Passive-blind Image Forensics

Tian-Tsong Ng, Shih-Fu Chang
Department of Electrical Engineering
Columbia University
New York, NY 10027
{ttng,sfchang}@ee.columbia.edu

Ching-Yung Lin
IBM T. J. Watson Research Center
Hawthorne, NY 10532
chingyung@us.ibm.com

Qibin Sun
Institute for Infocomm Research
Singapore 119613
qibin@i2r.a-star.edu.sg

June 23, 2006

Abstract

In this chapter, we will review the research area of passive-blind image forensics, i.e., an form of image analysis for finding out the condition of an image without relying on pre-registration or pre-embedded information. We consider the two main functions of passive-blind image forensics as being image forgery detection and image source identification. In this vein, we provide a detailed review of the prior work in these two areas of image forensics. Apart from this, we also give a brief history of image forgery creation, a brief review of the state-of-the-art image forgery creation techniques, the resources available to the researchers and the challenges facing the passive-blind image forensics research.

1 Introduction

One of the key characteristics of digital images with a discrete representation is its pliability to manipulation. Therefore, even back in 1989, the sesquicentennial year for photograph when digital images was gaining popularity, 10% of all color photographs published in United States were actually digitally altered and retouched, according to the Wall Street Journal [1]. The recent well-known tampered images are the Iraq soldier picture of the Los Angeles Times (March

2004) and the Internet image showing Jane Fonda and John Kerry sharing the same speaker platform (Feb 2004)[1]. The advanced digital image processing techniques provided by the image editing software like Adobe Photoshop are the catalyst for the prevalence of the manipulated digital images in the public domain, which is evident in web sites such as `www.worth1000.com`, which is a creative competition and Photoshop contest sites hosting as many as 178,582 Photoshop-created images as in Oct 2005. Besides image compositing, computer graphics nowadays can also produce image forgery of high photorealism. To showcase the high photorealism of computer graphics which rivals that of the real camera images, a 3D graphics company has setup a web site `www.fakeorfoto.com` for challenging viewers to distinguish computer graphics and camera images.

Traditionally, a photograph implies truth. However, a similar faith on digital images is diminished due the the ease of manipulation. Unlike text, images provide an effective and natural communication media for human, as human often need no special training to understand the image content. Therefore, being able to verify the credibility of digital images and perform image forensics can protect the truthfulness of digital images. Today, digital images have already been heavily used for news reporting, insurance claim investigation, forensic or criminal investigation, legal proceeding, and national intelligence analysis. As such, image forensic would have a great impact in the above-mentioned application domain.

The main function of image forensics is to assess the authenticity and the origin of images. Therefore, trustworthy digital image is a main concern for image forensics. Back in 1993, the idea of trustworthy camera [2] has been proposed as a way to make the trustworthiness of digital images accountable. A trustworthy camera embeds a digital watermark on an image at the instant of its acquisition and any later tampering of the image can be detected based on the changes on the digital watermark. However, the realization of the trustworthy camera idea requires the camera manufacturers to concur on a common standard protocol, while the consumers need to accept the reduced image quality due to the embedding of a digital watermark. Apart from that, the most basic worry lies on the fundamental security of digital watermarks, as being evident in the Secure Digital Music Initiative (SDMI) fiasco [3], where the proposed audio watermarking system was swiftly hacked by a coalition of cryptography and watermarking researchers from Princeton University, Xerox PARC and Rice University. Digital watermarking is considered an active approach as it requires a known signal to be embedded onto an image for image forensics to be possible. In contrast, *passive-blind image forensics* (PBIF) was proposed [4, 5, 6, 7, 8, 9, 10], with a goal of detecting image alteration or identifying the image source without any prior measurement and registration of the image including the availability of the original reference image. At this time when the digital image alteration techniques have become so versatile, the burgeoning of the passive-blind image forensics research is indeed timely.

[1]The L.A.Times image can be found at `http://www.sree.net/teaching/lateditors.html` and the John Kerry image can be found at `http://www.camerairaq.com/2003/02/john_kerry_and_.html`

We begin with an overview of PBIF in Section 2 where the two main functions of PBIF, i.e., passive-blind image forgery detection and passive-blind image source identification, are identified. The overview also covers a brief history and a general description of image forgery creation. Then, in Section 3, we provide a detailed review of the work in image forgery detection and image source identification. Before concluding, we describe the resources and the challenges for PBIF.

2 Overview of PBIF

In general, PBIF concerns with the following two main problems:

1. Image forgery (alteration) detection (Section 3.1)
2. Image source identification (Section 3.3)

As most of the image forgery detection techniques are associated to the specific image forgery creation techniques, we begin with a short history of image forgery creation in Sebsection 2.1, followed by a brief description of the image forgery creation process in Subsection 2.2 (More details in the Appendix).

2.1 The History of Image Forgery Creation

Just within a few decades from the birth of photography, various methods had already been invented for altering images. Combination print was one of the earliest form of image forgery creation techniques, where dark-room skills are used to print multiple fragments of image onto a single photograph paper. One of the earliest well-known combination prints was Oscar G. Reijlander's *The Two Ways of Life* (1857) which had used up to 30 images[2]. Later in the early twentieth century, photomontage, a cut-and-paste composite of image fragments, gained popularity, mainly for surreal art, political satires and many other purposes. Both combination print and photomontage are technically demanding and time consuming, and their application is often detectable.

2.2 Image Forgery Creation in Modern Time

With the wide availability of the powerful image editing tool, such Adobe Photoshop, the similar image alteration functions described in the previous section can be performed in the digital domain with a much easier process while resulting in a much higher verisimilitude. In general, the image forgery creation process involves selection, transformation, composition of the image fragments, and retouching of the final image as shown in Figure 1. The process often begins with extracting a fragment or a 3D object model from an image. The forgery creators can then fuse the transformed image fragment or the image portion

[2]The image of the combination print, the two ways of life, by Oscar G. Reijlander can be found in `http://www.bradley.edu/exhibit96/about/twoways.html`

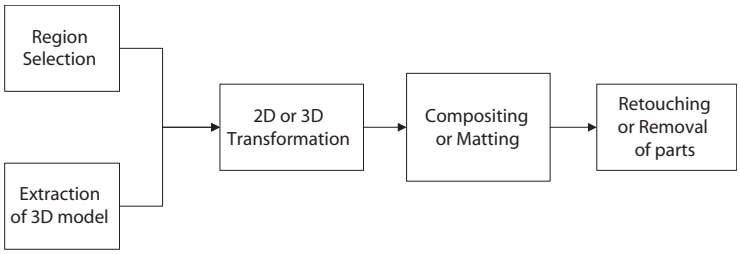

Figure 1: The process for image forgery creation.

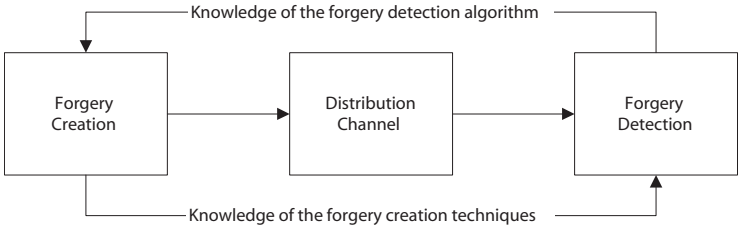

Figure 2: The setting for the passive-blind image forgery detection problem.

generated from the transformed 3D model into another image using techniques such as matting for coherent-looking composition. Finally, the composite image is retouched to remove the remaining artefact. This stage may involve the removal of certain objects from the image which is sometimes known as *reverse cropping*. The Appendix provides a glimpse of the state-of-the-art image editing techniques.

3 Forgery Detection and Source Identification

In this section, we provide a detailed description of the two main problems in PBIF, i.e., image forgery detection and source identification.

3.1 Passive-blind Image Forgery Detection

Just like the adversarial roles of the spy and the counter-intelligence in a espionage game, the forgery creators and the forgery detectors are opponents, as shown in Figure 2. The goal of the forgery creators is to create image forgery as a fabrication of the truth, while the forgery detectors try to uncover any possible act of the fabrication by assessing the *authenticity* of a given image. Examples of image forgery are the digital photomontage and the images with removed objects.

The concept of image authenticity is essentially based on the image characteristic. It is meaningless to talk about the authenticity of a random pixel image, as it has no meaningful characteristic. In contrast, natural-scene images

occupy a highly regularized subspace in the entire image space; if one tries to generate images using a random pixel generator, the chance of getting a natural-scene image is very small. Therefore, an authentic characteristic can be defined on natural-scene images. As the authenticity of natural-scene images is related to the natural scenes and the imaging process, we can define its authenticity based on two distinct qualities, i.e., the *the imaging-process quality* and the *the natural-scene quality*.

The imaging-process quality is due to the image acquisition devices. For instance, a typical CCD digital camera imposes effects of lens distortion, demosaicing, white-balancing, non-linear gamma correction and sensor noise on the images it produces. One can estimate the above-mentioned effects on an image in order to verify the authenticity of a camera image and distinguish it from a computer graphic image which has not undergone the camera acquisition process [11].

The natural-scene quality is entailed by the physics of the real-world light transport involved in the image formation process. An image is essentially a snapshot of the light field resulted from the complex interaction between the illumination sources and the objects. The physical process, for instance, imposes a relationship between the orientation of the shadow and the direction of the illumination sources. Therefore, by checking the consistency of the lighting directions estimated independently from the surface shading at two different locations, one can verify whether an image is produced by a composition [12].

The adversarial game between image forgery creators and image forgery detectors is made possible by the probabilistic nature of the natural-scene image authentic characteristic as well as the limited knowledge of its distribution. If the distribution of the natural-scene images is deterministic and fully known, then an image forgery creator can always produce perfect image forgery which is indistinguishable from an authentic natural-scene image. To construct a complete description of natural-scene images, one need to have a large amount of images to form an empirical distribution in the high-dimensional image space and this is difficult to achieve. As an approximation, one can model the marginal distribution or low-order joint distribution of the transformed images as the natural image statistics [13]. The resulting statistical image model is partial and incomplete. This is a good news for image forgery detectors is as it is hard for the opponent to check whether an image forgery is totally free of the forgery tell-tale signs. On the other hand, it makes passive-blind image forgery detection difficult. Without a complete model for the natural-scene images, the knowledge of the opponent's modus operandi would become a great advantage, as shown in Figure 2. This implies that image forgery detectors should prevent image forgery creators from having a full knowledge of the detection algorithm and at the same time they should understand the image forgery creation process well.

Although the above discussion focuses on natural-scene images, the same principle can be applied to other types of images, such as aerial images, x-ray images and microscopic images.

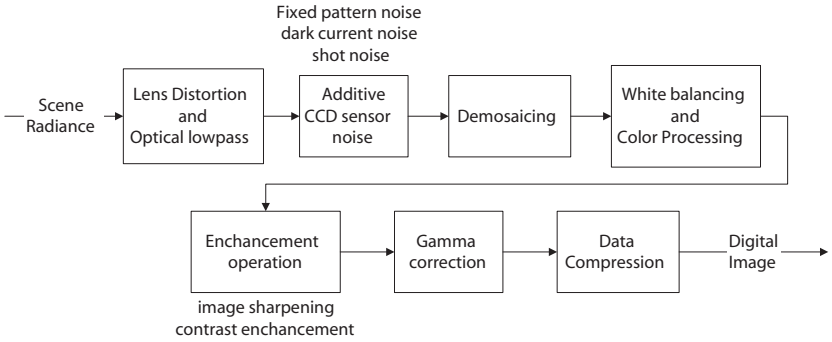

Figure 3: CCD camera imaging pipeline.

3.2 Passive-blind Image Forgery Detection Techniques

From the formulation of image forgery detection in Section 3.1, two approaches are possible for image forgery detection, i.e., detecting the authentic characteristics of images and detecting the tell-tale characteristics specific to the image forgery creation techniques.

3.2.1 Detecting Image Authenticity Quality

The authentic imaging-process quality is a characteristic of the imaging devices, such as digital cameras and scanners, and this quality can be different for various devices. We hereupon focus on the charged-couple device (CCD) digital camera, which is the most popular device for producing digital images. A CCD digital camera can be considered as a pipeline process, as shown in Figure 3. The following subsections review the work on image forgery detection using different characteristics of the digital camera as well as the natural-scene authentic characteristics.

3.2.2 Optical Low-pass of the Camera Lens

The work in [7] detects the presence of the abrupt discontinuities in an image or conversely the absence of the optical low-pass property as a tell-tale sign for identifying spliced images. The spliced images are produced by a simple cut-and-paste without any sophisticated matting or blending (refer to the Appendix) in the compositing step. For detecting the abrupt splicing discontinuity, a higher-order moment spectrum, bicoherence, is used. Bicoherence is a normalized third-order moment spectrum and its mathematical form for a 1-dimensional (1D) signal with a Fourier spectrum $X(\omega)$ is given by:

$$b(\omega_1, \omega_2) = \frac{E\left[X(\omega_1)X(\omega_2)X^{\star}(\omega_1+\omega_2)\right]}{\sqrt{E\left[\left|X(\omega_1)X(\omega_2)\right|^2\right] E\left[\left|X(\omega_1+\omega_2)\right|^2\right]}} \tag{1}$$

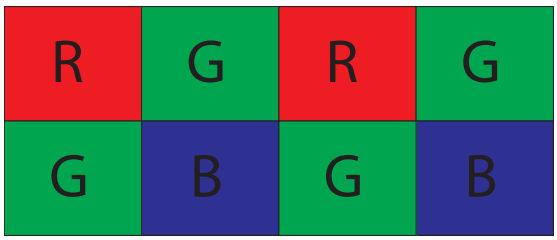

Figure 4: Bayer pattern for the camera sensor array

Note that the normalization factor is the upper bound for the Cauchy-Schwartz inequality, therefore $|b(\omega_1, \omega_2)|$ is between 0 and 1. Another important property of bicoherence is its sensitivity to a phenomena called *quadratic phase coupling* (QPC), i.e., the simultaneous presence of three frequency harmonics at ω_1, ω_2 and $\omega_1 + \omega_2$ respectively with a phase ϕ_1, ϕ_2 and $\phi_1 + \phi_2$ (the phases are coupled and hence not random). Note that at (ω_1, ω_2) which corresponds to a harmonic triplet with QPC, the bicoherence has a zero phase. However, it is shown in [6] that the bicoherence is sensitive to the splicing discontinuity due to a variant of quadratic phase coupling which induces a $\pm\frac{\pi}{2}$ phase for the bicoherence instead of the zero phase. This theoretical result is validated experimentally.

When only the bicoherence features are used for spliced image detection, the detection accuracy evaluated on the *Columbia Image Splicing Detection Evaluation Dataset* [14] is only 62% (50% for a random guess). To improve the detection performance, a functional texture decomposition method is used to decompose an image into a gross-structure component and a fine-texture component. The gross-structure component is used to approximate the authentic reference of an image (the hypothetically authentic image). By incorporating the discrepancy between an image and its authentic reference, the detection rate improves from 62% to 71%.

3.2.3 Demosaicing

Apart from the optical effect, the correlation between the image pixel values can also be useful for image forgery detection. The consumer CCD camera captures spectral energy corresponding to the red (R), blue (B) and green (B) color at the same time with a single CCD sensor array, by distributing the sensors in the array among the RGB color channels. The allocation of the sensors results in a partial sampling of the color signal in the sensor array. The process of designing the sensor allocation is likened to mosaicing and the most common sensor allocation pattern in the Bayer pattern as shown in Figure 4. To obtain a full array of pixel values for the RGB color channels, the missing samples are interpolated from the available samples and this operation called demosaicing. The interpolation process will inevitably introduce a statistical correlation between the interpolated pixels and the original pixels.

In [15], the demosaicing operation is modeled by a linear interpolation as

below:

$$I(x,y) = \sum_{u,v \in \Omega} \alpha_{u,v} I(x+u, y+v) \tag{2}$$

where Ω is a set of relative indices of a neighborhood, α's are the interpolation coefficients and the $I(x,y)$ is a 2D image. To evaluate the probability of a pixel as being an original pixel (or an interpolated one), an expectation-maximization (EM) algorithm is formulated, where the pixels type is the hidden variable and the linear interpolation coefficients are the model parameters. As the EM algorithm converges, a 2D probability map of the hidden variables is obtained and it shows the interpolation pattern of the image. The correctly estimated interpolation pattern of an authentic image would have a periodic pattern corresponding the sensor allocation pattern. The periodicity of the probability map leads to a set of the dominant frequency harmonics in the Fourier domain. As image compositing can disrupt the regular interpolation pattern, the estimated interpolation pattern can be used for detecting composite images. The experiments are conducted on the artificially generated demosaiced images as well as on images from three commercial cameras.

3.2.4 Camera Response Function

There are works which consider the camera response function for image forgery detection. The image irradiance (light energy incident on the image sensors) r is related to the image intensity (the final output image) R by a non-linear camera response function (CRF) f as in $R = f(r)$. This non-linear function is a characteristic of an image, which can be estimated from a single image [16, 17, 18]. The inconsistency in the estimated CRF over an image is a tell-tale sign for a composite image.

The work in [19] performs blind estimation of the CRF based on a gamma curve model, $f(r) = r^{\gamma}$, where γ is the gamma parameter. The estimation method is founded on the observation that the non-linear transform on the image irradiance introduces frequency harmonics with quadraticly coupled phases. This effect is due to the observation that a non-linear function can be approximated by the power series of a Taylor expansion as shown below:

$$f(r) = f(0) + \frac{r}{1!} f'(0) + \frac{r^2}{2!} f''(0) + \cdots \tag{3}$$

for the expansion of a function f at $r = 0$. The power series has a linear-quadratic function term (the linear combination of a linear and a quadratic term). The effect of a linear-quadratic function on a signal can be illustrated using a simple 1D signal with two frequency harmonics:

$$r(x) = a_1 \cos(\omega_1 x + \theta_1) + a_2 \cos(\omega_2 x + \theta_2) \tag{4}$$

The quadratic phase coupling phenomena is induced, when $r(x)$ passes through

a linear-quadratic operation:

$$\begin{aligned} r(x)+\alpha r(x)^2 &= a_1\cos(\omega_1 x+\theta_1)+a_2\cos(\omega_2 x+\theta_2) \\ &+\frac{1}{2}\alpha a_1^2\cos(2\omega_1 x+2\theta_1)+\frac{1}{2}\alpha a_2^2\cos(2\omega_2 x+2\theta_2) \\ &+a_1a_2\cos((\omega_1+\omega_2)x+(\theta_1+\theta_2)) \\ &+a_1a_2\cos((\omega_1-\omega_2)x+(\theta_1-\theta_2))+\frac{1}{2}\alpha a_1^2+\frac{1}{2}\alpha a_2^2 \end{aligned} \tag{5}$$

where α is an arbitrary constant. Note that there exists harmonics at ω_1, ω_2 and $\omega_1+\omega_2$ respectively with a phase θ_1, θ_2 and $\theta_1+\theta_2$. As QPC induces a (non-random) zero phase in bicoherence, the magnitude of bicoherence at the corresponding (ω_1,ω_2) takes a large value as an expectation of a constant-phase random variable. Therefore, bicoherence can be used to measure the amount of the QPC effect. As non-linear transform of an image increases the amount of QPC effect, it is reasonable to assume that the inverse transform of the image intensity by a correct gamma curve would correspond to a minimum for the bicoherence magnitude. As a result, the CRF can be estimated by searching for a curve which minimizes the bicoherence magnitude. In the paper, this idea is demonstrated using a simple image where the upper and the lower half of the image are separately transformed with gamma curves of a different gamma parameter.

A more realistic scenario for image forgery detection by the CRF characteristic is demonstrated in [9], where the CRF is estimated using the method proposed in [16]. This single-image CRF estimation method is based on the linear pixel blending property for the image irradiance at the edge pixels:

$$r_E=\alpha r_A+(1-\alpha)r_B \tag{6}$$

where α is the blending factor, and r_E, r_A and r_B are the image irradiance at the corresponding points E, A and B in Figure 5.

When the pixel blending factors α of an edge are uniform over the R, G and B color channels, there will exist a co-linear relationship between the values of the edge pixel and the values of the pixels in the adjacent homogenous regions (which is separated by the edge) in the RGB color space:

$$\begin{pmatrix} r_E^R \\ r_E^G \\ r_E^B \end{pmatrix}=\alpha\begin{pmatrix} r_A^R \\ r_A^G \\ r_A^B \end{pmatrix}+(1-\alpha)\begin{pmatrix} r_B^R \\ r_B^G \\ r_B^B \end{pmatrix} \tag{7}$$

where the upper index for r corresponds to the RGB color channels. As a non-linear transform distorts the co-linearity relationship, the CRF can be estimated from the form of the distortion.

3.2.5 Lighting Consistency

The work described above are all utilizing the camera authenticity quantity. However, the work in [12] demonstrates the novel idea of image forgery detection

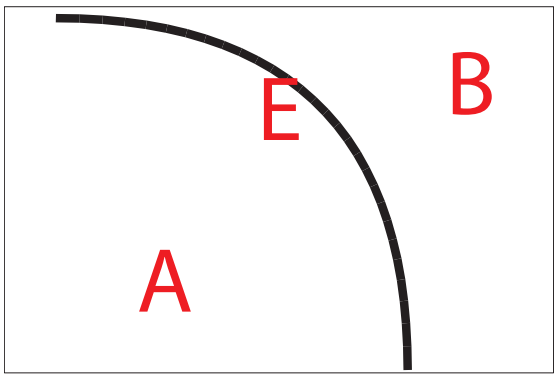

Figure 5: The curve line represents an image edge. There are three points at a local edge region: A and B are points at the homogenous intensity regions separated by the edge. E is the edge point.

by the natural-scene authenticity quality. The authors estimate the 2D lighting directions (in the image plane) from the occlusion edge. At the occlusion edge, the surface normal has a zero z-component, while the (x, y)-component is just the normal of the occlusion contour, which can be easily estimated. Under the assumption of the Lambertian surface, a constant reflectance and a single distant point light source, the image intensity (linear to the image irradiance) R is given by:

$$R(x, y) = \rho \mathbf{N}(x, y) \cdot \mathbf{L} + A \tag{8}$$

where $\mathbf{N}$ is the surface normal, the $\mathbf{L}$ is the point light source direction, A is the constant ambient light and ρ is the reflectance. As the surface normal at the occlusion edge is known, the light source direction in the x and y directions and the ambient light A can be recovered by the linear least square method when the surface normal and the image intensity at more than three points are available. When the 2D lighting directions is independently estimated at the occlusion edges of different objects, consistency checking of the lighting directions can be performed to verify whether an image is composite.

The authors further relax the above assumption by considering the case of locally constant reflectance and multiple local point light sources. However, the algorithm requires manually extracted edge points. Experiments using the real-world images and the synthetic images are shown and achieve promising results. In another ongoing work [20], it is shown that the lighting consistency can be examined without explicitly estimating the lighting. This preliminary theoretical result is based on the spherical frequency invariants and is currently assuming known object geometry.

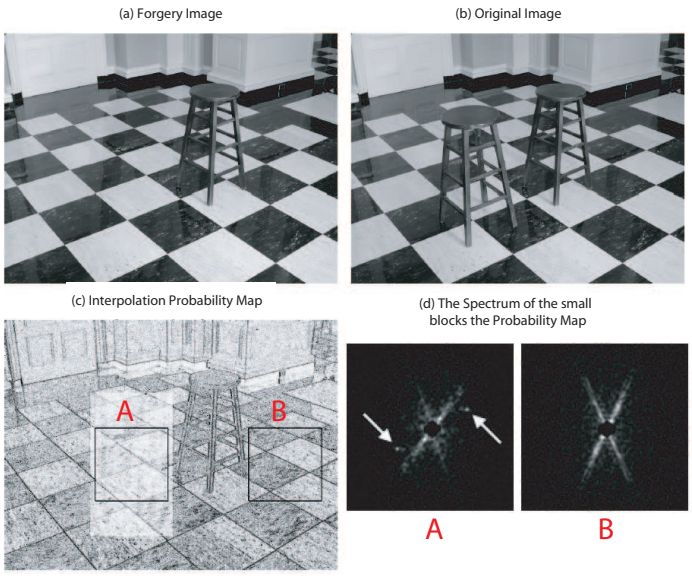

Figure 6: Image (a) is the forgery image created from image (b) by removed a stool from the image. The removed region is covered by the resized patch of the same background. (c) is the probability map output from the EM algorithm and (d) is the Fourier spectrum of the small blocks A and B. Dominant peaks in the Fourier spectrum indicate the periodicity of the probability map at the block region at a specific frequency. (Figure courtesy of Hany Farid as well as Springer Science and Business Media.)

3.2.6 Detecting Image Forgery Creation Artefact

Geometry transformation of a digital image potentially involves resampling or interpolation of some image pixels. To estimate the probability map of interpolation, the same EM algorithm as explained in subsection 3.2.1 is used [21]. In one scenario, during the process of image composition, an image fragment may undergo resampling as it is resized, before being spliced onto another image. The resampled region can be detected from the probability map if the host image is not similarly resampled. Figure 6[3] illustrates this scenario with a simple example. Note that the periodicity of the probability map at the spliced region manifests as peaks in its Fourier spectrum.

Apart from this, the presence of duplicate regions and the discrepancy in the signal-to-noise ratio (SNR) at different image regions can also be considered tell-tale signs for image compositing and techniques are proposed to detect these artefacts [19, 22]. When an object is removed from an image, one way to fill in the removed region is by example-based texture synthesis, i.e., to cover up the removed region using similar background patches. This method is especially effective for covering up the homogenous background region. The work in [22] proposes an effective way to detect duplicate image blocks (8×8 pixels) in a single image. The image blocks are reduced in dimension by using principal component analysis (PCA) and a lexicographic sort is applied to the PCA vectors for efficiently detecting the duplicate image blocks. The PCA vectors corresponding to the duplicate image blocks will be adjacent to each other in the lexicographically sorted list. The experiments show that the method is not only computationally efficient, it also works well even when the image is highly compressed and when there is additive noise in the image.

On the other hand, if the aussmption that image noise is uniform over an image, then the discrepancy of the noise variance at different regions of a same image would be a tell-tale sign for a composite image. A method for estimating noise variance, with the assumption of known signal kurtosis, is used to demonstrate the above image using a toy example.

Apart from the approaches that directly detect the artefacts closely linked to image forgery, there are approaches that detect the indirect evidences for the image forgery. The work in [19] proposes to consider JPEG double compression as an indirect evidence for image forgery. In the process of producing image forgery using the image editing software, it is likely that a JPEG image may be compressed once again at the end of the process with a different quality factor than the original one. Such JPEG double compression introduces a periodicity in the histogram of a JPEG Discrete Cosine Transform (DCT) coefficient. Figure 7[4] illustrates the effect of double JPEG compression using a sample sequence. Note that for both cases when the quantization step is increased or decreased at the second quantization, the histogram displays a periodic pat-

[3]Figure is extracted from [21], courtesy of Hany Farid as well as Springer Science and Business Media.

[4]Figure is extracted from [21], courtesy of Hany Farid as well as Springer Science and Business Media.

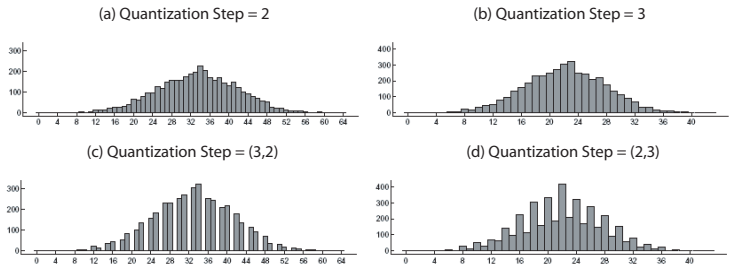

Figure 7: The double JPEG compression introduces an effect on the JPEG DCT coefficients, as a result of the double quantization. This figure illustrates the double quantization effect on a sample sequence. (a) The histogram of a sample sequence quantized with a quantization step 2 (b) The histogram of the same sequence quantized with a quantization step 3. (c) The histogram from a double quantization with a quantization step 3 followed by 2. (c) The histogram from a double quantization with quantization step 2 followed by 3. (Figure courtesy of Hany Farid as well as Springer Science and Business Media.)

tern. When such periodic pattern is observed for the DCT coefficients of a JPEG image, it indicates the act of JPEG double compression and calls for further examination on the image.

On the other hand, the work in [23] presents a method to detect the presence of camera pattern noise in an image for the purpose of image integrity verification or image forgery detection. The pattern noise is due to the non-uniform property of the individual camera sensors in terms of the dark current and the pixel responsivity. The absence of camera pattern noise in an image region may be a tell-tale sign of an image forgery. However, this method requires either the camera with which the image was produced or a set of images produced by the same camera.

Another type of indirect evidence for image forgery is the distortion resulted from the common post-processing operations on a composite image such as brightness adjustment, contrast adjustment and so on. A reference-free image quality/distortion measure is proposed for quantifying the quality of images. This objective image quality measure is used as features for training an image forgery detector [8], which achieves a detection rate of 69.2% for brightness adjustment, 74.2% for contrast adjustment and 80% for a mixed processing (i.e., a sequence of operations including scaling, rotation, brightness adjustment and contrast enhancement).

3.3 Passive-blind Image Source Identification

There are various devices from which digital images can be produced, examples are cameras, scanners, medical imaging devices and so on. Besides that, images can also be generated by computer graphic techniques. The goal of the passive-

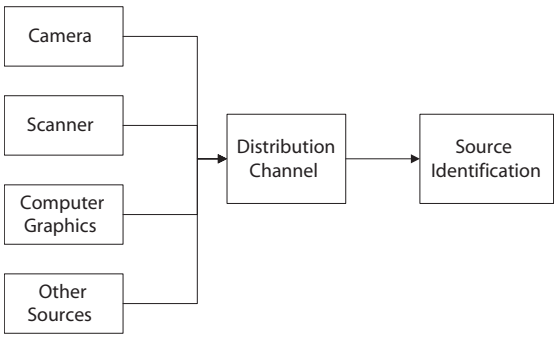

Figure 8: The setting for the passive-blind image source identification problem.

blind image source identification is to identify the type of image source, as shown in Figure 8. In a more complicated scenario, an image may be composed of fragments from multiple different sources.

Identification of the image source can help us in the decision of whether an image is acceptable for a specific application. For example, a computer graphic image is definitely unacceptable for news reporting and a human face image shown to a face biometric security system should not be mistaken by the authentication system as the actual presence of the person in front of the system.

3.4 Passive-blind Image Source Identification Techniques

One problem of concern in image source identification is the classification of photographic images (PIM) and photorealistic computer graphics (PRCG). Despite the fact that the classification which involves general computer graphics images (including drawing and cartoon) has already been applied for the purpose of improving the image and video retrieval performance [24, 25], the classification which involves photorealistic computer graphic is a new problem. The work in [26] uses the wavelet-based natural image statistics for the PIM and PRCG classification. The method extracts the first four order statistics (mean, variance, skewness and kurtosis) of the in-subband wavelet coefficients and also computes the first four order statistics of the linear prediction error for the wavelet coefficients using the coefficients from the neighboring spatial location, scale, orientation and the other color channels, as illustrated in Figure 9. The statistical features are used for classifying PIM and PRCG and achieve a PIM detection accuracy of 67% with 1% false alarm rate. As this technique is purely statistical, it provides little insight into the physical differences between PIM and PRCG.

In [11], the problem is approached by analyzing the physical differences between the image generative process for PIM and PRCG. This approach provides a physical explanation for the actual differences between PIM and PRCG, while the geometry features from this approach outperforms the features in the prior

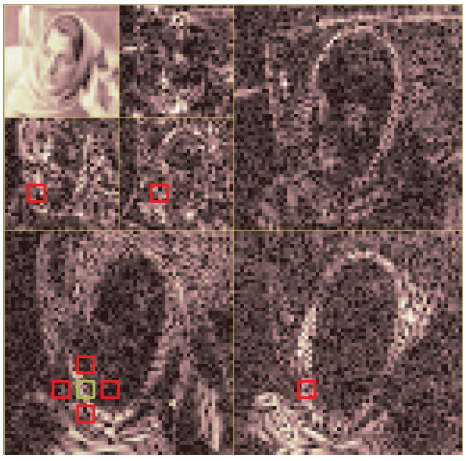

Figure 9: An illustration on how the linear prediction error of a wavelet coefficient (at the green box location) is computed using the wavelet coefficients (at the red box locations) in the neighboring spatial location, orientation and scale, for a grayscale image case. For a color image, the wavelet coefficients of other color channels can also be used for computing the prediction error.

work.

Specifically, the difference between PIM and PRCG can be briefly summarized below.

1. **Object Model Difference**: The surface of real-world objects, except for man-made objects, are rarely smooth or of simple geometry. Mandelbrot [27] has showed the abundance of fractals in nature and also related the formation of fractal surfaces to basic physical processes such as erosion, aggregation and fluid turbulence. Furthermore, surface such as human skin is full of subtleties and a result of the natural biological process. However, the computer graphics 3D objects are often represented by the polygonal models. Although the polygonal models can be arbitararily fine-grained, it comes with a higher cost of memory and computational load. Furthermore, such a polygonal model is not a natural representation for fractal surfaces [28]. A coarse-grained polygonal model may be used at the perceptually insignificant area for saving computational resources.

2. **Light Transport Difference** [29]: The physical light field captured by a camera is a result of the physical light transport from the illumination source, reflected to the image acquisition device by an object. The precise modeling of this physical light transport involves an 8D function of the object's reflectance property, hence its simulation requires substantial computational resources. Therefore, a simplified model based on the assumption of isotropy, spectral independence and parametric representation is often used.

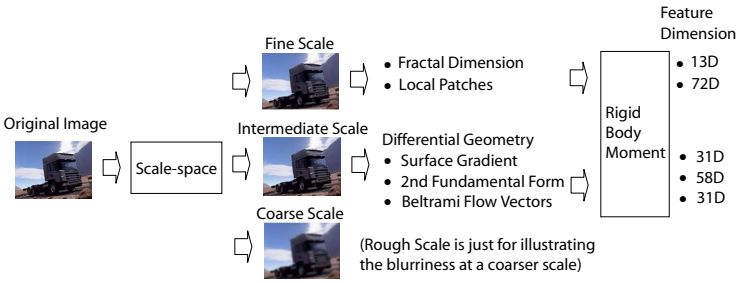

Figure 10: The geometry-based image description framework.

3. **Acquisition Difference**: PIM carry the characteristics of the imaging process, while PRCG may undergo different types of post-processing after the rasterizer stage. There is no standard set of post-processing techniques, but a few possible ones are the simulation of the camera effect, such as the depth of field, gamma correction, addition of noise, and retouching.

The above differences are captured using the geometry features derived from the differential geometry, the fractal geometry and the local patch statistics. Specifically, the authors propose a two-scale image description framework, as shown in Figure 10[5]. At the finest scale of the linear Gaussian scale-space, the geometry can be characterized by the local fractal dimension and also by the "non-parametric" local patches [30]. At an intermediate scale, when the fine-grained details give way to a smoother and differentiable structure, the geometry can be best described in the language of differential geometry, where the surface gradient, the second fundamental form and the Beltrami flow vectors are computed. While these fetaures are motivated by the physical properties of the image generative process, it provides a better classification performance compared to the techniques in prior work by at least 3%.

The image source identification problem includes the identification of the model of a camera which is useful for image forensics. The work in [31] exploits the characteristics of the in-camera color processing module to identify different models of digital cameras. The features related the color processing module are the average pixel value (motivated by the gray world assumption in white-balancing), and the pairwise correlation of the RGB color channels, the center of mass for the neighbor pixel distribution at different intensity values (related to the sensitivity of the camera at different intensity values), the pairwise energy ratio for the RGB color channels (related to the white point correction). They also use the mean of the wavelet subbands as additional features. Their experiment on the Sony DCS-P51, Nikon E-2100, Canon powershot S100, S110 and S200 obtains a classification accuracy of 88%.

On the other hand, a scanned image of a printed document can be ana-

[5]The truck image is from `http://www.realsoft.fi/gallery/vehicles/scania.jpg`

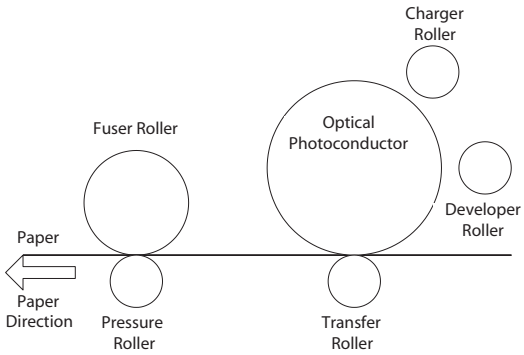

Figure 11: The cross-section view of a typical laser printer. The non-uniform movement of the optical photoconductor due to the variations at the gear train results in the banding artefact on a printout.

lyzed for identifying the printer from which the printed document is produced. In [32], the quasi-periodic banding artefacts in the process direction is used as the intrinsic printer signatures for identifying the laser printers. The banding artefacts often manifest as the non-uniform light and dark lines as the paper scrolls during the printing. The effect is due to the non-uniform movement of the optical photoconductor in the print mechanism, see Figure 11. The authors detect the effect by analyzing the pixel intensity co-occurrence matrix computed on the interior region of a specific printed character. The paper demonstrates the idea using the character "e" due to its high frequency of appearance. For printer classification, 22 features including the marginal mean, marginal variance, entropy, energy and so on are extracted from the co-occurrence matrix. Then, a 5-nearest neighbor classifier is trained to classify each feature vector corresponding to a single character "e". The printer model is identified through the majority vote from all the characters "e". The experiment involves 10 printers of different model and the test document are generated as random text. Promising result is obtained as the test documents from nine out of ten printer models are correctly classified.

4 Challenges and Resources for PBIF

PBIF is still a burgeoning research field and its advances depend on the carefully identified research directions and the availability of the required resources such as the experimental dataset. The general architecture of an image forensic engine, be it for image forgery detection or image source identification, is shown in Figure 12. In reference to the architecture, the elements for advancing the PBIF research is discussed in the following subsections.

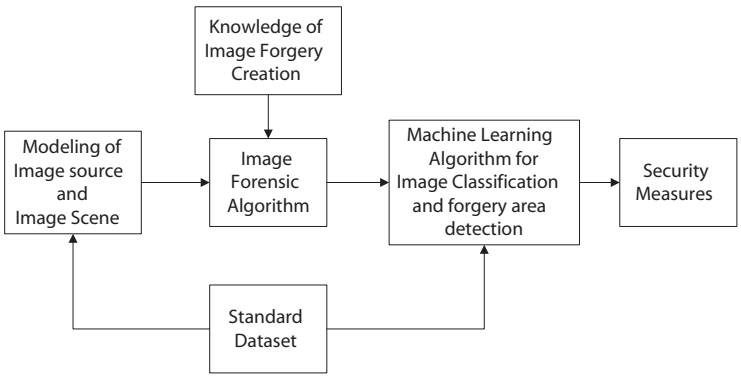

Figure 12: The basic architecture of an image forensic engine.

4.1 Image Modeling and Parameter Estimation

Image modeling is important for image forensics. Three types of image models relevant to image forensics are the natural image statistics, the physical image model based on the camera parameters and the model based on the scene constraints. A good review for natural image statistics can be found in [13]. Natural image statistics represents the statistical regularity in natural-scene images. The well-known natural image statistics are the power law of the natural image power spectrum, the sparse marginal distribution for wavelet coefficients, the non-trivial joint distribution of the wavelet coefficients and the higher-order statistics of images. For a physical model, images can be characterized by the parameters of the camera, such as the geometric lens distortion, the CCD sensor noise statistics, the camera response function, the demosiacing pattern and so on. Whereas at the scene level, a physical model can be based on the scene constraints such as the relationship between the shadow and the lighting, the consistency between the shading and the lighting, the consistency between the inter-reflection of light and the surface properties, and so on.

Once there is a good image model, the next concern would be the possibility to estimate the model parameters from a single image. Estimating the natural image statistics from a single image is not a problem, but it is a very difficult challenge for the physical model. The main reason is that an image is the combined effect of the various scene and camera factors. When attempting to factorize this combined effects or jointly estimate the multiple parameters, there exists multiple solutions and it has no unique solution. For instance, there is an inherent ambiguity in the estimation of the lighting, the reflectance property and the surface geometry from a single image, without any specific assumptions.

However, some progresses begin to be seen in the estimation of the camera parameters such as the camera response function from a single image [16, 17, 18]. Besides that, there are also some new semi-automatic methods for estimating the scene parameters such as the shadow [33]. Once the scene parameters are estimated, the consistency checking for the scene parameters is possible.

Interestingly, a parallel can be drawn between image forensics and face recognition in terms of the general approach towards image modeling in order to attain some kinds of invariance. As the image authenticity and the image source characteristics for image forensics are essentially independent of image content (e.g., lighting, the presence of objects and so on), the above-mentioned image model for image forensics are image content invariant. In face recognition, the face image model has to be pose and illumination invariant. The two general approaches to achieve pose and illumination invariance are the subspace-based model approach [34] and the physical geometric model-based approach [35]. These two general approaches correspond exactly to the natural image statistics approach and the physical model-based approach in image forensics.

4.2 Knowledge of Image Forgery Creation

For the image forgery detection techniques reviewed in Section 3.1, their evaluation on the state-of-the-art image forgery creation techniques (see Appendix) is still uncommon. For instance, the method of photomontage detection in [7] addresses only the simplest form image composition which is image splicing, the simple cut-and-paste of image fragments, without sophisticated matting or blending. The composite image detection method using the camera gamma curve in [19] demonstrates the idea only using a toy example. The main reason could be that the current detection techniques have not attained a level of sophistication which matches that of the image forgery creation techniques. However, if having no access to the image forgery creation system is one of the causes, a collaboration between the image forgery creation and image forgery detection research would be a good idea.

4.3 Full Automation and Fine-grained Analysis

An ideal image forensic system is one which is fully automated (i.e., requiring no human intervention) and provides a fine-grained analysis (e.g., at a local region). However, in most of the work discussed in Section 3.1, the proposed techniques are still either semi-automated or for a coarse-grained analysis (e.g., at the level of a large-size region or on the entire image). To replace the role of a human in the system, certain non-trivial tasks such as the detection of the object boundary in a composite image detection system need to be automated. To refine the analysis granularity, one needs to devise image forensic methods that relies on a smaller amount of data, while ensuring that the analysis remains reliable.

Despite the benefit of full automation, devising a fully automated yet sophisticated image forgery detection system is not always possible because of some fundamental limitations. For instance, as explained in subsection 4.1, the estimation of the physical model parameters from a single image without any user intervention is inherently impossible. If the level of automation has a tradeoff relationship with the detection accuracy and resolution, then a good system would have a simple but fully automatic module as a front end for pre-filtering

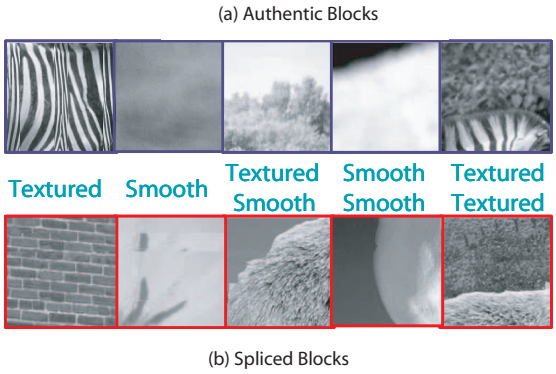

Figure 13: Examples from *Columbia Image Splicing Detection Evaluation Dataset*. The dataset has five subcategories with the textured, smooth, textured-smooth, smooth-smooth and textured-textured image blocks. (Image block courtesy of CalPhotos from the University of California at Berkeley and the individual photographers of the images.)

the potential image forgery and a semi-automatic but more comprehensive detection module as the back end for the final detailed analysis of the images.

4.4 Dataset

Dataset is important for image modeling and the evaluation of a proposed algorithm. Furthermore, a common dataset provides a common platform for the research community to compare various algorithms and thereby facilitates communications among researchers. To address this concern, the *Columbia Image Splicing Detection Evaluation Dataset* [14] and the *Columbia Photographic Images and Photorealistic Computer Graphics Dataset* [36] are made available to the research community. These two datasets can be downloaded from `http://www.ee.columbia.edu/trustfoto`.

The Columbia Image Splicing Detection Evaluation Dataset is for the image splicing detection experiments. It contains 933 authentic and 912 spliced image blocks of size 128×128 pixels. These image blocks are mainly extract from the Calphoto image set [37]. For image blocks of both classes, there are subcategories of the homogenous textured and smooth image blocks. There are also subcategories of image blocks with an edge or a splicing boundary which separates two textured, two smooth, and a textured with a smooth regions. Examples of the dataset is shown in Figure 13

The Columbia Photographic Images and Photorealistic Computer Graphics Dataset is for the PIM and PRCG classification experiments. There are four categories of images in the dataset, as described below and shown in Figure 14[6].

[6]The personal image at the second row is by the courtesy of Philip Greenspun. The

Figure 14: Examples from the dataset of photographic and computer graphic images. Note the photorealism of all images.

1. **800 photorealistic computer graphics from the Internet**: These images are categorized by content into architecture, game, nature, object and life. The PRCG are mainly collected from various 3D artists (more than 100) and about 40 3D-graphics websites, such as `www.softimage.com`, `www.3ddart.org`, `www.3d-ring.com` and so on. The rendering software used are such as 3ds MAX, softimage-xsi, Maya, Terragen and so on. The geometry modeling tools used include AutoCAD, Rhinoceros, softimage-3D and so on. High-end rendering techniques used include global illumination with ray tracing or radiosity, simulation of the camera depth-of-field effect, soft-shadow, caustics effect and so on.

2. **800 photographic images from a few photographers**: 400 of them are from the personal collection of Philip Greenspun, they are mainly travel images with content such as indoor, outdoor, people, objects, building and so on. The other 400 are acquired by the authors using the professional single-len-reflex (SLR) Canon 10D and Nikon D70. It has content diversity in terms of indoor or outdoor scenes, natural or artificial objects, and lighting conditions of day time, dusk or night time.

3. **800 photographic images from Google Image Search**: These images are the search results based on keywords that matches the computer graphics categories. The keywords are such as architecture, people, scenery, indoor, forest, statue and so on.

4. **800 re-photographed photorealistic computer graphics**: These are the photograph of the screen display of the mentioned 800 computer graphics. Computer graphics are displayed on a 17-inch (gamma linearized) LCD monitor screen with a display resolution of 1280×1024 and photographed by a Canon G3 digital camera. The acquisition is conducted in a dark room in order to reduce the reflections from the ambient scene.

Google image are from `http://www.geocities.com/nowarski7/ta/02110602.jpg` (first row) and `http://associate.com/photos/Samples-n-Things/fruit-bowl.jpg` (second row). The CG images are from `http://www.realsoft.fi/gallery/vehicles/scania.jpg` (first row) and `http://www.marlinstudios.com/gallery/cgallery/summerfun/sunshine.htm` (second row).

Despite the two datasets, there are many problems that also call for a benchmark dataset. For instance, the experiments involving the physical image model based on the camera characteristics require a dataset of images acquired by a diverse models of camera, at various acquisition settings. Furthermore, in order to facilitate the evaluation of the image forgery detection techniques using the images produced by the state-of-the-art image forgery creation techniques, a dataset of these images would be necessary. Therefore, further effort on producing and standardizing the additional benchmark dataset is needed.

4.5 Security Measure

Once a forgery creator has an unlimited access to the forgery detector, an oracle attack can be launched, where the forgery creator incrementally modifies the created forgery according to the detection results from the detector until it passes the detector. Such an attack is also a serious threat to the public watermarking system. For the incremental modification to be efficient and have a minimum distortion on the image, the attacker needs to identify the shortest path to the decision boundary of the detector. This is possible when the decision boundary is known to the attacker.

With an unlimited access to the detector, a parametric decision boundary can be estimated by the following procedure. The attacker first locates the sample points on the decision boundary by incrementally modifying some sample images until it just crosses the boundary. Then, the parameters of the decision boundary can be estimated by using the boundary points, so long as the number of the boundary points is equal or greater than the number of the parameters. In most cases, the number of the parameters is not too large and the estimation is feasible. To make the estimation more challenging, the work in [38] proposes a method of converting the parametric decision boundary of a detector into a fractal (non-parametric) one, so that an accuracy estimation of the boundary requires a much larger number of sample points on the decision boundary. However, there is a tradeoff where the fractal boundary should not be very well approximated by the original parametric decision boundary (ensuring high security), while the excessive deviation for the original boundary should be avoided (minimizing image distortion).

For another work in [39], the author addresses the oracle attack issue by modifying the temporal behavior of the detector such that the duration for returning a decision is lengthened when an oracle attack is suspected based on the sequence of input images. The hallmark of an oracle attack is the sequential input images with a similar content. The delay strategy for the lazy detector with memory can be designed so that the total time duration needed for an oracle attack to succeed is painfully long.

5 Summary and Conclusion

In this chapter, we have given a review for passive-blind image forensics, beginning with an overview and followed by a detailed review on the two main problems of image forensics, i.e., image forgery detection and image source identification. We also provide a description of our thoughts on the resources and the challenges concerning passive-blind image forensics.

Passive-blind image forensics is still a research area at its infancy. There are fundamental issues related to the physical model parameter estimation, the practical system design issues and the system security issues which remain to be addressed. For an effective solution to these issues, expertise from various domain such as expertise from various domains such as computer vision, signal processing, computer graphics, machine learning, imaging sensors, and even mechanical systems are needed. On the other hand, it is reasonable to envision that the digital watermarking techniques could be used in conjunction to the passive-blind image forensic methods. Therefore, the combined active and passive approach may be another future research direction.

References

[1] C. Amsberry, "Alterations of photos raise host of legal, ethical issues," *The Wall Street Journal*, Jan 1989.

[2] G. Friedman, "The trustworthy digital camera: Restoring credibility to the photographic image," *IEEE Transactions on Consumer Electronics*, vol. 39, no. 4, pp. 905–910, November 1993.

[3] S. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. Dean, and E. Felten, "Reading between the lines: Lessons learned from the sdmi challenge," in *Usenix Security Symposium*, Washington D.C., August 2001, pp. 353–363.

[4] H. Farid, "Detecting digital forgeries using bispectral analysis," MIT, MIT AI Memo AIM-1657, 1999. [Online]. Available: ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1657.pdf

[5] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Workshop on Statistical Analysis in Computer Vision*, Madison, Wisconsin, 2003.

[6] T.-T. Ng and S.-F. Chang, "A model for image splicing," in *IEEE International Conference on Image Processing*, Singapore, October 24-27 2004.

[7] T.-T. Ng, S.-F. Chang, and Q. Sun, "Blind detection of photomontage using higher order statistics," in *IEEE International Symposium on Circuits and Systems*, Vancouver, Canada, 2004.

[8] I. Avcibas, S. Bayram, N. Memon, M. Ramkumar, and B. Sankur, "A classifier design for detecting image manipulations," in *IEEE International Conference on Image Processing*, vol. 4, Singapore, Oct 2004, pp. 2645 – 2648.

[9] Z. Lin, R. Wang, X. Tang, and H.-Y. Shum, "Detecting doctored images using camera response normality and consistency," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 1087–1092.

[10] G. N. Ali, P.-J. Chiang, A. K. Mikkilineni, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, "Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices," in *IS&T International Conference on Digital Printing Technologies*, 2003, pp. 511–515.

[11] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.-P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *ACM Multimedia*, Singapore, November 2005.

[12] M. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *ACM Multimedia and Security Workshop*, New York, NY, 2005.

[13] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.

[14] T.-T. Ng and S.-F. Chang, "A data set of authentic and spliced image blocks," Columbia University, ADVENT Technical Report 203-2004-3, June 2004. [Online]. Available: http://www.ee.columbia.edu/trustfoto

[15] A. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, 2005.

[16] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric calibration from a single image," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, June 2004, pp. 938–945.

[17] S. Lin and L. Zhang, "Determining the radiometric response function from a single grayscale image," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, June 2005, pp. 66–73.

[18] T.-T. Ng, S.-F. Chang, and M.-P. Tsui, "Camera response function estimation from a single-channel image using differential invariants," in *review*, 2005.

[19] A. Popescu and H. Farid, "Statistical tools for digital forensics," in *6th International Workshop on Information Hiding*, Toronto, Canada, 2004.

[20] D. Mahajan, R. Ramamoorthi, and B. Curless, "Spherical harmonic convolution for inverse rendering, BRDF/lighting transfer and image consistency checking," in *review*, 2005.

[21] A. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 758–767, 2005.

[22] ——, "Exposing digital forgeries by detecting duplicated image regions," Computer Science, Dartmouth College, Tech. Rep. TR2004-515, 2004. [Online]. Available: http://www.cs.dartmouth.edu/~farid/publications/tr04.pdf

[23] J. Lukas, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *SPIE Electronic Imaging, Photonics West*, January 2006.

[24] T. Ianeva, A. de Vries, and H. Rohrig, "Detecting cartoons: A case study in automatic video-genre classification," in *IEEE International Conference on Multimedia and Expo*, vol. 1, 2003, pp. 449–452.

[25] J. R. Smith and S.-F. Chang, "Visually searching the web for content," vol. 4, no. 3, pp. 12–20, 1997.

[26] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 845–850, February 2005.

[27] B. B. Mandelbrot, *The fractal geometry of nature*. San Francisco: W.H. Freeman, 1983.

[28] A. Pentland, "On describing complex surface shapes," *Image and Vision Computing*, vol. 3, no. 4, pp. 153–162, November 1985.

[29] T. Akenine-Moller, T. Moller, and E. Haines, *Real-Time Rendering*. MA: A. K. Peters, Ltd., 2002.

[30] A. B. Lee, K. S. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *International Journal of Computer Vision*, vol. 54, no. 1, pp. 83–103, 2003.

[31] K. Mehdi, H. Sencar, and N. Memon, "Blind source camera identification," in *IEEE International Conference on Image Processing*, vol. 1, Singapore, Oct 2004, pp. 709–712.

[32] A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Printer identification based on texture features," in *IS&T International Conference on Digital Printing Technologies*, 2004, pp. 306–312.

[33] T.-P. Wu and C.-K. Tang, "A bayesian approach for shadow extraction from a single image," in *IEEE International Conference on Computer Vision*, Beijing, China, October 2005.

[34] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" in *IEEE Computer Vision and Pattern Recognition*, Washington, DC, USA, 1996, p. 270.

[35] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *ACM SIGGRAPH*, 1999, pp. 187–194.

[36] T.-T. Ng, S.-F. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," Columbia University, ADVENT Technical Report 205-2004-5, Feb 2005. [Online]. Available: http://www.ee.columbia.edu/trustfoto

[37] Calphoto, "A database of photos of plants, animals, habitats and other natural history subjects," University of Berkeley, 2000. [Online]. Available: http://elib.cs.berkeley.edu/photos/

[38] A. Tewfik and M. Mansour, "Secure watermark detection with non-parametric decision boundaries," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, pp. 2089–2092.

[39] I. Venturini, "Counteracting oracle attacks," in *ACM multimedia and security workshop on Multimedia and security*, Magdeburg, Germany, 2004, pp. 187–192.

[40] E. N. Mortensen and W. A. Barrett, "Intelligent scissors for image composition," in *ACM SIGGRAPH*, 1995, pp. 191–198.

[41] M. Gleicher, "Image snapping," in *ACM SIGGRAPH*, 1995, pp. 183–190.

[42] J. Reese and W. Barrett, "Image editing with intelligent paint," in *Eurographics*, vol. 21, no. 3, Saarbrcken, Germany, 2002, pp. 714–723.

[43] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *IEEE International Conference on Computer Vision*, vol. I, 2001, pp. 105–112.

[44] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM SIGGRAPH*, vol. 23, no. 3, pp. 303–308, 2004.

[45] P. E. Debevec, C. J. Taylor, and J. Malik, "Facade: Modeling and rendering architecture from photographs," in *ACM SIGGRAPH*, 1996.

[46] L. Zhang, G. Dugas-Phocion, J.-S. Samson, and S. M. Seitz, "Single view modeling of free-form scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[47] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[48] Y.-Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[49] H.-Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C.-K. Tang, "Pop-up light-field," *ACM Transaction on Graphics*, vol. 23, no. 2, 2004.

[50] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum, "Poisson matting," *ACM Transaction on Graphics*, vol. 23, no. 3, pp. 315–321, 2004.

[51] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transaction on Graphics*, vol. 2, no. 4, pp. 217–236, 1983.

[52] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transaction on Graphics*, vol. 22, no. 3, pp. 313–318, 2003.

[53] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *IEEE International Conference on Computer Vision*, 1999, pp. 1033–1038.

[54] M. Bertalmio, A. Bertozi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

A Forgery Creation Techniques

This section will provides a review of the automatic or the semi-automatic computer techniques for image forgery creation. The following subsections are according to the image forgery creation model shown in Figure 1.

A.1 Region Selection

While automatic image segmentation still leaves much to be desired, various interactive foreground-background image segmentation techniques have been invented. From the operational perspective, these techniques can be categorized into boundary-based methods [40, 41] and region-based methods [42, 43]. For boundary-based methods, users approximately trace the object boundary and the algorithm interactively refines the traced contour so that the background and the foreground are well segmented. Whereas in region-based methods, users mark the background and the foreground region, then the algorithm finds a contour to separate the two regions. The well-known Magic Wand and Intelligent Scissor in Photoshop are respectively a region-based method and a boundary-based method. There are techniques, such as Lazy Snapping [44] which combine the benefits of the two methods.

Figure 15: An example of single view modeling. (Figure courtesy of Li Zhang et al. and IEEE)

A.2 3D Model Extraction

Human has a good capability of extracting 3D scene structure from a single image, even under the condition of mono-vision. Mimicking such a human capability has been one of the focuses in computer vision research. A 3D morphable human face model [35] was used to extract the 3D human face model from a single image with some user's intervention for matching the correspondence points in the image. Once the 3D face model has been extracted, manipulation of the face feature such as changing the facial expression, altering the fullness and the gender characteristics of the face is possible. There are various other algorithms being proposed for the single-view 3D model reconstruction from a more generic scene with planar or other simple geometric primitives [45], such as a scene with buildings. For images of a more complex scene structure, a semi-automatic method for reconstructing the 3D structure of an arbitrary free-form curved surface from a single image using the sparse surface normals supplied by the user is demonstrated in [46] and shown in Figure 15[7].

A.3 Geometric Transformation

The common geometric transformation applied to an image fragment before being pasted onto another image includes translation, Euclidean transformation (translation and rotation), similarity transform (scaled rotation and translation), affine transform (a transform that preserves parallel lines) and projective transform (a transform that preserves straight lines). Mathematically, these transformation can be represented by a linear transform with a possibly constrained 3×3 matrices operating on 2D homogenous coordinate vectors [47]. These transformation takes the perspective projection of the camera into account. Interestingly, in his 1917 book *On growth and form*, biologist D'Arcy Thompson showed that different species of fish can be related by a simple geo-

[7]The image is extracted from `http://grail.cs.washington.edu/projects/svm/CGW/single_view_modeling.htm`, courtesy of Li Zhang et al. and IEEE

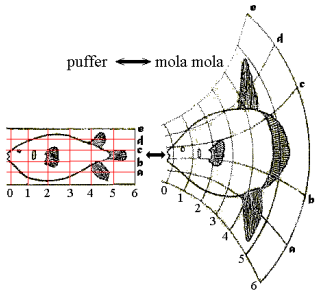

Figure 16: An example showed by D'Arcy Thompson for transforming a type of fish called puffer to another type called mola mola .

metric transformation, as shown in figure 16[8]. The similar type of transformation is also possible for human skull.

A.4 Compositing

Direct pasting of an image fragment onto another image would introduces visually perceptible seams. To produce a natural-looking composite image, matting or blending is usually performed. Matting is to mix the pixels near the fragment boundary by weighted sum of the pixels of the fragments and those of the original image. The weight is given by the matte which needs to be estimated. There are various ways to estimate the matte given a user-supplied *trimap*, which is a tri-region partitioning for "definitely foreground", "definitely background" and "unknown" regions. The examples for the matting methods are bayesian matting [48], coherence matting [49] and poisson matting [50]. The blending technique is more than just blending of near-boundary pixels, it has the capability of realigning the exposure differences and other misalignments between the pasted image fragments and the host image. This form of blending can be done by directly compositing of the multi-resolution version of the image fragments in a Laplacian pyramid [51] and the final image is recovered from the composite Laplacian pyramid. In another technique, direct compositing is performed in the gradient domain and the final composite image is recovered by solving a partial differential equation [52].

A.5 Retouch and Object Removal

At the final stage of the image editing pipeline, the composite image may be retouched by airbrushing to remove the remaining artefact or the minor/narrow objects like the overhead electrical wiring. However, removal of a larger-size foreground object is also possible and this practice is sometimes known as *reverse cropping*. After an object is removed, the resulted void needs to be filled in with the background pixels. This filling-in process is known as *image inpainting*. The

[8]The image is extracted from `http://www.bio.umass.edu/biology/kunkel/shape.html`

simple image inpainting technique would be to synthesize the background texture to fill in the empty region [53]. This technique works well for homogenous textured background. The more complex image inpainting algorithm takes the geometric structure at the surrounding of the void into account. One method employs the classical Navier-Stokes equation in fluid dynamics to propagate the geometric structure from the surrounding region into the void [54] and achieves promising inpainting results.