Detection of Image Compositing Based on a Statistical Model for Natural Images

SUN Shao-Jie[1] WU Qiong[1] LI Guo-Hui[1]

Abstract Nowadays, digital images can be easily tampered due to the availability of powerful image processing software. As digital cameras continue to replace their analog counterparts, the importance of authenticating digital images, identifying their sources, and detecting forgeries is increasing. Blind image forensics is used to analyze an image in the complete absence of any digital watermark or signature. Image compositing is the most common form of digital tampering. Assuming that image compositing operations affect the inherent statistics of the image, we propose an image compositing detection method on based on a statistical model for natural image in the wavelet transform domain. The generalized Gaussian model (GGD) is employed to describe the marginal distribution of wavelet coefficients of images, and the parameters of GGD are obtained using maximum-likelihood estimator. The statistical features include GGD parameters, prediction error, mean, variance, skewness, and kurtosis at each wavelet detail subband. Then, these feature vectors are used to discriminate between natural images and composite images using support vector machine (SVM). To evaluate the performance of our proposed method, we carried out tests on the Columbia Uncompressed Image Splicing Detection Dataset and another advanced dataset, and achieved a detection accuracy of 92 % and 79 %, respectively. The detection performance of our method is better than that of the method using camera response function on the same dataset.

Key words Image compositing, generalized Gaussian model (GGD), maximum-likelihood (ML), support vector machine (SVM), image forensics

Today, digital images have been used in a growing number of applications from news reporting and law evidences to forensic investigation and consumer photography. Due to the widespread popularity of digital images and availability of powerful image processing tools, it is important to authenticate digital images, identify their sources, and detect forgeries.

Blind image forensics research aims to solve these problems. Blind image forensics is a form of image analysis for finding out the condition of an image without relying on pre-extracted or pre-embedded information[1]. During the past few years, many blind forensics techniques for digital image authenticity have been developed. These approaches can be roughly divided into three categories:

1) Approaches based on the artifacts left by the process of image forgery: a forged image usually undergoes some common image processing operations. As a result, digital forgeries can be detected by tracing artifacts introduced by image operations. Different methods have been proposed to detect copy-move forgeries[2], double JPEG compression[3], and image lighting inconsistency[4].

2) Approaches based on the consistency of imaging equipment: natural images are usually obtained through data acquisition devices, which introduce uniform characteristics to the entire image, and then the variation in the local characteristics across the image can be used to detect tampering. Color filter array interpolation[3], consistency of camera response function[5], and sensor pattern noise[6] have been used to detect image forgeries.

3) Approaches based on the statistical characteristics of natural images: although good forgeries leave no visual artifacts, they may affect the underlying statistics of images more or less. In [7−8], several statistical approaches are proposed to detect image splicing.

In this paper, we aim at image compositing forgeries, and propose a detection method based on a statistical model for natural image in the wavelet transform domain.

Compositing is perhaps the most common operation when tampering with an image. It is a process of cropping and pasting regions from other images to a new image with necessary post-processing, such as blurring and scaling. Since compositing is often used for image manipulation, an efficient method is needed to detect image compositing. In this paper, we describe a statistical model for natural images that is built upon multiscale wavelet decomposition. The model consists of parameters estimated from generalized Gaussian model (GGD) and cumulant of features at each wavelet detail subband. Experimental results show that the proposed method can distinguish authentic images from composite images with a high accuracy, even when the images have undergone post-processing.

The rest of paper is organized as follows. The proposed natural image statistical model is described in Section 1. Section 2 describes the detection approach including feature extraction and classification. Section 3 presents and discusses the experimental results, and Section 4 contains conclusions.

1 Natural image statistical model in the wavelet transform domain

Due to the capability of multiscale and multiresolution analysis, discrete wavelet transform (DWT) has been used widely, as a powerful signal-processing tool. On the other hand, the statistic models for natural images have played an increasingly important role in the field of image processing and analysis. In the research of natural image statistics, it is found that the marginal distribution of the coefficients in individual detail wavelet subband can be well fitted with a generalized Gaussian density[9−10]. However, the image compositing operation changes the statistics of wavelet coefficients, leading to a change of their statistical distribution.

1.1 Generalized Gaussian density modeling of wavelet coefficients

Let $x = (x_1, \cdots, x_N)$ denote a set of N wavelet coefficients at a particular subband, and the marginal distribution of x can be well-fitted with two-parameter generalized Gaussian density model[9−11], which is defined as:

$$p_m(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\mathrm{e}^{-(|x|/\alpha)^\beta} \tag{1}$$

where $\Gamma(\cdot)$ is the Gamma function, i.e., $\Gamma(z) = \int_0^\infty \mathrm{e}^{-t}t^{z-1}\mathrm{d}t$, $z > 0$, α is the scale parameter, which determines the width of the model peak (standard deviation), and β is the shape parameter, which determines the decreasing rate of the peak. The Laplacian density and Gaussian density are the special cases of GGD model, when $\beta = 1$ and $\beta = 2$, respectively.

In this paper, the model parameters of GGD are estimated using maximum-likelihood (ML) estimator. We

Received May 21, 2008; in revised form December 11, 2008
1. College of Information System and Management, National University of Defense Technology, Changsha 410073, P. R. China
DOI: 10.3724/SP.J.1004.2009.01564

define the likelihood function of the wavelet coefficients $x = (x_1, \cdots, x_N)$ as:

$$L(x;\alpha,\beta) = \ln \prod_{i=1}^{N} p_m(x_i;\alpha,\beta) \tag{2}$$

According to the parameter estimation theory, α and β are the solutions of (3) and (4):

$$\frac{\partial L(x;\alpha,\beta)}{\partial \alpha} = -\frac{N}{\alpha} + \sum_{i=1}^{N} \frac{\beta\,|x_i|^{\beta}\,\alpha^{-\beta}}{\alpha} = 0 \tag{3}$$

$$\frac{\partial L(x;\alpha,\beta)}{\partial \beta} = \frac{N}{\beta} + \frac{N\Psi(1/\beta)}{\beta^2} - \sum_{i=1}^{N}\left(\frac{|x_i|}{\alpha}\right)^{\beta} \ln\left(\frac{|x_i|}{\alpha}\right) = 0 \tag{4}$$

where $\Psi(\cdot)$ is the digamma function, i.e., $\Psi(z) = \Gamma'(z)/\Gamma(z)$.

Fix $\beta > 0$, then (3) has a unique, real, and positive solution as

$$\hat{\alpha} = \left(\frac{\beta}{N}\sum_{i=1}^{N}|x_i|^{\beta}\right)^{1/\beta} \tag{5}$$

Substitute (5) into (4), then the shape parameter β is the solution of the following equation:

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum\limits_{i=1}^{N}|x_i|^{\hat{\beta}}\log|x_i|}{\sum\limits_{i=1}^{N}|x_i|^{\hat{\beta}}} + \frac{\ln\left(\frac{\hat{\beta}}{N}\sum\limits_{i=1}^{N}|x_i|^{\hat{\beta}}\right)}{\hat{\beta}} = 0 \tag{6}$$

The Newton-Raphson iterative procedure[12] is adopted to solve (6) and to get ML estimate $\hat{\beta}$ of the shape parameter β. Substituting $\hat{\beta}$ into (5), we can get ML estimate $\hat{\alpha}$ of the scale parameter α.

Fig. 1 shows a typical example of histograms of three-scale wavelet subband coefficients together with plots of the fitted GGD models. Fig. 1 shows that the histograms of wavelet coefficients of the natural image can be well fitted with GGD models and characterized by a sharp peak at zero and long symmetric tails. Therefore, we only need two model parameters to represent the marginal distribution of wavelet coefficients in a subband efficiently.

In addition to the two model parameters, we also include the prediction error as a third feature. Let us assume that $p(x)$ is the true probability density distribution of the wavelet subband coefficients $x = (x_1, \cdots, x_N)$, and the prediction error $d(p_m \| p)$ is defined as the Kullback-Leibler distance (KLD) between $p_m(x;\alpha,\beta)$ and $p(x)$:

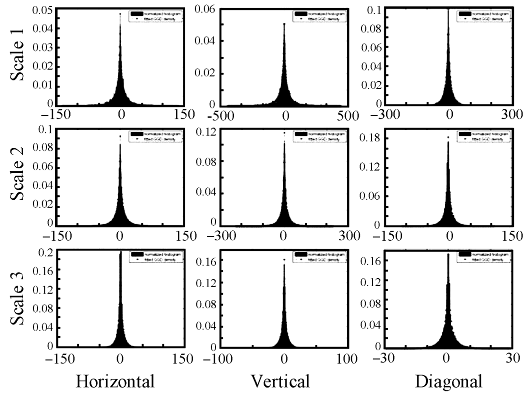
$$d(p_m \| p) = \int p_m(x;\alpha,\beta) \ln \frac{p_m(x;\alpha,\beta)}{p(x)} \mathrm{d}x \tag{7}$$

Practically, this quantity can be evaluated numerically using histograms:

$$d(p_m \| p) = \sum_{i=1}^{K} h_m(i) \ln \frac{h_m(i)}{h(i)} \tag{8}$$

where $h_m(i)$ and $h(i)$ are the normalized heights of the i-th histogram bins, and K is the number of bins in the histograms.

(a) A natural image



(b) The histograms of the coefficients of its three-scale DWT decomposition fitted with GGD models

Fig. 1 Histograms of wavelet subband coefficients together with plots of fitted GGD models

1.2 Cumulant-based features

In order to capture the characteristics and correlations of local image energy at different scales and different orientations, we also extract low-order and high-order statistical features from wavelet detail subbands, which are referred to as cumulant-based features. Given image DWT decomposition, the cumulant-features are composed of the mean, variance, skewness, and kurtosis of the subband coefficients at each orientation and at each scale. Assume that $x = (x_1, \cdots, x_N)$ denotes a set of N wavelet coefficients at a particular subband, we use μ and σ^2 to represent the mean and variance of x. We define:

$$\mu = \mathrm{E}(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{9}$$

$$\sigma^2 = \mathrm{E}[(x-\mu)^2] \tag{10}$$

$$S = \frac{\mathrm{E}[(x-\mu)^3]}{\sigma^3} \tag{11}$$

$$\kappa = \frac{\mathrm{E}[(x-\mu)^4]}{\sigma^4} \tag{12}$$

where S is the skewness and κ is the kurtosis.

2 Image compositing detection method

2.1 Feature extraction

In this subsection, the feature extraction procedure is described, as shown in Fig. 2. Firstly, DWT of L scales is applied to the given image, and there are $3L$ detail subbands. For each detail subband, two parameters of GGD model, the prediction error, and the four cumulant features are then extracted to form a feature vector with $7 \times 3L = 21L$ dimensions. Finally, the feature vector is fed into an SVM classifier.

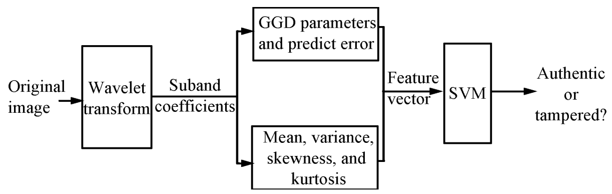


Fig. 2 The flow of our method

2.2 Feature validation analysis

To visualize the distribution of these image feature vectors, the high-dimensional feature vectors were projected onto a three-dimensional linear subspace spanned by the top three principal components obtained as a result of the principal component analysis (PCA) of all image feature vectors. The PCA is described as follows:

Denote column vectors $\boldsymbol{f}_i \in \mathbf{R}^n,\ i = 1, 2, \cdots, K$, as the original feature vectors, where n is the dimensionality of a column vector and K is the number of data points. The overall mean is as follows:

$$\boldsymbol{\mu} = \frac{1}{K}\sum_{i=1}^{K} \boldsymbol{f}_i \tag{13}$$

The zero-meaned data are packed into a $n \times K$ matrix:

$$M = (\boldsymbol{f}_1 - \boldsymbol{\mu} \quad \boldsymbol{f}_2 - \boldsymbol{\mu} \quad \cdots \quad \boldsymbol{f}_K - \boldsymbol{\mu}) \tag{14}$$

If the dimensionality n of $\boldsymbol{f}_i$ is smaller than the number of data points K, as in our case, then the $n \times n$ (scaled) covariance matrix is computed as follows:

$$C = MM^t \tag{15}$$

The principle components are the eigenvectors $\boldsymbol{e}_j$ of the covariance matrix (i.e., $C\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j$), where the eigenvalue, λ_j is proportional to the variance of the original data along the j-th eigenvector. The dimensionality of each $\boldsymbol{f}_i$ is reduced from n to p by projecting (via an inner product) each $\boldsymbol{f}_i$ onto the top p eigenvalue-eigenvectors. The resulting p-dimensional vector is the reduced-dimension representation.

Fig. 3 shows the projected feature vectors of 183 natural and 180 composite images (images come from Columbia Uncompressed Image Splicing Detection Dataset) onto the top three principal components for the $21 \times 3 = 63$ wavelet-domain statistics. In this case, the three top principal components capture over 70 % of the total variance in the original data set. By reducing the dimensionality, a significant fraction of information is discarded, but it allows us to inspect the distribution of the feature vectors for natural and composite images visually. As we have expected, the feature vectors from the proposed image statistics for natural images form a relatively tight cluster. More importantly, the composite images are well separated from the ensemble of natural images. This indicates that the proposed image statistics are capable of capturing statistical regularities in natural images, whereas composite images do not possess such regularities.

2.3 Classification

SVM is a popular technique for classification, which is widely used in machine learning applications.

The LibSVM toolbox[13] is used for our classification, which provides four common kernels: linear, polynomial, radial basis function (RBF), and sigmoid. The RBF kernel is chosen in our method. The RBF kernel nonlinearly maps samples into a higher dimensional space, so it can handle the case when the relation between class labels and attributes is nonlinear. To choose the penalty parameter C and kernel parameter gamma (symbol), we use 5-fold cross-validation to search for the best parameters.

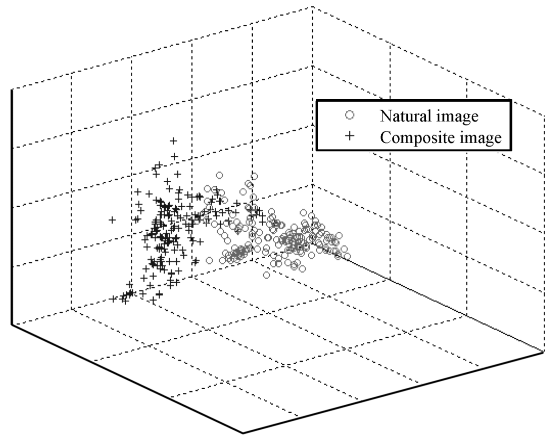


Fig. 3 Projection of the $21 \times 3 = 63$ wavelet-domain statistics for 183 natural images and 180 composite images onto the top three principal components

3 Experimental results

To evaluate the performance of our proposed method, we first carried out the test on the Columbia Uncompressed Image Splicing Detection Dataset[7], and then on an advanced image compositing detection dataset.

3.1 Basic image compositing detection

Columbia Uncompressed Image Splicing Detection Dataset consists of 183 authentic images and 180 spliced images. Authentic images are taken with four cameras: Canon G3, Nikon D70, Canon EOS 350D, and Kodak DCS330. Each spliced image has a "strange" region copied from another image taken by different camera, so each image contains contents from exactly two cameras.

Because no post-processing was performed during the generation of spliced images, this dataset is referred to as basic image compositing detection dataset.

The DWT of three scales is used. In each experiment, half of the authentic images and half of the spliced ones are chosen randomly for training, and the remaining for testing. In order to void the effect of the randomness, for each result, the experiment is conducted 20 times and the arithmetic average is recorded. The results of the experiment are shown in Table 1.

Table 1 Performance of the classifier in our experiment

Feature set	GGD feature	Cumulant feature	Proposed method
TP rate (Recall)	87.25 %	85.71 %	90.11 %
TN rate	90.31 %	76.67 %	94.44 %
Accuracy	88.76 %	81.22 %	92.27 %
Precision	90.06 %	78.72 %	94.25 %

1) The GGD parameters and prediction error features obtain 87.25 % true positive rate (TP rate, also referred to as Recall), 90.31 % true negative rate (TN rate), 88.76 % accuracy, and 90.06 % precision.

2) The cumulant features alone obtain a little worse performance, which are 1.54 %, 13.64 %, 7.54 %, and 11.34 %

lower than the GGD parameters and prediction error features, respectively.

3) The best results (last column) are obtained by incorporating all the proposed features, which improve the performance of the GGD parameters and prediction error features (first column) by 2.86 % in TP rate, 4.13 % in TN rate, 3.51 % in accuracy, and 4.19 % in precision, respectively.

It can be seen that although the performance of the cumulant features is lower than that of the GGD parameters and predict error features, combining all feature sets has further improved the detection rate.

In [8], Hsu proposed an automatic spliced image detection method based on consistency checking of camera characteristics in different areas of an image. The performances of their method and our method are compared on the same image compositing detection dataset, and the corresponding results are shown in Table 2, where $L=3$ in this paper, and $M=8$ in [8].

Table 2 Comparison of performance between the proposed method and Hsu's method

	The proposed method	Hsu method
Feature dimensions	$21 \times L$	$20 \times M$
Precision	94.25 %	70 %
Recall	90.11 %	70 %

As shown in Table 2, the proposed method not only reduces the feature dimensions by 97-D, but also improves the performance by 24.25 % in precision and 20.11 % in recall, respectively.

3.2 Advanced image compositing detection

In order to make the experiments more practical, we build another dataset of 100 authentic JPEG images and 100 composite JPEG images, which are manipulated by either inserting extra content or replacing the original content. To make them more realistic, some pre-processing operations, such as resizing, rotation, and brightness adjustment, are adopted to the inserted content before pasting it to the image. In some cases, we have to blur the block boundaries after pasting too. This test set is referred to as the advanced dataset. Note that, in order to evaluate the generalized capability of the proposed method, the experiments are conducted on the new advanced dataset without re-training the SVM classifier. The results are shown in Table 3. The results are still encouraging, even when the trained detector is applied to new composite images with pre- and post-processing operations.

Table 3 Performance of the trained classifier on the advanced dataset

Dataset	The advanced dataset
Recall	82.26 %
TN rate	75.18 %
Accuracy	78.73 %
Precision	80.28 %

4 Conclusion

In this work, we propose a method for image compositing detection, based on a statistical model for natural image in the wavelet transform domain. The algorithm has the following characteristics: 1) It is fully passive and automatic; 2) Its physical meaning is intuitive, and the feature dimensions are less than that of Hsu's method; 3) Due to simplified implementation and high detection accuracy rate, it has great potential for practical applications. The proposed method is tested over two datasets, and the results are satisfying. In the future work, the statistical models for natural images will be further studied, and more reliable and effective detection method for different image datasets and images with different types of manipulation will be explored.

References

1 Ng T T, Chang S F. Blind Detection of Digital Photomontage Using Higher Order Statistics, Technical Report #201-2004-1, Columbia University, USA, 2004

2 Fridrich J, Soukal D, Lukas J. Detection of copy-move forgery in digital images. In: Proceedings of Digital Forensic Research Workshop. Cleveland, USA: Springer, 2003. 1−10

3 Popescu A C. Statistical Tools for Digital Forensics [Ph. D. dissertation], Dartmouth College, USA, 2005

4 Johnson M K, Farid H. Exposing digital forgeries by detecting inconsistencies in lighting. In: Proceedings of the 7th Workshop on Multimedia and Security. New York, USA: ACM, 2005. 1−10

5 Lin Z C, Wang R R, Tang X O, Shum H Y. Detecting doctored images using camera response normality and consistency analysis. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 1087−1092

6 Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 2006, **1**(2): 205−214

7 Hsu Y F, Chang S F. Detecting image splicing using geometry invariants and camera. In: Proceedings of IEEE International Conference on Multimedia and Expo. Toronto, Canada: IEEE, 2006. 549−552

8 Hsu Y F, Chang S F. Image splicing detection using camera response function consistency and automatic segmentation. In: Proceedings of International Conference on Multimedia and Expo. Beijing, China: IEEE, 2007. 28−31

9 Mallat S G. Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1989, **37**(12): 2091−2110

10 Moulin P, Liu J. Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Transactions on Information Theory*, 1999, **45**(3): 909−919

11 Do M N, Vetterli M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 2002, **11**(2): 146−158

12 Kay S M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice-Hall, 1993

13 Chang C C, Lin C J. LIBSVM: a library for support vector machines [Online], available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm

SUN Shao-Jie Ph. D. candidate at the College of Information System and Management, National University of Defense Technology. His research interest covers image compression and processing. E-mail: sshj_mil@126.com

WU Qiong Ph. D. candidate at the College of Information System and Management, National University of Defense Technology. Her research interest covers digital watermark and image forensics. Corresponding author of this paper.
E-mail: wuqiong_nudt@126.com

LI Guo-Hui Professor in the College of Information System and Management, National University of Defense Technology. His main research interest is multimedia security.
E-mail: guohli@nudt.edu.cn