia information retrieval, pp 197–206

20. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol 2, pp 2169–2178

21. Li W, Chen C, Su H, Du Q (2015) Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans Geosci Remote Sens 53(7):3681–3693

22. Li W, Prasad S, Fowler JE (2014) Decision fusion in kernel-induced spaces for hyperspectral image classification. IEEE Trans Geosci Remote Sens 52(6):3399–3411

23. Liu C (2003) Maximum likelihood estimation from incomplete data via em-type algorithms. Advanced Medical Statistics 1051–1071

24. Lowe DG (2004) Distinctive image features from scale-invariant key points. Int J Comput Vis 60(2):91–110

25. Lu H, Plataniotis KN, Anastasios N (2008) Mpca: Multilinear principal component analysis of tensor objects. IEEE Trans Neural Netw 19(1):18–39

26. Margarit G, Milanes J, Tabasco A (2009) Operational ship monitoring system based on synthetic aperture Radar processing. Remote Sens 1(3):375–392

27. Ojala T, Pietikainen M, Maenpaa TT (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

28. Parameswaran S, Rainey K (2015) Vessel classification in overhead satellite imagery using weighted bag of visual words. In: Automatic target recognition XXV

29. Park S, Cho CJ, Ku B, Lee S, Ko H (2016) Simulation and ship detection using surface radial current observing compact HF radar. IEEE J Ocean Eng 99:1–12

30. Peng B, Li W, Xie X, Du Q, Liu K (2015) Weighted-fusion-based representation classifiers for hyperspectral imagery. Remote Sens 7

31. Perronnin F, Dance C (2007) Fisher kernels on visual vocabularies for image categorization. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8

32. Prasad S, Bruce LM (2008) Decision fusion with confidence-based weight assignment for hyperspectral target recognition. IEEE Trans Geosci Remote Sens 46(5):1448–1456

33. Prasad S, Li W, Fowler JE, Bruce LM (2012) Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification. IEEE Trans Geosci Remote Sens 50(9):3474–3486
34. Rainey K, Parameswaran S, Harguess J (2014) Maritime vessel recognition in degraded satellite imagery. Proc SPIE-Int Soc Opt Eng 9090(9):909004–909004–6
35. Rainey K, Parameswaran S, Harguess J, Stastny J (2012) Vessel classification in overhead satellite imagery using learned dictionaries. In: Proceedings of SPIE-the international society for optical engineering, vol 8499, pp 84992–12
36. Rainey K, Stastny J (2011) Object recognition in ocean imagery using feature selection and compressive sensing. In: 2011 IEEE applied imagery pattern recognition workshop, pp 1–6
37. Sanchez J, Perronnin F, Mensink T, Verbeek J (2013) Image classification with the fisher vector: theory and practice. Int J Comput Vis 105(3):222–245
38. Teng F, Liu Q (2015) Robust multi-scale ship tracking via multiple compressed features fusion. Signal Process Image Commun 31:76–85
39. Toh KA, Kim J, Lee S (2008) Biometric scores fusion based on total error rate minimization. Pattern Recogn 41(3):1066–1082
40. Verbancsics P, Harguess J (2015) Feature learning hyperneat: evolving neural networks to extract features for classification of maritime satellite imagery. Springer International Publishing, pp 208–220
41. Verbancsics P, Harguess J (2015) Image classification using generative neuro evolution for deep learning. In: 2015 IEEE winter conference on applications of computer vision, pp 1550–5790
42. Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2008) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
43. Xin H, Zhang L (2013) An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans Geosci Remote Sens 51(1):257–272
44. Xu Y, Lu Y (2015) Adaptive weighted fusion: a novel fusion approach for image classification. Neurocomputing 168:566–574
45. Xu Y, Zhang B, Zhong Z (2015) Multiple representations and sparse representation for image classification. Pattern Recogn Lett 68:9–14
46. Xu Y, Zhu X, Li Z, Liu G, Lu Y, Liu H (2013) Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition. Pattern Recogn 46(4):1151–1158
47. Yang G, Liu H, Yu X (2007) Hyperspectral remote sensing image classification based on kernel fisher discriminant analysis. In: 2007 international conference on wavelet analysis and pattern recognition, vol 3, pp 1139–1143
48. Zhang MM, Choi J, Daniilidis K, Wolf MT, Kanan C (2015) Vais: a dataset for recognizing maritime imagery in the visible and infrared spectrums. In: 2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 10–16

Longhui Huang received her B.S. degree from Beijing University of Chemical Technology, Beijing, China, in 2014. She is currently pursuing the M. S. degree at the same university. Her supervisor is Dr. Wei Li.

Wei Li received his Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, in 2012. Subsequently, he spent one year as a postdoctoral researcher at the University of California, Davis. He is currently with the College of Information Science and Technology at Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Chen Chen received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009 and the MS degree in electrical engineering from Mississippi State University, Starkville, in 2012 and the PhD degree in the Department of Electrical Engineering at the University of Texas at Dallas, Richardson, TX in 2016. He is currently a Post-Doc in the Center for Research in Computer Vision at University of Central Florida (UCF). His research interests include compressed sensing, signal and image processing, pattern recognition and computer vision. He has published over 40 papers in refereed journals and conferences in these areas.

Fan Zhang received his Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2008. He is currently an associate professor in the College of Information Science and Technology at Beijing University of Chemical Technology, Beijing, China. His research interests includes synthetic aperture radar, raw data simulation, and parallel processing.

Haitao Lang currently an associate professor within the Faculty of Science, Beijing University of Chemical Technology. His research interests includes synthetic aperture radar, and ship recognition.