

Blind Detection of Digital Photomontage using Higher Order Statistics

Tian-Tsong Ng and Shih-Fu Chang

Electrical Engineering Department,
Columbia University, New York.

{ttng, sfchang}@ee.columbia.edu

Abstract

The advent of the modern digital technology has not only brought about the prevalent use of digital images in our daily activities but also the ease of creating image forgery such as digital photomontages using publicly accessible and user-friendly image processing tools such as Adobe Photoshop. Among all operations involved in image photomontage, image splicing can be considered the most fundamental and essential operation. In this report, our goal is to detect spliced images by a passive-blind approach, which can do without any prior information, as well as without the need of embedding watermark or extracting image features at the moment of image acquisition. Bicoherence, a third-order moment spectra and an effective technique for detecting quadratic phase coupling (QPC), has been previously proposed for passive-blind detection of human speech splicing, based on the assumption that human speech signal is low in QPC. However, images originally have non-trivial level of bispectrum energy, which implies an originally significant level of QPC. Hence, we argue that straightforward applications of bicoherence features for detecting image splicing are not effective. Furthermore, the theoretical connection between bicoherence and image splicing is not clear. For this work, we created a data set, which contains 933 authentic and 912 spliced image blocks. Besides that, we proposed two general methods, i.e., characterizing the image features that bicoherence is sensitive to and estimating the splicing-invariant component, for improving the performance of the bicoherence technique. We also proposed a model of image splicing to explain the effectiveness of bicoherence for image-splicing detection. Finally, we evaluate the performance of the features derived from the proposed improvement methods by Support Vector Machine (SVM) classification on the data set. The results show a significant improvement in image splicing detection accuracy, from 62% to 72%.

1 Introduction

Photomontage refers to a paste-up produced by sticking together photographic images. While the term photomontage was first used for referring to an art form after the First World War, the act of creating composite photograph can be traced back to the time of camera invention.

Before the digital age, creating a good composite photograph required the sophisticated skill of darkroom masking or precise multiple exposure of a photograph negative. However, in the age where digital images are prevalent, creating photomontage can be as easy as performing a cut-and-paste with specific tools provided by image publishing software such as Adobe Photoshop, was reported to have 5 million registered users at 2004 [1]. Nowadays, this type of software is widely accessible to us. While such act is simple for naïve users, slightly more sophisticated users may employ other tools from the same software to apply additional tricks such as softening the outline of the pasted object, adjusting the direction of an object illumination and so on, in order to enhance the realism of the composite. All it takes to create a digital photomontage of fairly high quality is no more than a good image-publishing software and an average user of the software [2]. Image manipulation (mainly photomontaging) using Adobe Photoshop has actually become a pastime for certain users. For instances, sites like b3ta.com and Worth1000.com host weekly Photoshop challenges, where users, who may not be a professional, submit their work of photomontage to vie for the best. Up till May 2004, Worth1000.com site alone contains 85,464 photomontages and examples of these work are shown in Figure 1.





Figure 1: Examples of photomontages from the site Worth1000.com

The ease of creating digital photomontage, with a quality that could manipulate belief, would certainly make us to think twice before accepting an image as authentic [3]. This becomes a serious issue when it comes to photographic evidence presented in the court or for insurance claims. Image authenticity is also a critical for news photographs as well as for the scanned image of checks in the electronic check clearing system. Therefore, we need a reliable way to examine the authenticity of images, even at a situation where the images look real and unsuspecting to human.

1.1 Prior Work

One way to examine the authenticity of images is to check the internal consistencies within a single image, such as whether or not all the objects in an image are in correct perspective or the shadow is consistent with the lighting [3]. This technique inspects the minor detail of the image to locate the possible inconsistencies, which is likely to be overlooked by forgers. Coincidentally, Such general methodology finds parallel in the area of art connoisseurship and psychoanalysis [4]. However, unless there are major and obvious inconsistencies, minor or ambiguous inconsistencies can be easily argued away. Furthermore, creating a digital photomontage that is free from major inconsistencies is not difficult for professional forgers if internal consistencies are specifically taken into consideration.

Another approach of asserting image authenticity is through extracting digital signature [5] or content-based signature [6-10] from an image at the moment it is taken by a secure or trustworthy camera [5]. Alternatively, watermark data can be embedded into an image to achieve the same purpose. Such approach is considered to be an active approach because it requires a known pattern to be embedded in an image or the image features to be recorded before authenticity checking can be performed later. In general, watermarking techniques such as fragile watermarking [11-15], semi-fragile watermarking [16-19] or content-based watermarking [20, 21] are used for the image authentication application. The watermarking techniques have their own inherent issues. Fragile watermark is impractical for many real-life applications as compression and transcoding of images are very common in the chain of multimedia delivery. In this case,

fragile watermarking techniques will declare a transcoded or compressed image as inauthentic even though the transcoding or compression operation is a content preserving. Although semi-fragile watermark or content-based watermark can be designed to tolerate a specific set of content-preserving operations such as JPEG compression with adjustable degree of resilience well [17], to design a watermark that could meet the complex requirements of the real-life applications in terms of being resilient to a defined set of operations with adjustable degree of resilience while being fragile to another set of operations is in general challenging. As an example of the complex user requirements, watermark for facsimile should be resilient to the errors resulted from scanning and transmission and tolerant to intensity and size adjustment but not intentional change of the text or characters.

As the watermarking and the signature extraction technique for content authentication have to work together with a secure camera, the security of these approaches depends on the security of the camera as well as the security of watermarking or the signature extraction. Security issues facing a secure camera are such as how easy it is to hack the camera such the identification information of a watermark or signature can be forged, how easy it is to disable the embedding of watermark or how easy it is to embed a valid watermark onto a manipulated image or extract a valid signature from it using the camera itself or by other means. In short, a secure camera has to ensure that a watermark is only embedded or a signature is only extracted at the every moment an image is captured. Whereas the security of a watermarking scheme concerns with how easy it is to illegally remove, copy, forge and embed a watermark and the security a signature extraction scheme concerns with the ease of forging a signature.

Unfortunately, until today, there is still not a fully secure watermarking scheme; the watermarking secret for ensuring the security of a watermarking scheme can be hacked given sufficient conditions such as when sufficient number of images with the same secret watermark key are available or the embedded watermark can be removed by exploiting the weak points of a watermarking scheme [22]. Nevertheless, digital watermarking or signature extraction scheme make an attempt to deceive by means of digital photomontage more difficult, as forgers would not only need to avoid the suspicion from human inspectors, they also have to make an additional effort to fool the digital watermark detector. Above all, the prospect for the secure camera to become a common consumer product is still fairly uncertain. Other general issues of watermarking techniques include the needs for the both watermark embedder and the watermark extractor to stick to a common algorithm and degradation of image quality as a result of watermarking.

Setting aside the active approach, Farid [23] proposed a passive and blind approach to detect the splicing of human speech. The proposed technique uses bicoherence features, i.e. the mean of the bicoherence magnitude and the variance of the bicoherence histogram to detect the increase in the level of quadratic phase coupling (QPC) induced by splicing together two segments of human speech signal. When there exist three harmonically related harmonic at frequencies ω_1, ω_2 and $\omega_1 + \omega_2$ in a signal, quadratic frequency coupling (QFC) is considered happening at bi-frequency (ω_1, ω_2) . When the phase the

three harmonically related frequencies happen to be ϕ_1, ϕ_2 and $\phi_1 + \phi_2$ respectively, QPC happens. Although QFC alone may be coincidental, the occurrence of QPC is a strong indication for a quadratic non-linearity. The detectability of the increase in QPC induced by splicing relies significantly on the premise that human speech signal is originally low on QPC. In [24], Fackrell et al. empirically shows that the vowels, nasals as well as the voiced and the unvoiced fricatives, among the human speech sounds, do not exhibit significant level of QPC. However, in [25], Nemer et al. argued otherwise. In [25], bispectrum of speech linear predictive coding residual of short-term speech is used distinguish between speech activity and silence, and it is found that bispectrum of human speech is sufficiently distinct from the zero bispectrum of Gaussian noise.

It is easy to see that QPC arises when $x(t) = \cos(\omega_1 t + \phi_1) + \cos(\omega_2 t + \phi_2)$ goes through a linear-quadratic operation as shown in Figure 2, where the output is given by:

$$y(t) = \frac{1}{2} \cos(2\omega_1 t + 2\phi_1) + \frac{1}{2} \cos(2\omega_2 t + 2\phi_2) + \cos((\omega_1 + \omega_2)t + (\phi_1 + \phi_2)) \\ + \cos((\omega_1 - \omega_2)t + (\phi_1 - \phi_2)) + \cos(\omega_1 t + \phi_1) + \cos(\omega_2 t + \phi_2) + 1$$

It can be seen that there exists $\cos(\omega_1 t + \phi_1)$, $\cos(\omega_2 t + \phi_2)$ and $\cos((\omega_1 + \omega_2)t + (\phi_1 + \phi_2))$ which results in QPC at (ω_1, ω_2) .

In [23], it is argued that human speech signal splicing operation (which is followed by a smoothing post-processing) is a non-linear operation which contains the effect of a linear-quadratic operation, because a non-linear function contains a quadratic term in its Taylor series expansion.

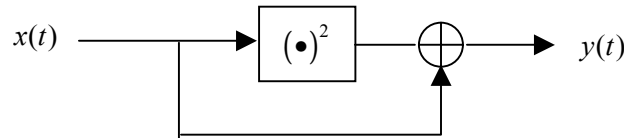


Figure 2: A linear-quadratic operation block diagram

Unlike human speech signal, the premise of low-valued bispectrum would not apply to image signal, which is often made up of edges and corners. In fact, it has been demonstrated that the two-dimensional bispectrum, $b(\omega_{x1}, \omega_{y1} ; \omega_{x2}, \omega_{y2})$, of natural images show a concentration of energy around regions where frequencies are aligned according to Equation (1.1) [26].

$$\frac{\omega_{x1}}{\omega_{y1}} = \frac{\omega_{x2}}{\omega_{y2}} \quad (1.1)$$

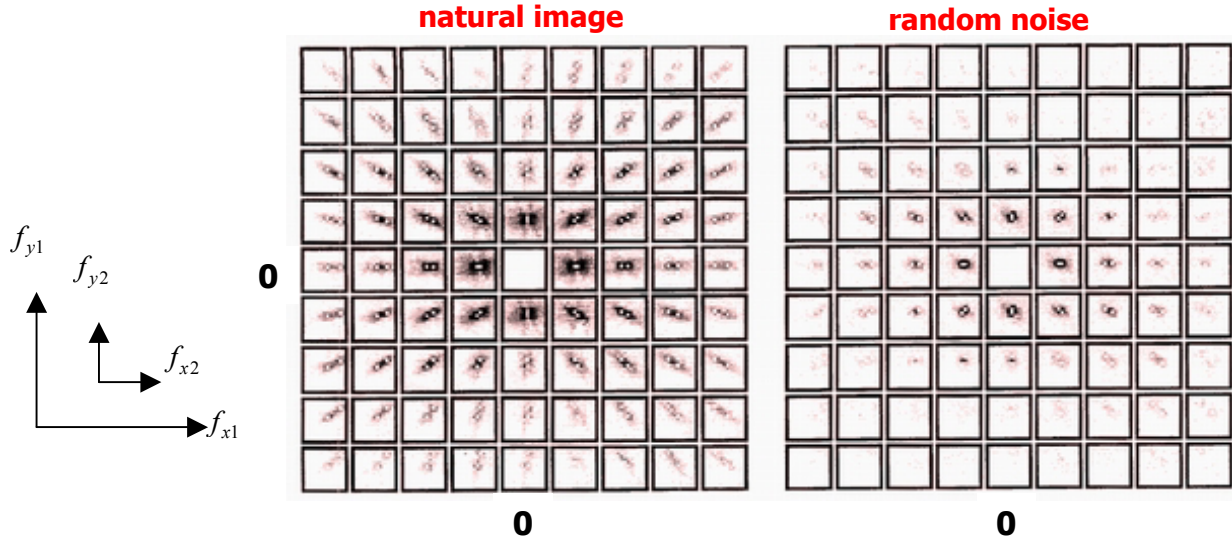


Figure 3: A side-by-side comparison for the magnitude of bispectrum for a natural image and that of a random noise. (source: extracted from [26])

It is also shown in [27] that, the concentration of energy in this particular regions is due to the intrinsically 0-D (i0-D) and 1-D (i1-D) local image features. i0-D local image features are the constant-valued patches and i1-D local image features are image features which can be described using a function of one variable (e.g. straight edges). As an aside, it is also shown that i0-D and i1-D local image features are the most common image features in natural images.

The relationship between the concentrations of bispectrum energy in natural images and the i1-D image features can be intuitively explained as follows:

A local image region with parallel straight lines oriented at an angle θ with respect to the horizontal axis can be described by a plane function as follows:

$$u(x, y) = \psi(x \sin \theta - y \cos \theta) \quad (1.2)$$

The Fourier transform of equation (1.2) is:

$$U(\omega_x, \omega_y) = \Phi(\omega_x \sin \theta - \omega_y \cos \theta) \cdot \delta(\omega_x \cos \theta + \omega_y \sin \theta)$$

where $\delta(\cdot)$ is a delta function. Hence, the non-zero frequency components are oriented at $90^\circ - \theta$ in the Fourier frequency domain, i.e. $\omega_x / \omega_y = \tan \theta$. Hence, quadratic frequency coupling (i.e., the occurrence of high Fourier frequency component at

ω_x , ω_y and $\omega_x + \omega_y$) due to the i0-D image features is only possible for frequencies along this orientation, i.e.,

$$\omega_{x1} / \omega_{y1} = \omega_{x2} / \omega_{y2} = (\omega_{x1} + \omega_{x2}) / (\omega_{y1} + \omega_{y2}) = \tan \theta .$$

Krieger's finding about the concentration of energy (magnitude) for the bispectrum of natural images directly implies that the bicoherence magnitude feature (i.e., the mean of the bicoherence magnitude) used by [23] would not be originally low for authentic images, as under the phase randomization assumption, the magnitude of bispectrum is an indication of QPC [28, 29] and the magnitude of bicoherence is a good estimator for the ratio of QPC energy [28]. As the level of the bispectrum energy is image feature dependent, detecting the increase of the magnitude feature value resulted by image splicing would be likened to a detection problem in low signal-to-noise environment, if the increase of magnitude feature value were relatively small compared to the original variance of the bicoherence features.

The assumption of phase randomization across the data segments refers to that each data segment can be considered as an independent phase realization when estimating a bispectrum through segment averaging. With phase randomization assumption, the magnitude of bispectrum could be high even in the absence of QPC; this could happen where is a deterministic coherent relationship between the phases of the harmonically related frequencies. In the case of coherent phase, the coherent-phased summation of the bispectrum across the data segment would amount to a large value without QPC actually happening. Table 1 shows the subtle distinction between the phase relationship of QPC and that of coherent-phased harmonics.

Table 1: Distinction between QPC with phase randomization and coherent-phased harmonics

QPC with phase randomization	Coherent phase but without QPC
$(f_1, \phi_1) \phi_1 \sim U(0, 2\pi)$	$(f_1, \phi_1) \phi_1 = c_1 + r$, and $r \sim U(0, 2\pi)$
$(f_2, \phi_2) \phi_2 \sim U(0, 2\pi)$	$(f_2, \phi_2) \phi_2 = c_2 + r$
$(f_3, \phi_3) f_3 = f_1 + f_2, \phi_3 = \phi_1 + \phi_2$	$(f_3, \phi_3) f_3 = f_1 + f_2, \phi_3 = c_3 + r$
U represents an uniform distribution	c_1, c_2 and c_3 is constant
	U represents an uniform distribution
$\phi_1 + \phi_2 - \phi_3 = 0$	$\phi_1 + \phi_2 - \phi_3 = c_1 + c_2 - c_3 = \text{constant}$

Since the phase of an image in Fourier domain is related to the location of edges [30] and, if the location of edges within an image segment are assumed to be random, then the phase randomization is valid for the estimation of an image bispectrum. In this case, the concentration of high energy for the bispectrum of natural images also implies that natural images originally have non-zero and high QPC. Hence, the bicoherence phase feature used by [23] (i.e., the variance of the bicoherence phase histogram) would face the same challenge as the above-mentioned bicoherence magnitude feature.

Recently, Farid [31] reported his recent system for detecting image manipulation based on a statistical model for natural images in wavelets domain. Manipulated images can be identified as it deviates from the eight defined statistical properties of natural images, which have been identified to be consistent across most natural images. One of the statistical properties mentioned in the report is the ratio of the number of the high-value wavelet coefficients between wavelet subbands of different scales. The system was claimed to be able to detect six types of tampering, i.e., splicing, resizing, printing and rescanning, double compression, artificial graphics and steganography. However, not much detail about the technical approaches and their performance was provided in the article. From the description of the article, it could be that it is the artifact from the post-processing operation such as airbrushing that follows the image splicing operation that the system is detecting and such artifact is considered as an indirect evidence of image splicing. However, such artifact does not necessarily imply image splicing. In contrast to this, our work described in this report would detect the direct artifacts given rise by image splicing. As photomontaging may not be followed by post-processing operation such as airbrushing, direct detection of image splicing is important.

1.2 The Important Role of Image Splicing

Creation of photomontages always involves image splicing, although additional tools can sometimes be applied to enhance the visual naturalness of the photomontages and most importantly to remove the rough edges of the cutout caused by unskillful cut-and-paste operation. Image splicing herein refers to a simple putting together of separate image regions, be they from the same or different images, without further post-processing steps. In a sense, image splicing can be considered the simplest form of photomontaging.

Common post-processing operations following image splicing are feathering effects on spliced edges (for creating a smooth blur edges) or airbrush style (with soft and smooth effect) erasing of rough object edges. These post-processing of an image will leave behind different traces, which can be detected using different techniques. For instance, the system of Farid [31] was reported to be able to detect the airbrush style effect and the detection of this artifact can be considered an indirect evidence of image splicing.

In contrary to the general belief, a good-quality photomontage can in fact be obtained by mere image splicing when the cut-and-paste operation is skillfully and carefully performed. The following figure shows examples of photomontages produced by a professional graphic designer with pure image splicing:

1.3 Outline of the Report

In this report, we intend to study the feasibility of applying higher order statistics techniques for image splicing detection and provide a model of image splicing to explain the response of bicoherence (i.e., a normalized version of the Fourier transform of a third-order moment known as bispectrum) magnitude and phase features to image splicing. Apart from image splicing being a basic operation in creating photomontage, another

main reason why we focus on the problem of image-splicing detection is based on our observation in a preliminary experiment described in section 1.4.



Figure 4: deceptively authentic looking photomontages with mere image splicing: (left) the golfer is a spliced object (right) the white truck is an spliced object

In Section 2, we will describe a data set consists of authentic and spliced image blocks, which will be used for all the experiments described in this report. Then, in Section 3, we will move on to provide an introduction to bicoherence and describes the two bicoherence features used in [23]. In Section 4, we describe a signal model from image splicing and investigate the effect of image splicing on the bicoherence features in accordance to the proposed image-splicing model. Then, we would show some empirical evidence supporting the proposed model.

From the experiment results described in Section 5, we will show some properties of the plain/baseline bicoherence features for different types of image blocks. The magnitude (i.e., the mean of the bicoherence magnitude) and the phase (the negative entropy of bicoherence phase histogram) features similar to that used by Farid [23] are considered as the plain/baseline features. Since the two baseline bicoherence features would not perform well for image splicing detection, we are motivated to propose two general methods (i.e., characterizing the image properties and the splicing-invariant features) to improve the performance of the bicoherence features in Section 6. As a result of the proposed methods, the new features are derived; those are the prediction residual for the plain bicoherence magnitude and phase features, and the edge percentage features. Finally, in Section 7, we evaluate the features derived from the proposed methods using SVM classification experiments. When combining the new features with the plain bicoherence features, the classification accuracy improves from the 62% obtained by the plain bicoherence features to 72%.

1.4 Related Preliminary Work

In a separate experiment (detailed description can be found in the Appendix of this report), we have used the higher-order statistics (HOS) (i.e., mean, variance, skewness and kurtosis) of both the wavelet coefficients and the linear prediction error of the wavelet coefficients [32] as features for distinguishing various types of image operations. Farid has used the wavelet features for classification of natural images versus steg image,

i.e., images with a hidden message (98.9% of the natural and 97.6% of the steg images are correctly classified), natural images versus photorealistic 3D computer graphics (99.5% of the natural and 36.9% of the computer graphics images are correctly classified), and natural images versus print-and-scanned images (99.5% of the natural and 99.8% of the print-and-scanned images are correctly classified). High classification accuracy is obtained except for the natural images versus computer graphics classification. However, our preliminary work of classification of a more comprehensive set of image operations, as listed below, using the wavelet features has not been reported before.

- Low pass (lp) - Gaussian low pass with kernel support size 5x5 and standard deviation 8
- JPEG (jpeg) – JPEG compression with quality factor 60
- Additive noise (noise) – Gaussian noise with mean 0 and standard deviation 18
- High pass (hp) – Gaussian high pass with cutoff frequency at 0.025 times sampling frequency, compensated with the original image
- Histogram equalization (histeq) – uniform histogram with 64 bins
- Median filtering (medfilt) – with filter kernel size 9x9
- Wiener filtering (wiener) – with filter kernel size 9x9
- Brightening (bright) – with 0.6 gamma pixel intensity mapping
- Simulated Image splicing (splicing) – swapping the upper-right quadrant and the lower-left quadrant.

Images used are of approximated size of 512x768 pixels. Classification of image operations are performed using Support Vector Machine (SVM). The classification performance of all operations are reasonably well except that the simulated cropping is doing just as good as random guessing. This shows that the wavelet-based multi-resolution features are inadequate for blind detection of image splicing. The failure of the wavelet features in detecting image splicing has motivated us to extend our effort in investigating other ways of detecting image splicing.

2 Data Set Description

We can imagine that every spliced image has a correspondent authentic counterpart, i.e. an image that is similar to the spliced image except that it is authentic or produced by a camera. If the authentic counterpart of spliced images does exist, the ideal image data set for experiments on detecting image splicing would be one, which comprises a set of spliced images and their authentic counterpart. However, constructing such ideal data set is highly difficult if not impossible. Therefore, instead of trying to construct such an ideal image data set, we try to populate our image data set with samples of diverse properties in terms of the orientation of splicing interface (i.e., vertical, horizontal or diagonal), the type of splicing (i.e., straight or arbitrary boundary splicing) and different properties of splicing regions (i.e., smooth or textured regions).


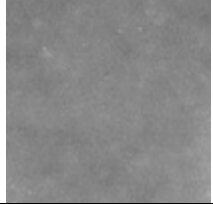
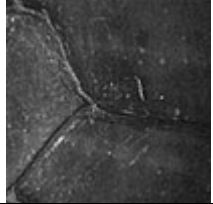







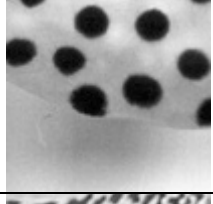
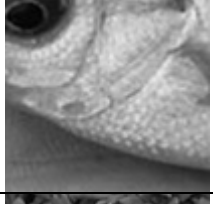

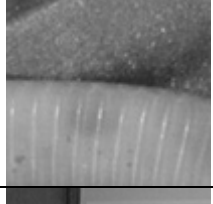


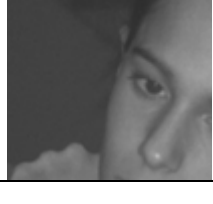

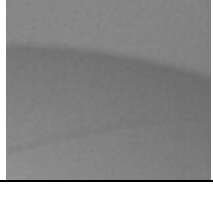
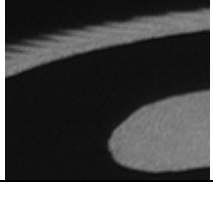

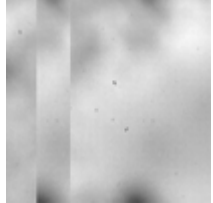
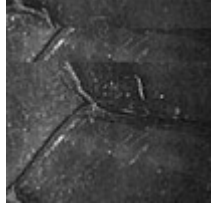

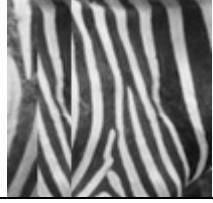

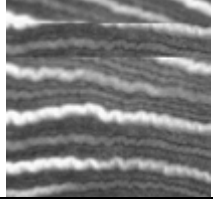

The data set has 933 authentic and 912 spliced image blocks of size 128 x 128 pixels. The image blocks are extracted from images in CalPhotos image set [33]. As the images are contributions from photographers, we assume that they can be considered as authentic i.e., not digital photomontages.

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, splicing boundary is either straight or according to arbitrary object boundary. The image block with arbitrary object boundaries 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 block.

In actual case, image splicing does not always introduce object boundary. For instance, when a forger wants to remove a region corresponding to an object from an image, he or she may attempt to fill up the removed region using patches similar to the background. Therefore, image splicing could take place at homogenous textured or smooth region too. This also forms the rationale for our technique of producing the spliced subcategories of an entirely homogeneous textured or smooth region.

Figure 5 shows the typical image blocks in each subcategory of the data set and Table 2 shows the number of image blocks within each subcategory.

Authentic Category				
Homogenous Smooth				
Homogenous Textured				
Textured- Smooth				
Textured- textured				
Smooth-smooth				
Spliced Category				
Homogenous Smooth				
Homogenous Textured				

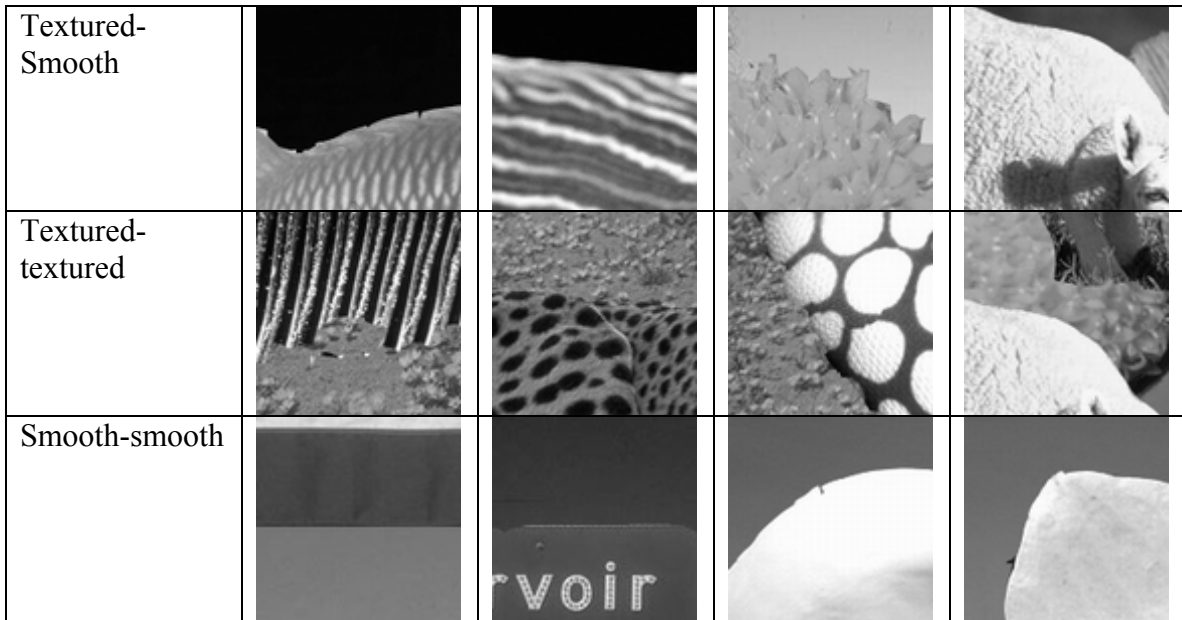


Figure 5 Typical images in the data set

Table 2 : The number of image blocks in each subcategory of the data set

Category	One Textured Background	One Smooth Background	Textured-Smooth Interface	Textured-texture Interface	Smooth-smooth Interface	Total
Authentic	126	54	409	179	165	933
Spliced	126	54	298	287	147	912

More details about the dataset can be found in [34]

3 Bicoherence

In this section, we will give a brief introduction of bicoherence and illustrate the way we estimate the 1-dimensional bicoherence from a 2-dimensional image.

3.1 Introduction to Bicoherence

Bispectrum is defined as the Fourier transform of the third order moment of a signal $x(t)$ and can be expressed as Equation (1.3) as the expected third-order or quadratic correlation of three harmonics from the Fourier transform of the signal, $X(\omega)$, at ω_1 , ω_2 and $\omega_1 + \omega_2$.

$$BIS(\omega_1, \omega_2) = E[X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2)] = |BIS(\omega_1, \omega_2)|e^{j\Phi(BIS(\omega_1, \omega_2))} \quad (1.3)$$

Bispectrum is often used for detecting the existence of the quadratic correlation within a signal, as being applied in oceanography [35], EEG signal analysis [36], manufacturing [37], non-destructive structural fatigue detection [38] and plasma physics [39] applications.

Whereas bicoherence is the normalized version of bispectrum and defined a below [39]:

Definition 1 (Bicoherence) *The bicoherence of a signal $x(t)$ with its Fourier transform being $X(\omega)$ is given by [39]:*

$$b(\omega_1, \omega_2) = \frac{E[X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2)]}{\sqrt{E[|X(\omega_1)X(\omega_2)|^2]E[|X(\omega_1 + \omega_2)|^2]}} = |b(\omega_1, \omega_2)|e^{j\Phi(b(\omega_1, \omega_2))} \quad (1.4)$$

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 a zero phase, we call such phenomena *quadratic phase coupling* (QPC). As such, the average bicoherence magnitude would increase as the amount of QPC grows. Besides that, signals satisfying the gaussianity property can be proved to have low bicoherence and thus bispectrum is often used as a measure of signal non-gaussianity [40].

The expression for bicoherence in Equation ((1.4)) is obtained by normalizing bispectrum with the Cauchy-Schwartz inequality upper bound on the magnitude of bispectrum; hence, its absolute value is bounded between 0 and 1. The Cauchy-Schwartz upper bound is achieved when the harmonically related frequencies in the numerator is perfectly phase coupled. In this case the magnitude of bicoherence is one and its phase, being $\phi_1 + \phi_2 - (\phi_1 + \phi_2)$, is zero. In general, for a signal

$$x(t) = \cos(\omega_1 t + \phi_1) + \cos(\omega_2 t + \phi_2) + A_C \cos((\omega_1 + \omega_2)t + (\phi_1 + \phi_2)) + A_{UC} \cos((\omega_1 + \omega_2)t + \phi_3)$$

where ϕ_1 , ϕ_2 and ϕ_3 are random and uncoupled. The squared magnitude of the bicoherence, $|b(\omega_1, \omega_2)|^2$, is a good measure for the fraction of QPC energy, i.e. $A_C^2 / (A_C^2 + A_{UC}^2)$. However, in this case, the phase of the bicoherence is still zero.

3.2 Estimation of Bicoherence Features

With limited data sample size, instead of computing 2-dimensional bicoherence features from an image, we compute 1-dimensional bicoherence magnitude and phase features (will be described in section 4.3) from N_v vertical and N_h horizontal image slices of an image and then combined as in equations (1.5) and (1.6). For the image blocks of 128×128 pixels in our data set, $N_v = N_h = 128$.

$$fM = \sqrt{\left(\frac{1}{N_h} \sum_i M_i^{Horizontal}\right)^2 + \left(\frac{1}{N_v} \sum_i M_i^{Vertical}\right)^2} \quad (1.5)$$

$$fP = \sqrt{\left(\frac{1}{N_h} \sum_i P_i^{Horizontal}\right)^2 + \left(\frac{1}{N_v} \sum_i P_i^{Vertical}\right)^2} \quad (1.6)$$

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

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

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. [38], 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 128×128 -pixel image block.

4 A Model for Image Splicing

In this section, we would like to propose a way to model image splicing based on the idea of signal perturbation by delta functions. The rationale of the proposed model will be explained. Before going into image-splicing model, let's us look at a related prior work on human speech splicing

4.1 Prior Work on Human Speech Splicing

In [23], bicoherence magnitude and phase features are applied for detecting human speech splicing and the approach is justified with the following arguments:

1. Human speech signal is originally weak in higher order correlation, reflecting on the low value of the bicoherence magnitude feature and a rather randomly distributed bicoherence phase.
2. A quadratic operation, by inducing a *Quadratic Phase Coupling* (QPC) phenomenon, increases the bicoherence feature values due to the quadratic harmonic relation and the 0° phase bias. A general non-linear operation, when expressed by a Taylor expansion, has a partial sum of low-order terms resembling a quadratic operation.

However, the arguments could not justify the use of bicoherence features for image splicing detection:

1. Image signal may not be originally weak in higher order correlation as demonstrated in [26]. With the assumption of phase randomization across data segments used for the estimation of bispectrum, bispectrum energy (magnitude) is an indication of the QPC. As we can consider bicoherence as a measure of QPC, this empirical observation implies that the detection of image splicing through the detection of the increase in the plain bicoherence magnitude and phase feature value would face a high level of noise if the increase of the feature value is relatively small.
2. In [23], detection of a cascaded splicing and smoothing (using a Laplacian pyramid) operation on fractal signal and human speech signal is demonstrated. However, a splicing operation, if were to be considered a function, is potentially discontinuous and has no Taylor expansion. Thus, the aforementioned argument about QPC cannot be applied. Besides that, the effect on the bicoherence features due to splicing is still unknown.

4.2 Image-Splicing Model

When an image is directly acquired by an image acquisition device, such as camera or scanner, we call this image an authentic image. Then, when an image created as a composite of multiple image regions, it is not considered as an authentic image. In the following discussion, we consider image splicing to be an act of putting together different image regions without other post-processing such as airbrush style softening of edges at the splicing interface.

While images are not just a collection of random pixels, neither is image splicing in real life likely to be just putting together of two or more random patches from different or the

same images. Spliced image is usually done with a purpose, such as for forging court evidence, for changing a figure in a digital check (scanned version of a real check) or for creating an image of a scene setting, which is dangerous or expensive to be set up physically. In all these cases, the resulting spliced image would highly resemble an authentic image if there is one truly exists. Therefore, we can imagine that there exists an image, which is visually similar to a spliced image, but is in fact an authentic image, hence without splicing effect. We call this possibly hypothetical image an *authentic counterpart* of a spliced image. Even for a spliced image, which looks totally unreal to human due to internal scene inconsistencies within the image, it is still in practice possible to find an authentic counterpart for such weird looking image. A way to find one that is close to the truly authentic counterpart is to recapture the resulting photomontage by a camera or a scanner after it has been printed out. Therefore the essential difference between a spliced image and its authentic counterpart lies in the fact that the former goes through a natural imaging process while the later does not. Figure 6 illustrates the concept of an authentic counterpart for a spliced image

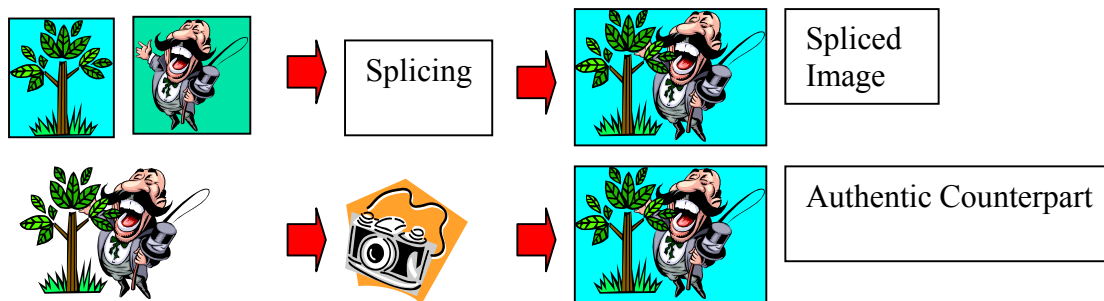


Figure 6 Illustration for the Concept of an Authentic Counterpart for a spliced image

As we compare a spliced image with its authentic counterpart, the most drastic difference between them is at the interface of the splicing regions. In this case, all the regions of the pair are authentic per se although they are from different sources. The interface may or may not correspond to an image edge, as image splicing can be done in such a way that it is meant to remove an object from a background and patching the removed region with patches similar to the background, rather than adding an image object to a background. However, such splicing interface may correspond to some form of discontinuity.

4.2.1 A general model

A general model for image splicing would be applying a pair of complementary window functions on a same or two different images for masking out the interested regions and then superpose them to form a composite image, as illustrated in Figure 7.

This model can be used to analyze the effect of combining regions from different sources as well as the effect of splicing at the splicing interface. In this report, we consider a simpler model, which only consider the effect of splicing at the splicing interface.

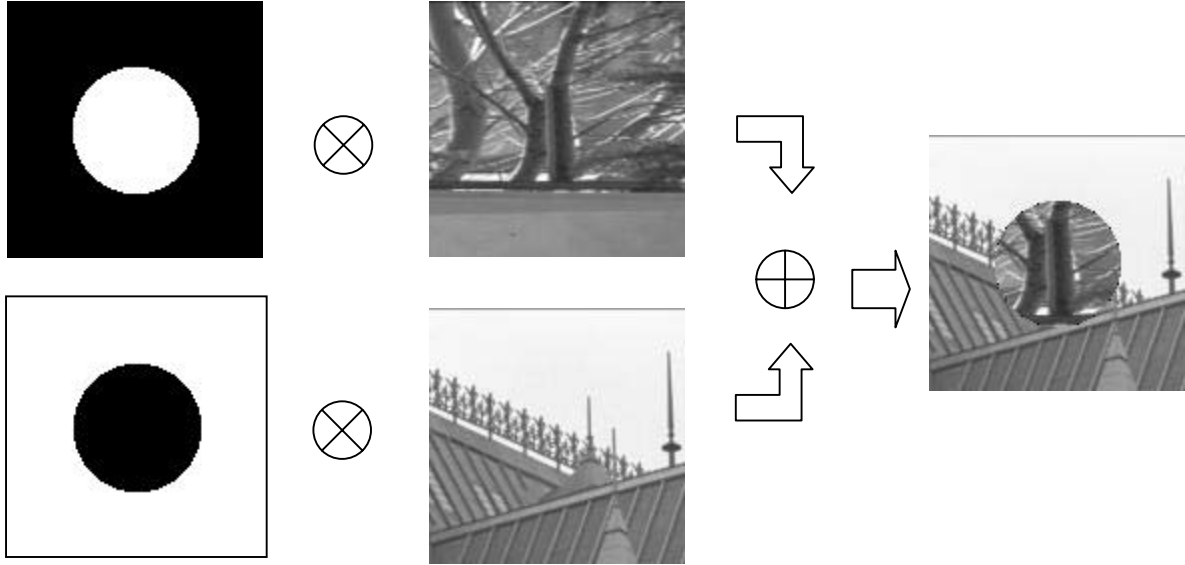


Figure 7 Illustration of a general splicing model

4.2.2 A Simpler Model

Although image splicing is performed with 2-dimensional regions, we detect the splicing through computing the bicoherence features of the vertical and horizontal 1-dimensional slices of a spliced image. Therefore, in this paper, we propose a 1-dimensional model for splicing, which is also applicable to the splicing of any 1-dimensional signal. Here, we consider splicing as a joining of signals without any post-processing of the spliced signal.

A composite signal, due to the splicing of two signal segments, is very likely to introduce a discontinuity or an abrupt change at the splicing point. The lack of smoothness can be thought of a departure from a smooth signal due to a perturbation of a bipolar signal (Figure 8), which is coincidentally similar to Haar high pass basis.

As almost every camera is equipped with an optical low pass filter and almost every scanner has a post-scanning low pass operation for avoiding aliasing effect which produces Moiré pattern, authentic images, being a direct output from image acquisition devices such as camera and scanner, could be modeled as a ‘smooth’ signal. With the idea of the authentic counterpart (i.e., a possibly hypothetical but authentic image that resembles the spliced image in every respect except for those properties induced by splicing), we can model image splicing as a perturbation of the authentic counterpart with a bipolar signal. Figure 8 illustrate the idea of bipolar signal perturbation.

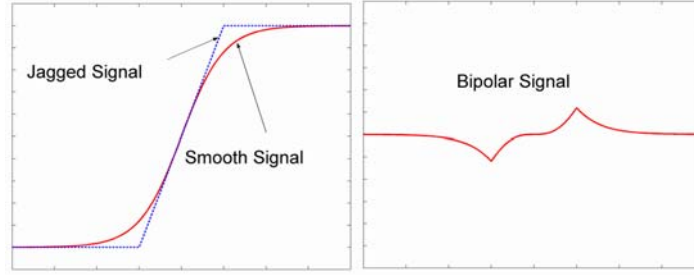


Figure 8 (Left) a jagged signal exhibits abrupt changes when compared to a smooth signal. (Right) the difference between the jagged and the smooth signals

Definition 2 (Bipolar signal) A bipolar at location x_o with the antipodal delta separated by Δ , and with k_1 and k_2 being of opposite sign, i.e., $k_1 k_2 < 0$, is represented as

$$d(x) = k_1 \delta(x - x_o) + k_2 \delta(x - x_o - \Delta), \delta(\cdot) \text{ being a delta function}$$

and its Fourier Transform is $D(\omega) = k_1 e^{-jx_o \omega} + k_2 e^{-j(x_o + \Delta)\omega}$

4.3 Definition for Bicoherence Features

Definition 3 (Bicoherence Phase Histogram) An N -bin bicoherence phase histogram given by:

$$p(\Psi_n) = (1/M^2) \sum_{\Omega} 1(\Phi(b(\omega_1, \omega_2)) \in \Psi_n), 1(\cdot) = \text{indicator function}$$

where

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1 = (2\pi m_1)/M, \omega_2 = (2\pi m_2)/M, m_1, m_2 = 0, \dots, M-1\}$$

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

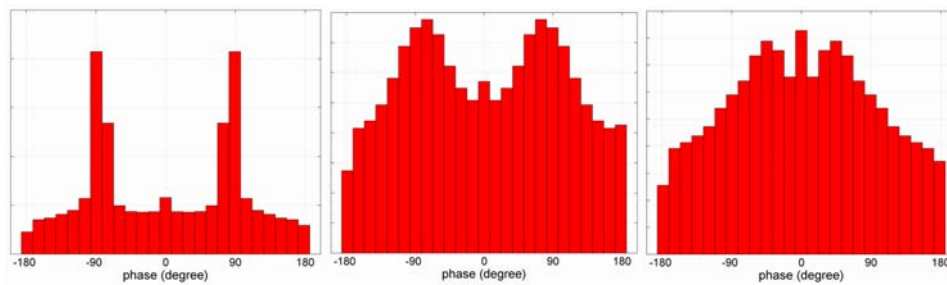


Figure 9 Typical examples of bicoherence phase histogram from spliced images: (Left) Strong $\pm 90^\circ$ phase bias (Middle) near $\pm 90^\circ$ phase bias (Right) non $\pm 90^\circ$ phase bias

Definition 4 (Bicoherence Magnitude Feature) The magnitude feature is the mean of the magnitude of the bicoherence:

$$f_M = (1/M^2) \sum_{\Omega} |b(\omega_1, \omega_2)|$$

Definition 5 (Bicoherence Phase Feature) *The phase feature which measures the non-uniformity or the bias of the bicoherence phase histogram:*

$$f_P = \sum_n p(\Psi_n) \log p(\Psi_n)$$

4.4 Response of Bicoherence Phase Feature

Proposition 1 (Symmetry of Bicoherence Phase Histogram) *For a real-valued signal, the N -bin bicoherence phase histogram is symmetrical: $p(\Psi_n) = p(\Psi_{-n})$ for all n*

Proof The Fourier transform of a real-valued signal is conjugate symmetric, i.e., $X(\omega) = X^*(-\omega)$, hence, from Definition 1, its bicoherence is also conjugate symmetric, i.e., $b(\omega_1, \omega_2) = b^*(-\omega_1, -\omega_2)$. Therefore, its bicoherence phase histogram is symmetrical.

Note that, from Proposition 1, it suffices to study the positive half of the bicoherence phase histogram (i.e. from 0° to 180°). Besides, as the phase of a bicoherence is equal to the phase of numerator of the expression in Equation (1.4), therefore, it suffices to examine the numerator as far as the response of the bicoherence phase feature is concerned.

Proposition 2 (Phase of Bipolar Signal Bicoherence) *Assuming $k_1 = -k_2 = k$, the phase of the bicoherence for a bipolar signal is concentrated at $\pm 90^\circ$.*

Proof When $k_1 = -k_2 = k$, the third-order moment of $D(\omega)$:

$$\begin{aligned} & D(\omega_1)D(\omega_2)D^*(\omega_1 + \omega_2) \\ &= k^3 - k^3 - k^3 e^{-j\Delta\omega_1} + k^3 e^{j\Delta\omega_1} - k^3 e^{-j\Delta\omega_2} + k^3 e^{j\Delta\omega_2} - k^3 e^{j\Delta(\omega_1 + \omega_2)} + k^3 e^{-j\Delta(\omega_1 + \omega_2)} \\ &= k^3 (e^{j\Delta\omega_1} - e^{-j\Delta\omega_1}) + k^3 (e^{j\Delta\omega_2} - e^{-j\Delta\omega_2}) - k^3 (e^{j\Delta(\omega_1 + \omega_2)} - e^{-j\Delta(\omega_1 + \omega_2)}) \\ &= 2k^3 j[\sin(\Delta\omega_1) + \sin(\Delta\omega_2) - \sin(\Delta(\omega_1 + \omega_2))] \end{aligned} \quad (1.7)$$

The term k^3 in Equation (1.7) is a real positive or negative number, independent of frequency, while the sign of the real-value term $[\sin(\Delta\omega_1) + \sin(\Delta\omega_2) - \sin(\Delta(\omega_1 + \omega_2))]$ is frequency-dependent, as shown in Figure 10. Hence, $D(\omega_1)D(\omega_2)D^*(\omega_1 + \omega_2)$ is an imaginary number with the a frequency-dependent positive or negative sign. The expectation of an imaginary random number is still an imaginary number with a phase at $\pm 90^\circ$.

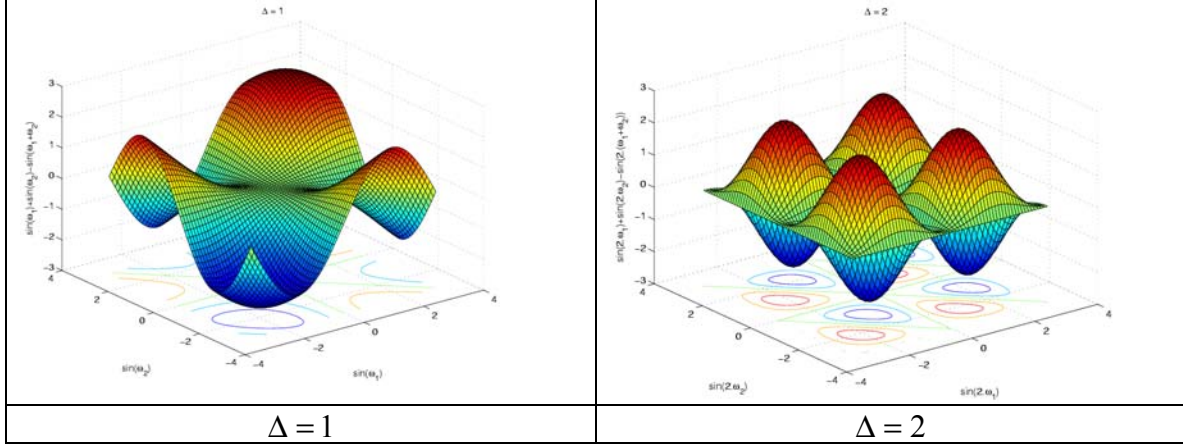


Figure 10: plot for $[\sin(\Delta\omega_1) + \sin(\Delta\omega_2) - \sin(\Delta(\omega_1 + \omega_2))]$

It is interesting to observe that the phase bias at $\pm 90^\circ$ is not due to QPC which could in turn give rise to 0° phase bias. QPC is due to the existence of harmonics with the same frequency and phase relationship, e.g., when there exists harmonics at ω_1 , ω_2 and $\omega_1 + \omega_2$ for the Fourier transform of a signal $S(\omega)$, the phase of the harmonics are ϕ_1 , ϕ_2 and $\phi_1 + \phi_2$ respectively, hence, $phase[S(\omega_1)S(\omega_2)] = phase[S(\omega_1 + \omega_2)]$. However, for the bipolar signal, the phase relationship is given by:

$$phase[D(\omega_1)D(\omega_2)] = phase[D(\omega_1 + \omega_2)] \pm \pi/2.$$

On the other hand, when $k_1 = k$ and $k_2 = -k + \varepsilon < 0$,

$$D(\omega_1)D(\omega_2)D^*(\omega_1 + \omega_2) = \varepsilon(3k^2 - 3k\varepsilon + \varepsilon^2) + k\varepsilon(\varepsilon - k)[\exp(j\Delta\omega_1) + \exp(-j\Delta\omega_2) + \exp(j\Delta(\omega_1 + \omega_2))] + 2k^2(k - \varepsilon)j[\sin\Delta\omega_1 + \sin\Delta\omega_2 - \sin\Delta(\omega_1 + \omega_2)] \quad (1.8)$$

Therefore, if ε is small relative to k , in equation (1.8), the last term becomes dominant and the phase remains concentrated around $\pm 90^\circ$. In other words, if the magnitudes of the opposite poles of the bipolar are approximately equal, the $\pm 90^\circ$ phase concentration occurs.

Before moving on to Proposition 3, please note that, as mentioned earlier, bicoherence is, in practice, computed by:

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

where the expectation terms are estimated by the average terms with a *set of signal segments* from the target 1-D signal [41].

When a signal $s(x)$, is perturbed by a bipolar signal, $d(x)$, the resulting perturbed signal and its Fourier transform is given by:

$$s_p(x) = s(x) + d(x) \leftrightarrow S_p(\omega) = S(\omega) + D(\omega)$$

Proposition 3 (Response of Bicoherence Phase Feature) *Perturbation of a signal with a bipolar contributes to a phase bias at $\pm 90^\circ$. The strength of the overall contribution is dependent on (1) the magnitude of the bipolar (2) the percentage of bipolar perturbed segments within the set of signal segments used for computing the bicoherence by averaging.*

Proof For simplicity, assume that $k_1 = -k_2 = k$ for the magnitude of the bipolar, the correlation of the Fourier transform of the perturbed signal is given by:

$$S_p(\omega_1)S_p(\omega_2)S_p^*(\omega_1+\omega_2) = S(\omega_1)S(\omega_2)S^*(\omega_1+\omega_2) + \text{cross terms} + 2k^3j[\sin\Delta\omega_1 + \sin\Delta\omega_2 - \sin\Delta(\omega_1 + \omega_2)] \quad (1.9)$$

where

$$\begin{aligned} \text{cross terms} = & kS(\omega_1) S^*(\omega_1 + \omega_2) \exp(-jx_o\omega_2)[1 - \exp(-j\Delta\omega_2)] + \\ & k S(\omega_2) S^*(\omega_1 + \omega_2) \exp(-jx_o\omega_1)[1 - \exp(-j\Delta\omega_1)] + k^2 S^*(\omega_1 + \omega_2) \\ & \exp(-jx_o(\omega_1 + \omega_2)) [1 - \exp(-j\Delta\omega_1)] [1 - \exp(-j\Delta\omega_2)] + \\ & k S(\omega_1) S(\omega_2) \exp(jx_o(\omega_1 + \omega_2)) [1 - \exp(j\Delta(\omega_1 + \omega_2))] + \\ & k^2 S(\omega_1) \exp(jx_o\omega_1) [1 - \exp(-j\Delta\omega_2)] [1 - \exp(j\Delta(\omega_1 + \omega_2))] + \\ & k^2 S(\omega_2) \exp(jx_o\omega_2) [1 - \exp(-j\Delta\omega_1)] [1 - \exp(j\Delta(\omega_1 + \omega_2))] \end{aligned}$$

We can see that the imaginary term due to bipolar perturbation contribute consistently at every (ω_1, ω_2) to the imaginary component of Equation (1.9). The strength of the contribution depends on k . The same argument is applicable to the case when $k_1 = k$ and $k_2 = -k + \varepsilon < 0$ with ε being small relative to k , but the strength of the contribution is lessened.

As numerator of bicoherence expression is estimated by an average of the third-order moment for the Fourier transform of a signal over a set of signal segments, the percentage of bipolar perturbed segment within the set affects the contribution to $\pm 90^\circ$ phase bias. In actual case, we estimate the bicoherence of a 1-D image slice of length 128 pixels with 3 overlapping segments of length 64 pixels [41]. The overlap of segments ensures a larger extent of the perturbation effect, as a splicing point is likely to be captured by two adjacent segments with a probability of 0.5, assuming uniformly distributed splicing point.

4.5 Response of Bicoherence Magnitude Feature

Proposition 4 (Response of Bicoherence Magnitude Feature) *Perturbation of a signal with a bipolar signal contributes to an increase in the value of the bicoherence magnitude feature. The amount of the increase depends on (1) the magnitude of the bipolar relative to the mean of the magnitude of the original signal Fourier transform (2) the percentage of bipolar perturbed segments within the set of signal segments used for computing the bicoherence by averaging.*

Proof For simplicity, we analyze the perturbation with a bipolar having $k_1 = -k_2 = k$. Note that the sign of equation (1.7) at a particular (ω_1, ω_2) is determined by the separation (denoted by Δ) and the orientation (denoted by the sign of k) of the poles of a bipolar

signal. With the following assumptions on the bipolar across the ensemble of signal used for estimating bicoherence, $D(\omega_1)D(\omega_2)$ would be equal to $D(\omega_1+\omega_2)$ within a multiplicative constant for a particular (ω_1, ω_2) .

- The orientation of the bipolars is the same. (This assumption is reasonable as the same bipolar can be captured by two different but overlapping windows.)
- The pole separation for the bipolar is the same (This assumption is also valid because the bipolar introduced by splicing is compact at the splicing interface)
- The magnitude of bipolar is the same. (This is also a reasonable assumption for a local region)

When $D(\omega_1)D(\omega_2) = c(\omega_1, \omega_2)D(\omega_1+\omega_2)$ with $c(\omega_1, \omega_2)$ being a constant for a particular (ω_1, ω_2) , the magnitude of the bicoherence is 1 at every frequencies (ω_1, ω_2) , as, in this case, the numerator of attains the Cauchy-Schwartz inequality upper bound. When a signal $s(x)$ with Fourier transform $S(\omega)$ is perturbed by a bipolar, the magnitude of the bicoherence is given by:

$$|b(\omega_1, \omega_2)| = \frac{\left| E\left[k^3 \left[\frac{S(\omega_1)}{k} + G(\omega_1) \right] \cdot \left[\frac{S(\omega_2)}{k} + G(\omega_2) \right] \cdot \left[\frac{S^*(\omega_1 + \omega_2)}{k} + G^*(\omega_1 + \omega_2) \right] \right] \right|}{\sqrt{E\left[k^4 \left| \left[\frac{S(\omega_1)}{k} + G(\omega_1) \right] \cdot \left[\frac{S(\omega_2)}{k} + G(\omega_2) \right] \right|^2 \right] E\left[k^2 \left| \frac{S(\omega_1 + \omega_2)}{k} + G^*(\omega_1 + \omega_2) \right|^2 \right]}} \quad (1.10)$$

where $G(\omega) = \exp(-jx_o\omega) [1 - \exp(-j\Delta\omega)]$

From Markov inequality, the term $|S(\omega)/k|$ in Equation (1.10) is upper-bounded in probability by $P(|S(\omega)/k| \geq \varepsilon) \leq E[S(\omega)]/(k\varepsilon)$, for any all $\varepsilon > 0$. Hence, for an energy signal, i.e., signal with finite energy such as normal image signal, $\lim_{k \rightarrow \infty} P(|S(\omega)/k| \geq \varepsilon) = 0$, for $E[S(\omega)]$ being finite. As a result, the magnitude of bicoherence $|b(\omega_1, \omega_2)|$ in equation (1.10) satisfies:

$$\lim_{k \rightarrow \infty} P \left(\left| |b(\omega_1, \omega_2)| - \frac{|E[D(\omega_1)D(\omega_2)D^*(\omega_1 + \omega_2)]|}{\sqrt{E[|D(\omega_1)D(\omega_2)|^2]E[|D^*(\omega_1 + \omega_2)|^2]}} \right| \geq \varepsilon \right) = 0$$

With the above assumptions: $\lim_{k \rightarrow \infty} P(|b(\omega_1, \omega_2)| - 1| \geq \varepsilon) = 0$

Therefore, the more frequency triplets which have a small $E[S(\omega)]$ relatively to k , the greater the contribution of the bipolar perturbation to an increase in the bicoherence magnitude feature.

Similar to the bicoherence phase feature, the extent of bipolar perturbation in the ensemble of signal is another factor affecting the contribution of bipolar perturbation to an increase in bicoherence magnitude feature.

4.6 Empirical Validation for the Proposed Model

4.6.1 Observations on Bicoherence Features

With the data set described in Section 2, by examining the difference of the mean of the phase histogram (Definition 3) for the authentic and spliced categories of our data set (Figure 11), a clear statistical difference of phase bias for the two categories at $\pm 90^\circ$ is observed. This observation supports the theoretical prediction of the $\pm 90^\circ$ phase bias (Proposition 2) based on the proposed image-splicing model. Note that, from Proposition 1, it suffices to study the positive half of the bicoherence phase histogram (i.e. from 0° to 180°). Figure 9 shows two examples of $\pm 90^\circ$ phase bias from spliced images.

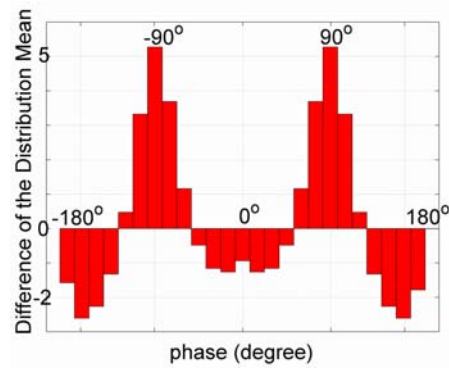


Figure 11 The mean of the authentic phase histogram minus the mean of the spliced phase histogram

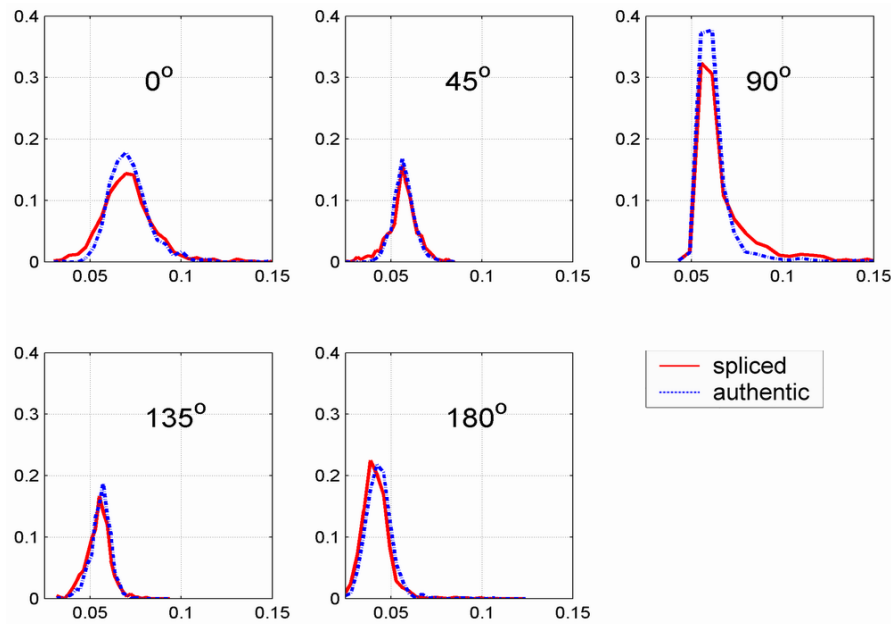


Figure 12 Distribution of the value of bicoherence phase histogram at a specific phase, ranging from 0° to 180° , for both the authentic and the spliced categories. (y-axis is sample count, x-axis is the value of bicoherence phase histogram at a specific phase)

Besides that, the distribution of the value of bicoherence phase histogram (Definition 3) at a specific phase for both the authentic and spliced categories in Figure 12 also shows a comparably greater difference of phase distribution between the authentic and spliced categories at 90° phase.

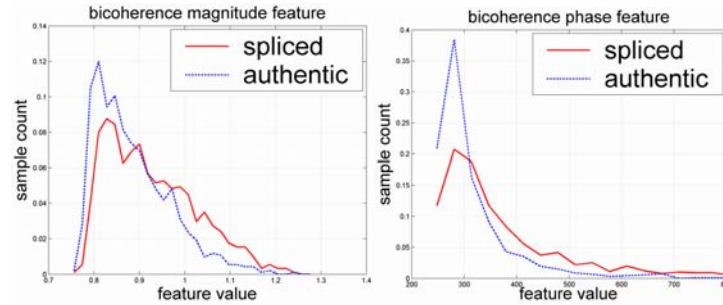


Figure 13 The histogram of the bicoherence features: (Left) magnitude feature (Right) phase feature

In addition, the histogram for bicoherence magnitude and phase features (Figure 13) for the spliced category is observed to have a larger mean and a heavier tail compared to that of the authentic category. This validates the Proposition 3 and Proposition 4.

4.6.2 90° Phase Bias as Prediction Feature

To evaluate the performance 90° phase bias as a feature for image splicing detection, we performed the same classification experiments as in [41] by replacing the negative phase entropy with the 90° phase bias, which is measured by the value of the bicoherence phase histogram at 90° . The results of detection accuracy over the same data set are comparable at about 70%. This indicates that the negative phase entropy, despite being a general measure of phase bias, has already captured the specific effect of 90° phase bias. The fact that the feature using the specific 90° phase bias fails to achieve noticeable improvement indicates the weakness of the phase bias effect, which is linked to the high estimation variance that commonly plagues the estimation of higher order statistics such as bicoherence.

5 Properties of Bicoherence Features on Different Image Features

In this section, we would like to explore the properties of bicoherence features in relation to various image features. We will first relate the bicoherence magnitude to the edge configuration or the texturedness of an image. Then, we will look at the empirical properties of the bicoherence features on three different types of object interface, i.e. smooth-smooth, textured-textured, and smooth-textured object interface, which are among the subcategories in the data set.

5.1 Relationship between Edge Configuration and Bicoherence Feature

One aspect of the edge configuration of an image is the edge sparsity. The relationship between the edge configuration of an image and its bicoherence features can be illustrated using an impulse train. Let $x(t)$ be a discrete-time signal of length N with K evenly-spaced impulses at every N/K samples:

$$x(t) = \sum_{i=0}^{K-1} h \delta(t - i \frac{N}{K}) \quad \text{where } t \in \{0, \dots, N-1\}$$

The DFT of this signal is given by:

$$X(\omega) = \sum_{j=0}^{\lfloor \frac{N-1}{K} \rfloor} h \delta(\omega - jK)$$

Then the bispectrum term would be:

$$X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2) = h^3 \sum_{i=0}^{\lfloor \frac{N-1}{K} \rfloor} \sum_{j=0}^{\lfloor \frac{N-1}{K} \rfloor} \delta(\omega_1 - iK, \omega_2 - jK)$$

Upon normalization, the h^3 will be divided off. Then, the bicoherence becomes:

$$b(\omega_1, \omega_2) = \sum_{i=0}^{\lfloor \frac{N-1}{K} \rfloor} \sum_{j=0}^{\lfloor \frac{N-1}{K} \rfloor} \delta(\omega_1 - iK, \omega_2 - jK)$$

$$\text{Mean}(b(\omega_1, \omega_2)) \propto \frac{1}{K^2}$$

Then, the bicoherence magnitude feature (Definition 4) satisfies the following relationship:

$$f_M \propto \frac{1}{K^2}$$

As the bicoherence magnitude feature is proportional to $1/K^2$, it implies that the denser the evenly spaced impulses, the lower the magnitude of the bicoherence, and vice versa. Note that the magnitude of the impulse, h , does not affect the magnitude of the

bicoherence. In this case, the phase of the non-zero bicoherence component is always zero, independent of the number of impulses, K .

However, the evenly spaced impulses with equal amplitude could not model the real world image signal well, but the above illustration provides a hint on the relationship between the sparsity of edge and the mean of the bicoherence magnitude. The relationship is observation in the empirical result shown below.

We use the edge percentage in an image block as a measure of edge sparsity, where edge pixels are detected using Canny edge detection algorithm. Edge percentage is just one of the many measures for edge density. Coincidentally, the different types of object interface (i.e. textured-textured, smooth-smooth and textured-smooth) can be characterized by edge percentage as shown in Section 5.2.

5.2 Properties of Bicoherence Features on Different Object Interface Types

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 14 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. The plots for bicoherence phase feature (fP) are qualitatively similar to **Figure 14**.

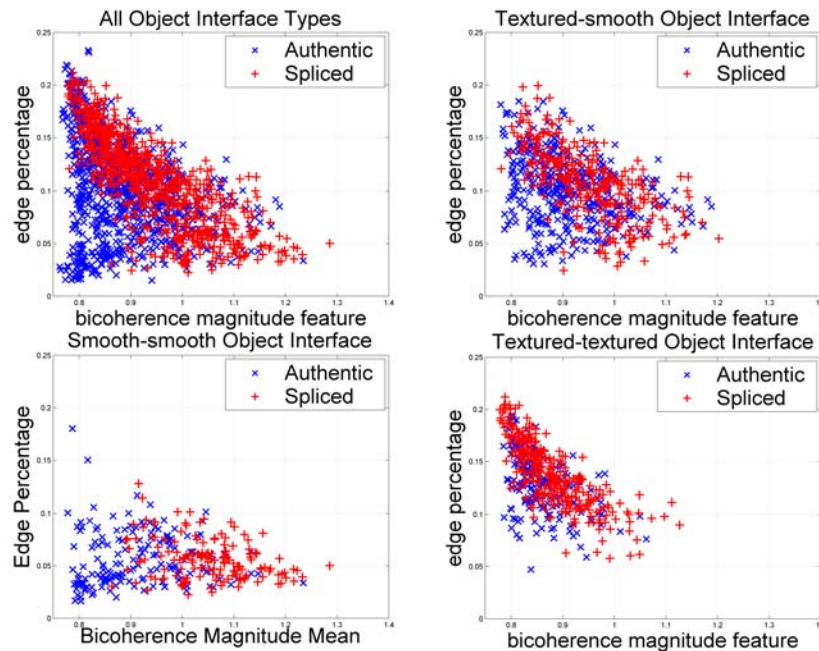


Figure 14: Bicoherence magnitude feature for different object interface types (x-axis is the bicoherence magnitude feature)

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 13 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.

6 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 measuring the discrepancy between a given image and its estimated authentic counterpart in terms of the bicoherence magnitude and phase features. In this case, authentic counterpart is the embodiment of the splicing-invariant features.
2. By incorporating image features that capture the image characteristics, which have an effect on the performance of the bicoherence features. For example, edge pixel percentage feature (f_E) was shown in Section 5.1 and 5.2 to have an effect on the bicoherence magnitude

6.1 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.

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

where f_{Bic} represents a bicoherence feature, 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 $f_{\text{Authentic}}$ and g_2 corresponds to the elevation of the bicoherence feature induced by splicing, denoted as $\Delta f_{\text{Splicing}}$. The splicing effect, $\Delta f_{\text{Splicing}}$, would be more observable if the significant interference from the splicing-invariant component, $f_{\text{Authentic}}$, can be removed. In this case, $\Delta f_{\text{Splicing}}$ can be estimated with $f_{Bic} - f_{\text{Authentic}}$, which we call *prediction residual*. The $f_{\text{Authentic}}$ 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 [42], which has a good edge preserving property, which is important for we have observed the sensitivity of the bicoherence features to edges. In other words, we want the estimated authentic counterpart to be as close to an image as possible in terms of the value of bicoherence features.

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

In functional representation, an image, f , can be decomposed into two functions, u and v , within a total variation minimization framework with a formulation [42]:

$$\inf_u \left\{ E(u) = \|u\|_{BV} + \lambda \|f - u\|_G ; f = u + v, f \in L_2(\mathbb{R}^2), u \in BV(\mathbb{R}^2), v \in G(\mathbb{R}^2), \|u\|_{BV} = \int_{\mathbb{R}^2} |\nabla u| \right\}$$

where the u component, also known as the 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. $\|\cdot\|_G$ 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.

For the texture decomposition, our assumption is that the splicing artifact, which can be modeled as a bipolar perturbation, would be captured by the fine-texture component, and the structure could be considered as being splicing-invariant. In general, there are two approaches for detecting the splicing artifact:

1. **Detecting the presence of splicing artifact in the fine-texture component:** Due to the noise-like characteristics of the fine-texture component, we empirically find that this approach does not work well because the value of the bicoherence features for the fine-texture component of both the authentic and spliced image block vary in a very narrow range and indistinguishably overlapped (figure not shown). Hence, these features are not discriminative at all.
2. **Detecting the absence of the splicing artifact in the structure component:** The absence of the splicing artifact can be detected by comparing the bicoherence features of the structure component against those of the undecomposed image. This approach exactly corresponds to the idea of detecting image splicing through its authentic counterpart as described in Section 6.1.

We adopt the second approach. In this case, the structure component can serve as an approximation for the authentic counterpart, hence, the estimator for the bicoherence magnitude feature of the authentic counterpart, $fM_{Authentic}$ and its phase features, $fP_{Authentic}$ are respectively $\hat{f}M_{Authentic} = fM_{structure}$ and $\hat{f}P_{Authentic} = fP_{structure}$.

Authentic Image block			Spliced Image block		
Original	Structure Component	Fine texture Component	Original	Structure Component	Fine texture Component

Figure 15: Examples of texture decomposition

For the linear prediction discrepancies between the bicoherence features of an image and those of its authentic counterpart, i.e., $\Delta fM = fM - \alpha \hat{f}M_{Authentic}$ and $\Delta fP = fP - \beta \hat{f}P_{Authentic}$, the scale parameters α and β , without being assumed to be unity, are learnt by Fisher Linear Discriminant Analysis in the 2-D space $(fM, \hat{f}M_{Authentic})$ and $(fP, \hat{f}P_{Authentic})$ 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. This idea of linear subspace projection is shown in Figure 16.

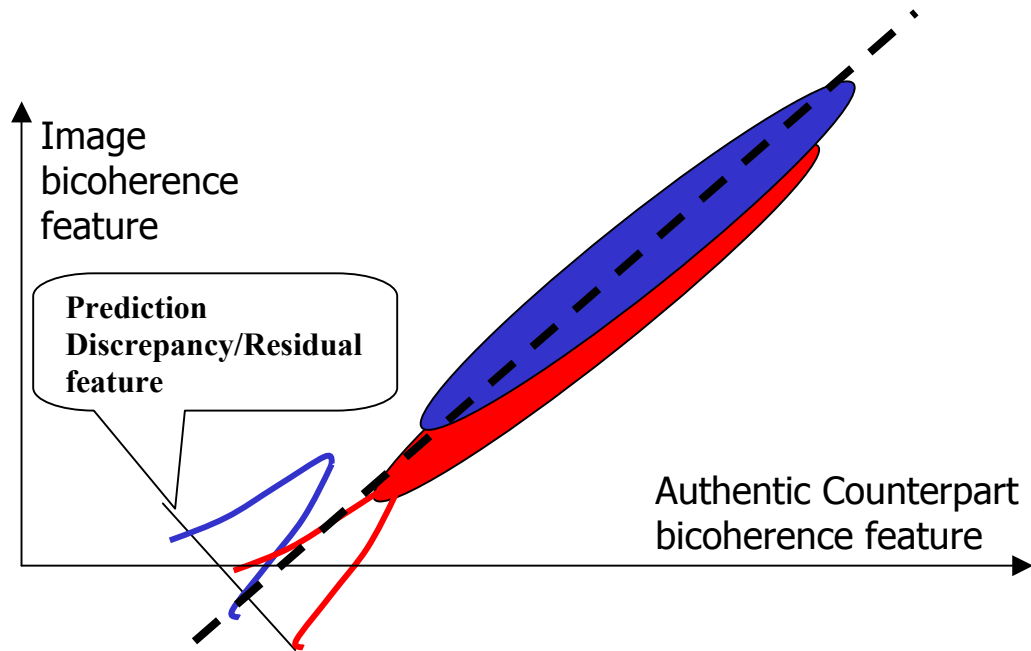


Figure 16 The illustration of linear subspace project using Fisher Discriminant Analysis

We would like to evaluate the effectiveness of the prediction residual features as shown in Figure 17. Our objective is to show that the prediction residual features $(\Delta fM, \Delta fP)$ 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 17 and Figure 13. (In Figure 17, two distributions are more separable)

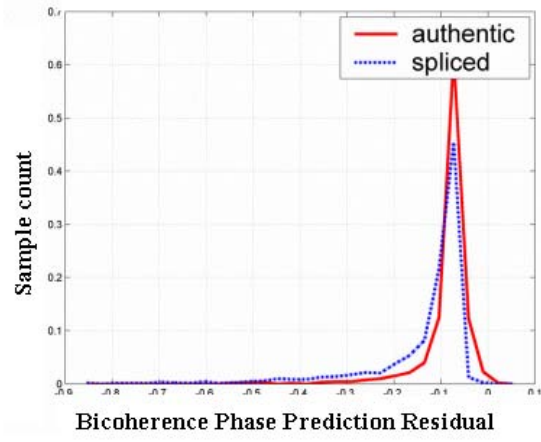
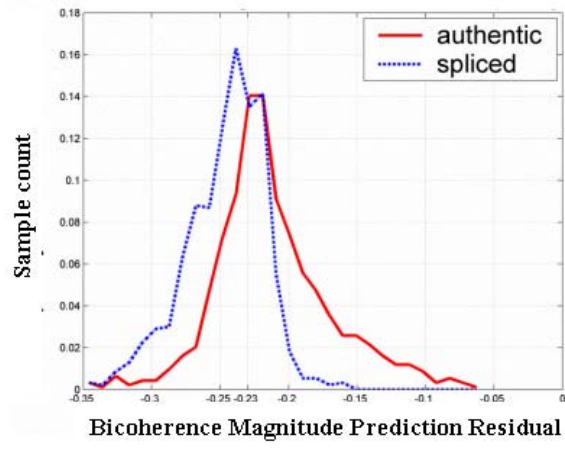


Figure 17: The distribution of prediction residual of the bicoherence magnitude (Left) and phase (Right) features

7 SVM Classification Experiments for Bicoherence Features Performance Evaluations

We herein evaluate the effectiveness of the features, which are derived from the proposed method (i.e., the prediction residual features and the edge percentage feature), by SVM classifications with radial basis function (RBF) kernel, performed our data set. We used SVM implementation from [43] and the SVM parameters (i.e., the error penalty parameter, C , and the RBF kernel parameter, γ) are chosen with an exhaustive grid search strategy as proposed by [44], such that the average classification accuracy (i.e., the mean of the true positive and negative rate) of the 5-fold cross-validation test results is the highest. Table 3 lists a summary of all features to be evaluated and Figure 18 shows the SVM classification results for different combination of features in the form of receiver operating characteristic (ROC) curve obtained from the 5-fold cross-validation test results. Table 4 shows the best average classification accuracy obtained by the above-mentioned grid search strategy.

Table 3 A summary of all features

Feature Name	Dimension	Legend
The bicoherence magnitude feature (i.e., the mean of the bicoherence magnitude)	1	PlainBIC
The bicoherence phase feature (i.e., the negative entropy of the bicoherence phase histogram)	1	
The prediction residual for the bicoherence magnitude feature	1	PredictionResidual
The prediction residue for the bicoherence phase feature	1	
Edge pixel percentage	1	EdgePercentage

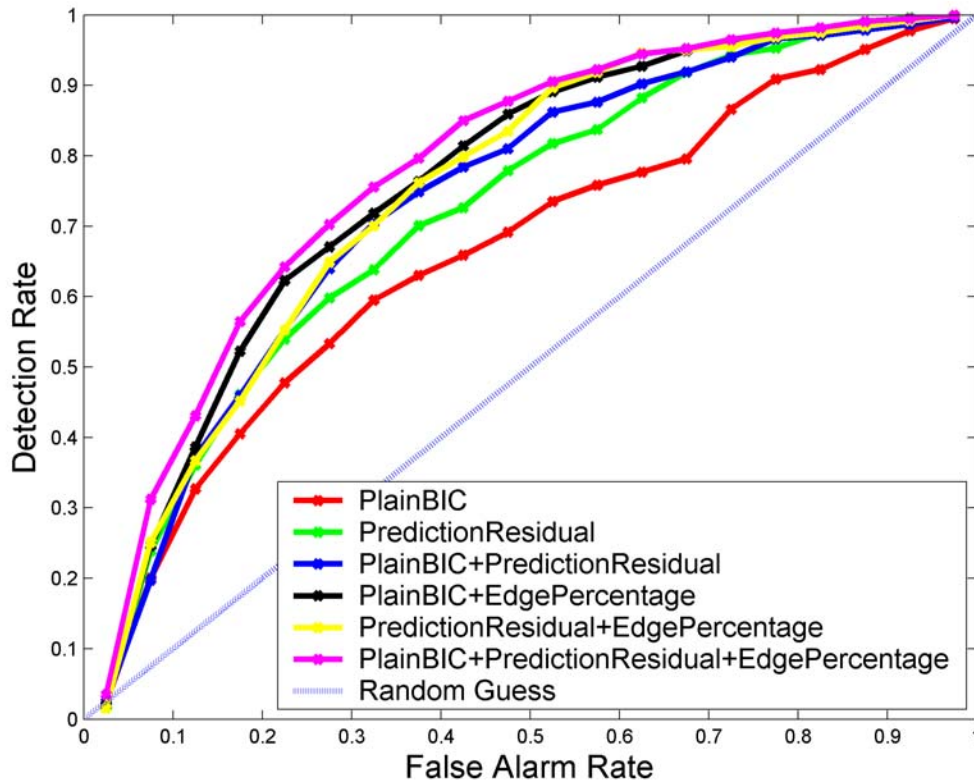


Figure 18: Receiver Operating Characteristic (ROC) curve for image splicing detection with different combination of features

Table 4: Best average classification accuracy (i.e., the average of the true positive rate and the true negative rate) for SVM classification with parameters obtained by grid search

Features	Best Average Classification Accuracy (0.5 for random guessing)
PlainBIC	0.6357
PredictionResidual	0.6644
PlainBIC+PredictionResidual	0.6866
PlainBIC+EdgePercentage	0.7041
PredictionResidual+EdgePercentage	0.6953
PlainBIC+PredictionResidual+EdgePercentage	0.7233

Below are the observations from the classification results in terms of the best average classification accuracy shown in Table 4:

1. Prediction residual features alone obtain 2.9% improvement the plain bicoherence features.
2. Edge percentage improves the performance of the bicoherence features by 6.8 %.
3. The best performance (last row) obtained by incorporating all the proposed features is 72%, which is 8.8% better than the baseline method (first row).

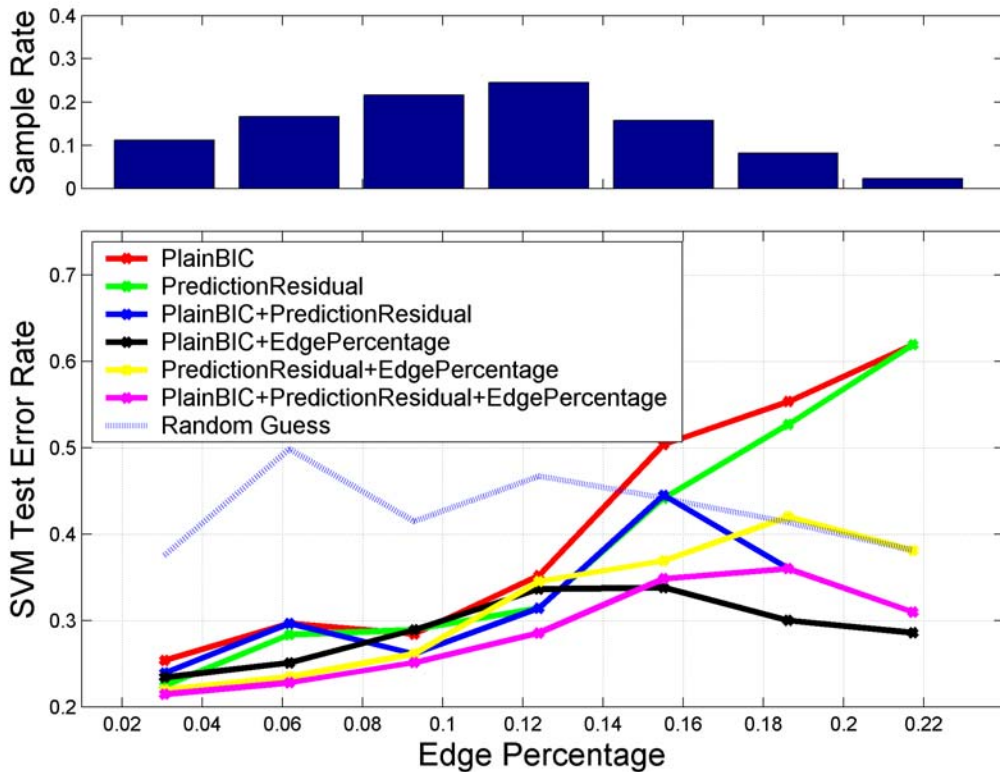


Figure 19: (Above) The fraction of image blocks over different edge percentage. (Below) The distribution of SVM 5-fold cross-validation test error over different edge percentage

As expected from the observation of Figure 14, Figure 19 shows that for the SVM classification error using plain bicoherence features is increasing with the edge percentage of the image blocks, with the error rate being greater than the random guessing error rate when the edge percentage is larger than 14%. The prediction residual features and the edge percentage feature help to greatly reduce the error rate at the high edge percentage region. Interestingly, although the prediction residual features do not perform well at the high edge percentage region but when it is combined with the plain bicoherence features, the error rate drops significantly at that particular region. It is also observed that the combination of all features obtain lower error rate compared to the plain bicoherence feature error rate at all edge percentage.

The results are encouraging as it shows the initial promise of the authentic counterpart estimation. 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 20 shows localizing the suspected splicing boundary.

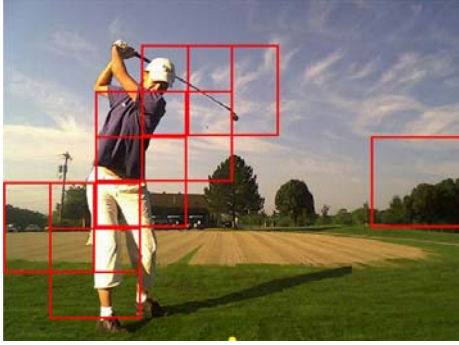


Figure 20: Examples of block-level detection on spliced images. The golfer with his shadow (in Left image) and the truck with shadow (in Right image) were cut-and-pasted from other images. The red boxes mark the suspicious blocks. The detection was controlled at a 10% false positive rate.

8 Conclusions and Further Work

8.1 Summary

In this report, we related our initial attempt on detecting image splicing using bicoherence features. We began with an introduction to bicoherence and then proposed an image-splicing model based on the idea of bipolar perturbation of an authentic signal. With the image-splicing model, we performed a theoretical analysis for the response of the bicoherence magnitude and phase features to splicing based on the proposed model. The analysis leads to the final propositions that image splicing increases the value of the bicoherence magnitude and phase features and a prediction of $\pm 90^\circ$ phase bias, which both are shown to be consistent with the empirical observations based on our data set. The proposed model has founded the use of bicoherence magnitude and phase features for image splicing detection on a sound theoretical ground.

We have also empirically shown how the performances of the bicoherence features depending on the different object interface types. Further on, we show that the plain bicoherence features do not serve well for image splicing detection. 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 percentage and the second offsets the effect of splicing-invariant component on bicoherence features. As a result, we obtain improved features in terms of discrimination power. Finally, we observe improvements in SVM classification after the derived features are incorporated.

In short, the plain/baseline bicoherence features do not perform well for image splicing detection and the proposal of incorporating image characteristics and the splicing-invariant (with respect to bicoherence) component has resulted in an improvement in the classification accuracy from the 62% obtained by using only the plain bicoherence features to 72% obtained by incorporating the three new features (i.e., the prediction residual for the plain magnitude and phase features, and the edge percentage feature).

8.2 Future Work

There is still extensive room for improvement based on this work. Possible directions could be to explore cross-block fusion and incorporate image structure in fusion.

The approach adopted in this report can be considered as a signal processing approach. We can combine the signal processing approach with the computer graphics/computer vision approach to automatically or semi-automatically examine the scene-level internal inconsistencies within an image. Lastly, we can explore beyond bicoherence for other discriminative features for image splicing (in particular) or photomontage (in general) detection.

9 Reference

- [1] K. Hafner, "The Camera Never Lies, but the Software Can," in *New York Times*. New York, 2004.
- [2] B. Daviss, "Picture Perfect," in *Discover*, 1990, pp. 55.
- [3] W. J. Mitchell, "When Is Seeing Believing?," in *Scientific American*, 1994, pp. 44-49.
- [4] C. Ginzburg, "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method," *History Workshop*, 1980.
- [5] G. L. Friedman, "The trustworthy digital camera: restoring credibility to the photographic image," *IEEE Transactions on Consumer Electronics*, vol. 39, pp. 905-910, Nov, 1993.
- [6] C.-Y. Lin and S.-F. Chang, "A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation," *IEEE Transactions on Circuits and Systems for Video Technology*, 2000.
- [7] M. Schneider and S.-F. Chang, "A Robust Content Based Digital Signature for Image Authentication," *IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept, 1996.
- [8] S. Bhattacharjee, "Compression Tolerant Image Authentication," *IEEE International Conference on Image Processing*, Chicago, Oct, 1998.
- [9] E.-C. Chang, M. S. Kankanhalli, X. Guan, H. Zhiyong, and W. Yinghui, "Image authentication using content based compression," *ACM Multimedia Systems*, vol. 9, pp. 121-130, Aug, 2003.
- [10] N. Memon, P. Vora, B.-L. Yeo, and M. Yeung, "Distortion Bounded Authentication Techniques," *SPIE Security and Watermarking of Multimedia Contents II*, Jan 24-26, 2000.
- [11] M. M. Yeung and F. Mintzer, "An Invisible Watermarking Technique for Image Verification," *IEEE International Conference on Image Processing*, Oct 26-29, 1997.
- [12] M. Wu and B. Liu, "Watermarking for Image Authentication," *IEEE International Conference on Image Processing*, Oct 4-7, 1998.
- [13] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*: Morgan Kaufmann, 2002.
- [14] J. Fridrich, M. Goljan, and B. A.C., "New Fragile Authentication Watermark for Images," *IEEE International Conference on Image Processing*, Vancouver, Canada, Sept 10-13, 2000.
- [15] P. W. Wong, "A Watermark for Image Integrity and Ownership Verification," *IS&T Conference on Image Processing, Image Quality and Image Capture Systems*, Portland, Oregon, May, 1998.
- [16] E. T. Lin, C. I. Podilchuk, and E. J. Delp, "Detection of Image Alterations Using Semi-Fragile Watermarks," *SPIE International Conference on Security and Watermarking of Multimedia Contents II*, San Jose, CA, Jan 23-28, 2000.

- [17] C.-Y. Lin and S.-F. Chang, "A Robust Image Authentication Method Surviving JPEG Lossy Compression," SPIE Storage and Retrieval of Image/Video Database, San Jose, Jan, 1998.
- [18] J. Fridrich, "Image Watermarking for Tamper Detection," IEEE International Conference on Image Processing, Chicago, Oct, 1998.
- [19] N. Memon and P. Vora, "Authentication Techniques for Multimedia Content," SPIE Multimedia Systems and Applications, Boston, MA, Oct, 1998.
- [20] J. Fridrich and M. Goljan, "Protection of Digital Images Using Self Embedding," Symposium on Content Security and Data Hiding in Digital Media, New Jersey Institute of Technology, May 14, 1999.
- [21] C. Rey and J.-L. Dugelay, "Blind Detection of Malicious Alterations on Still Images using Robust Watermarks," IEE Seminar Secure Images and Image Authentication, Apr, 2000.
- [22] S. A. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. S. Wallach, D. Dean, and E. W. Felten, "Reading between the lines: Lessons from the SDMI challenge," 10th USENIX Security Symposium, Washington, D.C., Aug 13–17, 2001.
- [23] H. Farid, "Detecting Digital Forgeries Using Bispectral Analysis," MIT, MIT AI Memo, AIM-1657, 1999.
- [24] J. W. A. Fackrell and S. McLaughlin, "Detecting Nonlinearities in Speech Sounds using the Bicoherence," *Proceedings of the Institute of Acoustics*, vol. 18, pp. 123-130, 1996.
- [25] E. Nemer, R. Goubran, and S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 217-231, Mar, 2001.
- [26] G. Krieger, C. Zetsche, and E. Barth, "Higher-order statistics of natural images and their exploitation by operators selective to intrinsic dimensionality," IEEE Signal Processing Workshop on Higher-Order Statistics, Banff, Canada, July 21-23, 1997.
- [27] G. Krieger and C. Zetsche, "Nonlinear Image Operators For The Evaluation Of Local Intrinsic Dimensionality," *IEEE Transactions on Image Processing*, vol. 5, pp. 1026-1042, Dec, 1996.
- [28] J. W. A. Fackrell, S. McLaughlin, and P. R. White, "Practical Issues Concerning the Use of the Bicoherence for the Detection of Quadratic Phase Coupling," IEEE-SP ATHOS Workshop on Higher-Order Statistics, Girona, Spain, June, 1995.
- [29] G. T. Zhou and G. B. Giannakis, "Polyspectral Analysis of Mixed Processes and Coupled Harmonics," *IEEE Transactions on Information Theory*, vol. 42, pp. 943-958, May, 1996.
- [30] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529-541, 1981.
- [31] H. Farid, "A Picture Tells a Thousand Lies," in *New Scientist*, vol. 179, 2003, pp. 38-41.
- [32] H. Farid and S. Lyu, "Higher-Order Wavelet Statistics and their Application to Digital Forensics," IEEE Workshop on Statistical Analysis in Computer Vision, Madison, Wisconsin, 2003.

- [33] CalPhotos, "A database of photos of plants, animals, habitats and other natural history subjects." Berkeley: Digital Library Project, University of California, 2000.
- [34] T.-T. Ng and S.-F. Chang, "A Data Set of Authentic and Spliced Image Blocks," Columbia University, New York, ADVENT Technical Report #203-2004 -3, June 8, 2004.
- [35] K. Hasselman, W. Munk, and G. MacDonald, "Bispectrum of Ocean Waves," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1963, pp. 125-139.
- [36] T. H. Bullock, J. Z. Achimowicz, R. B. Duckrow, S. S. Spencer, and V. J. Iragui-Madoz, "Bicoherence of intracranial EEG in sleep, wakefulness and seizures," *EEG Clin Neurophysiol*, vol. 103, pp. 661-678, Dec, 1997.
- [37] T. Sato, K. Sasaki, and Y. Nakamura, "Real-time Bispectral Analysis of Gear Noise and its Applications to Contactless Diagnosis," *Journal of the Acoustic Society America*, vol. 62, pp. 382-387, 1977.
- [38] 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, pp. 257-266, 1995.
- [39] Y. C. Kim and E. J. Powers, "Digital Bispectral Analysis and its Applications to Nonlinear Wave Interactions," *IEEE Transactions on Plasma Science*, vol. PS-7, pp. 120-131, June, 1979.
- [40] M. e. a. Santos, "An Estimate of the Cosmological Bispectrum from the MAXIMA-1 CMB Map," *Physical Review Letters*, vol. 88, 2002.
- [41] T.-T. Ng, S.-F. Chang, and Q. Sun, "Blind Detection of Photomontage Using Higher Order Statistics," IEEE International Symposium on Circuits and Systems, Vancouver, Canada, May 23 - 26, 2004.
- [42] 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.
- [43] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," 2001.
- [44] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," 2003.

APPENDIX – Classification of Image Processing Operations

In [32], Higher-order statistics (HOS) of both the wavelet coefficients and the linear prediction error of the wavelet coefficients are used for classification for

1. natural images vs. artificial images. Artificial images is composed of three categories, i.e. noise images (scrambled version of natural images – retaining intensity histogram), fractal images (have power spectrum similar to natural images) and disc images (have similar phase statistics as natural images)
2. between images with and without hidden message. Four different steganography algorithms are used: Jsteg (Information is hidden in the JPEG image by modulating the rounding choices either up or down in the DCT coefficients), EzStego (modulating the LSB of the sorted color palette index in 8-bit GIF images), OutGuess+/- (also modulating the DCT coefficients of JPEG images, with/without statistical correction)
3. natural images vs. computer graphics. The computer graphics are generated from 3D Studio Max, Maya, SoftImage 3D, Lightwave 3D, Imagine and Alias PowerAnimation.
4. original images vs. the recaptured version of the original images. Original images are printed on a laser printer and then scanned with a flat-bed scanner, both at 72dpi.

The higher-order statistics is herein referring to the mean, variance, skewness and kurtosis. The prediction error is the difference between the magnitude of the wavelet coefficients and prediction of them from a linear predictor based on the adjacent coefficients in the neighboring spatial, orientation and scale subbands.

The classification performance is impressive except for the classification between computer graphics and natural images.

We have using the some features for classifying images with have undergone different types of image processing operations:

- Low pass (lp) - Gaussian low pass with kernel support size 5x5 and standard deviation 8
- Jpeg (jpeg) – jpeg compression with quality factor 60
- Insertion of noise (noise) – Gaussian noise with mean 0 and standard deviation 18
- High pass (hp) – Gaussian high pass with cutoff frequency at 0.025 times sampling frequency, compensated with the original image
- Histogram equalization (histeq) – uniform histogram with 64 bins
- Median filtering (medfilt) – with filter kernel size 9x9
- Wiener filtering (wiener) – with filter kernel size 9x9
- Brightening (bright) – with 0.6 gamma pixel intensity mapping
- Simulated Image splicing (splice) – swapping the upper-right quadrant and the lower-left quadrant.

We selected 186 images with the approximated size of 512x768 pixels from the Calphotos image data set [33] and the Kodak research image data set. The set of images,

which is considered as original in our case, has already been JPEG compression with 80-quality factor. We performed the above-mentioned image processing operations on the 186 images independently. As a result, the number of images for each operation is 186.

We performed a 10-class SVM classification with the radial basis function (RBF) kernel. For this experiment, we used the OSU SVM Classifier Matlab Toolbox (ver 3.00) which makes use of the LIBSVM implementation [43]. The results shown below are the averaged testing error obtained from 10-folds cross-validation.

Table 5 shows the classification results. The blank cells represent zero. Each row of the table sums up to one. The number in each cell represent the ratio of images having undergone the operation of an row being classified as having undergone the operation of an column. From the table, we observe that:

1. The original image is mostly mistakenly classified as images of JPEG, brightening and image splicing. As the original are originally JPEG compressed, its being confused with images of JPEG compression is expected. Its confusion with the brightening and image splicing indicated that the classifier is having problems distinguishing these operations.
2. The rest of the operations are very well distinguished by looking at the small confusion among them.

Table 5 10-class SVM classification results for different types of image processing operations

	Classified As									
	Lp	Jpeg	Noise	Hp	Histeq	Medfilt	Wiener	Bright	Splice	Orig
Lp	1.0000									
Jpeg		0.9076		0.0111	0.0325				0.0211	0.0278
Noise			1.0000							
Hp		0.0056		0.8798	0.0164			0.0108	0.0599	0.0275
Histeq				0.0439	0.9126			0.0053	0.0272	0.0111
Medfilt						0.9836	0.0164			
Wiener							0.9836	0.0164		
Bright		0.0105		0.0164	0.0222		0.0053	0.5368	0.2401	0.1687
Splice		0.0901		0.0266	0.0371	0.0056	0.0053	0.1640	0.5272	0.1442
Orig		0.1047		0.0433	0.0380		0.0053	0.3491	0.2132	0.2465

This experiment leads to one important conclusion that the wavelet HOS features are very weak in distinguishing between spliced images and authentic images.