Single Image Camera Identification Using I-Vectors

Arash Rashidi
Faculty of Electrical and Computer Engineering
Islamic Azad University, Science and Research Branch,
Tehran, Iran
arasheri@gmail.com

Farbod Razzazi*
Faculty of Electrical and Computer Engineering
Islamic Azad University, Science and Research Branch,
Tehran, Iran
razzazi@srbiau.ac.ir

Abstract— **Recently, in the field of speech processing, I-Vector modeling has been appealed a great deal of interest. I-Vector has shown its benefits in modeling of intra and inter-domain variabilities to a single low dimension space for speaker identification tasks. This paper presents the usage of I-Vector in camera identification as a new approach in image forensics domain. In our approach, image texture is extracted from images as our features for the I-vector system. We have used 8 camera models in our work and the result shows 99.01% accuracy. We have also conducted attacks on the test images. We gained 99.01% accuracy for rotation attack and the average accuracy of 88.71% for three level brightness attack.**

Keywords— I-Vector, GMM, Forensics, JFA, Camera Identification

I. Introduction

In recent years, the use of digital media has become an everyday use in our lives. Digital cameras can be found almost in every digital device. Due to the existence of powerful image editing-software packages, it has become easy to manipulate the images. Hence the branch of image forensics is trying to track the changes and find the forgeries that are made in images [1]. Image forensics mainly considers two branches; source digital camera identification and Image forgery detection. Camera identification can be used in many areas that verification of an image is crucial like courts or security applications. Camera identification has valued great importance for detection of image origin in forensic applications and many researches have been conducted for the improvement of camera identification tools.

The physical specifications of each single camera can explicitly be said to be unique. Therefore, it is expected to be possible to track this unique characteristic to identify the camera [2].

In this study, a novel method has been proposed to statistically extract the intrinsic clues of camera identity based on i-vector extraction from texture of the image.

A. Previous Work

In recent years, according to the physical structures of digital cameras, many camera identification techniques have been introduced. Lens radial distortion, sensor imperfections, color filter array interpolation and inheriting image features are among these introduced methods.

Distortion in camera lens is a clue for finding unique structure of a camera. Choi et al believe that different lens producers have their own method in lens production. Therefore, the optical distortion for every single brand of camera will be almost unique [3]. By using this technique, an average accuracy of 91.53% was achieved.

Pixel defects and sensor pattern noise are two areas concerning sensor imperfection that is used for camera identification. Pixel defects is a common occurrence in the CCD of a camera, but fortunately, every single camera has its own pixel defect pattern. Geradts et al [4] have used this pattern for camera identification. The results for classification among three different camera models have shown an average accuracy of 91.39%.

Another approach for camera identification is the method of using sensor pattern noise. Since in the process of making sensors, imperfections in the sensor is inevitable. This causes the pixels to show diverse sensitivity to light which is the cause of sensor pattern noise. Lukas et al [5] have used this method for evaluating 9 different camera models. The researchers accomplish 100% accuracy in their own dataset.

Bayram et al [6] believe that CFA filter pattern and interpolation algorithms for camera are distinct from each other. Thus by using this factor, it is possible to classify the cameras. Bayram et al used the EM algorithm for feature extraction form images, for two different cameras and gained the accuracy of 95.71%. However, the result dropped to 83.33% when another camera was added to the dataset.

Other features for camera identification include color features, image quality matrix [7] and wavelet domain statistics [8]. Kharrazi et al [9] have employed image features to identify between two different cameras. The results claim an accuracy of 98.73% for uncompressed images. However, the result is dropped as the number of cameras increases.

A good feature in camera identification process is the texture information of an image. In this study, the texture is described by Local Binary Pattern (LBP) which is a common and widely used method for image texture description. LBP is a simple yet very efficient operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number [10]. Although LBP appears to perform very well for the stage of feature extraction, however unfortunately the number of extracted features is very high

978-1-5386-0804-3/17/$31.00 ©2017 IEEE

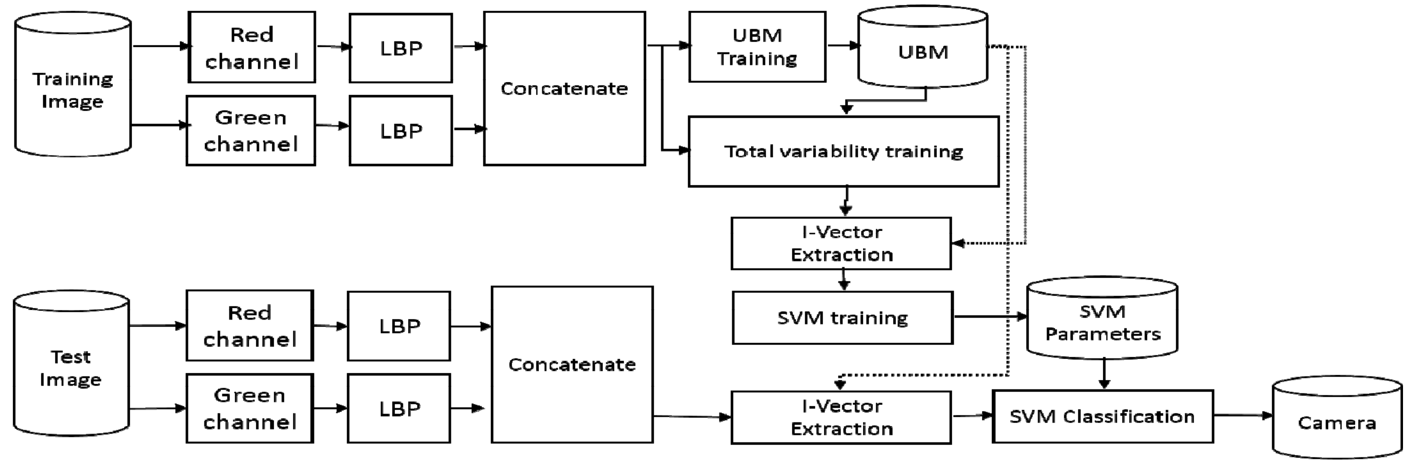

Fig. 1. Block diagram of proposed method

and not all of them is useful for the sake of camera identification. In fact, this feature represents both the camera and the image context simultaneously. Therefore an algorithm is needed to pick up the key features that are effective for camera identification.

The context information of an image may be regarded as session variability of a random process. Session variability has attracted great attention in the field of image processing specially in face authentication and object classification [11]. Session variability aims to model the variables that have a large impact on the specific classification task. In other words, session variability is anything that causes a mismatch between two images of the same group such as brightness, pose, angle, etc. in face authentication tasks. Two main methods can be used to tackle session variability, intersession variability (ISV) and joint factor analysis (JFA) [12]. In this paper, we used JFA for modeling the session variability in camera Identification application.

B. *Motivation of the study and paper structure*

Recently, in the domain of speaker verification, a method based on JFA theory, named as I-Vector, has been proposed. It introduces a new method for speaker authentication and verification which is more efficient than standard JFA. I-vector was proposed by Patrick Kenny [13] and has been given a lot of attention in many speech related applications. To our best knowledge, I-vector has never been used or adapted for neither camera identification, nor any other image classification problem. In contrast, in this study, we have presented camera identification using I-vector secondary features.

The outline of this paper is as follows. In section two, we present our proposed identification based on JFA and I-vector. Section three is devoted to present the experiments and results. Finally, in section four, we conclude the study and talk about the future works on using I-vector in this field.

II. PROPOSED METHOD

The flow of the proposed method used in our method is shown in Fig.1. As it is demonstrated, the red and green channels of training images are extracted simultaneously. The features from each channel are extracted by using LBP. The extracted features are concatenated as the single image primary feature vector and fed into the GMM training block for computing the UBM vectors. Following the UBM block, the total variability space is computed which is needed for the I-vector, the theory of total variability space and I-Vector will be explained in the section to come. From the total variability space the I-Vectors are gained, in the last block the I-Vectors from the training are fed into the SVM training block with the pre-set SVM parameters. The SVM classifier takes the I-Vectors from training and test data and classifies the image according to the camera model.

There is a rich literature in using GMM-UBM in speech processing applications. The structure of a GMM-UBM based speaker verification System can be shown in fig.2

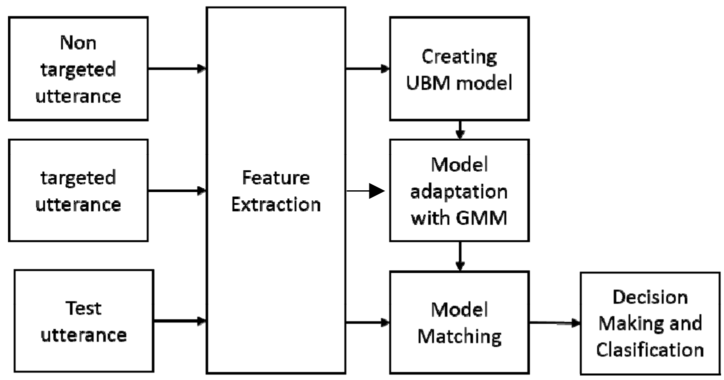

Fig. 2. GMM system block diagram

In the GMM system, a GMM model is trained using non targeted model known as Universal Background Model [14], next, the GMM model is adapted according to the features extracted from the targeted utterances. In the test phase, the extracted features from the test utterances are matched according to the adapted GMM model and in the decision making block, the test utterance is classified. The JFA model uses the GMM idea to extract new secondary features.

A. *Joint Factor Analysis*

JFA has recently become the center of attention in speaker verification [15], according to JFA, every single speaker is represented by a Supervector (M) [16]. The Supervector is defined as:

$$M = m + Vy + Ux + Dz \qquad (1)$$

Where m is the speaker-independent component, V is the speaker dependent component and U is the channel dependent component and D is the residual matrix. The vectors y, x and z are the speaker and session-dependent vectors in their respective subspaces and each is assumed to be a random variable with a normal distribution $N(0, I)$.

In our problem, camera plays the same role as the speaker and the content of the image is similar to the utterance content. Therefore, it is expected that this theory performs well for camera identification too.

In order to apply JFA to camera recognition system, at the beginning V, D, U have to be properly estimated from appropriate labeled data. Afterward, the camera and session factors (i.e. x, y, z) are estimated for a given camera. The scoring and verification is carried out by computing the likelihood ratio of the test features in UBM and GMM models.

B. I-vector Extraction

Our method uses i-vector secondary features for classification. The JFA system discussed in the previous part, evaluates the vector in two distinct subspaces: The camera sub-space which is defined by the matrix V, The channel subspace which is defined by the matrix U. In JFA system it is assumed that these two subspaces are totally independent of each other and when estimating one matrix, the other is assumed to be determined. Dehak et al proposed a total variability space in which only one single subspace is concerned. In other words, they stated that the channel and camera are not independent and the statistical behavior of the channel subspace leaks to speaker subspace and vice versa. This new space, contains both camera and channel variabilities simultaneously. The new camera and channel-dependent GMM super-vector, defined by (1), is rewritten as

$$M = m + Tw; \qquad (2)$$

where m is the camera-and channel-independent super-vector, T is a low rank rectangular matrix and w is a random vector having a standard normal distribution N(0,I). These new vectors are referred as identity vectors or I-Vectors. The application of training the total variability T is the same as training V.

The extracted secondary features are injected to a multi-class rbf-kernel SVM to classify the cameras.

III. Experiments and Results

For the evaluation of the proposed method, we used eight different cameras from Dresden image dataset. In this dataset, more than 14000 images from 73 different cameras have been gathered. These cameras consists of 25 different models [17]. We chose 8 different models of cameras. The allocation of images for training and test is presented in table 1.

Table. I Distribution of images for training and

Camera Model	Train image Quantity	Test image Quantity	Total
Agfa_DC-504	70	31	101
Canon_Ixus55	51	24	75
Casio_EX-Z150	31	12	43
FujiFilm_FinePixJ50	63	41	104
Nikon_CoolPixS710	62	26	88
Pentax_OptioW60	29	12	41
Ricoh_GX100	28	13	41
Samsung_L74wide	96	42	138
Total	403	201	631

At the beginning, we present the classification Result. Next we applied some attacks upon the test images in order to evaluate the robustness of the algorithm against session variabilities. For the attacks, we conducted brightness and rotation attacks upon the test images.

A. Image Classification

The first test concerns the classification of test raw images described in table1.

In total variability training step, we can adjust the dimension of the I-Vectors versus total variability dimension values. The result of classification according to the number of total variability (tv-dim) is presented in figure 3.

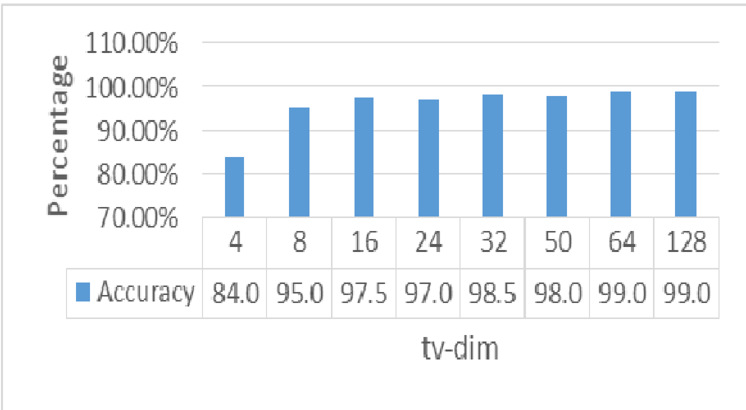

Figure 3 Effect of tv-dim on accuracy

We have gained an average accuracy of 98.5% for our classification. According to figure 3 as the number of tv-dim increases, the accuracy rate saturates to 99.1%. It is apparent

978-1-5386-0804-3/17/$31.00 ©2017 IEEE

Table. II confusion matrix for the average accuracy of 98.5%

Recognized	Agfa_DC-504	Canon_Ixus55	Casio_EX-Z150	FujiFilm_FinePixJ50	Nikon_CoolPixS710	Pentax_OptioW60	Ricoh_GX100	Samsung_L74wide
Agfa_DC-504	100%	0%	0%	0%	0%	0%	0%	0%
Canon_Ixus55	0%	95.84%	0%	0%	0%	0%	0%	4.16%
Casio_EX-Z150	0%	0%	100%	0%	0%	0%	0%	0%
FujiFilm_FinePixJ50	0%	0%	0%	100%	0%	0%	0%	0%
Nikon_CoolPixS710	0%	0%	0%	0%	100%	0%	0%	0%
Pentax_OptioW60	0%	0%	0%	0%	0%	100%	0%	0%
Ricoh_GX100	0%	0%	0%	0%	0%	15.38%	84.62%	0%
Samsung_L74wide	0%	0%	0%	0%	0%	0%	0%	100%

that the value of tv-dim plays an important role in the accuracy achieved which shows the important role of I-Vector in our Work for image classification. The confusion matrix for the average accuracy of 98.5% is shown in table 2.

To have a better understanding of the efficiency of the proposed method the result is compared with the method proposed by [18]. Similarly, we have used the same dataset as in [18] so it makes the comparison more realistic. The confusion matrix for 8 cameras of the proposed method by [18] is shown in table 3. By comparing the two confusion matrix of table 2 and figure 3 it can be seen that for two camera models of Fujifilm and Casio we have gained 100% accuracy whereas in table 3 it is evident that the latter authors had 43 and eight percent error for the mentioned camera models. Also for our method we have 4.16 and 15.38 percent of error for Canon and Ricoh cameras however the mentioned authors have gained 100% accuracy for the latter models. In general, it seems that for our method the errors show low fluctuations and in General our average accuracy is higher.

B. *Image Attacks*

To evaluate the robustness of our method against environmental and capture variabilities; we have conducted two test on the test data, Brightness test and Rotation test.

Brightness test: The Brightness of all test images were edited and we evaluated three different Brightness level for our test. In order the Brightness of images were increased with the values 10%, 30% and 50%. Next, the edited images were fed into the algorithm as the test images and the classification rate was assessed. The result is presented Fig 4. For three Brightneess levels of 10,30,50 percent we have gained an average of 88.83 % accuracy which shows our algorithm is very well resistant to Brightness changes of images , razzazi et al have also tested brightness changes and its effect on the images but the Brightness factor is not mentioned but in comparison, two algorithms show good resistance to brightness changes. When the brighness factor of image is effected some features When the brighness factor of image is effected some features which represent the charachteristics of camera are lost therefore as the brightmess Factor increases accuracy drops Bust still our method shows good resistance since the brightness factor drops by 50 % whereas the accurace drops only 21.5% which is the evidence of resistance to brightness variation.

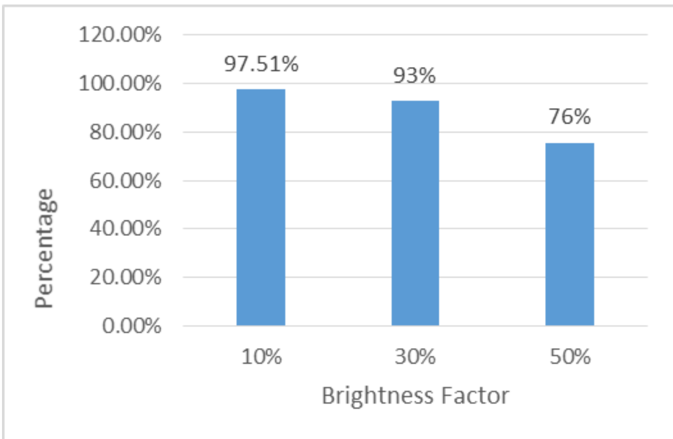

Figure 4 Effect brightness on accuracy

978-1-5386-0804-3/17/$31.00 ©2017 IEEE

Rotation test: In the next test, All of test images, were randomly rotated either ninety degrees clock wise and anti-clock wise and one hundred and eighty degrees. The aim of this test was to evaluate whether rotation has an effect upon the classification rate. For this test, we achieved 98.5075% accuracy ; this Results shows that our method is fully resistant to rotation of images.This result shows that bnoth I-Vector and LBP features provide the features that can modle the rotation variance therefore show good resistance for the rotation.

Table III. Accuracy Result for the previous method proposed by Razzazi et al

Camera Model	Accuracy of Proposed method by Razzazi et al
Agfa_DC-504	100%
Canon_Ixus55	100%
Casio_EX-Z150	91%
FujiFilm_FinePixJ50	56%
Nikon_CoolPixS710	100%
Pentax_OptioW60	100%
Ricoh_GX100	100%
Samsung_L74wide	100%

IV. Conclusion

In this paper we proposed a new method for camera identification using I-Vector, this is the first time that this method is used for image processing, and this method has demonstrated a new method which can be adapted to the applications where the images are altered. Our method proved to have high accuracy and it even showed good resistance against image brightness attack and complete resistance to rotation attack. This study will be continued on other statistical attacks to the test images.

A lot of work has been performed in the field of image processing but in this paper, it is the first time that I-Vector has been used for image processing applications. There are still many work that needs to be done for the employment of I-Vector in image processing. In our future work, we plan to conduct crop attack upon the test images. In addition, we would like to test the JPEG compression effect on classification rate. We will also use this feature for other image forensic applications.

Reference

[1] Birajdar, G.K. and Mankar, V.H., 2013. Digital image forgery detection using passive techniques: A survey. *Digital Investigation*, *10*(3), pp.226-245.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] Van Lanh, T., Chong, K.S., Emmanuel, S. and Kankanhalli, M.S., 2007, July. A survey on digital camera image forensic methods. In *Multimedia and Expo, 2007 IEEE International Conference on* (pp. 16-19).

[3] San Choi, K., Lam, E.Y. and Wong, K.K., 2006, February. Source camera identification using footprints from lens aberration. In *Digital Photography* (p. 60690J).

[4] Geradts, Z.J., Bijhold, J., Kieft, M., Kurosawa, K., Kuroki, K. and Saitoh, N., 2001. Methods for identification of images acquired with digital cameras. *Proc. of SPIE, Enabling Technologies for Law Enforcement and Security*, *4232*, pp.505-512.

[5] Lukas, J., Fridrich, J. and Goljan, M., 2006. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, *1*(2), pp.205-214.

[6] Bayram, S., Sencar, H., Memon, N. and Avcibas, I., 2005, September. Source camera identification based on CFA interpolation. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on* (Vol. 3, pp. III-69).

[7] Birajdar, G.K. and Mankar, V.H., 2013. Digital image forgery detection using passive techniques: A survey. *Digital Investigation*, *10*(3), pp.226-245.

[8] Orozco, A.S., González, D.A., Corripio, J.R., Villalba, L.G. and Hernandez-Castro, J.C., 2014. Source identification for mobile devices, based on wavelet transforms combined with sensor imperfections. *Computing*, *96*(9), pp.829-841.

[9] Kharrazi, M., Sencar, H.T. and Memon, N.D., 2004, October. Blind source camera identification. In *ICIP* (pp. 709-712).

[10] Xu, G. and Shi, Y.Q., 2012, July. Camera model identification using local binary patterns. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on* (pp. 392-397).

[11] Anantharajah, K., Ge, Z., McCool, C., Denman, S., Fookes, C., Corke, P., Tjondronegoro, D. and Sridharan, S., 2014, March. Local inter-session variability modelling for object classification. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on* (pp. 309-316).

[12] McCool, C., Wallace, R., McLaren, M., El Shafey, L. and Marcel, S., 2013. Session variability modelling for face authentication. *IET biometrics*, *2*(3), pp.117-129.

[13] Kenny, P., Ouellet, P., Dehak, N., Gupta, V. and Dumouchel, P., 2008. A study of interspeaker variability in speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(5), pp.980-988.

[14] Sizov, A., Khoury, E., Kinnunen, T., Wu, Z. and Marcel, S., 2015. Joint Speaker Verification and Antispoofing in the $ i $-Vector Space. *IEEE Transactions on Information Forensics and Security*, *10*(4), pp.821-832.

[15] Kenny, P., Ouellet, P., Dehak, N., Gupta, V. and Dumouchel, P., 2008. A study of interspeaker variability in speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(5), pp.980-988.

[16] Verma, P. and Das, P.K., 2015. i-Vectors in speech processing applications: a survey. *International Journal of Speech Technology*, *18*(4), pp.529-546.

[17] Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P. and Ouellet, P., 2011. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(4), pp.788-798.

[18] Gloe, T. and Böhme, R., 2010. The dresden image database for benchmarking digital image forensics. *Journal of Digital Forensic Practice*, *3*(2-4), pp.150-159.

[19] Razzazi, F. and Seyedabadi, A., 2014, December. A robust feature for single image camera identification using local binary patterns. in *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* pp. 000462-000467.

978-1-5386-0804-3/17/$31.00 ©2017 IEEE