Multimedia Tools and Applications (2021) 80:29719–29743
https://doi.org/10.1007/s11042-021-11201-7

Discriminative feature projection for camera model identification of recompressed images

Bo Wang[1] · Yue Wang[1] · Jiayao Hou[1] · Xue Sui[2] · Meineng Zhu[3]

Received: 19 July 2020 / Revised: 15 June 2021 / Accepted: 28 June 2021 /
Published online: 10 July 2021

Abstract

Camera model identification, which aims to identify the source camera model of the query images, has been well studied in the laboratory environment. Most existing methods regard it as a classification problem and rely on well-designed features that characterize the differences between various camera models. In these methods, however, the accuracy of identification results would suffer from the process of image recompression. The information loss reduces the discriminative ability of the designed identification features and results in a serious accuracy loss. To remedy this shortcoming, we investigate the handicap for accurate source identification when the query image is recompressed and creatively propose a new method Discriminative Feature Projection (DFP) to solve this problem. The proposed method learns a discriminative feature projection that projects the designed identification features into a new feature representation invariant to recompression by minimizing the divergence between recompressed and uncompressed images. We also incorporate two constraints that the discrepancy of different images sources should be large and the latent geometric relations of images neighbors should be preserved into our method to reinforce the discriminative ability. Moreover, we conduct extensive experiments over the public Dresden Image Database. Compared with several state-of-the-art methods on camera model identification, the experiment results verify that DFP can achieve significant accuracy promotion when identifying the recompressed images.

Keywords Image forensic · Camera model identification · Image recompression · Feature projection

✉ Yue Wang
31909072@mail.dlut.edu.cn

[1] School of Information and Communication Engineering, Dalian University of Technology, Dalian, People's Republic of China

[2] Liaoning Normal University, Dalian, Liaoning, 116029, People's Republic of China

[3] Beijing Institute of Electronics Technology and Application, Beijing, 100091, People's Republic of China

1 Introduction

Nowadays, digital images are everywhere in our daily life. These images are taken by digital cameras, camcorders and smart-phones and are utilized in many fields such as commercial advertisement, news media, military and even in a criminal investigation as evidence. However, the recent development of image tampering technology makes it easy to tamper an image without a visible trace left. There would be an extremely terrible social impact if tampered images are utilized in journalism or judicial evidence. As a result, it is essential to identify the authenticity and the source of the images.

In this circumstance, the technology of source camera identification develops rapidly [44, 45]. As a challenging branch of multimedia forensics, source camera identification aims to determine the sources of digital images to support forensics and more, specifically, to determine which camera a given image was taken by [43]. Generally, source camera identification can be divided into two groups: active source camera identification and blind source camera identification. Blind source camera identification which only uses the query image itself is more popular because it does not need to embed the source information, i.e. digital watermarking, into the original image in advance. This work focuses on blind source camera identification which has been well studied in the laboratory environment over the last decade, and a significant number of outstanding methods have been proposed.

Regarding the blind source camera identification, most existing works could be divided into two categories: camera individual identification and camera model identification. We briefly introduce these two technologies in the following subsections and the detailed description can be found in [24, 41, 43].

1.1 Camera individual identification

The objective of camera individual identification is to find out which specific imaging device the query image is taken by. Most researches are based on the matching of unique fingerprints extracted from a digital image namely sensor pattern noise (SPN). The internal defects, due to the shortage of raw material or the limitation of manufacturing techniques, are considered as an inevitability on camera lens for each imaging device. The defects would leave indelible mark in all images acquired by the sensor, and the mark is stable and unique of camera individual, so it could be used as the fingerprint to identify the camera individual. Also, SPN could be used for detection techniques such as forgery detection [9], hue modification [17], etc.

In general, SPN based camera individual identification mainly consists of three main stages: SPN extraction, reference SPN estimation and SPN matching. The process of SPN based methods is illustrated in Fig. 1.

Obviously, for SPN based methods, the extraction of SPN is a crucial step, more accurate extraction of SPN generally means higher identification accuracy. Unfortunately, SPN can easily be contaminated by image scene details, i.e. scene edges and textures [27] and this would directly cause misidentification. This problem has already attracted the attention of researchers and many attempts [7, 14, 18, 21, 25–27, 29, 32, 35, 39, 46, 57] have been made to extract purer SPN. Besides, the storage overhead of SPN and high computation complexity of SPN matching are also challenging problems for SPN based methods. Many efforts [1, 2, 19, 30, 52] have been made to lower the storage overhead and enable a more efficient process of camera individual identification.

When identifying the source of images, the image content is useless and will significantly reduce the purity of SPN. Many researchers regard this as the breakthrough point and tend

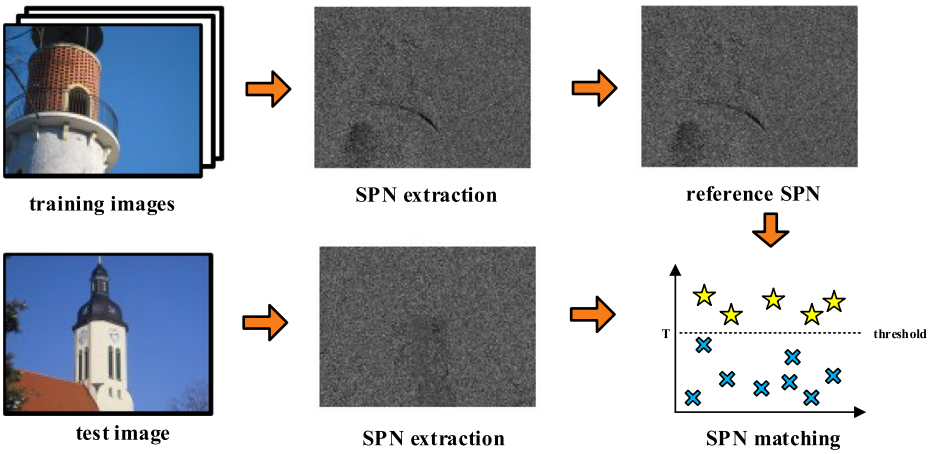


Fig. 1 The process of the SPN based camera individual identification

to choose images without many scene details, i.e. blue sky, to generate reference SPN in order to increase the reliability of reference SPN. That's the limitation of SPN based methods because the artificial selection of images samples is often not realistic in the practical forensic scenario.

1.2 Camera model identification

Different from individual identification and active forensics [6, 23], camera model identification aims to find the source camera model of the query image rather than the specific camera individual. To achieve this goal, design and extraction of distinctive characteristics (features) covering differences between different camera models while occurring very similar between devices of the same model, is considered as the priority task in camera model identification. Many sophisticated processing steps are carried out inside cameras during the imaging process, i.e. demosaicing, white balancing and JPEG compression. For a given camera model, these processing steps are exactly the same, but it is a rare case that different camera models use the very same set of algorithms and parameters [38]. Given the statistical traces left in images by the mentioned image processing steps, it is possible to derive the reliable identification of camera model. Thus, most researches take advantage of the well-designed statistical features and machine learning algorithms to identify camera model.

The flowchart of feature-based camera model identification also consists of three steps: feature extraction, classifier training and identification of query image. The features are specifically designed for capturing the distinctness of different camera models such as the ensemble of demosaicing features proposed in [8] and local binary pattern features proposed in [58], before being fed into support vector machine (SVM) or ensemble classifiers to train a classifier to identify the source model of the query image.

So far, feature-based methods have achieved satisfactory identification performance in a laboratory environment. An average identification accuracy of 99.2% over 12 camera models is reported in [8]. However, there is still one challenging problem that needs to be settled urgently, the robustness against double JPEG compression, and this is beyond the scope of existing methods. In real-world forensic applications, images usually have undergone JPEG

recompression when being shared through instant messaging software or uploaded to social networks. After the query image is recompressed, the feature extracted from the query image is no longer suitable for the model built on uncompressed images, and the identification result becomes no longer reliable either.

1.3 Goals and objectives

In our previous work [5], joint first and second-order statistics matching (JSM) algorithm was proposed to reduce the recompression divergence between training samples and query images caused by image recompression[1]. Although JSM has made great progress compared to previous works, it mainly focuses on reducing the recompression divergence and fails to fully exploit the class information, resulting in the loss of discriminative information. To solve this problem, we propose an iterative method namely discriminative feature projection (DFP) in this paper. The basic idea of our method is to learn a discriminative feature projection that projects the designed identification features into a low-dimensional and discriminative subspace that is invariant to recompression. Intuitively, there should be no divergence between the features of recompressed and original images any more in the learned representation. More importantly, it should be reinforced that the discriminative information between classes which is beneficial to the identification task. We also add a local geometric structure-preserving constraint when learning the new feature representation because samples lying close in the original feature space have large probability belonging to the same class and this should be preserved in the learned feature representation.

That is to say, not only do we seek a new feature presentation in minimizing the discrepancy between original images and recompressed images, but also incorporate the discriminative information from labeled data and the local geometric structure constraint into our method to enhance the classification performance. The latter makes our method more discriminative and effective when identifying the source camera model of recompressed images. We will demonstrate the benefits of our method both theoretically and experimentally in following sections.

Hence, our main contributions are summarized as follows:

1. Practical-oriented research focuses on the source camera model identification of recompressed images.
2. Find out a suitable measurement Maximum Mean Discrepancy to measure and reduce the divergence between recompressed images and original images.
3. Learn a discriminative feature projection to simultaneously reduce the side effect caused by recompression and reinforce the identification ability.

The rest of the paper is organized as the following. Section 2 describes the representative researches related to our work. Section 3 describes our proposed method and the corresponding optimization procedure in detail. Comprehensive experiments are carried out and comparisons of related works are presented in Section 4 to verify the superiority of the proposed method. Finally we draw our conclusion in Section 5.

[1]To prevent any potential confusion, we name the JPEG images directly exported from cameras the original images. The original images are all JPEG images and if they undergo another JPEG compression, we name them recompressed images.

2 Related work

In this section, we briefly introduce feature-based camera model identification methods and related feature projection based methods.

2.1 Feature-based camera model identification

Considering that the color filter array (CFA) and demosaicing algorithms may vary from different camera manufacturers and even different camera models with the same manufacturer, Bayram et al. [3] proposed to identify the source camera based on traces of the proprietary interpolation algorithms. Ho et al. [16] proposed to use the variance of color difference planes to measure the inter-channel correlation for source camera identification and achieved a promising result. Based on [16], Hu et al. [20] developed an improved algorithm using inter-channel demosaicing traces for camera model identification. The shape and texture features were fed into Stumps AdaBoost classifier and experimental results showed its superiority.

Xu et al. [58] proposed to use uniform gray-scale invariant local binary patterns (LBP) to capture the characteristics or artifacts generated by image processing algorithms which are block-wise implemented inside cameras, such as demosaicing, filtering and JPEG compression. Similarly, [59] also investigated the discriminative ability of local phase quantization (LPQ) to distinguish imaging devices. The combined texture features of LBP and LPQ were fed into a multi-class SVM classifier, and reached better detection accuracy than [58].

In addition, kinds of research aimed at using high dimensional features (more than 1000 dimensions) to classify the source camera model. Chen et al. [8] built a rich model of the demosaicing algorithm to identify the source camera model. The full feature dimension is 1372 and was fed into a multi-class ensemble classifier. The average identification accuracy over 12 camera models is 99.2%, as reported in their paper. Tuama et al. [47] extracted high order statistics features with 10932 dimensions consisting of co-occurrences matrix features, traces of color dependencies features and conditional probability statistics features. Recently, Roy et al. [42] proposed to use Discrete Cosine Transform Residual (DCTR) features combined with principal component analysis (PCA) and ensemble classifier for camera source identification. They proved DCTR features can capture the compression artifacts imposed by the camera model-dependent quantization tables and the average accuracy reported in their paper is 96.5%. Note that the camera models used in their experiments are all from different brands, it looks more like camera brand identification rather than camera model identification. The image feature vectors consist of structure element correlation (SEC), gradient value correlation (GVC) and gradient direction correlation (GDC), have also been proposed and evaluated in related applications, image retrieval for instance [48–51, 53–56].

In summary, there are kinds of research studying camera model identification, but existing methods can not handle the problem of double JPEG compression [3, 4, 8, 10–12, 16, 20, 22, 37, 38, 42, 47, 58, 59]. Actually, these statical feature-based methods suffer from image recompression severely. Our work aims to solve this problem by learning a discriminative feature projection that projects the training and test features into a low dimensional subspace to suppress the influence of image recompression while enhancing the discriminative ability.

2.2 Feature projection based methods

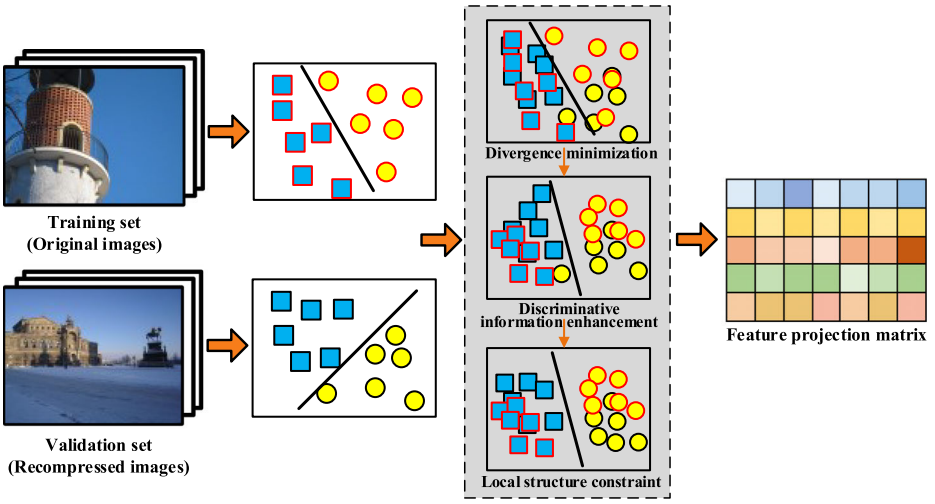
Identifying the source camera model of recompressed images with only original images as training data is a typical situation of domain shift problems. Domain shift exists in various applications, such as steganalysis and text classification [28, 33, 34, 36, 40, 60], and feature projection has been demonstrated significant successes in dealing with this problem. In these papers, the training and test sets are named as source and target domain respectively. The goal is to reduce the domain shift by projecting the original features into a new shared feature space with a small domain shift. We will briefly introduce several related representative feature projection based methods.

In [40], transfer component analysis (TCA) was proposed to learn a common feature transformation for source and target domain where the marginal distribution difference is minimized in projected feature space. Based on [40], Long et al. [33] took conditional distribution into consideration, and put forward joint distribution adaptation (JDA) to reduce both the marginal and conditional distribution differences between domains. Inspired by these studies, Li et al. [28] proposed generalized transfer component analysis (GTCA) to solve the mismatched problem in steganalysis. Using an intermediate domain before TCA, GTCA can learn more discriminate representations for mismatched steganalysis. In [36], Luo et al. also introduced a repulsive force term to drag the sub-domains with different labels far away from each other to increase the discriminative power of the adapted domain. They named it as close yet discriminative domain adaptation (CDDA).

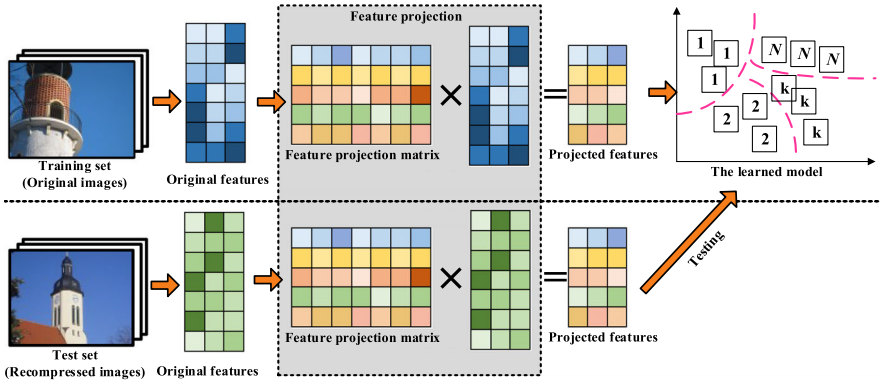
A recent feature projection based research studying source camera model identification of recompressed images is [61], Zhang et al. manually designed a feature transformation to minimize the distribution deviation between training and test set. The experimental results reported in their paper are quite encouraging. Note that the feature projection in [61] is manually set based on test data and the projection is only applied on training features, which is very different from our proposed method where the feature projection is automatically learned based on the well-designed objective function and both the training and test data are projected into the discriminative feature subspace to reduce the divergence caused by recompression.

3 Proposed method

The framework of camera model identification through DFP is mainly composed of two parts: discriminative feature projection matrix learning and classification through SVM under the new feature representation, as shown in Fig. 2. To improve the identification accuracy under recompression condition, we first learn the feature projection matrix by simultaneously minimizing the divergence which is caused by recompression, enhancing the discriminative information and constraining the local structure relations as illustrated in Fig. 2a. Once the projection matrix is available, we can get the new representation of training data through the learned feature projection matrix, and the features under this new representation are not only insensitive to image recompression but also more discriminative. Thus, the identification model could be trained under the new representation by SVM, and given a query image, we could project the identification feature of the query image using the same feature projection matrix and get a more reliable identification result compared with the non-projected feature as illustrated in Fig. 2b. As discussed above, step 1 is the key of DFP. In the following subsections, we will explain in detail the original intention to each item involved in the final objective function when learning the feature projection.



(a) Step 1:Discriminative feature projection matrix learning.



(b) Step 2:Classification under the new feature representation.

Fig. 2 The framework of camera model identification through DFP

3.1 Data properties preserving with dimensionality reduction

Suppose there are n_s labeled training samples (training set, original images) $\mathbf{X}_s = \{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, ..., \mathbf{x}_{s_{n_s}}\}$, $\mathbf{x}_{s_i} \in \mathcal{R}^d$, $i = 1, ..., n_s$, where d is the dimension of original features. Each column in $\mathbf{X}_s$ corresponds to one sample. In addition, $\mathbf{Y}_s = \{y_{s_1}, y_{s_2}, ..., y_{s_{n_s}}\}$ is the set of labels of $\mathbf{X}_s$ and $y_{s_i} \in \{1, ..., C\}$, where C is the number of classes in training set. Similarly, there are n_t validation samples (validation set, recompressed images) $\mathbf{X}_t = \{\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_{n_t}}\}$, $\mathbf{x}_{t_j} \in \mathcal{R}^d$, $j = 1, ..., n_t$. $\mathbf{Y}_t$ is the corresponding label set of $\mathbf{X}_t$.

We aim to learn a discriminative feature projection matrix $\mathbf{P}$ that projects the original features $\mathbf{X_s}$, $\mathbf{X_t}$ into a latent feature space which is insensitive to image recompression but more discriminative. Though projected into a new feature space, the data properties should

be preserved and the reconstruction error should be minimized. Methods like PCA to find a latent feature space with dimensionality reduction have shown their effectiveness. Here we maximize the data variance as performed by PCA to preserve the data properties. Denote $\mathbf{X} = \{\mathbf{x}_{s_1}, ..., \mathbf{x}_{s_{n_s}}, \mathbf{x}_{t_1}, ..\mathbf{x}_{t_{n_t}}\} \in \mathcal{R}^{d\times n}$ the input data matrix, where $n = n_s + n_t$, $\mathbf{XHX}^{\mathrm{T}}$ is the covariance matrix of $\mathbf{X}$, where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}$ is the centering matrix, $\mathbf{I} \in \mathcal{R}^{n\times n}$ is the identity matrix, and $\mathbf{1}$ is the $n \times n$ matrix of ones. Hence, the variance maximization of the projected data could be achieved while projecting the original data into the latent feature space $\mathbf{P}^{\mathrm{T}}\mathbf{X}$ as

$$\max_{\mathbf{P}} tr(\mathbf{P}^{\mathrm{T}}\mathbf{XHX}^{\mathrm{T}}\mathbf{P}) \tag{1}$$

where $tr(\bullet)$ denotes the trace of a matrix, $\mathbf{P} \in \mathcal{R}^{d\times k}$ is the projection matrix, and k is the dimension of the learned low dimensional subspace where $k < d$.

3.2 Divergence minimization by distribution matching

Only preserving the data properties is far from enough to identify the source camera model under recompression condition, because PCA could not explicitly reduce the distribution divergence caused by recompression between training set(original images) and validation set(recompressed images). Here we aim to reduce the distribution shift in the learned latent feature space by explicitly minimizing the distribution difference through a proper distance measurement. We adopt a non-parametric criteria Maximum Mean Discrepancy (MMD) as the distance measurement of distribution distance between two sets. MMD is introduced by Gretton et al. to measure the distribution difference in [15]. Using the k-dimensional embeddings mentioned in (1), the MMD distance could be formulated as

$$\begin{aligned} &\left\| \frac{1}{n_s} \sum_{\mathbf{x}_{s_i}\in\mathbf{X}_s} \mathbf{P}^{\mathrm{T}}\mathbf{x}_{s_i} - \frac{1}{n_t} \sum_{\mathbf{x}_{t_j}\in\mathbf{X}_t} \mathbf{P}^{\mathrm{T}}\mathbf{x}_{t_j} \right\|^2 \\ &= tr(\mathbf{P}^{\mathrm{T}}\mathbf{XM}_0\mathbf{X}^{\mathrm{T}}\mathbf{P}) \end{aligned} \tag{2}$$

where $\mathbf{M}_0$ is the MMD matrix and is computed as

$$(\mathbf{M}_0)_{ij} = \begin{cases} \frac{1}{n_s n_s}, & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_s \\ \frac{1}{n_t n_t}, & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_t \\ \frac{-1}{n_s n_t}, & \text{otherwise} \end{cases} \tag{3}$$

Since (2) measures the distribution difference of training set and validation set, we then achieve our goal of reducing the distribution difference by minimizing

$$\min_{\mathbf{P}} tr(\mathbf{P}^{\mathrm{T}}\mathbf{XM}_0\mathbf{X}^{\mathrm{T}}\mathbf{P}) \tag{4}$$

So the distribution shift is reduced under the new feature representation $\mathbf{Z} = \mathbf{P}^{\mathrm{T}}\mathbf{X}$ when (4) is minimized.

Considering there are C classes in training set, the distribution shift could be further reduced when the empirical distances over the same classes between training and validation set are minimized. An unavoidable problem is that we have no labels in validation set (The labels in validation set is only used for evaluating the performance of the learning of feature projection matrix, and can not directly used here). So we turn to utilize the pseudo labels of validation samples y'_{t_j}, which can be obtained by applying a base classifier such as nearest neighbor classifier (NN) or support vector machine (SVM) trained over the training data to

the unlabeled validation data. Here we use nearest neighbor classifier as the base classifier to predict pseudo labels in our experiment since it does not require tuning cross-validation parameters. Similar to (2), with pseudo labels in validation set and true labels in training set, the class-conditional distribution shift in projected feature space can be computed as

$$\begin{aligned} &\left\| \frac{1}{n_s^{(c)}} \sum_{\mathbf{x}_{s_i} \in \mathbf{X}_s^{(c)}} \mathbf{P}^{\mathrm{T}} \mathbf{x}_{s_i} - \frac{1}{n_t^{(c)}} \sum_{\mathbf{x}_{t_j} \in \mathbf{X}_t^{(c)}} \mathbf{P}^{\mathrm{T}} \mathbf{x}_{t_j} \right\|^2 \\ &= tr(\mathbf{P}^{\mathrm{T}} \mathbf{X} \mathbf{M}_c \mathbf{X}^{\mathrm{T}} \mathbf{P}) \end{aligned} \tag{5}$$

where $\mathbf{X}_s^{(c)} = \{\mathbf{x}_{s_i} : \mathbf{x}_{s_i} \in \mathbf{X}_s \ \wedge \ y_{s_i} = c\}$ and $\mathbf{X}_t^{(c)} = \{\mathbf{x}_{t_j} : \mathbf{x}_{t_j} \in \mathbf{X}_t \ \wedge \ y'_{t_j} = c\}$ represent the subset of samples belonging to class c in the training and validation set respectively. y_{s_i} and y'_{t_j} is the true label of $\mathbf{x}_i$ and pseudo label of $\mathbf{x}_j$ correspondingly. $n_s^{(c)}$ and $n_t^{(c)}$ is the number of samples in $\mathbf{X}_s^{(c)}$ and $\mathbf{X}_t^{(c)}$. $\mathbf{M}_c$ is the conditional MMD matrix for class c

$$(\mathbf{M}_c)_{i,j} = \begin{cases} \frac{1}{n_s^{(c)} n_s^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_s^{(c)} \\ \frac{1}{n_t^{(c)} n_t^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}_t^{(c)} \\ \frac{-1}{n_s^{(c)} n_t^{(c)}}, & \begin{cases} \mathbf{x}_i \in \mathbf{X}_s^{(c)}, \mathbf{x}_j \in \mathbf{X}_t^{(c)} \\ \mathbf{x}_i \in \mathbf{X}_t^{(c)}, \mathbf{x}_j \in \mathbf{X}_s^{(c)} \end{cases} \end{cases} \tag{6}$$

To further reduce the distribution shift caused by recompression, the class-conditional distribution shift computed by (5) for all the C classes should be minimized. So (4) and (5) could be incorporated as

$$\min_{\mathbf{P}} tr(\mathbf{P}^{\mathrm{T}} \mathbf{X} \sum_{c=0}^{C} \mathbf{M}_c \mathbf{X}^{\mathrm{T}} \mathbf{P}) \tag{7}$$

By minimizing the above equation such that (1) is maximized, the distribution between training and validation set are drawn closer under the new representation $\mathbf{Z} = \mathbf{P}^{\mathrm{T}} \mathbf{X}$. And the identification of source camera model for recompressed images in this new feature space becomes more reliable because the divergence caused by recompression has been decreased effectively through the feature projection.

3.3 Discriminative information enhancement

The distribution divergence has been reduced in the new feature representation by distribution matching, but the class information has not been fully exploited while learning the feature projection. Since the true labels of training samples are available, it is beneficial to encode the discriminative class information in the resulting feature representation. Here we aim to learn the projection that not only minimizes the distribution divergence between two sets, but also encourages the new representation to be discriminative and yields good classification performance. To reach this goal, the distance between the training samples in each class and the class mean should be minimized, which encourages the samples belonging to the same class to be compact by

$$\min_{\mathbf{P}} tr(\mathbf{P}^{\mathrm{T}} \mathbf{S}_w \mathbf{P}) \tag{8}$$

where $\mathbf{S}_w$ is the within-class scatter matrix of training set whose labels are available and is defined as

$$\mathbf{S}_w = \sum_{c=1}^{C} \sum_{\mathbf{x}_{s_i} \in \mathbf{X}_s^{(c)}} \left(\mathbf{x}_{s_i} - \mu^{(c)}\right) \left(\mathbf{x}_{s_i} - \mu^{(c)}\right)^{\mathrm{T}} \tag{9}$$

where $\mu^{(c)}$ is the mean of samples in training set of class c.

Minimizing (8) when learning the feature projection, the samples in the projected feature representation would become closer if they belong to the same source camera. Encoding the discriminative class information is a quite important improvement. With this improvement, the discriminative information is enhanced and the learned representation is more suitable for classification. In another word, enhancing the discriminative class information could lead to an obvious improvement of identification accuracy. And this will be verified thoroughly by the experiments.

3.4 Local structure constraint

With distribution matching and discriminative information enhancement, the samples have been projected into a new feature representation which has less distribution divergence but more discriminative ability. However, the local neighbor relations in sample level have not been taken into consideration. Intuitively the samples who are neighbors in the original feature space should also stay close to each other in the new feature representation.

The local structure construction method has been widely used to preserve the local geometry structure by constructing the relations of neighbors [31]. For our identification task of recompressed images, we incorporate the local structure constraint with class information while learning the feature projection. Specifically, neighbor samples only from the same class should keep close to each other and this will effectively enhance the similarities among the neighbor samples belonging to the same class.

With the true labels of training set and pseudo labels of validation set, the local structure constraint with class information is defined as

$$\begin{aligned} &\min_{\mathbf{P}} \sum_{i\ j} \left\| \mathbf{P}^{\mathrm{T}}\mathbf{x}_i - \mathbf{P}^{\mathrm{T}}\mathbf{x}_j \right\|^2 W_{ij} \\ &= \min_{\mathbf{P}} tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{X}\left(\mathbf{D} - \mathbf{W}\right)\mathbf{X}^{\mathrm{T}}\mathbf{P}\right) \\ &= \min_{\mathbf{P}} tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}}\mathbf{P}\right) \end{aligned} \tag{10}$$

where $W_{ij} = ||x_i - x_j||^2$ is the weight of the relation of two samples if they have the same class labels and among the nearest neighbors of the other. If two samples are not the nearest neighbors or have different class labels, we set $W_{ij} = 0$ by default. $\mathbf{D}$ is a diagonal matrix where $D_{ii} = \sum_i W_{ij}$, and $\mathbf{G} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix.

Both (8) and (10) utilize the class information to enhance the discriminative ability of the learned feature projection, but they are essentially different. In (8), the labels of training set are all correct and we make use of all samples in training set to compute the within-class scatter matrix S_w. But in (10) the pseudo labels of validation set which may be incorrect are involved, so we only constrain the local relations of samples if and only if they share the same label and are nearest neighbors.

3.5 Overall objective function and optimization

According to the above analysis, we aim to learn the feature projection by maximizing (1) while minimizing (7), (8) and (10). So we incorporate the above four equations to formulate our method as

$$\min_{\mathbf{P}} \frac{tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{X}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}+\alpha\left(\mathbf{S}_w+\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}}\right)\right)\mathbf{P}+\beta tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{P}\right)\right)}{tr\left(\mathbf{P}^{\mathrm{T}}\mathbf{X}\mathbf{H}\mathbf{X}^{\mathrm{T}}\mathbf{P}\right)} \tag{11}$$

where α is the trade-off parameter to balance the importance of distribution matching, discriminative information enhancement and local structure constraint. A constraint that $tr(\mathbf{P}^{\mathrm{T}}\mathbf{P})$ is small is further imposed to control the scale of $\mathbf{P}$ to guarantee the optimization problem to be well defined and β is the corresponding regularization parameter. So finding the optimal feature projection matrix $\mathbf{P}$ is transmuted into minimizing (11).

According to the generalized Rayleigh quotient, minimizing the numerator of (11) such that the denominator of (11) is maximized is equivalent to minimizing the numerator of (11) such that the denominator of (11) is fixed. So the constrained objective function is reformulated as:

$$\begin{aligned} &\min_{\mathbf{P}} tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{X}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}+\alpha(\mathbf{S}_w+\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}})+\beta\mathbf{I}\right)\mathbf{P}\right) \\ &s.t.\ \ \mathbf{P}^{\mathrm{T}}\mathbf{X}\mathbf{H}\mathbf{X}^{\mathrm{T}}\mathbf{P}=\mathbf{I} \end{aligned} \tag{12}$$

To optimize (12), we solve the optimization problem according to the constrained optimization theory. The Lagrange function of (12) is

$$\begin{aligned} L = &tr\left(\mathbf{P}^{\mathrm{T}}\left(\mathbf{X}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}+\alpha(\mathbf{S}_w+\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}})+\beta\mathbf{I}\right)\mathbf{P}\right) \\ &+tr\left(\left(\mathbf{I}-\mathbf{P}^{\mathrm{T}}\mathbf{X}\mathbf{H}\mathbf{X}^{\mathrm{T}}\mathbf{P}\right)\boldsymbol{\Phi}\right) \end{aligned} \tag{13}$$

where $\boldsymbol{\Phi}=\mathrm{diag}\,(\phi_1,...,\phi_k)\in\mathcal{R}^{k\times k}$ are the Lagrange multipliers. By setting $\frac{\partial L}{\partial \mathbf{P}}=\mathbf{0}$, we thus get a generalized eigendecomposition problem

$$\left(\mathbf{X}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}+\alpha(\mathbf{S}_w+\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}})+\beta\mathbf{I}\right)\mathbf{P}=\mathbf{X}\mathbf{H}\mathbf{X}^{\mathrm{T}}\mathbf{P}\boldsymbol{\Phi}. \tag{14}$$

As a consequence, the optimal projection matrix $\mathbf{P}$ can be learned by solving (14) for the k smallest eigenvectors of $(\mathbf{X}\mathbf{H}\mathbf{X}^{\mathrm{T}})^{-1}(\mathbf{X}\sum_{c=0}^{C}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}+\alpha(\mathbf{S}_w+\mathbf{X}\mathbf{G}\mathbf{X}^{\mathrm{T}})+\beta\mathbf{I})$. Once the projection matrix $\mathbf{P}$ is available, we can project the original data into the latent feature space $\mathbf{P}^{\mathrm{T}}\mathbf{X}_s$ and $\mathbf{P}^{\mathrm{T}}\mathbf{X}_t$ with distribution shift reduced, discriminative information enhanced and local structure constrained. Then traditional machine learning algorithms such as SVM can be applied to the projected data to identify the source camera model of recompressed images. Also, it is worth noticing that the optimization procedure is iterated 10 times and the pseudo labels in validation set are iteratively refined after each iteration in order to learn a better and stable feature projection.

4 Experiments

In this section, we conduct extensive experiments to evaluate the proposed method on source camera model identification of recompressed images. The detailed experimental setup and experimental results can be found in the following subsections.

4.1 Experimental setup

4.1.1 Database

We evaluate our proposed method on a widely used public JPEG image database 'Dresden Image Database'[13]. Some examples used in the experiments are shown in Fig. 3. The image database is specifically built for the purpose of development and benchmarking of camera-based digital forensic techniques [13], whose images are captured by different cameras indoor/outdoor with various camera settings.

As the detailed information shown in Table 1, we choose 16 camera models from well-known camera brands in the Dresden Image Database to form the dataset used in our experiments. For each camera model, we have 180 training samples and 180 test samples. What's more, in order to eliminate the influence of different individuals of the same camera model, the samples in training and test set of the same camera model are from different individuals. That is to say, the identification model is built on one camera individual of each camera model, and we test our method on another individual of the same camera model. This test setting is more practical in that we aim to detect the source model of the given

Fig. 3 Sample images used in the experiments

Table 1 The camera models used in the experiments

Camera model	Abbr.	Format	Size
anon_Ixus70	C1	JPEG	3072×2304
Casio_EX-Z150	C2	JPEG	3264×2448
FujiFilm_FinePixJ50	F1	JPEG	3264×2448
Kodak_M1063	K1	JPEG	3664×2748
Nikon_CoolPixS710	N1	JPEG	4352×3264
Nikon_D70	N2	JPEG	3008×2000
Nikon_D200	N3	JPEG	3872×2592
Olympus_mju_1050SW	O1	JPEG	3648×2736
Panasonic_DMC-FZ50	P1	JPEG	3648×2736
Praktica_DCZ5.9	P2	JPEG	2560×1920
Rollei_RCP-7325XS	R1	JPEG	3072×2304
Samsung_L74wide	S1	JPEG	3072×2304
Samsung_NV15	S2	JPEG	3648×2736
Sony_DSC-H50	SD1	JPEG	3456×2592
Sony_DSC-T77	SD2	JPEG	3648×2736
Sony_DSC-W170	SD3	JPEG	3648×2736

images and also more difficult compared with the situation that the samples in training and test set are all from the same individual of each camera model.

4.1.2 Baselines

Our proposed method DFP is compared with several representative source camera model identification methods and several related feature projection based methods in the experiments. The representative forensic methods are: three SVM based methods with color filter array features (CFA) [3], local binary pattern features (LBP) [58], and texture features (TF) [59], two ensemble classifiers based methods with ensemble of demosaicing features (EDF) [8] and discrete cosine transform residue features (DCTR) [42]. The related feature projection based methods are: transfer component analysis (TCA) [40], joint distribution matching (JDA) [33], generalized transfer component analysis (GTCA) [28], close yet discriminative domain adaptation (CDDA) [36], cross-class alignment (CCA) [61] and joint first and second order statistics matching (JSM) [5]. Moreover, two subversions of DFP namely DFPv1 (DFP without discriminative information enhancement) and DFPv2 (DFP without local structure constraint) are also compared in the experiments in order to demonstrate the effectiveness of discriminative information enhancement and local structure constraint separately.

4.1.3 Implementation details

In the experiments, we use the LBP features proposed in [58] as the original features for feature projection based methods to identify the source camera model. As shown in Table 2, the bold font indicates the highest identification accuracies when training samples are not recompressed. In the left half of Table 2, the identification accuracies are obtained through traditional forensics methods. Compared with other methods, LBP features have the highest basic identification accuracies, which is the reason why we choose it as the original features

Table 2 Identification accuracy (%) when training samples are not recompressed

	Traditional forensic methods					Feature projection based methods								
Train→Test	LBP	LPQ	CFA	DCTR	EDF	CCA	TCA	GTCA	JDA	CDDA	JSM	DFPv1	DFPv2	DFP
SOri→TOri	96.25	93.26	92.53	92.85	**97.33**	90.97	92.05	89.24	94.44	93.99	95.31	96.32	96.46	96.46
SOri→T100	85.42	34.86	81.70	92.88	73.85	89.55	89.79	87.33	92.50	92.33	94.44	95.00	95.56	**95.83**
SOri→T95	62.95	36.35	61.32	63.78	38.96	66.25	68.58	67.47	85.94	83.33	87.33	89.90	89.90	**90.83**
SOri→T90	37.78	22.57	35.14	28.37	20.21	36.04	48.26	44.72	65.66	65.10	68.26	68.96	67.26	**69.27**
SOri→T85	23.23	18.78	14.86	13.78	13.61	29.03	32.81	41.11	52.57	51.63	**59.20**	57.12	56.04	58.89
SOri→T80	18.96	13.30	11.70	11.15	11.22	22.29	25.73	31.35	44.62	44.31	**52.40**	46.76	46.60	49.93
SOri→T75	11.98	13.13	10.52	9.38	11.11	16.74	21.25	25.10	42.36	39.62	**46.91**	40.42	41.46	44.83
SOri→T70	10.49	10.49	8.44	8.02	10.87	15.17	18.54	20.10	32.26	32.88	39.44	36.01	**39.44**	39.34
Ave.	43.38	30.34	39.53	40.03	34.35	45.76	49.63	50.80	63.79	62.90	67.91	66.31	66.59	**68.18**

for feature projection. The LBP features are extracted in the 1024×1024 subimage cropped from the upper left corner of each image, because the image size of the 16 camera models are different as shown in Table 1. Other compared methods also have the same setting mentioned above in order to have a fair comparison.

We conduct two experiments with different settings to evaluate the performance of our proposed method in recompression condition. For Experiment I, there are only original images in our training set but recompressed images in test set, and you can find the detailed description and experimental results in the next subsection. We also conduct Experiment II with mixed training set consisting of images recompressed with various quality factors to prove that our proposed method is suitable for more complicated forensic scenario. In this paper, the recompression of the original image is implemented by MATLAB function.

Note that there are 3 parameters involved in the objective function (12), we tune them through cross validation by empirically searching the parameter space $\alpha, \beta \in \{0.001, 0.01, 0.1, 1\}$ for the optimal parameter setting. For TCA, GTCA, JDA, CDDA and JSM, we also tune their parameters using the same way with DFP. As for subspace dimension k, we empirically fix $k = 50$ for all the experiments because it have little influence of the experimental results.

4.2 Results and discussion

In this section, we conduct extensive experiments under different situations as mentioned above to evaluate the performance of our proposed method. The classification accuracies of DFP and other compared methods are shown in Tables 2–5. The highest accuracy of each group of results is highlighted in bold.

4.2.1 Experiment I

In this experiment, in order to get the recompressed test images, we firstly compress the images in test set using standard quantization tables with quality factors from 70 to 100 with an interval of 5. We name the original training and test set as SOri and TOri, and the corresponding recompressed test sets are named as T70, T75, T80, T85, T90, T95 and T100.

For each test case, we randomly select 30% samples of each class in SOri and then compress them using the same quantization table of the test set to form the corresponding validation set. The regularization parameters $\alpha = 0.01$ and $\beta = 0.001$ are tuned by cross validation using the validation set. Taking T70 as an example, we compress the selected images with quality factor of 70 as the validation set, and learn the feature projection matrix between the remaining samples in SOri and validation set using DFP method. Then we evaluate the performance on the new feature representation using SVM and the experimental results are listed in Table 2. Note that this validation set construction strategy is unpractical because we have no priori knowledge about the test set in real forensic scenario. This experiment just aims to prove the effectiveness of DFP which can obviously eliminate the distribution divergence caused by recompression and yields good performance even there is a huge gap between training and test set.

From Table 2, we notice that traditional forensic methods performs very well when test set is TOri, but poorly when recompressed especially when the quality factor is quite low. The identification accuracy of LBP is 96.25% for SOri→TOri while EDF achieves a better result of 97.33%. When the test images are recompressed, there is a remarkable downward tendency of identification accuracy for all traditional forensic methods. The lower the recompression quality factor, the worse the classification performance. And the decreasing

rate of EDF is much more obvious compared with LBP as illustrated in Table 2. Though performing worse when test samples are recompressed, LBP [58] is still the best traditional forensic method in terms of average identification accuracy. LBP depicts an average accuracy of 43.38% which is much higher than other traditional forensic methods, and that's why we choose LBP features as the original features for feature projection based methods.

As for feature projection based methods, their identification accuracies are higher than LBP in almost all the test situations except for SOri→TOri. And it is obvious that our proposed method is the best among them. DFP achieves much better performance than the compared methods (both traditional forensic methods and feature projection based methods) in each mismatched case when test samples are recompressed. Taking SOri→T95 as an example, DFP shows an identification accuracy of 90.83%, gaining a significant improvement compared with 62.95% of LBP, 63.78% of DCTR, 85.94% of JDA, 83.33% of CDDA, 87.33% of JSM. When training set and test set are matched (case SOri→TOri), the result of DFP is also slightly higher than LBP which means our DFP method is still effective when there is no clear divergence between training and test set, and that is reasonable because the discriminative information in the training set is enhanced while the local structure is constrained. When the quality factor of test images decreases, the identification accuracy of all the methods has declined due to the growing divergence between training and test set, but DFP is still the best among the methods. DFP depicts an overall average accuracy of 68.18%, with an improvement of 24.8% compared with LBP and 0.27% compared with the best baseline method JSM.

Note that DFPv1 and DFPv2 are two subversions of DFP without discriminative information enhancement and local structure constraint separately. They both perform better than compared baseline feature projection based methods, which provides a good demonstration for the effectiveness of the proposed two components separately. Moreover, our proposed method DFP simultaneously enhances the discriminative information and constrains the local relations while minimizing the recompression divergence and leads to better performance than DFPv1 and DFPv2. And that is why DFP is more powerful in identifying the source camera model of recompressed images.

4.2.2 Experiment II

As mentioned in Experiment I, we always know nothing about the compression history of the query image in prior. So we can not construct the validation set according to the compression quality factor of test set as we did in Experiment I. In practical forensic scenario, the model trained on SOri in Experiment I is not suitable for complicated forensic demand, especially when we have several different compression quality factors in test set. To solve this problem, an intuitive solution is to retrain a classifier using a mixed training set consisting of images recompressed with different quality factors. So we also compress SOri using standard quantization tables with quality factors from 70 to 100 with an interval of 5. Thus, we get several sets named S70, S75, S80, S85, S90, S95 and S100. And we mix up these compressed training sets with SOri to get a new mixed training set named SMixed. The corresponding mixed test set is named as TMixed.

Then, we rebuild the identification model using SMixed as training set and report the experimental results in Tables 3–5. Similar to Experiment I, we should firstly learn the discriminative feature projection matrix for feature projection based methods. So we randomly choose 30% samples of each class in SMixed as a validation set to learn the feature projection matrix **P**. Since the mixed training set SMixed consists of several kinds of recompressed images, the selected validation set also consists various recompressed images. This

validation construction strategy no longer relies on the prior information of the test samples and is more practical in real forensic scenario. With the validation set, the regularization parameters α and β are tuned by cross validation as mentioned in implementation details, and we get $\alpha = 0.01, \beta = 0.001$ in this experiment. Once the feature projection matrix is available, we can project the original features into the new feature representation and train the identification model by SVM. Thus, given a query image, we should firstly project the feature of the test image using the same feature projection matrix and then get the identification result using the model learned above. It is worth noting that the feature projection matrix various test sets used when testing is exactly the same, we need only to learn the feature projection matrix once using the mixed training set and validation set and this can handle various recompressed situations.

The identification accuracies of each method using the mixed training set SMixed when test images are recompressed with aforementioned quality factors are shown in Table 3. For the sake of clarity, we list the results of different recompression cases separately. Each row shows the identification results of different methods on the same test case while each column shows the results of one method on various test cases. As shown in Table 3, all the traditional forensic methods perform barely satisfactory, LBP achieves an average accuracy of 89.06% and DCTR achieves a better result of 89.25% while LPQ, CFA and EDF performs worse with average accuracy lower than 89%. The identification accuracy of all the methods has declined with the decrease of the quality factor of test images, but the decreasing degree of identification accuracy is smaller compared with Experiment I. For SMixed→T70, the identification accuracy of LBP is 81.25% while DCTR achieves 84.38%. And our DFP method achieves 87.60%with an improvement of 6.35% and 3.22% compared with LBP and DCTR.

For feature projection based methods, TCA, JDA, JSM and DFP all performs better than LBP, especially our proposed method DFP. DFP performs better than the basic LBP method on every test case with an improvement of 1.8%-6.35%. In general, DFP depicts an average identification accuracy of 92.63% which is 3.57% higher than the basic LBP method and 2.26% higher than the best result of baseline method JSM. The promotion illustrates that divergence minimizing, discriminative information enhancement and local structure constraint help a lot in identifying the source camera when the test images are recompressed and this is a strong demonstration of the effectiveness of the proposed method.

To intuitively display the identification results of each camera model, we list the confusion matrix of the average result of DFP in Table 4. For the sake of simplicity, the accuracy lower than 0.5% is replaced with '-'. The average accuracy is 92.63% for 16 camera models which is quite encouraging for the identification of recompressed images with 8 different test conditions. From the confusion matrix, we can see that the highest identification accuracy is 98.68% recorded by Rollei RCP-7325XS while the least identification accuracy is 70.35% recorded by Sony_DSC-W170. One might notice that the accuracy of two Sony camera models (Sony_DSC-H50 and Sony_DSC-W170) is significantly lower than the average accuracy. And these two models are always misclassified into the other model. Similarly, we also find the same phenomenon in the results of all other compared methods, which means it is not a specific problem just happened in DFP. The reason for this interesting phenomenon is probably that these two models may share the same post-processing algorithms. And it is difficult to distinguish the images from these two models of the same brand especially when they share strong similarity of features. For other camera models, the identification accuracies are mostly higher than 93%.

Table 3 Identification accuracy (%) when training samples are mixed (SMixed)

	traditional forensic methods					feature projection based methods								
Train→Test	LBP	LPQ	CFA	DCTR	EDF	CCA	TCA	GTCA	JDA	CDDA	JSM	DFPv1	DFPv2	DFP
SMixed→TOri	94.62	91.39	90.03	92.01	92.78	94.38	94.24	91.22	94.10	91.74	94.93	94.72	96.46	**96.53**
SMixed→T100	94.58	90.63	88.51	92.33	92.64	94.41	93.75	90.87	93.89	91.01	94.58	94.90	96.15	**96.46**
SMixed→T95	92.92	89.90	87.01	91.32	91.56	92.85	93.23	89.83	93.40	87.99	94.31	94.03	94.86	**95.21**
SMixed→T90	91.08	88.44	80.35	90.45	89.51	90.42	90.83	84.31	90.42	85.87	91.91	91.74	92.88	**93.31**
SMixed→T85	87.95	86.63	72.43	89.06	86.18	88.37	88.58	79.93	89.13	84.34	90.10	90.24	91.56	**91.98**
SMixed→T80	85.83	85.56	64.65	88.02	82.40	86.35	87.36	77.08	87.33	84.13	87.60	88.44	89.93	**90.69**
SMixed→T75	84.27	83.58	58.85	86.42	78.19	84.03	85.14	72.92	85.35	83.33	85.73	87.19	88.09	**89.31**
SMixed→T70	81.25	81.53	51.42	84.38	71.08	81.22	83.26	69.41	82.81	81.22	83.82	84.55	86.01	**87.60**
Ave.	89.06	87.20	74.16	89.25	85.54	89.00	89.55	81.94	89.55	86.20	90.37	90.72	91.99	**92.63**

Table 4 Average confusion matrix for experiment II. '-' means less than 0.5%

Ave. 92.63%		Predicted															
		C1	C2	F1	K1	N1	N2	N3	O1	P1	P2	R1	S1	S2	SD1	SD2	SD3
Actual	C1	**97.71**	–	–	–	0.83	–	–	–	–	0.56	–	–	–	–	–	–
	C2	0.76	**93.61**	–	0.56	2.08	–	–	–	–	0.69	–	0.63	–	–	0.56	–
	F1	–	–	**98.13**	–	–	–	–	–	–	–	0.76	–	–	–	–	–
	K1	–	–	–	**97.01**	–	1.53	–	0.35	–	0.42	0.21	–	–	–	–	–
	N1	0.69	–	–	–	**95.28**	–	0.76	–	–	–	0.63	1.04	–	–	–	–
	N2	0.83	0.69	–	–	–	**89.51**	7.15	–	–	–	–	–	–	–	–	–
	N3	–	–	–	–	–	2.36	**93.26**	2.78	–	–	–	0.56	–	–	–	–
	O1	–	–	–	–	0.97	–	–	**96.81**	–	–	–	1.11	–	–	–	–
	P1	–	–	–	0.56	–	–	–	–	**94.65**	–	–	–	–	0.56	–	1.74
	P2	–	–	–	–	–	–	0.56	0.14	0.28	**97.78**	–	0.63	–	–	–	–
	R1	–	–	–	–	–	–	–	–	–	–	**98.68**	–	–	–	–	–
	S1	–	–	–	–	0.63	–	0.69	1.74	–	1.11	0.97	**94.31**	–	–	–	–
	S2	–	1.39	–	–	0.63	–	–	2.22	–	–	–	2.15	**91.74**	–	–	–
	SD1	0.83	–	–	–	–	–	–	–	–	–	–	–	0.56	**84.38**	2.92	10.63
	SD2	–	–	–	–	–	–	–	–	–	–	–	–	–	4.79	**88.96**	4.93
	SD3	0.56	–	–	–	–	–	–	1.11	1.25	–	–	–	1.04	20.00	3.82	**70.35**

As mentioned before, the practical forensic demand is quite complicated. It is a quite common situation that the compression quality factors of query images may even not included in the mixed training set and this situation needs to be taken into consideration in the experiments. To evaluate the performance of our proposed method under this situation, we also compress the original test images TOri with quality factor 72, 77, 82, 87, 92 and 97, and report the identification results in Table 5 using the same model learned above. As one might expect, the identification accuracies of all the 11 methods have decreased compared with the results in Table 3, but DFP is still the best compared with other methods. The average identification accuracy of DFP has decreased to 89.99% from 92.63% while LBP has decreased to 85.83% from 89.06%. Under this situation, the promotion compared with the basic LBP method is 4.16% while the promotion compared with JSM is 2.65%. Though the average accuracy has declined, the average promotion compared with LBP and JSM has increased, which proves the effectiveness of our proposed method under this complicated situation.

4.2.3 Computational cost analysis

For a fair comparison, we analyze quantitatively the computational cost by the example of SOri→T95 in Experiment I. The details of experimental platform are as follows: Windows 10 64-bit, Intel(R) Core(TM) i7-6700 CPU @3.40GHz, 64G RAM, MATLAB R2014a. The results of computational cost (time consuming) are shown in Table 6. Costs are mainly divided into two aspects: the cost of learning the feature projection and the cost of training the classifier. The cost of learning the feature projection is much higher than other feature projection based methods, which is the drawback and also the price for identification accuracy improvement of the proposed method. Moreover, since the feature projection is learned in a principled dimensionality reduction procedure, the dimension of identification features in the new feature space is reduced to k, which is 50 as mentioned in above section, and that is why the cost of training classifier of the proposed method is much lower than the basic LBP method. And EDF is a quite outstanding method whose performance is good while the computational cost is low. Generally, the computational cost of the proposed method is a bit high but within an acceptable range, it is reasonable to improve the identification accuracy as the cost of time as long as the computational cost is acceptable.

4.2.4 Generalization performance analysis

In order to illustrate the generalization performance of the method proposed in this paper, we conducted an extended experiment on the VISION dataset. The VISION dataset contains images and videos from 35 phones from 11 different brands, which includes 11,732 original images, 7,565 high-quality and low-quality images from Facebook and 7,565 images from WhatsApp for a total of 34,427 images. As a training set, we randomly selected 10 types of images from VISION, 200 images for each type. From Table 7, we notice that traditional forensic LBP methods performs poorly when recompressed especially when the quality factor is quite low. However, the DFP method proposed in this article has a certain improvement in accuracy compared to the original method before transfer learning. In total, for high-quality images, this method does not improve much, and for low-quality images, there is a more obvious improvement.

Table 5 Identification accuracy (%) when training samples are mixed (SMixed)

	Traditional forensic methods					Feature projection based methods								
Train→Test	LBP	LPQ	CFA	DCTR	EDF	CCA	TCA	GTCA	JDA	CDDA	JSM	DFPv1	DFPv2	DFP
SMixed→T97	91.42	89.83	84.17	82.43	91.81	91.81	91.74	85.87	92.12	87.60	93.33	93.37	94.06	**94.44**
SMixed→T92	87.36	88.13	80.45	75.97	90.97	87.01	90.03	83.44	89.76	84.55	90.80	91.32	91.97	**92.95**
SMixed→T87	87.95	87.40	73.96	68.37	87.47	88.09	88.68	80.03	88.30	83.51	89.58	90.42	91.18	**91.42**
SMixed→T82	85.17	85.28	64.31	75.59	84.24	85.17	85.03	77.15	86.08	81.84	86.60	87.26	88.06	**88.72**
SMixed→T77	82.36	83.51	59.90	74.58	79.58	82.92	82.99	72.74	81.46	78.65	81.63	85.28	85.80	**86.84**
SMixed→T72	80.69	82.57	54.24	75.59	73.37	80.59	82.29	68.30	81.39	79.58	82.12	81.91	85.31	**85.56**
Ave.	85.83	86.12	69.50	75.42	84.57	85.93	86.79	77.92	86.52	82.62	87.34	88.26	89.40	**89.99**

Table 6 Computational time (s) of SOri→T95 in Experiment I

	Traditional forensic methods					Feature projection based methods					
Methods	LBP	LPQ	CFA	DCTR	EDF	CCA	TCA	GTCA	JDA	CDDA	DFP
Feature projection learning	0	0	0	0	0	2.3	4.4	4.6	135.2	303.3	**1404.3**
Classifier training	1583.9	**6799.8**	2884.5	34.9	47.9	1497.1	400.5	451.2	441.4	416.6	435.1
Overall cost	1583.9	**6799.8**	2884.5	34.9	47.9	1499.4	404.9	455.8	576.6	719.9	1839.4

Table 7 Identification accuracy (%) on the VISION dataset

Train→Test	LBP	DFPv1	DFPv2	DFP
SMixed→T100	79.50	79.94	**80.00**	**80.00**
SMixed→T95	79.33	**80.00**	79.83	79.94
SMixed→T90	78.67	**79.89**	79.83	**79.89**
SMixed→T85	79.50	**79.94**	79.72	**79.94**
SMixed→T80	76.11	79.17	78.50	**79.44**
SMixed→T75	71.00	77.28	76.67	**77.39**
SMixed→T70	68.61	75.33	74.78	**75.94**
Ave.	76.10	78.79	78.48	**78.93**

5 Conclusion

In this paper, we propose a method discriminative feature projection to identify the source camera model of recompressed images. In the proposed method, the projection matrix which projects the original features into a new feature representation that is invariant to recompression is firstly learned by simultaneously minimizing the divergence caused by recompression, enhancing the discriminative information and constraining the local structure relations. Then the training and test data are projected into the learned feature representation and the identification is conducted under the new feature space using SVM. The cooperation of divergence minimization, discriminative information enhancement and local structure constraint makes the identification results under the learned feature representation more reliable. The experimental results demonstrate the effectiveness of the proposed method and show that the performance of DFP is superior to existing methods when identifying the source camera model of recompressed images.

Acknowledgments This work is supported by the National Natural Science Foundation of China (No. U1936117, No. 62076052, No. 61772111), and the Fundamental Research Funds for the Central Universities (DUT21GF303, DUT20RC(3)088).

References

1. Bayram S, Sencar HT, Memon N (2012) Efficient sensor fingerprint matching through fingerprint binarization. IEEE Trans Inf Forensics Secur 7(4):1404–1413
2. Bayram S, Sencar HT, Memon N (2015) Sensor fingerprint identification through composite fingerprints and group testing. IEEE Trans Inf Forensics Secur 10(3):597–612
3. Bayram S, Sencar H, Memon N, Avcibas I (2005) Source camera identification based on cfa interpolation. In: Proc IEEE Int Conf Image process, vol 3, Genova, Italy, pp 69–72
4. Bo W, Guo Y, Kong X, Meng F (2009) Source camera identification forensics based on wavelet features. In: Proc IEEE Fifth Int Conf Intelligent inf. Hiding and multimedia signal process, Kyoto, Japan, pp 702–705
5. Bo W, Li Y, Sui X, Li M, Guo Y (2019) Joint statistics matching for camera model identification of recompressed images. Math Biosci Eng 16(5):5041–5061
6. Calmon FdP, Wei D, Vinzamuri B, Ramamurthy KN, Varshney KR (2018) Data pre-processing for discrimination prevention: Information-theoretic optimization and analysis. IEEE J Select Topics Signal Process 12(5):1106–1119
7. Chen M, Fridrich J, Goljan M, Lukás J (2008) Determining image origin and integrity using sensor noise. IEEE Trans Inf Forensics Security 3(1):74–90

8. Chen C, Stamm MC (2015) Camera model identification framework using an ensemble of demosaicing features. In: Proc IEEE international workshop on information forensics and security (WIFS), pp 1–6
9. Chierchia G, Parrilli S, Poggi G, Sansone C, Verdoliva L (2010) On the influence of denoising in prnu based forgery detection. In: Proceedings of the 2Nd ACM Workshop on Multimedia in Forensics, Security and Intelligence, MiFor '10. ACM, New York, pp 117–122
10. Choi KS, Lam EY, Wong KKY (2006) Source camera identification by jpeg compression statistics for image forensics. In: Proc IEEE Region 10 Conf, Hong Kong, China, pp 1–4
11. Gao S, Xu G, Hu R-M (2011) Camera model identification based on the characteristic of cfa and interpolation. In: Proc International Workshop on Digital Watermarking (IWDW). Springer, Berlin, pp 268–280
12. Gloe T (2012) Feature-based forensic camera model identification. Trans Data Hiding and Multimedia Security, pp 42–62
13. Gloe T, Böhme R (2010) The dresden image database for benchmarking digital image forensics. J Digital Forensic Practice 3(2-4):150–159
14. Goljan M, Fridrich JJ (2013) Sensor fingerprint digests for fast camera identification from geometrically distorted images. In: Proc SPIE, Burlingame, CA, USA, pp 86650B–86650B10
15. Gretton A, Borgwardt KM, Rasch M, Schölkopf B, Smola AJ (2007) A kernel method for the two-sample-problem. In: Proc Advances neural inf Process Syst, Vancouver, BC, Canada, pp 513–520
16. Ho JS, Oscar CA, Zhou J, Guo Y (2010) Inter-channel demosaicking traces for digital image forensics. In: Proc IEEE Int Conf Multimedia expo, Suntec, Singapore, pp 1475–1480
17. Hou J-U, Lee H-K (2016) Detection of hue modification using photo response nonuniformity. IEEE Trans Circ Syst Video Technol 27(8):1826–1832
18. Hu Y, Jian C, Li CT (2010) Using improved imaging sensor pattern noise for source camera identification. In: Proc IEEE Int Conf Multimedia expo, Suntec, Singapore, pp 1481–1486
19. Hu Y, Li CT, Lai Z (2015) Fast source camera identification using matching signs between query and reference fingerprints. Multimedia Tools and Applications 74(18):7405–7428
20. Hu Y, Li CT, Lin X, Liu B-b (2012) An improved algorithm for camera model identification using inter-channel demosaicking traces. In: 8Th Int Conf Intelligent Inform Hiding and multimedia signal process, Piraeus, Greece, pp 325–330
21. Kang X, Li Y, Qu Z, Huang F (2012) Enhanced source camera identification performance with a camera reference phase sensor. IEEE Trans Inf Forensics Security 7(2):393–402
22. Kharrazi M, Sencar HT, Memon N (2004) Blind source camera identification. In: Proc IEEE Int Conf Image Process., Vol 1, Singapore, pp 709–712
23. Khosravi MR, Samadi S (2019) Data compression in visar sensor networks using non-linear adaptive weighting. EURASIP J Wirel Commun Netw 2019(1):1–8
24. Kirchner M, Gloe T (2015) Forensic camera model identification. In: Handbook of digital forensics of multimedia data and devices. Wiley, Chichester, pp 3299–374
25. Lawgaly A, Khelifi F (2017) Sensor pattern noise estimation based on improved locally adaptive dct filtering and weighted averaging for source camera identification and verification. IEEE Trans Inf Forensics Security 12(2):392–404
26. Lawgaly A, Khelifi F, Bouridane A (2014) Weighted averaging-based sensor pattern noise estimation for source camera identification. In: Proc IEEE Int Conf Image process, Pari, France, pp 5357–5361
27. Li CT (2010) Source camera identification using enhanced sensor pattern noise. IEEE Trans Inf Forensics Security 5(2):280–287
28. Li X, Kong X, Bo W, Guo Y, You X (2013) Generalized transfer component analysis for mismatched jpeg steganalysis. In: Proc IEEE Int Conf ssImage process, Melbourne, VIC, Australia, pp 4432–4436
29. Li C, Li Y (2012) Color-decoupled photo response non-uniformity for digital image forensics. IEEE Trans Circ Syst Video Technol 22(2):260–271
30. Li R, Li CT, Guan Y (2018) Inference of a compact representation of sensor fingerprint for source camera identification. Pattern Recogn 74((Supplement C)):556–567
31. Li Z, Liu J, Tang J, Lu H (2015) Robust structured subspace learning for data representation. IEEE Trans Pattern Anal Mach Intell 37(10):2085–2098
32. Lin X, Li CT (2016) Enhancing sensor pattern noise via filtering distortion removal. IEEE Signal Process Lett 23(3):381–385
33. Long M, Wang J, Ding G, Sun J, Yu PS (2013) Transfer feature learning with joint distribution adaptation. In: Proc IEEE Int Conf Comput Vision, San Francisco, CA, USA, pp 2200–2207
34. Long M, Wang J, Ding G, Sun J, Yu PS (2014) Transfer joint matching for unsupervised domain adaptation. In: Proc IEEE Conf Comput Vision pattern recognit, Columbus, OH, USA, pp 1410–1417
35. Lukáš J, Fridrich J, Goljan M (2005) Determining digital image origin using sensor imperfections. In: Proc SPIE, San Jose, CA, USA, pp 249–260

36. Luo L, Wang X, Hu S, Wang C, Tang Y, Chen L (2017) Close yet distinctive domain adaptation. arXiv:1704.04235
37. Marra F, Poggi G, Sansone C, Verdoliva L (2015) Evaluation of residual-based local features for camera model identification. In: Proc International Conference on Image Analysis and Processing. Springer International Publishing, Genoa, pp 11–18
38. Marra F, Poggi G, Sansone C, Verdoliva L (2017) A study of co-occurrence based local features for camera model identification. Multimedia Tools and Applications 76(4):4765–4781
39. Mo, Fridrich JJ, Goljan M (2007) Digital imaging sensor identification (further study). In: Proc SPIE, San Jose, CA, USA, pp 65050P–65050P13
40. Pan SJ, Tsang IW, Kwok JT, Yang Q (2011) Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 22(2):199–210
41. Piva A (2013) An overview on image forensics. Isrn Signal Process., 2013(1)
42. Roy A, Chakraborty RS, Sameer U, Naskar R (2017) Camera source identification using discrete cosine transform residue features and ensemble classifier. In Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops
43. Stamm MC, Wu M, Liu KJR (2013) Information forensics: an overview of the first decade. IEEE Access 1:167–200
44. Tang Z, Chen L, Zhang X, Zhang S (2019) Robust image hashing with tensor decomposition. IEEE Trans Knowl Data Eng 31(3):549–560
45. Tang Z, Zhang X, Li X, Zhang S (2016) Robust image hashing with ring partition and invariant vector distance. IEEE Trans Inform Forensics Security 11(1):200–214
46. Tomioka Y, Ito Y, Kitazawa H (2013) Robust digital camera identification based on pairwise magnitude relations of clustered sensor pattern noise. IEEE Trans Inf Forensics Security 8(12):1986–1995
47. Tuama A, Comby F, Chaumont M (2016) Camera model identification based machine learning approach with high order statistics features. In: Proc 24th European signal processing conference (EUSIPCO), pp 1183–1187
48. Unar S, Wang X, Wang C, Wang M (2019) New strategy for cbir by combining low-level visual features with a colour descriptor. IET Image Process 13:1191–1200
49. Unar S, Wang X, Wang C, Wang Y (2019) A decisive content based image retrieval approach for feature fusion in visual and textual images. Knowl-Based Syst 179:8–20
50. Unar Sn, Wang X, Zhang C (2018) Visual and textual information fusion using kernel method for content based image retrieval. Inform Fusion 44:176–187
51. Unar S, Wang X, Zhang C, Wang C (2019) Detected text-based image retrieval approach for textual images. IET Image Process 13:515–521
52. Valsesia D, Coluccia G, Bianchi T, Magli E (2015) Compressed fingerprint matching and camera identification via random projections. IEEE Trans Inf Forensics Secur 10(7):1472–1485
53. Wang X, Chen Z (2009) A fast fractal coding in application of image retrieval. Fractals 17(04):441–450
54. Wang X, Chen Z, Yun J (2012) An effective method for color image retrieval based on texture. Computer Standards & Interfaces 34(1):31–35
55. Wang X, Wang Z (2013) A novel method for image retrieval based on structure elements' descriptor. J Vis Commun Image Represent 24(1):63–74
56. Wang X, Wang Z (2014) The method for image retrieval based on multi-factors correlation utilizing block truncation coding. Pattern Recogn 47(10):3293–3303
57. Wu G, Kang X, Ray Liu KJ (2012) A context adaptive predictor of sensor pattern noise for camera source identification. In: Proc IEEE Int Conf Image process, Orlando, FL, USA, pp 237–240
58. Xu G, Shi YQ (2012) Camera model identification using local binary patterns. In: Proc IEEE Int Conf Multimedia expo, Melbourne, VIC, Australia
59. Xu B, Wang X, Zhou X, Xi J, Wang S (2016) Source camera identification from image texture features. Neurocomputing 207:131–140
60. Zeng L, Kong X, Li M, Guo Y (2015) Jpeg quantization table mismatched steganalysis via robust discriminative feature transformation. In: Proc SPIE Media Watermarking, Security, and Forensics, San Francisco, CA, USA, pp 94090u–94090u9
61. Zhang G, Bo W, Li Y (2017) Cross-class and inter-class alignment based camera source identification for re-compression images. In: Proc IEEE Int Conf Image graphics, Shanghai, China