Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

DIGITAL IMAGE FORENSICS USING STATISTICAL FEATURES AND NEURAL NETWORK CLASSIFIER

WEI LU[1], WEI SUN[2], JI-WU HUANG[1], HONG-TAO LU[3]

[1]School of Information Science and Technology and Guangdong Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou 510275, China
[2]School of Software and Guangdong Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou 510275, China
[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
E-MAIL: {luwei3,sunwei,isshjw}@mail.sysu.edu.cn, lu-ht@cs.sjtu.edu.cn

Abstract:

Digital image forensics is a new topic in recent years, which deals with the authenticity and credibility of digital images. How to recognize fake images is still a problem. This paper presents a fake image classification scheme using higher order image statistics and RBF neural networks. The features constructed on the higher order statistics reveal the intrinsic statistical features between fake images and real images. Then a classifier based on RBF neural networks is used to classify the fake and real images using these features. Experimental results demonstrated the effectiveness of the proposed scheme.

Keywords:

Digital Image Forensics; Higher Order Autocorrelation Statistics; RBF Neural Network

1. Introduction

Nowadays, digital images are almost parts of our daily lives. However, the truth of digital images becomes a serious ethical issue as they can be altered everywhere at any moment using advanced digitization and image processing techniques. Human has puzzled that which images are believable. For example, a fake photo will be a perjury in court, which will badly influence the justice. Image fakery was developed early last century [4]. Although early fake images are created using darkroom techniques, it is obvious that almost all of the fake images are created using computer graphic software today, i.e., digital fakery is in majority. Fig. 1 shows the examples of real images and fake images. Generally, the content of fake images is inconsistent with the real world, which, however, usually leaves no visual clews of artificial manipulations. Computable methods should be developed to recognize this class of digital images, i.e. digital image forensics [1], which is a main branch of information forensics.

Some methods have been devoted to detect digital fakery [2,5,7,8,10]. The composition of two or more people into a single image is a common form of manipulation. In [6], the authors described how such composites can be detected by estimating a cameras intrinsic parameters from the image of a persons eyes. Differences in these parameters across the image are used as evidence of tampering. In [5], a fully automatic spliced image detection method is proposed based on consistency checking of camera characteristics among different areas in an image, which use camera response function to estimate from each area using geometric invariants from locally planar irradiance points. In [3,11], the higher order statistics in DWT domain were used to develop a fake and real image classification scheme. In [13], a blind detection scheme of photomontage was proposed, which used a higher order statistics, i.e., biocoherence, to introduce the fake characteristics, which were designed to detect human speech signals originally. As is well known the statistical distribution of audio signals is very different from that of digital images.

(a) (b)

Figure 1. (a) A real image. (b) A fake image.

978-1-4244-2096-4/08/$25.00 ©2008 IEEE

In this paper, we present a detection scheme to reveal digital fakery based on multiresolution decomposition and higher order local autocorrelations. Then, we employ Neural Network based classifier to classify real images and fake images. The rest parts of this paper are organized as follows. Section 2 gives the construction of the higher order image statistical features. Section 3 describes the RBF neural network classifier. Section 4 shows some experimental results and discussion on the proposed scheme. Finally, conclusions are given in section 5.

(a) (b)

Figure 2. An illustration of 3-level DWT using symlets wavelet, a fake version of Lenna image (a) and its scaled images based on DWT (b).

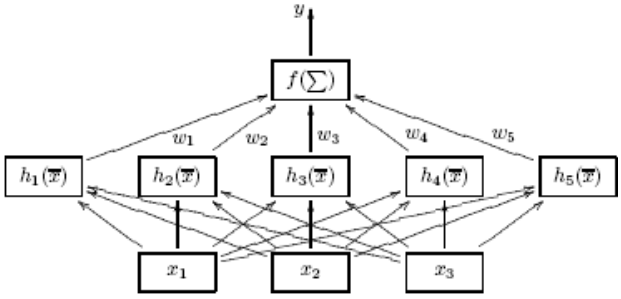

Figure 3. The 3-5-1 structure of RBF neural network.

2. Higher Order Statistical Features

Autocorrelation is a powerful shift-invariant signal statistic, which has been extended to higher orders in [12], where the n-th order autocorrelation function with n displacements $\tau_1, \tau_2 \cdots \tau_n$ is defined by:

$$r_x^{(n)}(\tau) = \int_\Omega f(x) f(x+\tau_1) \cdots f(x+\tau_n) dx \qquad (1)$$

where $\tau = \{\tau_1, \tau_2 \cdots \tau_n\}$, n denotes the order of the autocorrelation function, Ω is the two-dimensional image coordinate plane, and $x \in \Omega$ is the image coordinate vector. In order to decrease the computation complexity, some restrictions are given in [9], where the order n is restricted up to 2, i.e., n=1,2, and the range of the displacements are restricted within a local 3x3 window, the center of which is the reference point. Then based on the shift-invariance, the number of the patterns of the displacements is reduced to 25, which forms the feature vector using higher order autocorrelation statistics.

Two-dimensional discrete wavelet transformation (2DDWT) is one of the most useful image analysis tools in many applications, e.g., image compression, image coding, etc., which decomposes an image into 3 components. Given an image I, DWT splits the frequency into 3 scales and orientations, i.e., $V_i(x)$, $H_i(x)$ and $D_i(x)$, where i = 1,2,…l denotes the decomposition scale. Fig. 2 shows the 3 level of 2D DWTs using symlets wavelet with kernel filter length 10 for an fake lenna image. Generally speaking, there is a local energy in DWT for salient image features, such as edge, corner, etc., which can be measured by a large coefficient in the sequential decompositions. The detail coefficients contain the classifiable statistical features between fake image and real images.

Generally there is a boundary between the fake area and the real area for a fake image. Here, post processing can be done whose purpose is to decrease the effect of the boundary. These characters can be described by different coefficient distributions in the corresponding DWT subbands. Furthermore, without loss of generality, consider the vertical band $V_i(x)$, if there is a large coefficient $V_i(x)$ at scale i, it is more likely that it is also large for $V_{i+1}(x/2)$ at scale i+1. In order to extract the primary characters, we give a higher order statistic similar with Eq. (1), which is defined by:

$$v_i^{(n)}(\tau) = \int_{P_i^v} V(x) V(x+\tau_1) \cdots V(x+\tau_{n-1}) dx \qquad (2)$$

Where $\tau = \{\tau_1, \tau_2 \cdots \tau_{n-1}\}$ are the displacements with a given neighborhood, and P_i^v is the i-th level vertical subband coefficient plane. Here, we restrict the displacements within a 3x3 window, the central point of which is the reference point, the order of n-1 is restricted up to 2, i.e., n = 1; 2; 3. Thus, the number of the patterns of the displacements is reduced to 25. By computing the sums over P_i^v, we obtain 25 features. The process can be repeated for the subbands Hi(x) and Di(x) as follows:

$$h_i^{(n)}(\tau) = \int_{P_i^h} H(x) H(x+\tau_1) \cdots H(x+\tau_{n-1}) dx \qquad (3)$$

$$d_i^{(n)}(\tau) = \int_{P_i^d} D(x)D(x+\tau_1)\cdots D(x+\tau_{n-1})dx \qquad (4)$$

Thus, we can obtain 75xl features totally. In the cases of color images, the whole process can be repeated in red, green and blue channel separately, where 225xl features are obtained.

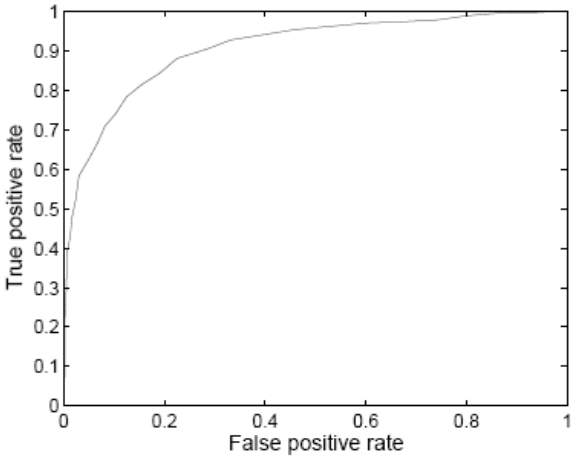

Figure 4. The ROC curves between the false positive rate and the true positive rate.

3. Classification

The extracted features are not proper for the classification and detection, although they contain many information for recognized objects. Here, we use a 3-layer RBF neural network as the classifier to detect the fake and real images, as it recombines the features using trained nonlinear mapping before classification. Fig. 3 shows an illustration of a 3-5-1 structure of RBF neural networks, where there are 3 feature inputs, 5 hidden neurons, and 1 output. We use the extracted features in section 2 as the input features as follows:

$$\begin{aligned} F = \; & v_1^1, v_1^2, v_1^3, v_2^1, v_2^2, v_2^3, \cdots, v_l^1, v_l^2, v_l^3, \\ & h_1^1, h_1^2, h_1^3, h_2^1, h_2^2, h_2^3, \cdots, h_l^1, h_l^2, h_l^3, \\ & d_1^1, d_1^2, d_1^3, d_2^1, d_2^2, d_2^3, \cdots, d_l^1, d_l^2, d_l^3 \end{aligned} \qquad (5)$$

To the training stage, if the tested image is fake, the output is set to -1, and if the tested image is real, the output is set to 1. Then, to the classification, if the output of the neural network classifier is less than 0, then the input image is labeled as a fake one, otherwise a real one.

4. Simulations and Discussions

In our experiments, a image database of 1000 fake images and 1000 real images is used to train and test the proposed scheme, where half of these images are used to train the RBF neural network classifier, and the others for test. Two examples taken from the database are shown in Fig. 1. The decomposition parameter l is set to 3. Table 1 shows the classification accuracy using the proposed classifier. Under the false negative rate 3.3%, about 50.2% real images are classified correctly, where false negative refers that a fake image is classified as a real image. Note that the testing accuracy is close to the training accuracy, which shows that the proposed classifier is general.

Table 1. Classification accuracy (percent) usingRBF neural network classifier.

database	training	testing
real images	57.3	50.2
fake images	97.4	96.7

Table 2. Classification accuracy (percent) for the schemes proposed in [11] using RBF neural network classifier.

database	training	testing
real images	53.1	47.5
fake images	95.1	93.3

(a) (b)

In order to further evaluate the proposed scheme, it is compared with the previous fake image detection scheme proposed in [11]. Table 2 shows the classification accuracies for the proposed scheme in [11] using the same test condition with that in Table 1. Note that the testing accuracies of the classifier in [11] are lower than our classifier in Table 1. This indicates that the performance of our scheme is better than that of the scheme in [11].

In Fig. 4, the ROC curve between the false positive rate and the true positive rate for the RBF neural network classifier is shown. Again, the false positive rate is the percentage of real images that are incorrectly detected as fake images, and the true positive rate is the percentage of real images that are correctly classified as real images. It can be concluded that the performances is good for the proposed classification scheme based on higher order image statistics and neural network classification.

To further evaluate the performance of the proposed

scheme, some image processing methods, including JPEG compression and low-pass filtering, were applied to the images in the database. Fig. 5 shows the detection results for the images which have been processed with random JPEG compression of quality 80 or 60 and Gaussian zero-mean low-pass filtering with size 5x5 and σ =0.5. We can see that the classification accuracy decreases less. Under the false negative rate 3.5%, about 45.7% real images are classified correctly. It can be concluded that the proposed classification method is robust to variations caused by common signal processing techniques.

We also assigned the training images with random outputs of 1 and -1, where half of the images are randomly assigned to the fake images and the others are the real images. Then we trained the RBF neural network classifier using the data set and then tested it. We found that the classification accuracy is 35.4% for the real images with the false negative rate 6.3%, which is badly worse than the case when the correct outputs are assigned. This indicates that the constructed features and RBF neural network classifier are based on rational higher order statistics for real and fake images.

5. Conclusions

Digital image forensics is one of hot research topic in information security in recent years. In this paper, we have proposed a digital fake image classification scheme using higher order statistics in DWT domain, which uses RBF neural networks to classify the fake and real images. Simulation results demonstrate that the proposed detection scheme is effective. Future works are to further the application of the proposed scheme.

Acknowledgements

This work is supported by the Scientific Research Foundation for the Young Teachers in Sun Yat-sen University, NSFC under project no. 60573033, and Program for New Century Excellent Talents in University (no. NCET-05-0397).

References

[1] H. Farid. Digital doctoring: How to tell the real from the fake. Significance, 3(4):162–166, 2006.

[2] H. Farid. Exposing digital forgeries in scientific images. In ACM Multimedia and Security Workshop, pages 29–36, Geneva, Switzerland, 2006.

[3] H. Farid and S. Lyu. Higher-order wavelet statistics and their application to digital forensics. In IEEE Workshop on Statistical Analysis in Computer Vision (in conjunction with CVPR), 2003.

[4] R. D. Fiete. Photo fakery. http://oemagazine.com/fromTheMagazine/jan05/photofakery.html

[5] Y.-F. Hsu and S.-F. Chang. Image splicing detection using camera response function consistency and automatic segmentation. In International Conference on Multimedia and Expo (ICME), pages 28–31, Beijing, China, 2007.

[6] M. Johnson and H. Farid. Detecting photographic composites of people. In 6th International Workshop on Digital Watermarking, Guangzhou, China, 2007.

[7] M. Johnson and H. Farid. Exposing digital forgeries in complex lighting environments. IEEE Transactions on Information Forensics and Security, 3(2):450–461, 2007.

[8] M. Johnson and H. Farid. Exposing digital forgeries through specular highlights on the eye. In 9th International Workshop on Information Hiding, Saint Malo, France, 2007.

[9] T. Kurita, N. Otsu, and T. Sato. A face recognition method using higher order local autocorrelation and multivariate analysis. In 11th IAPR Int. Conf. on Pattern Recognition, volume 2, pages 213–216, 1992.

[10] W. Lu, F.-L. Chung, and H. Lu. Blind fake image detection scheme using SVD. IEICE Trans. Communications, E89-B(5):1726–1728, May 2006.

[11] S. Lyu and H. Farid. How realistic is photorealistic. IEEE Trans. Signal Processing, 53(2):845–850, Feb. 2005.

[12] J. A. Mclaughlin and J. Raviv. Nth-order autocorrelations in pattern recognition. Information and Control, 12:121–142, 1968.

[13] T.-T. Ng, S.-F. Chang, and Q. Sun. Blind detection of photomontage using higher order statistics. In IEEE Int. Symposium on Circuits and Systems (ISCAS), volume 5, pages 688–691, May 2004.