BLIND DETECTION OF PHOTOMONTAGE USING HIGHER ORDER STATISTICS

Tian-Tsong Ng, Shih-Fu Chang
Department of Electrical Engineering
Columbia University, New York, NY
{ttng,sfchang}@ee.columbia.edu

Qibin Sun
Institute of Infocomm Research
Singapore
qibin@i2r.a-star.edu.sg

ABSTRACT

In this paper, we investigate the prospect of using bicoherence features for blind image splicing detection. Image splicing is an essential operation for digital photomontaging, which in turn is a technique for creating image forgery. We examine the properties of bicoherence features on a data set, which contains image blocks of diverse image properties. We then demonstrate the limitation of the baseline bicoherence features for image splicing detection. Our investigation has led to two suggestions for improving the performance of the bicoherence features, i.e., estimating the bicoherence features of the authentic counterpart and incorporating features that characterize the variance of the feature performance. The features identified through the suggestions are evaluated using Support Vector Machine (SVM) classification and shown to be encouraging.

1. INTRODUCTION

Photomontage refers to a paste-up produced by sticking together photographic images. In olden days, creating a good composite photograph required sophisticated skills of darkroom masking or multiple exposures of a photograph negative. In today's digital age, however, the creation of photomontage is made simple by the cut-and-paste tools of the popular image processing software such as Photoshop. With such an ease of creating good digital photomontages, we could no longer take image authenticity for granted especially when it comes to legal photographic evidence [1] and electronic financial documents. Therefore, we need a reliable and objective way to examine image authenticity.

Lack of internal consistency, such as inconsistencies in object perspective, in an image is sometimes a telltale sign of photomontage [1]. However, unless the inconsistencies are obvious, this technique can be subjective. Furthermore, forgers can always take heed of any possible internal inconsistencies.

Although image acquisition device with digital watermarking features could be a boon for image authentication, presently there still is not a fully secured authentication watermarking algorithm, which can defy all forms of hacking, and the hardware system has to secure from unauthorized watermark embedding. Equally important are the issues such as the need for both the watermark embedder and detector to use a common algorithm and the consequence of digital watermarks degrading image quality.

On the premise that human speech signal is highly Gaussian in nature [2], a passive approach was proposed [3] to detect the high level of non-gaussianity in spliced human speech using bicoherence features. Unlike human speech signal, the premise of high guassianity does not hold for image signal. It was shown [4] that bispectrum and trispectrum of natural images have a concentration of high values in regions where frequency components are aligned in orientation, due to image features of zero or one intrinsic dimensionality such as uniform planes or straight edges. As images originally have high value in higher order spectrum, detecting image splicing based on the same principle of increased non-gaussianity would be a very low signal-to-noise problem, not to mention the possible complex interaction between splicing and the non-linear image features.

Recently, a new system for detecting image manipulation based on a statistical model for 'natural' images in the wavelet domain is reported [5]. Image splicing is one kind of image tampering the system takes on; however, no further detail about the technical approach is provided in the article.

Image splicing is defined as a simple joining of image regions. We currently do not address the combined effects of image splicing and other post-processing operations. Creation of digital photomontage always involves image splicing although users could apply post-processing such as airbrush style edge softening, which can potentially be detected by other techniques [5]. In fact, photomontages with merely image splicing, as in Figure 1, can look deceivingly authentic and each of them only took a professional graphic designer 10-15 minutes to produce.

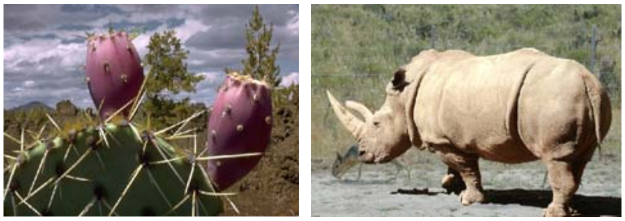

Figure 1: Spliced images that look authentic subjectively

In this paper, we pursue the prospect of grayscale image splicing detection using bicoherence features. We first examine the properties of the proposed bicoherence features [3] in relation to image splicing and demonstrate the insufficiency of the features. We then propose two new methods on improving the performance of the bicoherence features for image splicing detection. Lastly, we evaluate the methods using SVM classification experiments over a diverse data set of 1845 image blocks. More details about this work are included in [6].

2. BICOHERENCE

Bicoherence is a normalized bispectrum, i.e., the third order correlation of three harmonically related Fourier frequencies of a signal, *X(ω)* [7]:

$$b(\omega_1,\omega_2)=\frac{E[X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2)]}{\sqrt{E[\left|X(\omega_1)X(\omega_2)\right|^2]E[\left|X(\omega_1+\omega_2)\right|^2]}}=\left|b(\omega_1,\omega_2)\right|e^{j\Phi(b(\omega_1,\omega_2))}$$

When the harmonically related frequencies and their phase are of the same type of relation, i.e., when there exists (ω_1, ϕ_1), (ω_2, ϕ_2) and $(\omega_1+\omega_2, \phi_1+\phi_2)$ for $X(\omega)$, $b(\omega_1,\omega_2)$ will have a high magnitude value and we call such phenomena *quadratic phase coupling* (QPC). As such, the average bicoherence magnitude would increase as the amount of QPC grows. Besides that, bicoherence is insensitive to signal gaussianity and bispectrum is often used as a measure of signal non-gaussianity [8].

2.1. Bicoherence Features

Motivated by the effectiveness of the bicoherence features used for human-speech splicing detection [3], similar features are extracted from a bicoherence with

- Mean of magnitude: $M = |\Omega|^{-1}\sum_{\Omega}|b(\omega_1, \omega_2)|$
- Negative phase entropy: $P=\Sigma_n p(\Psi_n)\log p(\Psi_n)$

where

$\Omega=\{(\omega_1, \omega_2)|\ \omega_1=(2\pi m_1)/M,\ \omega_2=(2\pi m_2)/M,\ m_1, m_2= 0,\ldots,.M\text{-}1\}$

$p(\Psi_n)= |\Omega|^{-1}\sum_{\Omega} 1(\Phi(b(\omega_1, \omega_2))\in \Psi_n)$, $1(\cdot)$=*indicator function*

$\Psi_n=\{\phi|\text{-}\pi+(2\pi n)/N\leq\phi< \text{-}\pi+2\pi(n+1)/N\}$, $n=0,\ldots, N\text{-}1$

2.2. Estimation of Bicoherence Features

With limited data sample size, instead of computing 2-D bicoherence features from an image, 1-D bicoherence features can be computed from N_v vertical and N_h horizontal image slices of an image and then combined as follows:

$$fM=\sqrt{(\tfrac{1}{N_h}\sum_i M_i^{Horizontal})^2+(\tfrac{1}{N_v}\sum_i M_i^{Vertical})^2}$$

$$fP=\sqrt{(\tfrac{1}{N_h}\sum_i P_i^{Horizontal})^2+(\tfrac{1}{N_v}\sum_i P_i^{Vertical})^2}$$

In order to reduce the estimation variance, the 1-D bicoherence of an image slice is computed by averaging segment estimates:

$$\hat{b}(\omega_1,\omega_2)=\frac{\frac{1}{k}\sum_k X_k(\omega_1)X_k(\omega_2)X_k^*(\omega_1+\omega_2)}{\sqrt{\left(\frac{1}{k}\sum_k\left|X_k(\omega_1)X_k(\omega_2)\right|^2\right)\left(\frac{1}{k}\sum_k\left|X_k(\omega_1+\omega_2)\right|^2\right)}}$$

We use segments of 64 samples in length with an overlap of 32 samples with adjacent segments. For lesser frequency leakage and better frequency resolution, each segment of length 64 is multiplied with a Hanning window and then zero-padded from the end before computing 128-point DFT of the segment.

In Fackrell et al. [9], it is suggested that N data segments should be used in the averaging procedure for estimating a N-point DFT bispectrum of a stochastic signal. Overall, we use 768 segments to generate features for a 128x128-pixel image block.

3. IMAGE DATA SET

Our data set [10] is collected with sample diversity in mind. It has 933 authentic and 912 spliced image blocks of size 128 x 128 pixels. The image blocks are extracted from images in CalPhotos image set [11]. As the images are contributions from photographers, in our case, they can be considered as authentic i.e., not digital photomontages.

The authentic category consists of image blocks of an entirely homogenous textured or smooth region and those having an object boundary separating two textured regions, two smooth regions, or a textured regions and a smooth region. The location and the orientation of the boundaries are random.

The spliced category has the same subcategories as the authentic one. For the spliced subcategories with object boundaries, image blocks are obtained from images with spliced objects; hence, the splicing region interface coincides with an arbitrary-shape object boundary. Whereas for the spliced subcategories with an entirely homogenous texture or smooth region, image blocks are obtained from those in the corresponding authentic subcategories by copying a vertical or a horizontal strip of 20 pixels wide from one location to another location within a same image.

4. PROPERTIES OF BICOHERENCE FEATURES

We are interested in investigating the performance of bicoherence features in detecting spliced images on the three object interface types for which such performance varies over, i.e. smooth-smooth, textured-textured, and smooth-textured. Figure 2 shows the scatter plot of the bicoherence magnitude feature (*fM*) of the authentic and spliced image blocks with a particular object interface type. The plots also show how well the edge percentage (y-axis) captures the characteristics of different interface types. The edge pixels are obtained using Canny edge detector. The edge percentage is computed by counting the edge pixels within each block. As the plots for bicoherence phase feature (*fP*) are qualitatively similar, they are omitted due to space constraints.

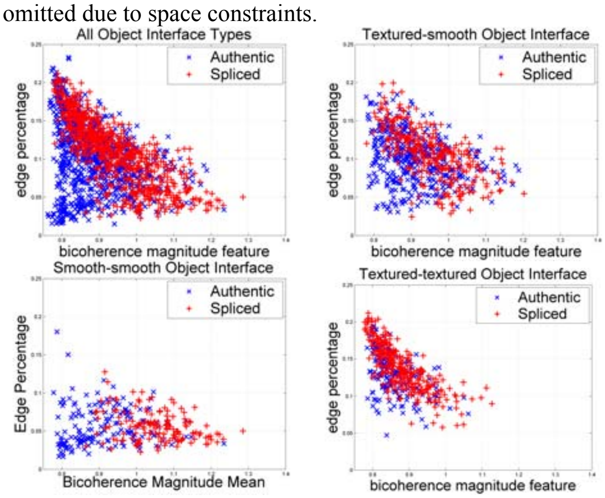

Figure 2: Bicoherence magnitude feature for different object interface types

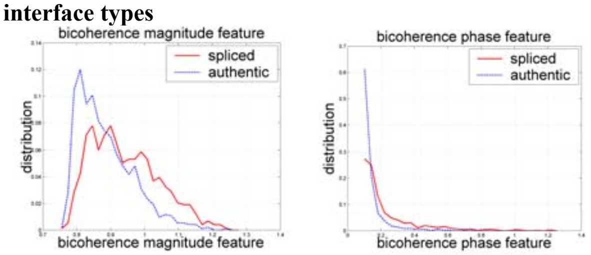

Figure 3: Distribution of the bicoherence magnitude feature, *fM*, (left) and the phase feature, *fP* (right)

We observe that the performance of the bicoherence feature in distinguishing spliced images varies for different object interface types, with textured-textured object interface type being the worst case. Figure 3 shows the distribution of the features for the authentic and spliced image categories. We can observe that the distributions of the two image block categories are greatly overlapped, although there are noticeable differences in the peak locations and the heavy tails. Hence, we would expect poor classification between the two categories if the features were to be used directly.

5. METHODS FOR IMPROVING THE PERFORMANCE OF BICOHERENCE FEATURES

Our investigation on the properties of bicoherence features for images leads to two methods for augmenting the performance of the bicoherence features in detecting image splicing:

1. By estimating the bicoherence features of authentic images.
2. By incorporating image features that capture the image characteristics on which the performance of the bicoherence features varies, e.g., edge pixel percentage feature (*fE*) capture the characteristics of different object interface.

5.1. Estimating Authentic Counterpart Bicoherence Features

Assume that for every spliced image, there is a corresponding authentic counterpart, which is similar to the spliced image except that it is authentic. The rationale of the approach, formulated as below, is that if the bicoherence features of the authentic counterpart can be estimated well, the elevation in the bicoherence features due to splicing could be more detectable.

$$\begin{aligned} f_{Bic} &= g(\Lambda_I(image), \Lambda_S(image, s), s) + \varepsilon \\ &\approx g_1(\Lambda_I(image)) + g_2(\Lambda_S(image, s), s) + \varepsilon \\ &\approx f_{Authentic} + \Delta f_{Splicing} + \varepsilon \end{aligned}$$

where Λ_I is a set of splicing-invariant features while Λ_S is a set of features induced by splicing, s is a splicing indicator and ε is the estimation error. In this formulation, g_1 corresponds to an estimate of the bicoherence feature of the authentic counterpart, denoted as $f_{Authentic}$ and g_2 corresponds to the elevation of the bicoherence feature induced by splicing, denoted as $\Delta f_{Splicing}$. With $\Delta f_{Splicing}$, splicing would be more detectable after the significant interference from the splicing-invariant component, g_1, is removed. $\Delta f_{Splicing}$ can be estimated with $f_{Bic} - f_{Authentic}$, which we call *prediction discrepancy*. The $f_{Authentic}$ estimation performance would be determined by two factors, i.e., how much we capture the splicing-invariant features and how well we map the splicing-invariant features to the bicoherence features.

A direct way to arrive at a good estimator is through an approximation of the authentic counterpart obtained by depriving an image of the splicing effect. As a means of 'cleaning' an image of its splicing effect, we have chosen the texture decomposition method based on functional minimization [12], which has a good edge preserving property, for we have observed the sensitivity of the bicoherence features to edges.

5.2. Texture Decomposition with Total Variation Minimization and a Model of Oscillating Function

In functional representation, an image, f defined in $\Omega \subset R^2$, can be decomposed into two functions, u and v, within a total variation minimization framework with a formulation [12]:

$$\inf_u \left\{ E(u) = \int_\Omega |\nabla u| + \lambda \|v\|_*, f = u + v \right\}$$

where the u component, a structure component of the image, is modeled as a function of bounded variation while the v component, representing the fine texture or noise component of the image, is modeled as an oscillation function. $\|\cdot\|_*$ is the norm of the oscillating function space and λ is a weight parameter for trading off variation regularization and image fidelity.

The minimization problem can be reduced to a set of partial differential equations known as Euler-Lagrange equations and solved numerically with finite difference technique. As the structure component could contain arbitrarily high frequencies, conventional image decomposition by filtering could not attain such desired results. In this case, the structure component will serve as an approximation for the authentic counterpart, hence, the estimator for $fM_{Authentic}$ and $fP_{Authentic}$ are respectively $\hat{fM}_{Authentic} = fM_{structure}$ and $\hat{fP}_{Authentic} = fP_{structure}$.

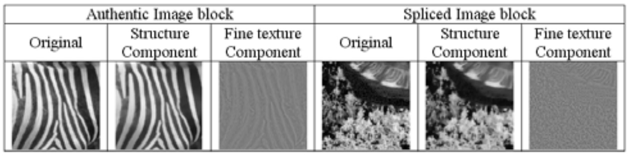

Figure 4: Examples of texture decomposition

For the linear prediction discrepancies between the bicoherence features of an image and those of its authentic counterpart, i.e., $\Delta fM = fM - \alpha \hat{fM}_{Authentic}$ and $\Delta fP = fP - \beta \hat{fP}_{Authentic}$, the parameters α and β, without being assumed to be unity, are learnt by Fisher Linear Discriminant Analysis in the 2-D space (fM, $\hat{fM}_{Authentic}$) and (fP, $\hat{fP}_{Authentic}$) respectively, to obtain the subspace projection where the between-class variance is maximized relative to the within-class variance, for the authentic and spliced categories.

We evaluate effectiveness of the estimator, as shown in Figure 5 using the prediction discrepancy for the magnitude and phase features. Our objective is to show that the new features (*ΔfM, ΔfP*) have a stronger discrimination power between authentic and spliced compared to the original features (*fM, fP*). This objective is partially supported by observing the difference between Figure 5 and Figure 3 (In Figure 5, two distributions are more separable)

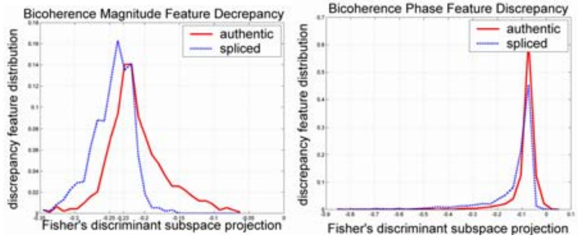

Figure 5: Distribution of prediction discrepancy

6. SVM CLASSIFICATION EXPERIMENTS

We herein evaluate the effectiveness of the features, which are derived from the proposed method, i.e., prediction discrepancy and edge percentage using our data set. SVM classifications with RBF kernel are performed with parameters chosen for ensuring no overfitting as verified by 10-fold cross-validation. Three statistics obtained from 100 runs of classification are used to evaluate the performance of feature sets:

- Accuracy mean: $M_{accuracy} = \frac{1}{100}\sum_i (N^i_{S|S} + N^i_{A|A}) \big/ (N^i_{\bullet|S} + N^i_{\bullet|A})$
- Average precision: $M_{precision} = \frac{1}{100}\sum_i N^i_{S|S} / N^i_{S|\bullet}$
- Average recall: $M_{recall} = \frac{1}{100}\sum_i N^i_{S|S} / N^i_{\bullet|S}$

where S and A represents Spliced and Authentic respectively and $N^i_{A|B}$ denotes the number of samples B detected as A in the ith run. The results of the experiment are shown below:

Feature Label	Feature Name
Orig	magnitude and phase features { *fM*, *fP* }
Delta	Prediction discrepancy { *ΔfM*, *ΔfP* }
Edge	Edge percentage *fE*

Feature Set	$M_{accuracy}$	$M_{precision}$	M_{recall}
Orig	0.6259	0.6354	0.5921
Delta	0.6876 (+6.2 %)	0.6685	0.7477
Orig+Delta	0.7028 (+7.7 %)	0.6725	0.7925
Orig+Edge	0.7005 (+7.5 %)	0.6780	0.7667
Delta+Edge	0.6885 (+6.3 %)	0.6431	0.8517
Orig+Delta+Edge	0.7148 (+8.9 %)	0.6814	0.8098

Note: Statistical t-tests for classification results using feature set {*fM, fP*} against all other results are performed. The null hypothesis (i.e., the mean of the two results are the same) is rejected at a 0.05 significance level for all tests.

Below are the observations from the classification results:

1. Prediction discrepancy features alone obtain 6.2 % improvement in $M_{accuracy}$ over the original bicoherence features.
2. Edge percentage improves the performance of the bicoherence features by 7.5 % in $M_{accuracy}$.
3. Prediction discrepancy and edge percentage are redundant with respect to each other.
4. The best performance (last row) obtained by incorporating all the proposed features is 71 % in $M_{accuracy}$, which is 8.9 % better than the baseline method (first row).

The results are encouraging as it shows the initial promise of the authentic counterpart estimation. The third observation may be an indication that the prediction discrepancy features are less affected by image texturedness. Hence, if the estimation of the authentic counterpart bicoherence features can be further improved, it may help in the classification of the toughest case where the object interface type is textured-textured.

The block level detection results can be combined in different ways to make global decision about the authenticity of a whole image or its sub-regions. For example, Figure 6 illustrates the idea of localizing the suspected splicing boundary.

7. CONCLUSIONS

In this paper, we have shown the difficulties of image splicing detection using bicoherence features, despite the technique being effective on human speech signals. We have also empirically shown how the performances of the bicoherence features depending on the different object interface types. Two methods are proposed for improving the capability of the bicoherence features in detecting image splicing. The first exploits the dependence of the bicoherence features on the image content such as edge pixel density and the second offsets the splicing-invariant component from bicoherence and thereby obtains better discriminative features. Finally, we observe improvements in SVM classification after the derived features are incorporated.

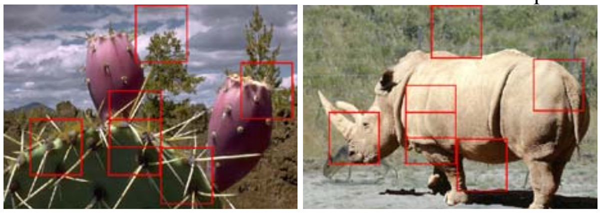

Figure 6: Spliced image blocks (marked with a red box)

This is the first step of our effort in using bicoherence features for image splicing detection. We will next seek a model to get an insight on why bicoherence is sensitive to splicing, from which other effective features can be derived.

8. ACKNOWLEDGEMENTS

Gratitude to A*STAR, Singapore for sponsoring the first author, to Lexing Xie, Shahram Ebadollahi and Anita Huang for their discussions and helps, and to Huei-Sim Tang for the professionally produced photomontages.

9. REFERENCES

[1] W. J. Mitchell, "When Is Seeing Believing?", *Scientific American*, pp. 44-49, February 1994.

[2] J.W.A. Fackrell and S. McLaughlin. “Detecting nonlinearities in speech sounds using the bicoherence”, *Proc. of the Institute of Acoustics*, 18(9), pp. 123–130, 1996.

[3] H. Farid, “Detecting Digital Forgeries Using Bispectral Analysis”, *Technical Report*, AIM-1657, MIT AI Memo, 1999.

[4] Krieger, G., Zetzsche, C. and Barth, E., “Higher-order statistics of natural images and their exploitation by operators selective to intrinsic dimensionality”, *Proc. of IEEE Signal Processing Workshop on HOS*, pp. 147-151, 21-23 July 1997.

[5] H. Farid, “A Picture Tells a Thousand Lies”, *New Scientist*, 179(2411), pp. 38-41, Sept. 6, 2003.

[6] T.-T. Ng and S.-F. Chang, “Blind Detection of Photomontage Using Higher Order Statistics”, *ADVENT Technical Report*, #201-2004-1, Columbia University, http://www.ee.columbia.edu/dvmm/, Jan 2004.

[7] Y. C. Kim and E. J. Powers, “Digital Bispectral Analysis and its Applications to Nonlinear Wave Interactions”, *IEEE Trans. on Plasma Science,* vol. PS-7, No.2, pp. 120-131, June 1979.

[8] M. Santos et al., “An estimate of the cosmological bispectrum from the MAXIMA-1 CMB map”, Physical Review Letters, 88, 241302, 2002

[9] J. W. A. Fackrell, P. R. White, J. K. Hammond, R. J. Pinnington and A. T. Parsons, "The interpretation of the bispectra of vibration signals-I. Theory", *Mechanical Systems and Signal Processing*, Vol. 9(3), pp. 257-266, 1995.

[10] Data set of authentic and spliced image blocks, DVMM, Columbia Univ., http://www.ee.columbia.edu/dvmm/researchProjects/AuthenticationWatermarking/BlindImageVideoForensic/

[11] CalPhotos: A database of photos of plants, animals, habitats and other natural history subjects. Digital Library Project, University of California, Berkeley.

[12] L. A. Vese and S. J. Osher, “Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing”, *UCLA C.A.M. Report,* 02-19, May 2002.