Camera Source Identification with Limited Labeled Training Set

Yue Tan[1], Bo Wang[1(✉)], Ming Li[1], Yanqing Guo[1], Xiangwei Kong[1], and Yunqing Shi[2]

[1] Dalian University of Technology, Dalian, Liaoning, China
yue985@mail.dlut.edu.cn, {bowang,mli,guoyq,kongxw}@dlut.edu.cn
[2] New Jersey Institute of Technology, Newark, NJ, USA
shi@njit.edu

Abstract. This paper investigates the problem of model-based camera source identification with limited labeled training samples. We consider the realistic scenario in which the number of labeled training samples is limited. Ensemble projection (EP) method is proposed by introducing prototype theory into semi-supervised learning. After constructing sub-sets of local binary patterns (LBP) features, several pre-classifiers are established for all labeled and unlabeled samples. According to the ranking of posterior probabilities, several prototype sets are constructed for the ensemble projection. Combining the outputs of all labeled samples from classifiers trained by prototype sets, a new feature vector is generated for camera source identification. Experimental results illustrate that the proposed EP method achieves a notable higher average accuracy than previous algorithms when labeled training samples is limited.

Keywords: Camera source identification · Limited labeled training samples · Ensemble projection · LBP features

1 Introduction

The advances of digital technologies, including low-cost digital cameras, sophisticated image editing software and internet techniques, bring us convenience, as well as a new issue and challenge with the integrity and authenticity of digital images. Developing reliable and accurate algorithms to verify the trust worthiness of digital images becomes an urgent need for law enforcement authorities, legal affairs, etc. Passive digital image forensics, which is considered as a promising solution of digital image authentication, has attracted more and more research interests. As an important branch, source camera identification focuses on the authentication of the originality of digital images and has significant potential in the applications of digital image forensics.

In recent years, various techniques have been proposed to solve the problem of source camera identification. These approaches can be categorized into two classes. The first one is tracing a unique intrinsic fingerprint of a specific device. Lukas *et al.* first utilized the photo response non-uniformity noise of

Y.-Q. Shi et al. (Eds.): IWDW 2015, LNCS 9569, pp. 18–27, 2016.
DOI: 10.1007/978-3-319-31960-5_2

imaging sensors as a device fingerprint for camera-based source identification [1]. Later on, a series of improved algorithms were proposed [2–7]. Another branch of image source identification is model-based source identification. A typical solution of model-based camera source identification is based on extracting multi-dimensional statistical characteristics for classification. For example, Swaminathan *et al.* [8] proposed a method for source camera identification by the estimation of CFA pattern and interpolation kernel, and gained an overall average accuracy of 90 % for 19 camera brands. Kharrazi *et al.* [9] proposed 34 features which can be categorized into three types: color features, image quality measurement (IQM), and high order wavelet characteristics (HOWS). They considered 16 models of cell-phone cam-eras. Then a classifier based on the features achieved an average accuracy of 88.02 %. Recently, Xu and Shi used the uniform gray-scale invariant local binary patterns (LBP) [10] and received an average classification accuracy of 98.0 % for 18 camera models from "Dresden Image Database".

In the previous works, a sufficient number of labeled training samples is indispensable to construct an accurate classifier. For example, a training set consisting of 150 images for each camera brand was used in [9]. And similarly in [10], the author used around 150 to 300 labeled images as training samples for each class. For camera model identification, it is possible to obtain enough labeled image samples to train a sophisticated classifier. While considering the time and manpower cost, collecting amount of labeled images as training set is usually not a simple task. For realistic scenario, reducing the labeled training samples meanwhile keeping the identification accuracies should be a practical and important problem for moving camera model identification from the laboratory into the real world. In this paper, we consider the realistic scenario in which the number of labeled training samples is limited. We propose to use prototype theory and ensemble projection method to achieve camera source identification with limited labeled training samples. Through constructing a series prototype set using LBP features, we obtain multiple sample sets. It means we include more information in training set from limited labeled samples. Ensemble learning features are proposed in our paper, and we use this feature to train SVM and identify the camera model source.

The rest of the paper is organized as follows. In Sect. 2, we introduce local binary patterns method for camera source identification. In Sect. 3, we proposed the EP method to solve the camera source identification with limited samples. In Sect. 4, experimental results demonstrate the efficiency of the proposed method. Finally, conclusions are drawn in Sect. 5.

2 Local Binary Patterns

The uniform gray-scale invariant local binary patterns [11] can be described by

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{1}$$

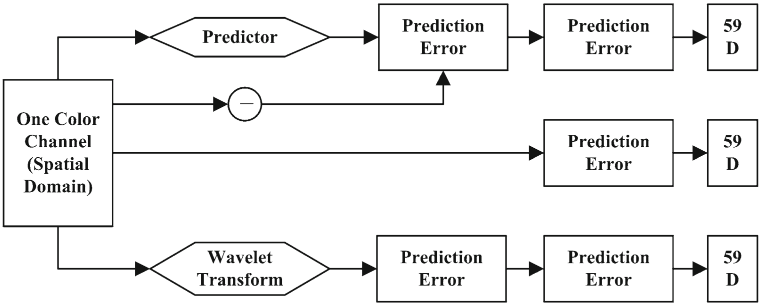

Fig. 1. Feature extraction framework for one color channel (Color figure online).

where g_c represents a gray level of the center pixel, g_p, $p = 1, \ldots, P$, represent its neighbor pixels which are located on a circle with center at g_c and a radius R. Then we can use the P samples to calculation local binary pattern. In this paper, we set $P = 8$, $R = 1$. The function $s(x)$ can be defined as

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} . \tag{2}$$

From Eqs. (1) and (2), we can calculate the gray-level difference between center pixel and its eight neighbors. The difference value s between each couple point has totally $2^8 = 256$ patterns.In [10], the “uniform” local binary pattern and “non-uniform” local binary pattern were used, inspired by [11]. While “uniform” local binary pattern occupies majority of the total patterns, the authors only consider 58 “uniform” patterns and all of the “non-uniform” patterns are merged to one pattern. Thereby, the number of effective patterns is reduced from 256 to 59.

Finally, from each color channel, we extract LBP features from (i) original image, (ii) its prediction-error 2D array, and (iii) its 1^{st}-level diagonal wavelet subband, resulting in a total of $59 \times 3 = 177$ features. Since red and blue color channels usually share the same image processing algorithms, only red and green channels are considered. Therefore, the final dimensions of the feature extracted from a color image is $177 \times 2 = 354$. The feature extraction framework is shown in Fig. 1.

3 Proposed Method

To our best knowledge, most of model-based camera source identification algorithms need large-scale labeled samples databases to sophisticatedly train a model and construct a classifier. Unfortunately, in many realistic cases, the labeled samples may be limited for model training but unlabeled samples are always sufficient. Our goal is to construct a reliable model for camera source identification with a rather limited number of labeled samples in conjunction with lots of unlabeled samples. Semi-supervised learning and ensemble learning

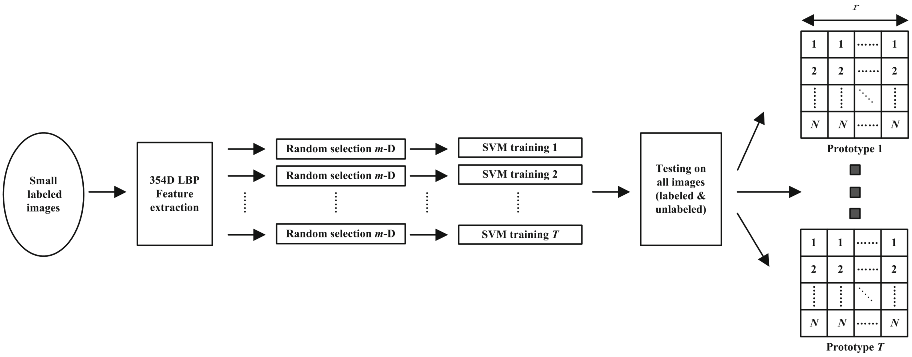

Fig. 2. Pipeline of constructing prototype set.

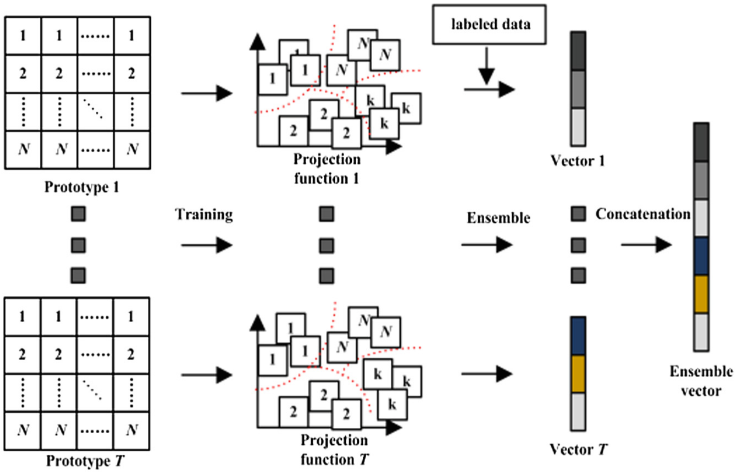

Fig. 3. Ensemble projection.

have been proposed to solve similar problem. However, the accuracy of classification is unsatisfactory when the number of the labeled samples is very low. To solve this problem, in this paper, we consider to construct a series of prototypes set by multiple random selected dimensional from LBP feature. And we propose a new ensemble projection algorithm to make combination of the information from all prototypes. The proposed method consists of two components, constructing prototype set and ensemble projection. The framework of these two steps are shown in Figs. 2 and 3.

3.1 Constructing Prototype Set

Eleanor Rosch's prototype theory [12] is a mode of graded categorization in cognitive science, where some members of a category are more central than others.

For example, when asked to give an example of the concept furniture, chair is more frequently cited than stool. This means if we can find some unlabeled samples with more probability belong to a certain class, and using the information in these samples, we can get a well result for classification. This procedure is just like the first process in semi-supervised learning method, but we don't consider the circulation process in semi-supervised learning method. Peter Gärdenfors [13] has elaborated a possible partial explanation of prototype theory in terms of multi-dimensional feature spaces called Conceptual Spaces, where a category is defined in terms of a conceptual distance. More central members of a category are "between" the peripheral members. He postulates that most natural categories exhibit a convexity in conceptual space, in that if x and y are elements of a category, and if z is between x and y, then z is also likely to belong to the category. According this, we believe if we can construct the projection set by different ways, and making combination of these result we can get a higher accuracy for classification.

Since the number of training samples is sometimes much less than feature dimension, under-training is a common problem in limited labeled samples scenario. In an effort to mitigate the impact of the under-training problem, we can construct several prototype sets. A dataset D is given, which consisted by labeled and unlabeled samples, so we divided it into two subsets D_l and D_u. And the dataset contain N-class samples. The goal is using the labeled samples to classify the unlabeled samples. We first extract 354 dimensions LBP features for all samples according to the method proposed in [10]. Then, from the labeled samples in set D_l, we randomly select m-dimensional features from 354 dimensions LBP features [10], and use m-dimensional features to train an N-class classifier. Then, we can predict labels for the all samples in D. Through this procedure each sample get an posterior probabilities belong to each class. By sorting the posterior probabilities from large to small for all samples, top r sample images with higher posterior probabilities in each class are selected to construct a prototype. This procedure is repeated T times, then T prototypes are constructed subsequently, as shown in Fig. 2. In this figure, the "small" labeled images refer to the labeled samples is limited, and the m-D represents m-dimensional features.

The samples, which have equal probability of belonging to each class, may influence the classification accuracy. We named those ambiguous samples as noise samples. According to information theory, if an image has equal probability of belonging to each category, it has the largest entropy. The entropy can explicitly be written as:

$$H = -\sum_{i=1}^{N} p(c_i)\log_2 p(c_i) \tag{3}$$

where $p(c_i)$ represents the probability of belonging to c_i class. So we can set a threshold value e to exclude those samples from prototype set. If an image's entropy is less than the threshold e, it will be used for classifier training; otherwise, we consider it as a noise sample and ignore.

3.2 Ensemble Projection

Inspired by the ensemble learning [14], each prototype set, which represents partial classification information, can be seen as a new training set and T classifiers could be trained. For each labeled image, we can obtain T projection vectors from T classifiers. Each projection vector is assembled by all of the posterior probabilities belong to N classes, and the dimension of projection vector is $N \times 1$. By combining all T vectors, an $NT \times 1$ dimensional feature for a labeled image is obtained. For all labeled images, the feature vectors are extracted and then fed to SVM to construct a final classifier to identify the camera model, as shown in Fig. 3.

4 Experimental Studies

We carry out our experiments using "Dresden Image Dataset" used in [15]. In this dataset, 18 camera models (see Table 1) are employed and 350 JPEG images are captured by each camera model with varying settings. The LBP features used in our method are extracted in the central block with size of 512×512 and SVM classifier [16] is employed in our experiment. For comparison purpose, the LBP algorithm [10] is also evaluated.

Table 1. Dataset in Experiments.

Camera model	Resolution	Abbr
Casio_EX_Z150	3264 × 2448	CEZ
Kodak_M1063	3664 × 2748	KM1
Nikon_CoolPixS710	4352 × 3264	NCP
Olympus_mju	3648 × 2736	OMJ
Panasonic_DMC	3264 × 2736	PDM
Praktica_DCZ5.9	2560 × 1920	PDC
Nikon_D200	3872 × 2592	ND1
Ricoh_GX100	3648 × 2736	RGX
FujiFilm_FinePixJ50	3264 × 2448	FFP
Pentax_OptioA40	4000 × 3000	POA
Rollei_RCP_7325X	3072 × 2304	RRC
Samsung_L74wide	3072 × 2304	SLW
Samsung_NV15	3648 × 2736	SNV
Sony_DSC_H50	3456 × 2592	SD1
Sony_DSC_T77	3648 × 2736	SD2
Agfa_Sensor530s	2560 × 1920	AFS
Canon_Ixus70	3072 × 2304	CI7
Nikon_D70	3008 × 2000	ND2

Table 2. Average comfusion matrix obtained by svm classification over 20 iterations.

Average TP=90.2		Predicted																	
		CEZ	KM1	NCP	OMJ	PDM	PDC	ND1	RGX	FFP	POA	RRC	SLM	SNY	SD1	SD2	AFS	CI7	ND2
Actual	CEZ	87.7	*	1.3	*	2.3	*	*	*	*	3.7	*	*	3.0	*	*	*	*	*
	KM1	*	90.0	*	*	1.3	*	*	*	*	1.3	1.0	*	*	4.0	*	*	*	*
	NCP	*	*	92.7	*	*	2.7	*	*	*	1.7	*	*	*	*	*	*	1.0	*
	OMJ	*	*	*	92.0	*	*	*	*	*	*	*	4.0	2.7	*	*	*	*	*
	PDM	*	1.3	1.0	1.0	90.0	2.3	*	*	*	1.0	*	*	*	1.3	*	*	*	*
	PDC	*	*	*	*	1.7	95.3	*	*	*	*	*	*	2.3	*	*	*	*	*
	ND1	*	*	*	*	1.7	2.0	90.0	*	*	*	*	1.0	3.7	*	*	*	1.0	*
	RGX	*	5.0	*	*	*	*	*	85.0	*	*	*	*	1.0	4.3	2.7	*	*	1.0
	FFP	*	*	*	*	*	1.0	*	*	90.0	*	2.3	*	2.0	*	*	*	*	2.7
	POA	*	*	*	1.7	*	*	*	*	*	89.3	*	1.0	*	*	*	1.0	*	4.3
	RRC	*	*	*	*	*	*	2.7	*	*	*	92.3	*	2.0	*	*	*	*	*
	SLM	*	*	1.3	*	*	*	*	*	*	*	1.3	93.7	3.0	*	*	*	*	*
	SNY	*	*	*	*	*	4.0	1.0	*	1.0	*	*	*	89.0	1.0	*	*	*	*
	SD1	*	1.3	*	*	*	*	*	6.7	*	*	*	*	*	88.7	2.3	*	*	*
	SD2	*	*	*	*	*	*	*	2.0	*	*	*	*	*	9.0	88.0	*	*	*
	AFS	*	*	*	*	*	*	*	*	*	1.3	1.0	*	*	*	1.0	91.0	*	5.0
	CI7	1.0	*	*	*	1.0	*	*	*	*	*	*	*	1.3	*	*	*	94.3	*
	ND2	*	*	*	*	*	*	1.0	3.7	*	3.7	*	*	*	1.0	*	5.3	*	83.7

To prove the effectiveness of the method, the labeled images of each camera model are randomly selected from the image dataset, and the rest of images are considered as unlabeled and test samples. Experimental results are shown in Table 2, which give the average accuracy over 20 iterations. The asterisks in the table represent the classification probability below one percent. As demonstrated, our EP method achieves a highest accuracy of 95.3 % for Praktica_DCZ5.9 and a lowest accuracy of 83.7 % for Nikon D70. The overall average classification accuracy for 18 camera model is 90.2 %, when the number of labeled samples is $L = 50$, the number of prototype sets is T=200, and the number of the samples of each class in the prototype sets is $r = 50$.

Then, we attempt to evaluated the accuracy performance of the proposed algorithm, under different parameter L, T, r, condition. We first investigate the influence on parameter L,the numbers of labeled images, to the performance. We carry out experiment under the condition that the number of prototype sets is $T = 50$, and the number of samples of each class in the prototype sets is $r = 50$. As shown in Table 3, our EP method achieves average classification accuracy of 90.2 %, when the number of labeled samples is $L = 50$, while the LBP algorithm can only obtain the accuracy of 36.0 %. When the number of labeled samples decrease to an extremely low level $L = 10$, the classification accuracy of LBP algorithm is as low as 8.4 %, but our EP method can still maintain 74.5 %.

Table 3. Average accuracy of camera source identification with different number of labeled image samples.

Algorithm	$L = 50$	$L = 40$	$L = 30$	$L = 20$	$L = 10$
LBP	36.0 %	26.7 %	29.3 %	20.9 %	8.4 %
EP	**90.2** %	**88.3** %	**85.0** %	**82.6** %	**74.5** %

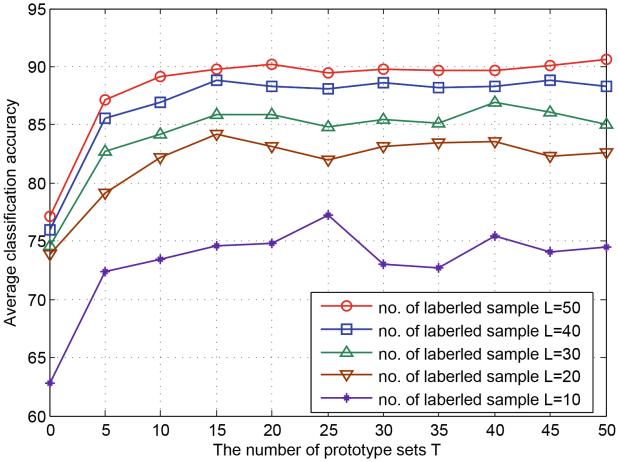

Fig. 4. Accuracy rate versus the number of prototype sets T.

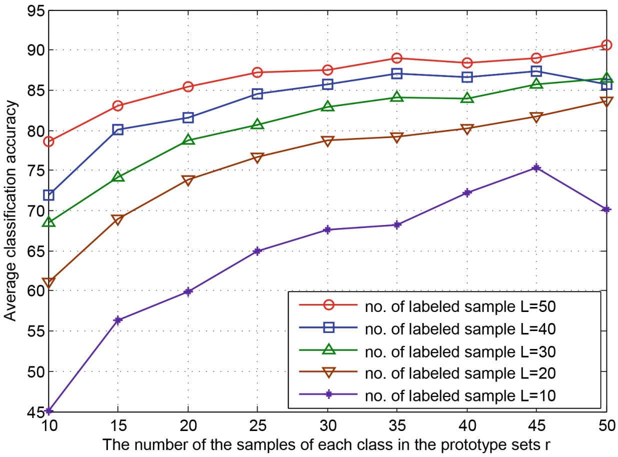

Fig. 5. Accuracy rate versus the number of the samples of each class in the prototype sets r.

The number of prototype sets T is also an important parameter in our method. Figure 4 illustrates the average classification accuracy as a function of T. It can be observed that the accuracy can be always maintained at a level close to 90 % as long as T is greater than 15, under the condition $L = 50$. And the number of labeled samples L may influence the stability of performance. When the parameter L drops down to 10, the stability decreases and the result presents a lot of volatility.

Besides the parameter T, we also try to investigate whether the number of the samples of each class in the prototype sets r has influences to the average accuracy rate. The result is shown in the Fig. 5.

5 Conclusions

In this paper, we proposed to utilize the ensemble projection vector as features for camera model identification with limited labeled samples and amount unlabeled samples. We carried out experiment to compare LBP algorithm and EP algorithm, and the result demonstrated that our proposed method EP has better performance when the labeled samples is limited. At the same time, the proposed method has robustness to parameter T and r, when the L surpass one value. In a future work, we will focus on improving the classification accuracy rate, and consider introduce new feature into our scheme.

Acknowledgments. This work is supported by the Open Fund of Artificial Intelligence Key Laboratory of Sichuan Province (Grant No. 2012RZJ01), and also the Fundamental Research Funds for the Central Universities (Grant No. DUT14RC(3)103).

References

1. Fridrich, J., Lukáš, J., Goljan, M.: Digital camera identifica-tion from sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **12**, 205–214 (2006)
2. Yu, B., Hu, Y., Jian, C.: Source camera identification using large components of sensor pattern noise. In: Proceedings of International Conference on Computer Science and its Applications, pp. 1–5 (2009)
3. Hu, Y., Li, C.-T., Jian, C.: Building fingerprints with infor-mation from three color bands for source camera identification. In: Proceedings of ACM Workshop on Multimedia in Forensics, Security and Intelligence, New York, pp. 111–116 (2010)
4. Li, C.-T.: Source camera identification using enhanced sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **5**(2), 280–287 (2010)
5. Hu, Y., Jian, C., Li, C.-T.: Using improved imaging sensor pattern noise for source camera identification. In: Proceedings of IEEE International Conference on Multimedia & Expo (ICME), Singapore, pp. 1481–1486 (2010)
6. Kang, X., Li, Y., Qu, Z.: Enhanced source camera identifi-cation performance with a camera reference phase sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **7**(2), 393–402 (2012)
7. Fridrich, J.: Sensor defects in digital image forensic. In: Sencar, H.T., Memon, N. (eds.) Digital Image Forensics, pp. 179–218. Springer, New York (2013)
8. Swaminathan, A., Wu, M., Liu, K.J.R.: Nonintrusive com-ponent forensics of visual sensors using output images. IEEE Trans. Inf. Forensics Secur. **2**(1), 91–105 (2007)
9. Celiktutan, O., Sankur, B., Avcibas, I.: Blind identification of source cell-phone model. IEEE Trans. Inf. Forensics Secur. **3**(3), 553–566 (2008)
10. Xu, G., Shi, Y.Q.: Camera model identification using local binary patterns. In: Proceedings of IEEE International Conference on Multimedia & Expo (ICME), Melbourne, Australia, pp. 392–397 (2012)

11. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
12. Roach, E., Lloyd, B.B.: Cognition and Categorization, Mul-timedia Systems. Hillsdale, New Jersey (1978)
13. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. England, London (2000)
14. Dai, D., Gool, L.V.: Ensemble projection for semi-surpervised image classification. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, pp. 2072–2079 (2013)
15. Gloe, T., Bohme, R.: The 'Dresden Image Database' for benchmarking digital image forensics. In: Proceedings of ACM Symposium on Applied Computing, Sierre, Switzerland, vol. 3(2–4) (2010)
16. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machine. ACM Trans. Intell. Syst. Technol. N.Y. **2**(3), 27:1–27:27 (2011). Software, [Online]. Avaliable: http://www.cise.edu.tw/ cjlin/libsvm