New Developments in Color Image Tampering Detection

Patchara Sutthiwan, Yun-Qing Shi
Dept. of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ, USA
{ps249, shi}@njit.edu

Jing Dong, Tieniu Tan
Institute of Automation
Chinese Academy of Sciences
Beijing, China
{jdong, tnt}@nlpr.ia.ac.cn

Tian-Tsong Ng
Institute for Infocomm Research
Singapore
tng@2r.a-star.edu.sg

Abstract—In this paper, an efficient framework for passive-blind color image tampering detection is presented. Statistical features are extracted from a given test image and a set of 2-D arrays derived by applying multi-size block discrete cosine transform to the given test image. Image features are extracted from Cr channel, a chroma channel in YCbCr color space, because of its observed sensitivity to color image tampering. A support vector machine is employed to evaluate the effectiveness of image features over a color image dataset recently established for tampering detection. Boosting feature selection is applied to having feature dimensionality reduced so as to make detection accuracy generalizable and computational complexity decreased. Experimental results have demonstrated that the proposed framework applied to the aforementioned dataset outperforms the state of the arts by distinct margins.

Keyword: color image tampering detection, multi-block discrete cosine transform, moments of characteristic functions, Markov process, support vector machine, boosting feature selection

I. INTRODUCTION

An old saying, “Don't judge a book by its cover,” has its root dating back before the Digital Revolution. Undeniably, however, image tampering is emphasizing the importance of this everlasting adage in modern society as splicing is a common way to distort semantic content of an image, which could mislead the public into misbelieving the veracity behind the scene.

Image splicing is a fundamental operation in image forgery and is characterized by a naive cut-and-paste operation, taking a portion of an image and putting it onto either the same or another image without any post-processing such as matting for perfect blending [1]. The term “image tampering” is practically used in place of “image splicing” when any post-processing is done in addition to image splicing to make the artifacts from image forgery less perceptible. Image tampering may make tampered images difficult for ordinary people to differentiate them from authentic images; therefore, an automatic classification system would be able to handle image tampering detection, more reliably than human inspection.

Image forgery detection can be generally classified as active [2] one and passive [3] one. The former detects the integrity of images by checking the change in digital watermark embedded either at the instant of image acquisition or before image distribution; unfortunately, this method requires that image capturing devices be equipped with standardized watermarking functionality, which has not fulfilled yet. On contrary, the latter exploits only the knowledge of images themselves for forgery detection; undoubtedly, passive detection methods have overshadowed active ones. Under a machine learning framework, this paper introduces an effective approach to passive-blind image tampering detection.

Over the past few years, there have emerged several techniques of image splicing/tampering detection. Ng et al. [4] considered the presence of the abrupt changes of pixel values as a key observation for image splicing detection and achieved 72% detection accuracy over the gray image dataset [5] with the use of features formulated by a higher order moment spectrum and bicoherence of images. Johnson and Farid [6] devised a scheme to detect image forgery which checks the lighting inconsistency in an image; however, their method fails in detecting spliced images created from two images taken approximately under the same level of ambient light. Hsu and Chang [7] exploit geometric invariants and camera characteristic consistency to interactively detect spliced images. Fu et al. [8] proposed to use Hilbert-Huang transform and moment of characteristic function (MCF) of the wavelet subbands as features for splicing detection with a reported accuracy of 80% over [5]. Chen et al. [9] proposed features derived from 2-D phase congruency and MCF of wavelet subbands for image splicing detection which achieved 82% accuracy over [5]. Shi et al. [10] adopted a natural image model consisting of statistical features extracted from a test image and 2-D arrays generated by applying multi-size block discrete cosine transform (MBDCT) to the test image. The statistical features include MCF of wavelet subbands and Markov transition probabilities of difference 2-D arrays. This method attains a detection rate of 91.9% over [5]. Dong et al. [11] analyzed the discontinuity of image pixel correlation and coherency caused by splicing. Their statistical features are extracted from image run-length representation and image edge statistics. Qu et al. [12] proposed a scheme to detect image splicing based on the assumption that image splicing leaves sharp boundary. Therefore their method may not work well for detecting forged images with tampering boundaries being blurred such as those demonstrated in [1].

Most of the aforementioned schemes are based on grayscale images or the luminance component of color images, which disregard other meaningful information for detecting color spliced images. Wang et al. [13] employed image chroma to enhance color image tampering detection. Their features are

The authors gratefully acknowledge the support of K.C.Wong Education Foundation, Hong Kong and the National Science Foundation of China (Grant No. 60603011).
derived from gray level co-occurrence matrix (GLCM) of thresholded edge image of image chroma and perform well over an early version of color image tampering detection dataset [14].

In this paper, effective statistical features for passive-blind color image tampering detection are proposed based on a natural image model which stems from the analysis of the changes of frequency distribution of a host image inherited from image tampering. This natural image model is derived from a given test image and its corresponding 2-D arrays generated by MBDCT. In total, 266 image features are extracted from Cr channel, a chroma channel in YCbCr color system, which is observed to be the most sensitive to color image tampering. This observation is also consistent with that made in [13]. The performance of features from Y and that from any chroma channel differ considerably in [13], but only slightly in this work; however, the selection of Cr channel herein does point out the importance of a chroma channel to image color tampering detection. The effectiveness of features is assessed by support vector machines (SVM) over a color image dataset [14]. According to the rule of thumb in pattern recognition theory [15], the ratio of the number of training images in a class to the number of image features in SVM training is deemed too low; boosting feature selection (BFS) [16] is thus applied to selecting an optimal feature subset of a reasonable dimension, 50, to increase the ratios. Not only does a resulting feature dimension reduction bring about a more reliable evaluation on the performance of the proposed statistical model over the foresaid dataset, it also reduces the computational complexity of training and testing classifiers. Experimental results have shown that the performances of the proposed features are distinctively better than that of the state of the art [13].

The rest of this paper is organized as follows: Section II presents the proposed features; the experimental circumstance and results are presented in Section III; conclusions are drawn in Section IV.

II. PROPOSED IMAGE FEATURES

The study on the effectiveness of image chroma brought to color image splicing detection in [13] encourages us to apply this knowledge into our feature set. That is, our features are extracted from one image chroma channel. The original study on the difference and similarity between steganalysis and splicing detection [17] has opened the opportunity to utilize features used in universal steganalysis to tackle tampering detection. The image model in [17] achieves 92.2% accuracy over [5]. This leads to the concept of natural image model which distinctly represents an original image in a high-dimensional space.

Our natural image model, based on that in [10], is derived from the following two combinations: (1) that of features derived from the image pixel 2-D array and those derived from the MBDCT coefficient 2-D arrays; (2) that of moments of characteristic functions based (moment-based, in short) features and Markov-based features.

It is well known that the block discrete cosine transform (BDCT) possesses superior capability in decorrelation and energy compaction. Image tampering changes the local frequency distribution of the host images and such changes can be reflected by the coefficients of a BDCT 2-D array. Wide varieties of host images, pasted image fragments and image forgery techniques are important key factors which make the changes introduced to a tampered image various and complicated. Therefore, it is expected that, with various block sizes, the coefficients of MBDCT 2-D arrays can more efficiently capture such diverse and complicated changes than those of a single-block-size BDCT array. In [10] it has been shown that each of BDCT 2-D arrays generated with block sizes 2×2, 4×4 and 8×8 complements one another in terms of tampering detection rate.

A. Moment-Based Features

The moment-based features combine the information from first-order statistic (the characteristic functions of the 1-D histogram) and second-order statistic (the characteristic functions of the 2-D histogram) to capture tampering artifacts. Pasting image fragments onto a host image is analogous to adding independent noise onto the host image. Assuming that the magnitude of the characteristic function of the noise is non-increasing, it can be shown by using the discrete Chebyshev inequality [18] that such moments never increase due to image tampering.

B. Prediction-error 2-D Array

A prediction-error 2-D array reduces the influence caused by diversity of image content; as a result, high frequency contents, e.g., edge, where tampering artifacts reside in, are enhanced.

\[
\Delta x = x - \overline{x} = x - \text{sign}(x) \cdot \|x\| \cdot |x| \quad (1)
\]

C. Discrete Wavelet Transform (DWT)

Discrete wavelet transform (DWT) is suitable to catch transient or localized changes in discrete spatial and frequency domains, so it is favorable to tampering detection. Three-level Haar wavelet transform is previously reported and justified in [8]. In this implementation, for simplicity, one-level Haar wavelet transform is applied to decomposing the image pixel 2-D array, its corresponding MBDCT coefficients, and prediction-error 2-D arrays into several wavelets subbands which are then used in order to compute statistical moments.

D. Moments and Marginal Moments

The 1-D characteristic function (CF) is the DFT of the 1-D histogram of each wavelet subband. In this sequel, the image, coefficient arrays and their prediction error 2-D arrays are considered as the low-low subbands at level 0 denoted by \( LL_0 \). \( H(u) \) denotes the CF component at a discrete frequency \( u \) and \( K \) is its total number of different coefficient values. The \( l \)-th absolute moment of 1-D CF is defined by
As a second-order statistic, a 2-D histogram can further enhance tampering detection capability and such a histogram is a measure of the joint occurrence of pairs of pixels separated by a specified distance and orientation. The unit distances between two pixels along horizontal and vertical directions are concerned herein, resulting in two second-order histograms per image. Like two pixels along horizontal and vertical directions are concerned specified distance and orientation. The unit distances between measured moments of 2-D CF are constructively blended.

\[
M_{ij} = \frac{\sum_{j=1}^{K} \sum_{j=1}^{K} |H(u,v)|}{\sum_{j=1}^{K} |H(u,v)|}
\]

(2)

E. Markov-Based Features

In [10], it has been demonstrated that, characterized by transition probability matrix (TPM), Markov process (MP) may model the correlation among elements of a 2-D array; hence it is expected that the statistical changes caused by image tampering can be reflected by the Markovian TPM.

A JPEG 2-D array is defined as a 2-D array which consists of the magnitudes of all of Huffman-decoded JPEG coefficients, \(F(u,v)\), \(u \in [0,S_u-1]\) and \(v \in [0,S_v-1]\), of a given image. Difference JPEG 2-D arrays along horizontal \((h)\) and vertical \((v)\) axes are defined as

\[
\begin{align*}
D_h(u,v) &= F(u,v) - F(u+1,v), \\
D_v(u,v) &= F(u,v) - F(u,v+1)
\end{align*}
\]

(4)

According to our statistical study on some large image dataset, the distribution of elements of difference JPEG 2-D array is Laplacian-like, thus making thresholding technique feasible. All the elements in (5) are truncated such that their elements whose values are larger than \(T\) or smaller than \(-T\) will be represented by \(T\) or \(-T\), respectively. This procedure makes the dimensionality of each TPM equal to \((2T+1)\times(2T+1)\). MP is then applied to difference JPEG 2-D arrays and transition probabilities can be calculated as follows.

\[
p[D_{h}(u,v) = m | D_h(u,v) = n] = \frac{\sum_{i=0}^{K-2} \sum_{j=0}^{K-2} \delta(D_h(u,v) = m, D_h(u,v+1) = n)}{\sum_{i=0}^{K-2} \sum_{j=0}^{K-2} \delta(D_h(u,v) = m)}
\]

\[
p[D_{v}(u,v) = m | D_v(u,v) = n] = \frac{\sum_{i=0}^{K-2} \sum_{j=0}^{K-2} \delta(D_v(u,v) = m, D_v(u+1,v) = n)}{\sum_{i=0}^{K-2} \sum_{j=0}^{K-2} \delta(D_v(u,v) = m)}
\]

where \(m, n \in \{0, T, \ldots, T\}\), and \(\delta()\) and \(\delta()\) are respectively 1-D and 2-D dirac delta functions. Our study indicates that when \(T = 3\), about 90% of such elements are captured. Hence, TPM can be calculated with reduced dimensionality. Markov-based features are formed from all the elements in the TPMs. A concrete justification in [10] reveals that Markov features enhance moment-based features.

G. Boosting Feature Selection

Boosting feature selection (BFS) [16] is an iterative algorithm which selects an optimal set of \(D\)-dimensional features from \(M\)-dimensional features. Less than or equal to \(M\), selected feature dimension \(D\) is user-defined and corresponds to the number of iterations in the algorithm. BFS can not only reduce the feature dimension and the computational complexity of training and testing SVM classifiers, but also improve the detection accuracy under some circumstances as verified by [13]. A proper value of \(D\) may be determined by the relationship between \(D\) and its SVM classification accuracy. Unfortunately, in our proposed features, no \(D\) can raise the detection rate, so a \(D\) is selected to manage the low ratios of number of training images available in a class to the dimension of image features in SVM training and reduce computational complexity of training and testing classifiers.

III. EXPERIMENTAL RESULTS

A. Image Dataset

The Institute of Automation at Chinese Academy of Sciences (CASIA) recently released a color image tampering detection evaluation dataset [14] consisting of 800 authentic and 921 tampered images of sizes \((384 \times 256, 256 \times 384, 256 \times 256)\) close to that of the luminance frame of CIF video sequences. Most of authentic images are from CorelDraw Image Database, and the others are either downloaded from the Internet or captured using digital cameras. With anti-aliasing filter enabled to smoothen the boundaries of tampering region by color interpolation, Adobe Photoshop was used to create the tampered images in the dataset. Therefore, the boundaries of the tampered region(s) of the resulting tampered images are not as sharp as those created by naïve cut-and-paste operation. Fig. 2 depicts samples of authentic images in the top row and their forgery counterparts in the bottom row.

In this work, at the time of the experiments, only 800 authentic and 626 tampered color images of those in [14] were available. In [13], 500 authentic and 448 tampered images of those in [14] were available.

Figure 2. Examples of authentic images (all in the top row) and their forgery counter parts (all in the bottom row).
B. Classification

The support vector machine (SVM) with degree 2 polynomial kernel is employed. SVM, the Matlab codes of which are available at [20], has independently been trained and tested for 20 times, each by 5/6 randomly selected authentic and tampered images for training and the leftover for testing. Reliable performance assessments were carried out on a set of 50 selected features. In Table 1, the detection rates of the proposed features are shown in comparison with the best detection rates under two different scenarios reported in [13] in which features derived from gray level co-occurrence matrix (GLCM).

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Color Channel</th>
<th>Feature Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-BFS</td>
<td>Cr</td>
<td>324</td>
</tr>
<tr>
<td>MBDCT</td>
<td>Cr</td>
<td>50</td>
</tr>
<tr>
<td>BFS</td>
<td>Cr</td>
<td>50</td>
</tr>
<tr>
<td>MBDCT</td>
<td>Y</td>
<td>266</td>
</tr>
<tr>
<td>BFS</td>
<td>Y</td>
<td>266</td>
</tr>
<tr>
<td>AR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>89.9</td>
<td>94.1</td>
</tr>
<tr>
<td>TN</td>
<td>90.5</td>
<td>94.5</td>
</tr>
<tr>
<td>AR</td>
<td>94.8</td>
<td>98.0</td>
</tr>
<tr>
<td>AR</td>
<td>98.0</td>
<td>97.8</td>
</tr>
<tr>
<td>AR</td>
<td>97.5</td>
<td></td>
</tr>
</tbody>
</table>

IV. CONCLUSION

This paper presents a set of features, moments of 1-D and 2-D characteristic functions derived from multi-size block discrete cosine transform (MBDCT) and Markov transition probability matrix derived from difference arrays of magnitudes of Huffman-decoded JPEG coefficients. Features are extracted in a chroma channel for color image tampering detection. Although the performance of features from Cr is the best, it is only slightly better than those from Y (as shown in Table 1). Nevertheless, the performance gap between features from the luminance and those from a chroma channel is huge in [13]. Therefore, we conclude that the performance gain of the exploitation of a chroma channel in color image tampering detection depends on the nature of image statistical model. In this paper, the rationale behind formulating image features from Cr channel is not to emphasize any dramatic increase in performance but to stress the importance of chroma channels to color image tampering detection. Features are extracted by using HP Pavilion dv6930us with un-optimized Matlab codes over [14] and the average computational time for feature extraction is 26.4 seconds per image.

Although the detection rate of the 266 features is very high, 98.0%, we do not suggest it as a reliable result because of the insufficient generalization properties [15] of the classifiers owing to the low ratios of the number of training images to the feature dimension in SVM training. Boosting feature selection (BFS) circumvents such a limitation by reducing the feature dimension without serious deterioration in detection accuracy. A dimension reduction caused by BFS can also reduce the computational complexity of training and testing SVM. By BFS, 50 selected features perform better than the best detection rates reported in [13] over the fundamentally same dataset [14].

As a summary, the new developments on color image tampering detection are listed below. 1) CASIA dataset V1.0 is a color image dataset designed and available for color image tampering detection, where Adobe Photoshop 7.0 was used in image tampering. 2) Proposed MBDCT based feature set is effective. 3) Features from Cr channel play an important role for color image tampering detection. 4) BFS benefits tampering detection by reducing number of features without any serious deterioration in detection performance.

REFERENCES