Multimedia Tools and Applications (2019) 78: 8397–8422
https://doi.org/10.1007/s11042-018-6835-2

Ensemble classifier based source camera identification using fusion features

Bo Wang[1] · **Kun Zhong**[1] · **Ming Li**[1]

Received: 25 July 2018 / Revised: 12 October 2018 / Accepted: 30 October 2018 /
Published online: 6 November 2018

Abstract
Source camera identification, which means identifying the camera source of a given image, has become one of the most important branches of digital image forensics. In order to improve the detection accuracy, the feature dimensions used in existing methods are increasing, and consequently Support Vector Machine (SVM) seems no longer applicable. In this paper, an ensemble classifier is introduced into to source camera identification, which uses the fusion features to capture software-related, hardware-related, and statistical characteristics left on the images by digital camera. Experimental results indicate that the proposed method can achieve near 100% accuracy for camera brand and model identification, and also outperforms the baseline methods in identifying different camera individuals.

Keywords Source camera identification · Ensemble classifier · Fusion features

1 Introduction

With the development of Internet technologies and the increasing popularization of digital cameras, digital image has become one of the main communication modes, which greatly promotes the development of society and facilitates the people's lives. However, with the emergence of a variety of powerful image editing software, people can use these software to tamper and fake the images arbitrarily, which makes the authentication and verification of a given digital content become vital. And as a silent witness, digital images have played an important role in forensic evidence and criminal investigation. Therefore, the study of digital image forensics technology is of great significance to social stability and justice.

According to whether the information is embedded in advance, digital image forensics can be divided into passive image forensics and active image forensics. Active image forensics embeds the identification information, i.e. digital watermarking, into the original image in advance, and uses the information to determine whether the image content has been modified [25]. Image passive forensics only uses the test image itself, without access to the

✉ Bo Wang
bowang@dlut.edu.cn

[1] School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, Liaoning, China

original information related to the image. Digital image forensics can also be divided into source camera identification, forgery detection [34, 35] and steganalysis [19, 36] according to its purpose. This paper focuses on blind-passive source camera identification, which aims to identify the source cameras of the given images. Further, digital image source identification includes three aspects:

1) The device-based identification: For a digital image, it is necessary to first identify the device type that generates the image. The common types of imaging devices include digital cameras, mobile phones, scanners, and computers.
2) The model-based identification: For imaging devices such as digital cameras, mobile phones and scanners, there are a wide variety of device manufacturers and models. Therefore, after determining the type of imaging device, identification and acquisition of its collection equipment model is an important part of digital image source identification.
3) The camera individual based identification: In some cases, it is important to identify the different device individuals of the same brand and model. For example, the forensic party has to identify certain device individual of a given image to determine whether it really comes from the digital camera held by a suspect.

2 Related work

Most of the state-of-art methods can be categorized into the following three classes: the methods based on the inherent characteristics of hardware caused by defects in manufacturing processes, the methods based on the software-related characteristics left by different image processing algorithms, the methods based on image statistical features.

(1) Intrinsic hardware characteristics.

The cameras inevitably have some internal imperfections because of the limitations of manufacturing process and the materials used. Lukas et al. [17] pioneered a source camera identification method based on sensor pattern noise in 2006. Sensor pattern noise, which is mainly composed of fixed pattern noise (FPN) and photo-response non-uniformity noise (PRNU), is caused by the inhomogeneity of the image sensor and the defects of the sensor manufacturing process. The main component of sensor pattern noise is photo-response non-uniformity noise, which was found to be one of the distinctive characteristics to identify the images captured from different cameras [12]. In [17], the given images firstly were denoised using wavelet filter, which was presented by Mhak [21]. Then the residual noise images can be obtained by subtracting the denoised images from the original images, the main component of which is the pattern noise. In order to judge the source of the given images, it is necessary to calculate the correlation between pattern noise of the images and the reference pattern noise of different cameras, where the reference pattern noise can be obtained by calculating the average of the multiple residual noise images taken by the camera. Finally, the camera with the highest correlation is the source of the test images. Although the method is simple and effective, it heavily relies on the purity of the extracted pattern noise. Unfortunately the pattern noise is generally affected by image content, JPEG compression, random noise and other factors. To solve this problem, many researchers have proposed different methods to obtain more accurate pattern noise. A method was proposed to enhance the pattern noise by weighting the noise components to reduce the interference of the scene information [15]. Kang et al. [13] enhanced source camera identification performance with a camera reference phase sensor. Chan et al. [5] proposed a method based

on confidence map and pixel-based weighted correlation to get pure pattern noise. Wu et al. [32] proposed a spatial adaptive noise reduction method based on wavelet transform. In this work, the effect of spatial complexity on pattern noise was considered to reduce the effect of content information on the extracted pattern noise and the experimental results proved that the method achieves better performance compared to other existing methods. However, since JPEG compression will affect the correlation between pattern noise and reference pattern noise, the method based on pattern noise correlation is not robust to the images after JPEG compression. Besides, in order to calculate the reference pattern noise of the tested camera, it is essential to obtain a large number of flat images taken by the camera, which is very difficult in most cases. Therefore Akshatha et al. [1] proposed a novel source camera identification method based on PRNU feature. After the pattern noise was obtained by using the method [32], the high frequency wavelet coefficients of the residual noise image were extracted to compose the feature vectors for classification. Experiments show that the method has a higher accuracy even if the test images are JPEG compressed. In addition to using sensor pattern noise, other hardware characteristics of the camera can also be used for source camera identification. Choi et al. [26] used the intrinsic lens radial distortion and Van et al. [30] used chromatic aberration to identify the source camera of the given images.

(2) Software-related characteristics.

In addition to the inherent hardware characteristics of the camera, the software-related characteristics left in images in the process of image processing also can be used for source camera identification. When the camera shutter is pressed, the natural light gathers to the optical sensor through the lens and optical filter for imaging, and then a series of image processing algorithms are carried out to obtain the output image. The common image processing algorithms are color filter array (CFA) interpolation algorithm, white balance, camera response function, JPEG compression and so on. Any camera model has a unique image processing algorithm, even if the cameras made by same manufacturer do not have the same image processing algorithms and algorithm parameters. Therefore, it is possible to identify the source camera using the traces left by camera image processing algorithms. Among these methods, the most widely used are the algorithms based on CFA interpolation. Because of the high cost of production, most of the consumer-level digital cameras use a single sensor, which can only get a single color image. In order to obtain a color image, it is necessary for each camera to carry out the CFA interpolation. However, the CFA interpolation algorithms used by different camera manufacturers are generally different, so the traces left by different interpolation algorithms can be used to identify source camera. This idea was first proposed by Bayram et al. [3]. The basic idea of this method is that the CFA interpolation kernel and the interpolation algorithm are different for different camera, which will make the related structure of the spatial pixel values different. Then, the Expectation-Maximization (EM) algorithm is used to extract the peak location, magnitudes and the image weighting coefficients in the spectrum as the eigenvectors. Finally, these features are sent to the SVM to get the results. Since severe compression will destroy the spatial correlation properties between pixels caused by CFA interpolation, the method can not apply to severely compressed images. In order to solve the problem, Bayram improved the method [27] by extracting the periodicity of the second-order derivatives. Long et al. [16] proposed a second-order model based on pixel correlation, and then estimated the CFA interpolation coefficients for each color channel and sent them to the BP neural network for classification. In [28], the interpolation coefficients of small blocks were estimated by texture classification and linear approximation, and then the CFA pattern of digital camera with minimum interpolation error was obtained. Although the methods based on CFA

interpolation have achieved good results, these methods have some limitations. First of all, the complexity of these methods is usually high, which is time-consuming. Secondly, these methods are not suitable for severely compressed images. Many researchers have also proposed source camera identification methods based on other software-related characteristics. For example, Swaminathan et al. proposed a method based on white balance [29], in which, the white balance parameters are estimated as features of the source camera identification. Hany Farid et al. [7] proposed a method based on the blind inverse gamma correction.

(3) Image statistical characteristics.

As mentioned above, even if the image signal acquisition and internal processing process of digital camera system are basically the same, different digital camera algorithms and parameters design will make the image statistical characteristics different, which can turn the identification problem into a simple feature classification problem. Kharrazi et al. [14] proposed a blind source camera identification method based on features in 2004, which is one of the first papers to focus on differences in statistical characteristics of images taken by different cameras rather than the specific camera artifacts. The 34-dimensional features that can reflect the characteristics of the internal signal processing process of digital cameras were used in this method. These features were consisted of average pixel value, color channel correlation, energy ratio, Image Quality Metrics (IQM) (firstly proposed in [2]) and so on, and then were fed to a SVM classifier for classification. Xu et al. [33] proposed a source camera identification method based on the texture features, which mainly refer to Local Binary Pattern (LBP) features and Local Phase Quantization (LPQ) features. Ojala et al. [22] first proposed the LBP operator to describe the local texture of the image, which has many advantages such as rotation invariance, gray scale invariance, and so on. The LPQ operator was first presented by Ojansivu et al. [23] to classify the blur and insensitive texture images. In [33], the original images were denoised to obtain the residual noise images, and the contourlet transformation was performed on the residual noise images. Finally, the LBP features and LPQ features were extracted from original images, the residual noise images, and the contourlet transformation, respectively and were fed to a SVM classifier for source camera identification. The experimental results shown that the combination of LBP and LPQ features can achieve higher accuracy and more robust than single feature. Francesco et al. [20] studied the co-occurrences based local features, which were proposed originally for steganalysis, and used them to train a classifier to identify source camera.

From the brief review above, it can be seen that more and more methods select high-dimensional features or fusion features for source camera identification to improve the detection accuracy. The superiority of the feature fusion is obvious. Different features extracted from the test image always can capture the different characteristics left by the camera. By combining these different features, more available discriminant information can be used, which is important for classification. Most of the existing algorithms based on fusion features selected a variety of image statistical features such as color features, Image Quality Matrices, LBP and Co-occurrence Matrix to form the feature set for training. These methods, however, can not effectively identify different camera individuals of the same model. In this paper, we propose a source camera identification method based on fusion features. We believe that the use of fusion features can capture more traces left in the captured images, which can get higher detection accuracy. Besides, most of the existing source camera identification methods are based on SVM classifier. Although this classifier is convenient and its detection accuracy is high, the complexity of the SVM classifier increases with the feature dimensionality and the size of training set, which is time-consuming. In this paper, we extract high-dimensional feature set to get a high accuracy, in which case, SVM classifier

seems no longer appropriate. Inspired by [8], in the paper, the features extracted are fed to the ensemble classifier implemented as random forest to achieve higher accuracy with less time.

The rest of the paper is as follows. The proposed method and features used in this paper will be described in detail in Section 3. A large number of comparative experiments will be carried out to verify the performance of the method in Section 4. The last are the conclusions.

3 Proposed method

According to the above analysis, in this section we propose a source camera identification method based on fusion features, which can capture software-related, hardware-related and statistical characteristics left by different cameras. The fusion features used in this paper are composed of CFA features, HOWS features, LBP, LPQ features and co-occurrence matrix features. Firstly, CFA features can capture artifacts left by different CFA interpolation algorithms and have been widely used in source camera identification. The methods based on artifacts left by other post-processing operations such as JPEG compression, white balance, etc. are not mature enough. Therefore CFA features are selected in this paper to extract software-related artifacts. Secondly, LBP features, LPQ features and co-occurrence matrix features are all image texture features, which can be obtained by counting the relationship between adjacent pixels, but they are different from each other. LBP features and LPQ features describe the relationship between neighboring pixel size and center pixel size in the spatial domain and frequency domain respectively, while co-occurrence matrix features describe the histogram information of four adjacent pixels in the horizontal and vertical directions. The statistical calculation of adjacent pixels makes these texture features robust to noise. Therefore, these three different texture features are combined in this paper to extract image statistical characteristics. In addition, capturing camera-specific hardware characteristics is essential for the feature set used to improve the detection accuracy for different camera individuals of the same brand. In this paper, a high-order statistical model based on wavelet decomposition, HOWS is extracted from the PRNU to form the complete feature set because of its better performance than the basic statistical features, such as color features and IQM features [11]. The proposed method can be divided into two parts: fusion features extraction and classification. Firstly, the hardware-related features, software-related features and image-related features are extracted from the given images. Then, these extracted features are fused into the ensemble classifiers to train a model to identify the source camera. Next, I will introduce the features used and their extraction process in detail.

3.1 Feature set

In this paper, the hardware-related features mainly refer to the higher order wavelet statistics (HOW) features based on PRNU noise, the software-related features mainly refer to the CFA features and the image-related features are consisted of three different image texture features. These features are extracted to form a 3410-dimensional feature set for classification.

3.1.1 HOWS features based on PRNU noise

Photo response non-uniformity (PRNU) noise is the main component of the pattern noise, and it is impossible for different cameras to have the same PRNU noise. Therefore, the

statistical features of PRNU noise have strong resolving capability for images captured by different cameras. Higher order wavelet statistics (HOWS) is a statistical model for the analysis of natural images based on the multi-scale wavelet decomposition, which was first proposed by Lyu et al. [18]. In this paper, we extract the HOWS features of PRNU noise to capture the traces left by inherent hardware characteristics of different cameras. The extraction process of the features is mainly composed of the following parts: PRNU noise estimation and feature extraction. The block diagram of the process is shown in Fig. 1.

The most important part of this process is PRNU noise estimation. Firstly, the given images are denoised using a wavelet-based filter, which was first proposed by Celiktutan in work [4]. In this paper, the script provided by Binghampton University is used to denoise the given images, which can reduce the influence of image scene information on PRNU noise get the better estimation of PRNU noise. The acquisition of residual noise images can be represented by the following formula:

$$I^k_{noise} = I^k - I^k_{denoised} \tag{1}$$

where I^k represents the original images, $I^k_{denoised}$ represents the images after wavelet denoising and I^k_{noise} denotes the residual noise images, the main component of which is PRNU noise.

Then the HOWS features are extracted to capture the statistical characteristics of PRNU noise existing in pattern noise. The process of HOWS features extraction is shown in Fig. 2.

For the residual noise images, the multi-scale wavelet decomposition is carried out for each color channel by Quadrature Mirror Filters (QMF), so that the frequency space can be divided into horizontal (LH), vertical (HL) and diagonal (HH) sub-band at each level. Then the statistical characteristics such as skewness, means, variance and kurtosis of each sub-band coefficients are calculated to form the HOWS features.

In this work, three-levels wavelet decomposition is carried out, and nine sub-bands are obtained, which brings about 108-dimensional HOWS features.

3.1.2 LBP, LPQ features

Texture is an important visual clue, which widely exists in images and is hard to describe. Image texture features can be divided into many kinds, such as local binary pattern (LBP), local phase quantization (LPQ), co-occurrence matrix and so on. As one of the image statistical features, the image texture features can be used to identify the source camera and achieve high accuracy [33]. In this section, the LBP features and LPQ features are extracted as part of the fusion features. LBP features, LPQ features, and their extraction method are briefly described below.

LBP operator was first proposed by Ojala et al. [22] to describe the local texture features of the images, which has many advantages such as rotation invariance, gray scale invariance, etc. The original LBP operator is defined in a 3×3 window, where the center pixel is the threshold and then compare the surrounding pixels with the threshold. If the surrounding pixel value is greater than the threshold, the position of the pixel is marked as 1, otherwise

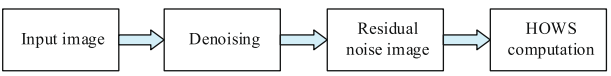

Fig. 1 The block diagram of feature extraction

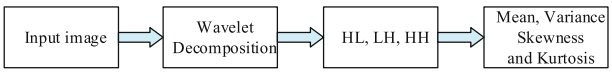

Fig. 2 The block diagram of HOWS computation

0. In this way, a string of 8-bit binary number can be obtained, which is the LBP value of the center pixels. The encoding process is shown in Fig. 3.

The LPQ operator was first presented by Ojansivu et al. [23] to classify the blur and insensitive texture images and it has the following characteristics: (1) preferable ability of classification for blur-insensitive image; (2) ideal rotational invariance. These characteristics make the LPQ operator have good robustness, which is important for source camera identification. In the process of image processing, the spatial ambiguity is usually expressed by the convolution between the image intensity and the point spread function (PSF), which can be expressed by the formula $G = O \cdot H$ in the frequency domain. Where G is the Discrete Fourier Transforms (DFT) of fuzzy image, O is the original image, H is the point spread function (PSF). The LPQ operator can be obtained by quantifying the phase spectrum of DFT.

For example, the 2-dimensional DFT of the image $f(x)$ can be expressed by the following formula:

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u T_y} \tag{2}$$

where u is the frequency. The Fourier coefficients of the following four frequency points are calculated separately: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$, $u_4 = [a, -a]^T$. Then the real parts and imaginary parts of the Fourier coefficients can be quantized:

$$q_j = \begin{cases} 1, & g_j \geq 0 \\ 0, & g_j < 2 \end{cases} \tag{3}$$

where g_j represents the j-th part of $G(x) = [\mathrm{Re}\{F(x)\}, \mathrm{Im}\{F(x)\}]$. Then the LPQ operator can be obtained:

$$f_{LPQ}(x) = \sum_{j=1}^{8} q_j 2^{j-1} \tag{4}$$

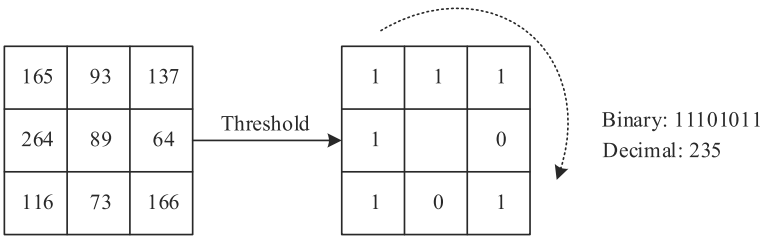

Fig. 3 The encoding process of LBP

Firstly, the original images are denoised to obtain the residual noise images, and then the contourlet transformation is performed on the residual noise images. The reason for selecting contourlet transform is that the features obtained from contourlet transformation generally perform better than the features obtained from wavelet transformation due to the diversity of directions [24]. Finally, the LBP features and LPQ features are extracted from the original images, the residual noise images, and the contourlet transformation, respectively. The eigenvector is extracted from H and V color channels, which has been proved to have a better performance in [33]. The final eigenvector includes 708-dimensional LBP features and 236-dimensional LPQ features.

3.1.3 CFA features

CFA interpolation is a very important part of the imaging process, the interpolation algorithm also directly affects the statistical characteristics of digital images. Since the different CFA interpolation algorithms are used by different brands of cameras, CFA features are an important part of the feature set used. In this paper, the method based on covariance matrix proposed in our previous work [31] is used to estimate the CFA interpolation coefficients. Taking the G color channel as an example, the estimation process of CFA interpolation coefficients is as follows. In the $(2k+1)\times(2k+1)$-neighborhood, (m,n) is the position of the point to be interpolated. The general model of CFA interpolation can be expressed as:

$$G_{m,n}=\sum_{i=-k}^{k}\sum_{j=-k}^{k}a_{i,j}^{g}g_{m+i,n+j}\Big|_{a_{0,0}^{g}=0}+\sum_{i=-k}^{k}\sum_{j=-k}^{k}a_{i,j}^{r}r_{m+i,n+j}+\sum_{i=-k}^{k}\sum_{j=-k}^{k}a_{i,j}^{b}b_{m+i,n+j} \tag{5}$$

where $a_{i,j}^{g}$, $a_{i,j}^{r}$ and $a_{i,j}^{b}$ represent the weight coefficient of interpolation, respectively. g, r and b represent the pixel values of the green, red and blue channels, respectively. For N pixels to be interpolated, the raster line scanning are carried out on all pixels in the neighborhood, then the formula (5) can be represented by the following vector:

$$\begin{aligned}\begin{pmatrix}G_1 & G_2\cdots G_N\end{pmatrix}^T=&\begin{pmatrix}a_1^g g_1^1+\cdots+a_{(2k+1)^2-1}^g g_{(2k+1)^2-1}^1\\ a_1^g g_1^2+\cdots+a_{(2k+1)^2-1}^g g_{(2k+1)^2-1}^2\\ \vdots\\ a_1^g g_1^N+\cdots+a_{(2k+1)^2-1}^g g_{(2k+1)^2-1}^N\end{pmatrix}\\&+\begin{pmatrix}a_1^r r_1^1+\cdots+a_{(2k+1)^2}^r r_{(2k+1)^2}^1\\ a_1^r r_1^2+\cdots+a_{(2k+1)^2}^r r_{(2k+1)^2}^2\\ \vdots\\ a_1^r r_1^N+\cdots+a_{(2k+1)^2}^r r_{(2k+1)^2}^N\end{pmatrix}\\&+\begin{pmatrix}a_1^b b_1^1+\cdots+a_{(2k+1)^2}^r b_{(2k+1)^2}^1\\ a_1^b b_1^2+\cdots+a_{(2k+1)^2}^r b_{(2k+1)^2}^2\\ \vdots\\ a_1^b b_1^N+\cdots+a_{(2k+1)^2}^r b_{(2k+1)^2}^N\end{pmatrix}\end{aligned} \tag{6}$$

where g_s^t represents the s-th G-channel coefficients of the raster scanning performed on the neighborhood of the t-th pixel to be interpolated, and the definition of r_s^t and b_s^t are similar to g_s^t. The formula (6) also can be expressed as:

$$\begin{aligned}\left(G_1 \cdots G_N\right)^T = & a_1^g \left(g_1^1 \cdots g_1^N\right)^T + \cdots + a^g_{(2k+1)^2-1} \begin{pmatrix} g^1_{(2k+1)^2-1} \\ \vdots \\ g^N_{(2k+1)^2-1} \end{pmatrix} \\ & + a_1^r \begin{pmatrix} r_1^1 \\ \vdots \\ r_1^N \end{pmatrix} + \cdots + a^r_{(2k+1)^2} \begin{pmatrix} r^1_{(2k+1)^2} \\ \vdots \\ r^N_{(2k+1)^2} \end{pmatrix} \\ & + a_1^b \begin{pmatrix} b_1^1 \\ \vdots \\ b_1^N \end{pmatrix} + \cdots + a^b_{(2k+1)^2} \begin{pmatrix} b^1_{(2k+1)^2} \\ \vdots \\ b^N_{(2k+1)^2} \end{pmatrix} \end{aligned} \tag{7}$$

$$\begin{aligned}\mathbf{G} = & a_1^g \mathbf{g}_1 + \cdots + a^g_{(2k+1)^2-1} \mathbf{g}_{(2k+1)^2-1} + a_1^r \mathbf{r}_1 + \cdots + a^r_{(2k+1)^2} \mathbf{r}_{(2k+1)^2} \\ & + a_1^b \mathbf{b}_1 + \cdots + a^b_{(2k+1)^2} \mathbf{b}_{(2k+1)^2} \end{aligned} \tag{8}$$

where $\mathbf{g}_i$, $\mathbf{r}_i$ and $\mathbf{b}_i$ represent the vectors that consists of the pixels in the fixed scan positions of G, R and B channels respectively. Theoretically, $3 \times (2k+1)^2 - 1$ pixels in the same channel can be used to establish the equations to solve $3 \times (2k+1)^2 - 1$ interpolation coefficients. However, due to the diversity of the digital image content, the influence of other subsystems in the imaging process and the lossy compression in the image post-processing, there will be inevitable errors in the CFA interpolation coefficient estimation. Therefore the formula (8) can be modified as:

$$\begin{aligned}\mathbf{G} = & a_1^g \mathbf{g}_1 + \cdots + a^g_{(2k+1)^2-1} \mathbf{g}_{(2k+1)^2-1} + a_1^r \mathbf{r}_1 + \cdots + a^r_{(2k+1)^2} \mathbf{r}_{(2k+1)^2} \\ & + a_1^b \mathbf{b}_1 + \cdots + a^b_{(2k+1)^2} \mathbf{b}_{(2k+1)^2} + \boldsymbol{\varepsilon} \end{aligned} \tag{9}$$

where $\boldsymbol{\varepsilon}$ is the interference term vector in the interpolation process. Considering the robustness and stability of the results, the covariance is used to establish the equations. According to the homogeneity of covariance, the following formula can be obtained:

$$\begin{pmatrix} \operatorname{cov}\left(\mathbf{G}, \mathbf{g}_1\right) \\ \operatorname{cov}\left(\mathbf{G}, \mathbf{g}_2\right) \\ \vdots \\ \operatorname{cov}\left(\mathbf{G}, \mathbf{g}_{(2k+1)^2-1}\right) \end{pmatrix} = \begin{pmatrix} a_1^g \operatorname{cov}\left(\mathbf{g}_1, \mathbf{g}_1\right) + \cdots + a^g_{(2k+1)^2-1} \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_1\right) \\ a_1^g \operatorname{cov}\left(\mathbf{g}_1, \mathbf{g}_2\right) + \cdots + a^g_{(2k+1)^2-1} \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_2\right) \\ \vdots \\ a_1^g \operatorname{cov}\left(\mathbf{g}_1, \mathbf{g}_{(2k+1)^2-1}\right) + \cdots + a^g_{(2k+1)^2-1} \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_{(2k+1)^2-1}\right) \end{pmatrix} \tag{10}$$

Finally, the estimation of the interpolation weight coefficient $\mathbf{a}^g$ can be expressed as:

$$\begin{aligned}
&\left(a_1^g \quad a_2^g \cdots a_{(2k+1)^2-1}^g\right)^T \\
&= \begin{pmatrix}
\operatorname{cov}(\mathbf{g}_1, \mathbf{g}_1) & \operatorname{cov}(\mathbf{g}_2, \mathbf{g}_1) & \cdots & \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_1\right) \\
\operatorname{cov}(\mathbf{g}_1, \mathbf{g}_2) & \operatorname{cov}(\mathbf{g}_2, \mathbf{g}_2) & \cdots & \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_2\right) \\
\vdots & \ddots & \ddots & \vdots \\
\operatorname{cov}\left(\mathbf{g}_1, \mathbf{g}_{(2k+1)^2-1}\right) & \operatorname{cov}\left(\mathbf{g}_2, \mathbf{g}_{(2k+1)^2-1}\right) & \cdots & \operatorname{cov}\left(\mathbf{g}_{(2k+1)^2-1}, \mathbf{g}_{(2k+1)^2-1}\right)
\end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix}
\operatorname{cov}(\mathbf{G}, \mathbf{g}_1) \\
\operatorname{cov}(\mathbf{G}, \mathbf{g}_2) \\
\vdots \\
\operatorname{cov}\left(\mathbf{G}, \mathbf{g}_{(2k+1)^2-1}\right)
\end{pmatrix}
\end{aligned} \tag{11}$$

The estimation of the weight coefficient $\mathbf{a}^r$, $\mathbf{a}^b$ of neighborhood pixels in R and B channels can be obtained by a similar way. In this paper, we fix $k = 3$. So the number of interpolation coefficients to be estimated for each color channel is $(2 \times 3 + 1)^2 - 1 = 48$. Since Bayer CFA is the most popular CFA model at present, we uses it as the analytical model. According to the basic structure of 2×2 Bayer CFA, we estimate the interpolation coefficients of the following 5 kinds of interpolation points: G interpolation coefficients of R and B sampling points and R, B interpolation coefficients of two different G sampling points in Bayer CFA, so that $48 \times 5 = 240$ interpolation coefficients can be obtained. In order to capture more discriminant information of the CFA interpolation coefficient estimation, the variance of the estimation is also taken as another group of features. Therefore, a total of 480-dimensional CFA features are extracted in this paper.

3.1.4 Co-occurrence matrix features

Co-occurrence matrix features were proposed originally by Jessica et al. for steganalysis in [9], which is based on co-occurrences of residual image. The local features extracted from the neighborhood of the pixel of the images play an important role in source camera identification. In [20], the co-occurrence was computed on residual image as discriminative features for classification and a series of experiments had proved the feasibility of the feature. In this paper, the co-occurrence based local features are extracted to form the feature set. The extraction process can be divided into the following steps: calculation of residuals, quantization and truncation, calculation of co-occurrence [20]. In the following, the each step of the extraction process will be described in detail.

1) Calculation of residuals.

Image residuals are widely used in most existing methods of source camera identification. In fact, by doing so, the impact of image content on the results will be minimized, which make it easier to detect artifacts. In work [9], many linear and non-linear filters have been proposed to calculate the image residuals, and each filter is named according to its own construction. The naming rule of these filters are as follows:

$$\text{name} = \{type\}\{f\}\{\sigma\}\{scan\} \tag{12}$$

where type is the filter type, which can be spam and minmax, spam refers to the linear filter and minmax is a combination of minimun and maximum operators, f denotes the number of filters, σ is the index of symmetry, and scan is the direction of filters used, which can be h (horizontal), v (vertical) and so on. For example, spam12hv denotes a high-pass filter consisting of a SPAM type filter with a symmetry index of 2 and two scanning directions: horizontal and vertical. Among these filters, we can select some filters to calculate the image residuals for the source camera identification, such as second order, third order linear filter and minmax filter. The calculation of the residuals using high-pass filters is as follows:

$$r_{i,j} = \hat{x}_{i,j}(N_{ij}) - cx_{i,j} \tag{13}$$

where c is the residual order, $x_{i,j}$ denotes the pixel value of the image, N_{ij} denotes the local neighborhood of $x_{i,j}$ and $\hat{x}_{i,j}(.)$ denotes the predictor of $cx_{i,j}$. For example, the calculation of the residuals for second order filter is as follows:

$$r^h_{i,j} = x_{i,j-1} - 2x_{i,j} + x_{i,j+1} \tag{14}$$

$$r^v_{i,j} = x_{i-1,j} - 2x_{i,j} + x_{i+1,j} \tag{15}$$

where $r^h_{i,j}$ denotes the residual image corresponding to the horizontal direction, $r^v_{i,j}$ denotes the residual image corresponding to the vertical direction. However, due to symmetry, $r^h_{i,j}$ and $r^v_{i,j}$ have the same statistical properties [9], so they can be concatenated in residual r to simplify the method.

2) Quantization and truncation.

In order to reduce the dynamic range of the image residuals and obtain feasible co-occurrence matrices, it is necessary to quantize and truncate on the image residuals as shown below:

$$\hat{r}_{i,j} = trunc_T(round(\frac{r_{i,j}}{q})) \tag{16}$$

where $q > 0$ is the quantization step and $trunc_T$ is the truncation function, which is defined as:

$$trunc_T = \begin{cases} x & |x| \le T \\ sign(x) \cdot T & |x| > T \end{cases} \tag{17}$$

3) Calculation of co-occurrences.

Then the co-occurrences can be obtained by the following formula:

$$C^{\mathrm{h}}(k_0, k_1, k_2, k_3) = \sum_{i,j} I(\hat{r}_{i,j} = k_0, \hat{r}_{i,j+1} = k_1, \hat{r}_{i,j+2} = k_2, \hat{r}_{i,j+3} = k_3) \tag{18}$$

$$C^{\mathrm{v}}(k_0, k_1, k_2, k_3) = \sum_{i,j} I(\hat{r}_{i,j} = k_0, \hat{r}_{i+1,j} = k_1, \hat{r}_{i+2,j} = k_2, \hat{r}_{i+3,j} = k_3) \tag{19}$$

where $k_0, k_1, k_2, k_3 \in \{-T, ..., T\}$ and we set $T = 2$ in this work. $I(X)$ is the function of event X, and defined as:

$$I(X) = \begin{cases} 1 & true \\ 0 & false \end{cases} \tag{20}$$

The co-occurrence matrix features are calculated on each color channel (R, G and B) of the images, respectively. For horizontal co-occurrence C^{h} and vertical co-occurrence C^{v}, each co-occurrence have 625 elements, which can be reduced to 313 because of the symmetry. The residual calculation of other filters is similar to the second order filter s2_spam12hv, and we only consider the residuals obtained from s2_spam12hv in this paper. So the final feature dimension is $3 \times 2 \times 313 = 1878$.

3.2 Ensemble classifier

As mentioned above, the most accurate source camera identification methods are built as supervised classifiers on the features extracted from the images, where support vector machine (SVM) seems to be the most popular choice. This is due to the fact that SVM is a classification model based on statistical learning theory and has good generalization ability. However, with the increase of the feature dimension and the training samples, the training time of SVM will be greatly increased because the training complexity of SVM is directly proportional to the square of the product of the training samples and the feature dimension [8]. In this paper, a high dimensional feature set is extracted to improve the detection accuracy, in which case the use of SVM is time-consuming. Inspired by [8], the ensemble classifier instead of SVM is used in this paper to get a high detect accuracy with less time. The main idea of the ensemble classifier is to combine the base classifiers, which have the independent decision-making ability. The conceptual diagram of the ensemble classifier is shown in Fig. 4. The ensemble classifier proposed is composed of many base learners, which can save training time. For a sample of test set, the final result is obtained by fusing the result of each base learner. The ensemble learning strategy can work only if each base learner has the sufficient diversity. In order to increase the diversity of each base learner, two strategies are used in this paper.

Firstly, for a given dataset D with M samples, we sample it to generate another dataset $D^{'}$: We randomly select a sample from D, put a copy of it into $D^{'}$, and then put the sample back into the original dataset D, so that the sample can still be acquired at the next sampling. This process is repeated M times, then you can get a dataset $D^{'}$ containing M samples. This strategy called 'bootstrap sampling' is first proposed in work [6]. Suppose the number of base learners is T. The process of bootstrap sampling will be repeated T times on the whole training set to generate the different training sets for each base learner, and by doing so, the diversity of each base learner will be increased. This strategy called 'bagging' is the most famous representative of parallel ensemble learning methods. Since each base learner uses only about 63.2% of the samples in the whole training set, the remaining 36.8% of the samples can be used as test set to estimate the generalization performance of base learners, which is known as 'out-of-bag' (OOB) estimate. The OOB estimate will plays an important role in the selection of optimal parameters, which will be introduced later. The OOB estimate can be expressed as:

$$E^{oob} = \frac{1}{|D|} \sum_{(x,y)\in D} I(H^{oob}(x) \neq y) \tag{21}$$

where $D = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$ is the whole training set and x, y represent the features and labels, respectively. $|D|$ represents the number of elements in set D. $I(a)$ is the

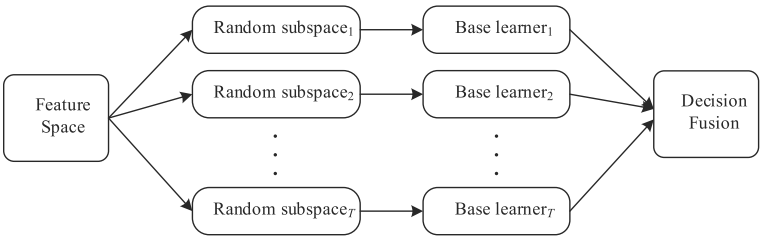

Fig. 4 Diagram of the ensemble classifier

index function of event a, equal to 1 if a is true and 0 otherwise. $H^{oob}(x)$ is the prediction of the base learners that do not use x for training and can be expressed as:

$$H^{oob}(x) = \arg\max_{y \in Y} \sum_{t=1}^{T} I(h_t(x) = y) \cdot I(x \notin D^{'}) \tag{22}$$

where y is the set of training set labels, $h_t(x)$ represents the prediction of the t base learner.

Secondly, in order to further increase the diversity of base learners, the original feature space is randomly divided into various subspaces with the same dimension d_{sub}, which are used to train the base learners to get the models for classification. The difference between the training samples and the feature space used by the base learners can make the base learner be diverse enough to ensure the effectiveness of the ensemble learning. In [8], the Fisher Linear Discriminant (FLD) was selected as base learner to solve binary classification problem because of its low complexity. However, the training complexity of FLD greatly increases when dealing with the multi-classification problem. In order to reduce the training complexity of the ensemble classifier, we choose the decision tree as the base classifiers. The decision tree can randomly select a small set of the original attribute as splitting attribute, which brings additional diversity to the base learners and reduces the training complexity. Then the final decision can be obtained by fusing the decisions of every base learner. The most common fusion strategy for classification is the majority voting, which can be expressed as:

$$H(x) = \begin{cases} y_j & if \sum_{i=1}^{T} h_i^j(x) > 0.5 \sum_{k=1}^{N} \sum_{i=1}^{T} h_i^k(x); \\ reject & otherwise \end{cases} \tag{23}$$

where $\{y_1, y_2, ..., y_N\}$ is the set of labels, $h_i^j(x)$ is the prediction of i-th base learner on y_j. There are many other fusion strategies can be used in ensemble learning such as weighted voting, stacking and so on. However, we observe that the three fusion strategies give the comparable results. So, considering the trade-off between training complexity and accuracy, we choose the simplest fusion strategy-majority voting.

In the construction of ensemble classifier, we add two random processes to increase the diversity of the base learners, which bring two parameters to be determined: The number of base learners T and subspace dimension d_{sub}. As mentioned above, the OOB estimate is an unbiased estimation of generalization error and can be used to determine the optimal parameters. The OOB estimate will change with the change of T and d_{sub}, so we design two stopping criterions to find T and d_{sub} corresponding to the minimum OOB estimate.

4 Experimental results

In this section, we will introduce the experimental set and results of our work. In order to verify the performance of the proposed method, we select a set of camera models to carry out the experiment from 'Dresden Image Dataset' [10], which is one of the most widely used database in digital image forensics. As shown in Table 1, the camera group used for the experiment consists of 12 different cameras, including 4 different brands of cameras such as Agfa, Canon, Nikon and Sony, each brand has two different camera models. In addition, the different camera individuals from same brand and same model such as two individuals of Sony DSC-H50, two individuals of Sony DSC-W170 and so on are also included to test the effectiveness of the proposed method in the worst situation. For each camera model we

Table 1 The parameters of cameras and images

Make	Model	Image resolution	Format	Notation
Agfa	DC-504	4032×3024	JPEG	AD1
Agfa	DC-733	3072×2304	JPEG	AD2
Canon	Ixus55	2592×1944	JPEG	C1
Canon	PS640	3648×2736	JPEG	C2
Nikon	D70(1)	3008×2000	JPEG	ND1
Nikon	D70(2)	3008×2000	JPEG	ND2
Nikon	D200(1)	3872×2592	JPEG	ND3
Nikon	D200(2)	3872×2592	JPEG	ND4
Sony	DSC-H50(1)	3456×2592	JPEG	SD1
Sony	DSC-H50(2)	3456×2592	JPEG	SD2
Sony	DSC-W170(1)	3648×2736	JPEG	SD3
Sony	DSC-W170(2)	3648×2736	JPEG	SD4

used 150 full size images, of which 90 images for training and 60 images for testing. The proposed method is implemented in MATLAB 2014a, and our experiments are carried out by using the computer with Inter (R) Xeon (R) 2.4 GHz CPU, 48 GB RAM. We designed several experiments to test the performance of our method, the details of the experimental set are as follows:

4.1 Experimental sets

As described, the method proposed in this paper can identify different camera brands, different camera models of the same brand, and different camera individuals of same brand and same model with higher accuracy. In order to demonstrate the superiority of the method in this field, we design four groups of experiments as shown below:

Group 1: Camera brand identification.

In this group, twelve different cameras listed in Table 1 are divided into four groups according to their brands, namely Agfa, Canon, Nikon and Sony. For each camera, 90 images are used for training, and the remaining 60 images are used for testing.

Group 2: Camera model identification.

Table 2 The comparison results of Group1

Camera brand	Proposed method	HOWS [1]	LBP,LPQ [33]	Co-occurrence [20]
Agfa	100	99.3	100	100
Canon	100	99.4	98.64	98.1
Nikon	100	100	100	100
Sony	100	99.7	100	100
Average accuracy	100	99.6	99.7	99.6

Table 3 Confusion matrix between two camera models of Agfa

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		AD1	AD2	AD1	AD2	AD1	AD2	AD1	AD2
Actual	AD1	100	0	100	0	100	0	100	0
	AD2	0	100	0	100	0	100	0.6	99.4
Average accuracy		100		100		100		99.7	

Table 4 Confusion matrix between two camera models of Sony

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		SD1	SD3	SD1	SD3	SD1	SD3	SD1	SD3
Actual	SD1	97.6	2.4	89.3	10.7	92.1	7.9	96.0	4.0
	SD3	2.0	98.0	4.5	95.5	1.1	98.9	2.8	97.2
Average accuracy		97.8		92.4		95.5		96.6	

Table 5 Confusion matrix between two camera models of Canon

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		C1	C2	C1	C2	C1	C2	C1	C2
Actual	C1	100	0	99.4	0.6	99.4	0.6	100	0
	C2	0	100	3.9	96.1	7.3	92.7	0	100
Average accuracy		100		97.7		96.1		100	

Table 6 Confusion matrix between two camera models of Nikon

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		ND1	ND3	ND1	ND3	ND1	ND3	ND1	ND3
Actual	ND1	100	0	95.9	4.1	100	0	99.4	0.6
	ND2	0	100	0.5	99.5	2.4	97.6	0	100
Average accuracy		100		97.7		98.9		99.7	

Table 7 Confusion matrix between two camera individuals of Nikon D70

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		ND1	ND2	ND1	ND2	ND1	ND2	ND1	ND2
Actual	ND1	62.6	37.4	59.2	40.8	41.5	58.5	33.7	66.3
	ND2	37.7	62.3	48.6	51.4	31.6	68.4	27.9	72.1
Average accuracy		62.5		55.3		54.9		52.9	

Table 8 Confusion matrix between two camera individuals of Nikon D200

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		ND3	ND4	ND3	ND4	ND3	ND4	ND3	ND4
Actual	ND3	95.8	4.2	88.3	11.7	85.6	14.4	88.3	11.7
	ND4	10.6	89.4	11.5	88.5	10.1	89.9	12.3	88.7
Average accuracy		92.3		88.4		87.7		88.0	

Each of the four different brands of cameras used in the experiment contained two different models of cameras. We separately classify different models of cameras for each brand to verify the performance of the method in identifying different models of cameras.

Group 3: Camera individual identification.

It is very difficult to identify different camera individuals of the same brand and model because of their similarities. In this group, different camera individuals of Sony DSC-H50, Sony DSC-W170, Nikon D70 and Nikon D200 are used for experiments.

Group 4: Classification of twelve cameras.

In this group, twelve different cameras are used, including two cameras of Agfa, two cameras of Canon, four cameras of Nikon and four cameras of Sony. Finally, the confusion matrix of the classification results of twelve cameras will be obtained, which can effectively prove the comprehensive performance of the method.

For the guarantee of the universality of the experimental results, each experiment are repeated 20 times. The training samples and test samples chosen for each experiment are randomly selected and there is no overlap between them. The final accuracy of the experiment is the average of 20 experimental results. Besides, in order to prove the superiority of the proposed method compared with other existing methods, the methods proposed in work [1, 33] and [20] are used as comparison methods. Work [1] is an effective method based on PRNU features, work [33] is one of the most advanced methods based on LBP features and work [20] is a recently proposed method based on co-occurrence matrix features.

4.2 Accuracy comparison

Group 1 The purpose of this set of experiments is to verify the performance of the proposed method in identifying different brands of cameras. And for that, the twelve cameras listed in Table 1 are divided into four groups according to their brands, namely Agfa, Canon, Nikon and Sony. As mentioned above, we choose 90 images from each camera for training, and the

Table 9 Confusion matrix between two camera individuals of Sony DSC-H50

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		SD1	SD2	SD1	SD2	SD1	SD2	SD1	SD2
Actual	SD1	91.9	8.1	84.0	16.0	73.2	26.8	68.0	32.0
	SD2	8.7	91.3	15.2	84.8	10.4	89.6	14.4	85.6
Average accuracy		91.6		84.4		81.4		76.8	

Table 10 Confusion matrix between two camera individuals of Sony DSC-W170

		Proposed method		Hows [1]		LBP, LPQ [33]		Co-occurrence [20]	
		SD3	SD4	SD3	SD4	SD3	SD4	SD3	SD4
Actual	SD3	81.2	18.8	52.8	47.2	62.0	38.0	75.4	24.6
	SD4	16.2	83.8	18.0	82.0	13.2	86.8	20.8	79.2
Average accuracy		82.5		67.4		74.4		77.3	

remaining 60 images are used for testing. For each image, we extract the HOWS features, LBP and LPQ features, CFA features and co-occurrence matrix features separately to form the feature set, and then send them to the ensemble classifier for training or testing. For comparison, the experimental setup of the methods proposed in work [1, 33] and [20] are exactly the same as that mentioned above.

The comparison results of four different brands of cameras are shown in Table 2. It can be seen that the identification accuracies of the proposed method are 100% for all camera brands used in experiment, which are superior to the methods proposed in [1, 33] and [20]. It can also be seen from the experimental results that different features have their own advantages when identifying camera brands, such as HOWS features have the highest detection accuracy for Nikon, while LBP, LPQ and co-occurrence matrix features are easier to identify other cameras. Therefore, the combination of these features makes the proposed method have better identification ability for all camera brands.

Group 2 The purpose of this set of experiments is to verify the performance of the proposed method in identifying different models of cameras. And to do this, we select 8 different camera models from 4 different brands, including two camera models of Agfa, two camera models of Canon, two camera models of Nikon and two camera models of Sony. For each camera, the numbers of training images and testing images are same as Group1. Similarly, we extract the HOWS features, LBP and LPQ features, CFA features and co-occurrence matrix features separately and feed them to the ensemble classifier. The method in work [1, 33] and [20] are baselines.

As can be seen from Tables 3, 4, 5 and 6, the average accuracies of the proposed method are up to 100% for different camera models of Agfa, Nikon and Canon, which are superior to that of the methods [1, 33] and [20]. In this group of experiments, the image texture features (LBP, LPQ and co-occurrence) perform better than the HOWS features, which greatly improve the ability of the proposed method to detect camera models. Besides, it is noteworthy that the average accuracy of the proposed method is reduced when identifying different models of Sony. This is a common phenomenon in most methods, which means that the two models of Sony have great similarities and are difficult to identify.

Group 3 The proposed method not only can identify different brands, different models, but also can effectively identify different camera individuals of the same brand and model. In order to verify the performance of the method in identifying different camera individuals, we take two camera models of Sony and Nikon respectively, in which each camera model contains two different individuals. The experimental setup and baselines are the same as the first two groups.

The experimental results are shown in Tables 7, 8, 9 and 10. It can be concluded that the performance of the proposed method is better than that of method [1, 20] and [33] when

Table 11 Confusion matrix of the Group4 by using proposed method

Average accuracy	Predicted(%)											
=89.5%	AD1	AD2	C1	C2	ND1	ND2	ND3	ND4	SD1	SD2	SD3	SD4
AD1	99.6	1.2	0	0	0	2.2	0	0	0	0	0	0
AD2	0	100	0	0	0	0	0	0	0	0	0	0
C1	0	0	100	0	0	0	0	0	0	0	0	0
C2	0	0	0	100	0	0	0	0	0	0	0	0
ND1	0.4	0	0	0	64.9	34.7	0	0	0	0	0	0
ND2	0	0	0	0	36.0	64.0	0	0	0	0	0	0
ND3	0	0	0	0	0	0	95.4	4.6	0	0	0	0
ND4	0	0	0	0	0	0	8.4	91.6	0	0	0	0
SD1	0	0	0	0	0.4	0	0	0	99.2	0.4	0	0
SD2	0	0	0	2.2	0	0	0	0	4.2	91.0	0	2.6
SD3	0	0	0	0	0	0	0	0	0.9	0	87.3	11.8
SD4	0	0	0	0	0	0	0	0	0	2.3	16.3	81.4

Table 12 Confusion matrix of the Group4 by using method [1]

Average accuracy	Predicted(%)											
=78.4%	AD1	AD2	C1	C2	ND1	ND2	ND3	ND4	SD1	SD2	SD3	SD4
AD1	96.6	1.2	0	0	0	2.2	0	0	0	0	0	0
AD2	0	97.2	0.4	0.2	2.0	0.2	0	0	0	0	0	0
C1	0	0	94.6	5.4	0	0	0	0	0	0	0	0
C2	0	0.4	3.6	95.6	0.2	0.2	0	0	0	0	0	0
ND1	0	0.8	0	0	58.2	40.8	0.2	0	0	0	0	0
ND2	0	1.0	0	0	37.6	61.4	0	0	0	0	0	0
ND3	0	0	0.4	0	1.6	0.4	86.6	11.0	0	0	0	0
ND4	0	0	0	0	0.2	1.4	5.0	93.4	0	0	0	0
SD1	0	0	0	0	0	0.6	0	0	78.2	6.0	9.2	6.0
SD2	0	0.8	0	0	0	0	0	0	7.4	61.4	6.6	23.8
SD3	0	0	0	0	0	0	0	0	6.6	5.8	61.8	25.8
SD4	0	0.8	0	0	0	0	0	0	6.4	20.6	17.0	55.2

Table 13 Confusion matrix of the Group4 by using method [33]

Average accuracy	Predicted(%)											
=87.8%	AD1	AD2	C1	C2	ND1	ND2	ND3	ND4	SD1	SD2	SD3	SD4
AD1	96.0	0	0	0	0	2.8	0	0	1.2	0	0	0
AD2	0	100	0	0	0	0	0	0	0	0	0	0
C1	0	0	100	0	0	0	0	0	0	0	0	0
C2	0	0	3.6	95.6	0	0.4	0	0	0	0.4	0	0
ND1	0	0	0	0	67.2	32.8	0	0	0	0	0	0
ND2	1.6	0	0	0	34.8	63.6	0	0	0	0	0	0
ND3	0	0	0	0	0	1.2	90.8	8.0	0	0	0	0
ND4	0	0	0	0	0	0	7.2	92.8	0	0	0	0
SD1	0	0	0	0	0	0	0	0	89.6	8.0	2.4	0
SD2	0	0	0	0	0	0	0	0	2.8	87.6	2.0	7.6
SD3	0	0	0	0	0	0	0	0	0	3.6	88.4	8.0
SD4	0	0	0	0	0	0	0	0	0	16	2.4	81.6

Table 14 Confusion matrix of the Group4 by using method [20]

Average accuracy	Predicted(%)											
=87.5%	AD1	AD2	C1	C2	ND1	ND2	ND3	ND4	SD1	SD2	SD3	SD4
AD1	98.0	0	0	0	0	0	2.0	0	0	0	0	0
AD2	0	100	0	0	0	0	0	0	0	0	0	0
C1	0	0	100	0	0	0	0	0	0	0	0	0
C2	1.2	0	0	94.0	0	0	4.8	0	0	0	0	0
ND1	0	0	0	0	62.8	37.2	0	0	0	0	0	0
ND2	0	0	0	0	38.4	61.2	0.4	0	0	0	0	0
ND3	0	0	0	0	0	0	90.8	9.2	0	0	0	0
ND4	0	0	0	0	0	0	6.8	93.2	0	0	0	0
SD1	0	0	0	0	0	0	2.0	0	92.4	5.6	0	0
SD2	0	0	0	0	0	0	0	0	4.4	92.4	0.8	2.4
SD3	0	0	0	0	0	0	1.2	0	2	1.6	80.8	14.4
SD4	0	0	0	0	0	0	0	0	0	1.2	14.8	84.0

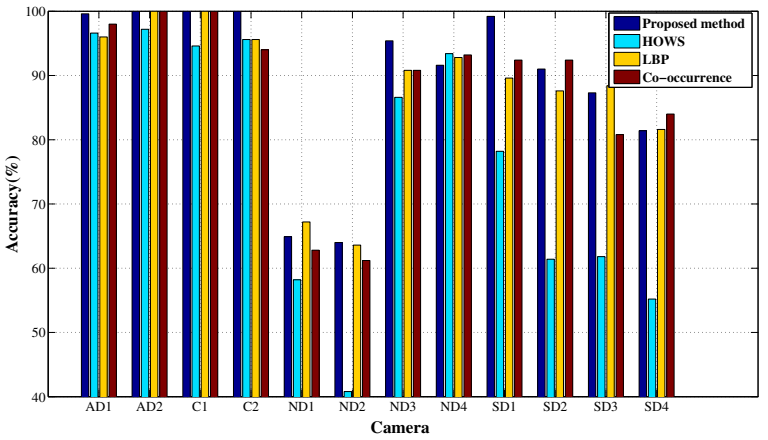

Fig. 5 Comparison of the identification results

identifying the different camera individuals. Further, among the features used, the PRNU-based HOWS features have higher detection accuracy, which greatly contribute to the high accuracy of the proposed method in identifying camera individuals.

Group 4 In order to test the performance of the proposed method as the number of cameras increases, we take 12 cameras to carry out the experiment, which include two cameras of Agfa, two cameras of Canon, four cameras of Nikon and four cameras of Sony.

The confusion matrixes of 12 different cameras are shown in Tables 11, 12, 13 and 14. It can be seen that in the identification of different brands and different models, the four methods all show a high accuracy rate, but the accuracy rate would decrease when identifying different camera individuals, which is also the reason leading to the final average accuracy declines. The bar graph in Fig. 5 shows the accuracy comparison of the four methods when identifying twelve different cameras. For most cameras used in the experiment, the detection accuracy of the proposed method is higher than that of the method [1, 20] and [33], the average accuracy of the twelve cameras obtained by the proposed method is the highest among the four methods.

4.3 Complexity comparison

As mentioned above, most of the existing source camera identification methods, such as method [1, 20, 33] and so on, are based on support vector machine (SVM) classifier. However, the complexity of the SVM classifier increases with the feature dimensionality and the training set size. In this paper, we extract the HOWS features, LBP and LPQ features, CFA features and co-occurrence matrix features separately to form high dimensional feature set, which help to improve the detection accuracy. In this case, the use of SVM is time-consuming. So the ensemble classifier is used in this paper to get a high detect accuracy with less time. In order to demonstrate the performance of the ensemble classifier used in this paper, we compare the feature extraction time, classification time and accuracy of different methods in Group 4, the results is shown in Table 15. Where feature extraction time refers to the average extraction time of an image, classification time is the total classification time under the experimental sets of Group 4, and accuracy refers to the classification accuracy

Table 15 Comparison of complexity and accuracy for various methods

	Feature extraction time (each image)	Classification time		Accuracy	
		SVM	Ensemble	SVM	Ensemble
HOWS	<1 min	11.2 min	1.06 min	78.4%	80.8%
LBP,LPQ	1.2 min	146.6 min	13.9 min	87.8%	87.4%
Co-occurrence	21.3 min	303.7 min	30.9 min	87.5%	86.5%
Proposed method	79 min	498.5 min	50.9 min	85.3%	89.5%

of the two classifiers when the same features are used. It can be seen that while using high dimensional features to improve detection accuracy, the feature extraction time of proposed method will increase, which is inevitable. In order to reduce the algorithm complexity, it is necessary to use the ensemble classifier instead of SVM to reduce the classification time. What's more, the classification accuracy of the ensemble classifier is comparable to that of SVM when using the same features.

5 Conclusions

We have proposed a novel method based on the fusion features of software-related, hardware-related and image-related features to identify the source camera. The combination of the high dimensional feature set and ensemble classifier can get higher detect accuracy with less time. The effectiveness of the proposed method has been proved by numerous experiments. The average accuracy of proposed method is close to 100% when identifying the camera brand and model, and the accuracy is higher than the existing algorithm when identifying the different camera individuals with same brand and model. Our future work will focus on combining some new descriptors, such as transition local binary pattern (tLBP) and center-symmetric local binary patterns (CSLBP) to further improve the performance of the method.

Acknowledgements This work is supported by the National Science Foundation of China (No. 61502076) and the Scientific Research Project of Liaoning Provincial Education Department (No. L2015114).

References

1. Akshatha KR, Karunakar AK, Anitha H, Raghavendra U, Shetty D (2016) Digital camera identification using PRNU: a feature based approach. Digit Inv 19:69–77
2. Avcibas I, Memon N, Sankur B (2003) Steganalysis using image quality metrics. IEEE Trans Image Process 12(2):221–229
3. Bayram S, Sencar HT, Memon N, Avcibas I (2005) Source camera identification based on CFA interpolation. In: IEEE conference on image processing (ICIP), pp 1–4
4. Celiktutan O, Avcibas I, Sankur B, Ayerden NP, Capar C (2006) Source cell-phone identification. In: Proceedings of the IEEE 14th signal processing and communications applications, pp 1–3
5. Chan LH, Law NF, Siu WC (2013) A confidence map and pixel-based weighted correlation for PRNU-based camera identification. Digit Inv 10(3):215–225
6. Efron B, Tibshirani R (1993) An introduction to the bootstrap, vol 57. CRC Press, Boca Raton

7. Farid H (2001) Blind inverse gamma correction. IEEE Trans Image Process 10(10):1428–1433
8. Fridrich J, Kodovsky J (2012) Rich models for steganalysis of digital images. IEEE Trans Inf Forensic Secur 7(3):868–882
9. Fridrich J, Kodovsky J (2012) Rich models for steganalysis of digital images. IEEE Trans Inf Forensic Secur 7(3):868–882
10. Gloe T, Bohme R (2010) Dresden image database for benchmarking digital image forensics. In: Proceedings 2010 ACM symposium on applications computing, vol 22–26. Sierre, pp 1584–1590
11. Hu Y, Li CT, Zhou C (2010) Selecting forensic features for robust source camera identification. In: Computer symposium (ICS), pp 506–511
12. Janesick JR, Blouke M (1995) Scientific charge-coupled devices: past, present, & future. Opt Photon News 6(4):16–20
13. Kang X, Li Y, Qu Z, Huang J (2012) Enhancing source camera identification performance with a camera reference phase sensor pattern noise. IEEE Trans Inf Forensic Secur 7(2):393–402
14. Kharrazi M, Sencar H, Memon N (2004) Blind source camera identification. In: IEEE international conference on image processing (ICIP), p 712
15. Li CT (2010) Source camera identification using enhanced sensor pattern noise. IEEE Trans Inf Forensic Secur 5(2):280–287
16. Long YJ, Huang YZ (2006) Image based source camera identification using demosaicking. In: IEEE 8th workshop multimedia signal process, pp 419–424
17. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Secur 1(2):205–214
18. Lyu S, Farid H (2006) Steganalysis using higher-order image statistics. IEEE Trans Inf Forensics Secur 1(1):111–119
19. Ma Y, Luo X, Li X, Bao Z, Zhang Y (2018) Selection of rich model steganalysis features based on decision rough set |Á-positive region reduction. IEEE Transactions on Circuits and Systems for Video Technology
20. Marra F, Poggi G, Sansone C, Verdoliva L (2017) A study of co-occurrence based local features for camera model identification. Multimed Tools Appl 76(4):4765–4781
21. Mihcak M, Kozintsev I, Ramchandran K (1999) Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In: IEEE international conference on acoustics, speech and signal processing, pp 3253–3256
22. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24:971–987
23. Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. In: International conference on image and signal processing. Springer, pp 236–243
24. Ozparlak L, Avcibas I (2011) Differentiating between images using wavelet-based transforms: a comparative study. IEEE Trans Inf Forensic Secur 6(4):1418–1431
25. Qin C, Ji P, Zhang X, Dong J, Wang J (2017) Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy. Signal Process 138:280–293
26. San Choi K, Lam EY, Wong KKY (2006) Automatic source camera identification using the intrinsic lens radial distortion. Opt Express 14(24):11551–11565
27. Sutcu Y, Bayram S, Sencar HT, Memon N (2007) Improvements on sensor noise based source camera identification. In: IEEE international conference on multimedia and expo, pp 24–27
28. Swaminathan A, Wu M, Liu KJR (2006) Non-intrusive forensic analysis of visual sensors using output images. In: IEEE international conference on acoustics, speech and signal processing, pp V–V
29. Swaminathan A, Wu M, Liu KJR (2007) Optimization of input pattern for semi non-intrusive component forensics of digital cameras. In: IEEE international conference on acoustics, speech and signal processing, pp II–225
30. Van L, Emmanuel S, Kankanhalli M (2007) Identifying source cell phone using chromatic aberration. In: IEEE international conference on multimedia and expo, pp 883–886
31. Wang B, Kong X, You X, Fu H (2009) Blind CFA interpolation detection based on covariance matrix. Trans J Electron Inform Technol (Chin) 31:1175–1179
32. Wu G, Kang X, Liu KJR (2012) A context adaptive predictor of sensor pattern noise for camera source identification. In: IEEE conference on image processing (ICIP), pp 237–240
33. Xu B, Wang X, Zhou X, Xi J, Wang S (2016) Source camera identification from image texture features. Neurocomputing 207:131–140

34. Yao H, Wang S, Zhang X, Qin C, Wang J (2017) Detecting image splicing based on noise level inconsistency. Multimed Tools Appl 76(10):12457–12479
35. Yao H, Cao F, Tang Z, Wang J, Qiao T (2018) Expose noise level inconsistency incorporating the inhomogeneity scoring strategy. Multimed Tools Appl 77(14):18139–18161
36. Zhang Y, Qin C, Zhang W, Liu F, Luo X (2018) On the fault-tolerant performance for a class of robust image steganography. Signal Process 146:99–111

Bo Wang received his B.S. degree in Electronic and Information Engineering, M.S.degree and Ph.D. degree in Signal and Information Processing from Dalian University of Technology, Dalian, China, in 2003, 2005 and 2010, respectively. From 2010 to 2012, he was a post-doctoral research associate in Faculty of Management and Economics in Dalian University of Technology. He is currently an Associate Professor in School of Information and Communication Engineering in Dalian University of Technology. His current research interests focus on the areas of multimedia processing and security, such as digital image processing and forensics.

Kun Zhong received his B.S. degree in Electrical and Electronic Engineering from from Dalian University of Technology, China, in 2017. He is currently pursuing a master's course in School of Information and Communication Engineering at the university. His research interests include digital image processing and forensics.

Ming Li received his B.S. degree in Electrical Engineering from Harbin Engineering University in 2000 and M.S. degree, Ph.D. degree in Communications and Signal Processing from The State University of New York at Buffalo. He is currently an Associate Professor in School of Information and Communication Engineering in Dalian University of Technology. His current research interests focus on Millimeter wave communications, Physical-Layer Security and so on.