Discrimination of Computer Synthesized or Recaptured Images from Real Images

Tian-Tsong Ng and Shih-Fu Chang

Abstract An image that appears to be a photograph may not necessarily be a normal photograph as we know it. For example, a photograph-like image can be rendered by computer graphics instead of being taken by a camera or it can be a photograph of an image instead of a direct photograph of a natural scene. What is really different between these photographic appearances is their underlying synthesis processes. Not being able to distinguish these images poses real social risks, as it becomes harder to refute claims of child pornography as non-photograph in the court of law and easier for attackers to mount an image or video replay attack on biometric security systems. This motivates digital image forensics research on distinguishing these photograph-like images from true photographs. In this chapter, we present the challenges, technical approaches, system design, and other practical issues in tackling this multimedia forensics problem. We will also share a list of the open resources and the potential future research directions in this area of research which we hope readers will find useful.

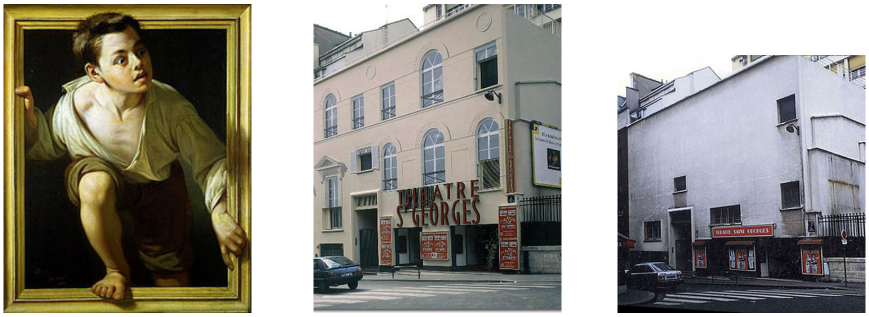
1 Motivations

Since the ancient times of the Greeks and Romans, artists have been playing with special painting techniques for inducing visual illusion where objects in a painting appear to be real and immersed in the real surrounding. *Trompe l'oeil*, the name for such visual artistry, literally means *deceiving the eye*. For example, the painting entitled *Escaping Criticism* created by Pere Borrell del Caso in 1874 depicts a person climbing out of the painting with a make-believe quality (Figure 1a); the mural painting on the facade of the Saint-Georges Theater created by a painter Dominique

Tian-Tsong Ng
Institute for Infocomm Research, Singapore 138632. e-mail: ttng@i2r.a-star.edu.sg

Shih-Fu Chang
Columbia University, New York, NY 10027, USA. e-mail: sfchang@ee.columbia.edu

Antony induces an impression of balcony (Figure 1b), while the facade is in fact just a flat wall (Figure 1c).



(a) A painting, *Escaping Criticism* created by Pere Borrell del Caso in 1874.

(b) The facade of Saint-Georges Theater in Paris, France, created by the mural painter Dominique Antony

(c) The facade of Saint-Georges Theater before the mural painting

Fig. 1: The art of visual deception, *trompe l'oeil*.

Trompe l'oeil makes believe with not only the photorealistic quality in the painting, it also exploits the visual gullibility of human observers through immersion into the real surrounding. Such adversarial nature of *trompe l'oeil* aptly mirrors that of digital image forensics where the intention to deceive is present. While the deceptive intent is from humans, the photorealism of an image which takes many years of practice for conventional artists to master can be easily produced by laymen with modern technology. Physics-based computer graphics is capable of rendering photorealistic images that emulate images of real three-dimensional scenes, a great advancement from the two-dimensional graphics like cartoon which was popular in the early days of computer graphics. It has been shown that a scene with diffuse reflectance can be realistically rendered to the extent that the rendered scene radiance is close to that of a real scene and perceptually indistinguishable for humans [43].

A photograph is photorealistic by definition. Recapturing a photograph with a camera when displayed in good quality on a paper or screen preserves the photorealism of the photograph if perspective distortion is minimized [1]. We refer to such types of images as *recaptured or rephotographed image*. The modern artist Richard Prince pioneered a unique art form of rephotographing advertisements from magazines where the artistic expression was conveyed through intensifying certain distinct characteristics of the reproduced image with various photographing techniques such as blurring, cropping, and enlarging, while preserving the photorealism of the original images.

[1] The perspective distortion, in the form of planar homography, due to image recapturing can result in a non-zero skew in the camera internal parameters [24] and could appear visually unnatural to human observers

As synthesizing photorealism gets easier, seeing a red apple can no longer immediately imply the actual presence of the apple, i.e., the link between the image of an object and the presence of the object is weakened. The weakened link poses real security risks. In the US, possession of child pornography is punishable as it implies abuse of minors. However, establishing the presence of minors from the child pornography is challenging on legal ground, as owners of child pornography can proclaim the images to be computer generated [13] [2]. An important implication of the weakened link is the need for technology to recognize the underlying formation process of an image.

On the other hand, a photorealistic rephotographed image is instrumental for *image/video replay attack* on biometric authentication systems [7]. In 2008, the Vietnamese Internet Security Center BKIS demonstrated the ease of breaching the face authentication login in commercial laptop computers using printouts of face images of a legitimate user [54]. Similar vulnerability is also shown for the face authentication login in a version of Andriod operating system known as Ice-cream Sandwich introduced in October 2011 [33]. The ease of accessing someone's face images on the Internet has made the image replay attack problem loom larger.

Although computer graphics and rephotographed images can be as perceptually photorealistic as real photographs, their underlying image formation processes have distinctive characteristics. Such differences, though subtle, can be used for distinguishing these images as described in Section 4 and 5. Being able to recognize the underlying image formation process has far-reaching impacts. Such capability will make it harder for child pornography owners to get away with the computer graphic claim and harder for image/video replay attack to succeed. It will also make it harder for doctored images to escape detection through rephotographing, which can turn a doctored image into a quintessential photograph.

In essence, detecting rephotographed image is related to a fundamental problem in computer vison: *monocular depth perception*. It is important that a robot with monocular vision [45] does not confuse objects in a poster as real, as the two have greatly distinct semantics in a physical scene. Such monocular depth perception is also useful for algorithms that convert the conventional 2D movies into 3D content for display on 3D televisions.

General-class or specific object recognition has been widely researched for high-level computer vision. Object recognition is generally approached through appearance-based recognition, hence it does not differentiate a red apple for example from a red apple in a picture found in a scene. This is evidenced from the form of the public benchmark data sets for object recognition such as Caltech 101 [16, 39, 12, 17, 70], there is no object class called poster for example. However, a

[2] The ruling of the United States Supreme Court in 2002 on a clause in the 1996 Child Pornography Prevention Act (CPPA) defines child pornography as *any* visual depiction of explicit sexual conduct that involves a child. However, in 2002, the United States Supreme Court considered the broad definition of child pornography in CPPA that includes "virtual imagery" as unconstitutional and violating the freedom of speech as enacted in the First Amendment. As a result, the Court ruled that computer generated images including those with child pornography content are to be protected constitutionally.

versatile object recognition system ideally should be capable of recognizing a scene picture such as a poster in the scene. The capability of detecting rephotographed images will augment the functionality of the current object recognition algorithms.

In Section 2, we show the evidence that distinguishing photorealistic computer graphics and rephotographed images from true photographs is challenging. In Section 3, we describe the desired characteristics in an algorithm for recognizing the underlying image formation of images when considering various fundamental and application-related issues. In Section 4 and 5, we survey the various approaches for distinguishing photorealistic computer graphics and rephotographed images from photographs. It is common for a security system to face threat of attacks. In Section 6, we describe the potential attacks on a computer graphics or recaptured image detector and the corresponding counterattack measures. In Section 7, we give a list of resources useful for researchers, including a few open benchmark data sets and a public evaluation system. Finally, we describe the open issues and future direction for this area of research in Section 8, before concluding in Section 9.

2 Challenges

Visual realism is no longer a hallmark exclusive for common photographs. We will look into the level of visual realism achievable by computer graphics rendering and image rephotographing, and the challenges for humans to discern them perceptually. We also explore the possibility of discerning these image formative processes using the computer.

2.1 Visual Realism of Computer Graphics

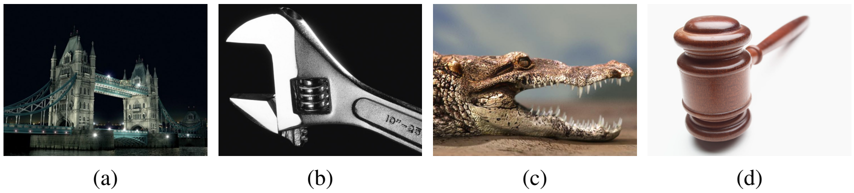

(a) (b) (c) (d)

Fig. 2: Visually challenging images from the Fake or Foto website. The true label of these images, photographic or computer generated, can be found in Appendix.

One of the important goals of computer graphics rendering is to produce photorealistic imageries that are perceptually close to real-scene images. Real-scene radiance induces visual stimuli that constantly impacts the human visual system. Through biological evolution, the human visual system is adapted to such visual stimuli [65] and hence develops a keen sensory for real-scene images. Ferwerda [18] defined three varieties of realism: *physical realism* which provides the same visual stimulation as the real-world scene, *photorealism* which produces the same visual response in humans as the scene, and *functional realism* which allows humans to receive the same visual information, e.g., object shape and scene depth, as from the real scene. From the definition, achieving photorealism does not require faithful reproduction of real-scene radiance, although the physics-based computer graphics based on Kajiya's rendering equation [27] is capable of simulating real-scene radiance or achieving physical realism.

Studies of computer graphics photorealism and its perception are of interest to the computer graphics community, as the level of achievable photorealism represents a measure of success for computer graphics research. Knowing the visual elements of photorealism provides a guide for trading off rendering accuracy for efficient rendering without compromising photorealism. Such studies on perception of photorealism offer some clues about the perceptual differences between photographs and computer graphics.

Meyer *et al.* [43] asked human observers to label two images of the same scene displayed side-by-side on a monitor display. One of the images is photograph while the other is a computer graphics image rendered with a radiosity algorithm. The human observers found these images perceptually indistinguishable. McNamara [44] conducted a similar experiment with a diffuse scene that is more complex and found that computer graphics rendering that simulates up to second bounce reflection is sufficient to achieve photorealism for such types of scenes.

Rademacher *et al.* [59] performed a set of experiments on human perception to study what visual clues contribute to photorealism. They found that the softness of shadow and surface roughness correlate positively with photorealism perception, while scene complexity and the number of scene lightings do not show such correlation. This result implies that computer graphics can be made more believable to humans by manipulating its soft shadow and rough surface despite the fact that hard shadow and smooth surface do exist in real scenes. Such a trick has been employed in the *Fake or Foto* [3] visual quiz, where human observers are asked to label ten images as real or computer generated purely through visual inspection. Four images from the quiz are shown in Figure 2.

Farid and Bravo [14] conducted a series of psychophysical experiments that used images of varying resolution, JPEG compression, and color to explore the ability of human observers to distinguish computer generated upper body images of people from photographic ones. The computer graphics images used in the experiments are downloaded from the Internet. The experiments provide a probability that an image that is judged to be a photograph is indeed a true photograph, which has 85% reliability for color images with medium resolution (between 218 × 218 and 436 × 436 pixels in size) and high JPEG quality. The reliability drops for lower resolution

[3] `http://area.autodesk.com/fakeorfoto`

and grayscale images. This work shows that the computer graphics of human images in the Internet are quite distinguishable for human observers. This may indicate the level of difficulty for rendering highly photorealistic human images. However, this may change as computer graphics progresses while the ability of human observers remains unchanged.

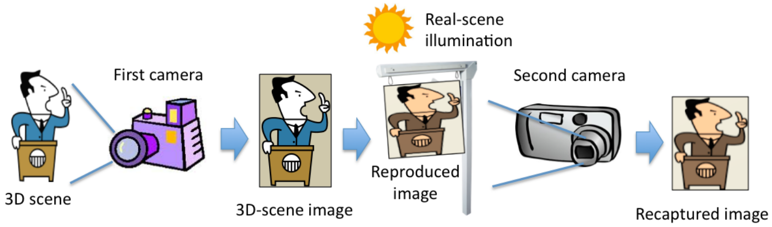
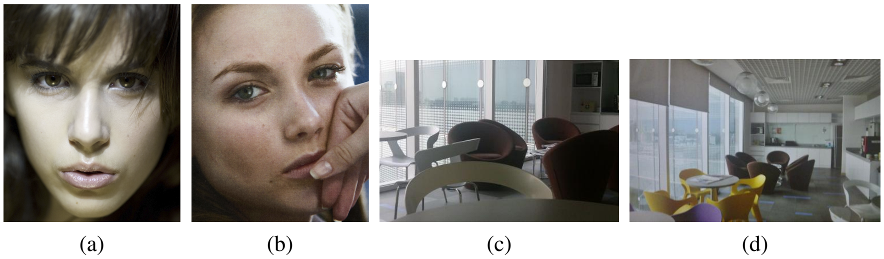


Fig. 3: An image recapturing pipeline.

2.2 *Visual Realism of Recaptured Image*

The process of rephotographing in general involves steps as shown in Figure 3. A 3D scene is first captured as an image and reproduced on a physical surface such as a printing paper or an LCD display before it is recaptured again under a different illumination. Under control environment, perspective distortion can be minimized by placing the image reproduction surface such that it is parallel to the camera's sensor plane, in order to reduce the effective skew in the internal parameters of the recaptured image [24]. In general, the rephotographing process is pure image-based and involves no graphics models or rendering, unless the first image is computer graphics. Recaptured images are also different from the common photographs in that what being captured is an image reproduction surface instead of a general scene.

Human observers may expect specific color tints associated with an image reproduction surface to be found in a recaptured image. However, such shades of color can exist in a real scene due to illumination. Figure 4 shows four images where two are produced by rephotographing a color laser printout on an ordinary office paper. Readers would notice how misleading the color clue could be for distinguishing recaptured images.



(a) (b) (c) (d)

Fig. 4: One of the human face images (a) and (b) and one of the office scene images (c) and (d) were obtained by rephotographing a color laser printout on an ordinary office paper. The true label of these images, captured from true 3D scene or recaptured from 2D color laser printout, can be found in Appendix.

2.3 Statistics of Computer Graphics and Recaptured Images

Natural images, as opposed to microscopic, astronomic, aerial, or X-ray images, are images of the everyday scenes which serve as natural stimuli to the human visual system. Natural images are believed to live in a very small subspace within the large image space. To characterize this subspace, researchers have proposed various statistics which demonstrate regularity over nature images [66]. One of the important natural image statistics is the sparse distribution of the wavelet coefficients of natural images that are aptly modeled by a generalized Laplacian density [40].

As a crude evaluation, we show in Figure 5 the wavelet coefficient distributions of the second-level horizontal subbands, respectively, for a photograph, a computer graphics, a non-recaptured photograph, and its corresponding recaptured image. Their distributions appear to be visually similar and modeled well by a generalized Laplacian density. This experiment indicates the statistical similarity of these different types of images at a crude level. More detailed statistics has been considered for distinguishing different image types [56, 47].

2.4 Automatic Detection using Computer

By design, the current computer vision systems are rarely equipped with the capability to recognize an image beyond its appearance. The plain vulnerability of the laptop's face authenticator is a good example [54] and the various object recognition data sets [16, 39, 12, 17, 70] remain restricted to the pure appearance-based approach. The main reason is that the awareness about the potential security risks due to the inability to distinguish the various underlying image formation processes remains low.

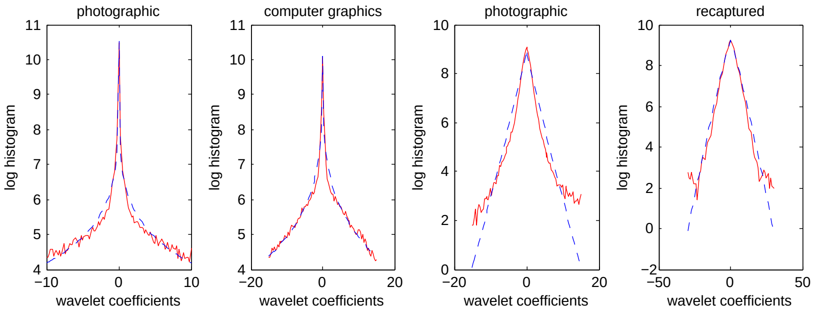


Fig. 5: The log-histogram of the first-level detail wavelet coefficients computed using Daubechies 8 filters for the photographic image in Figure 2b, computer graphics image in Figure 2a, photographic face image in Figure 4a, and recaptured face image in Figure 4b (from left to right). The dashed is the least-squared fitted generalized Laplacian density.

How does the computer perform as compared to human observers in recognizing an image beyond image appearance? When it comes to making visual judgement, human observers suffer from various forms of subjective bias and are insensitive to the low-frequency differences in visual signals. In contrast, the computer is objective and particularly good at picking up fine features in signals. For example, a computer algorithm is capable of extracting the camera curve property from a single image [53] while this low-frequency and global signal is largely imperceptible to humans. As shown in Section 4.6, computer-based detection meets with some success in classifying computer graphics and recaptured images from photographs. However, its performance will be much lower in an adversarial setting when the intention to deceive is considered.

3 Pattern Recognition System and Design Issues

A pattern recognition system computes a pattern from an input and matches it to a pattern model to form a decision [11]. This process in general involves *sensing*, *preprocessing*, *feature extraction*, *classification*, and *post-processing* as shown in Figure 6. A recognition candidate is sensed and the input is preprocessed to remove the irrelevant information, e.g., segmenting the foreground objects from an input image. Pattern recognition is mainly based on distinguishing features of the targets, e.g., the distinctive color offers a good feature for distinguishing apples and oranges. The classifier then assigns the input pattern to one of the pattern models which are obtained from a set of previously observed data known as training samples before the decision is mapped into a recommended action.

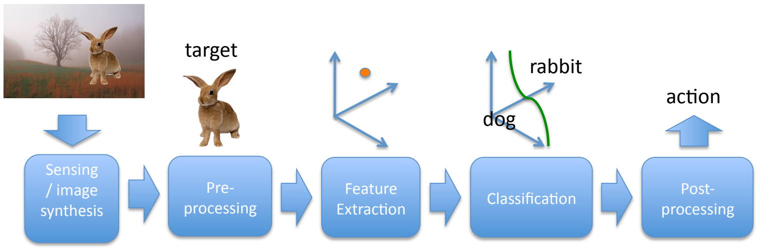


Fig. 6: The typical process in a pattern recognition system.

Although computer graphics and recaptured image recognition can be approached through pattern recognition as in object recognition, these are new problems with unique characteristics as below:

Recognizing Image Formation. For typical pattern recognition such as object recognition, it is the sensor's output (e.g., image appearance) but not the sensor per se (e.g., image formation process) that is of interest. In contrast, for computer graphics and recaptured image recognition, the image formation process is the object of recognition and the image appearance per se is largely immaterial. Therefore, these problems call for features that are beyond visual appearance such as those inspired by natural image statistics and steganalysis (see Section 4 and 5).

Vague Definition of Classes. In most pattern recognition problems, the target of recognition is well-defined. For example, gender recognition is a two-class recognition problem with well-defined male and female classes. However, the computer graphics images that we may encounter in real life may not be as well-defined in that the images may have been textured-mapped with photographic inputs or they may be images of a computer graphics foreground object composed with a photographic background scene. On the other hand, a recaptured image may be an image of a picture set against a natural-scene background. Ideally, instead of discrete classes, there should be a continuous measure for the level of computer-graphics-ness or recaptured-ness in an image.

Dynamic Definition of Classes. If there are distinctive features among photorealistic computer graphics, recaptured images, and photographic images, these features are technology dependent. With the advancement in graphics rendering techniques and image reproduction devices, some of the distinctive features may disappear while new ones may surface. This points to the dynamic nature of the image class model that evolves with time. To maintain the effectiveness of the pattern recognition system, adaptiveness may be essential. Furthermore, in an adversarial setting, the system also needs to adapt to the attack patterns.

Potential Application in the Court of Law. For computer graphic detector to be useful for forensics, it has to comply with the forensics procedure and requirements. As forensics is an endeavor to use scientific methods to gain probative facts for criminal investigations, ensuring the objectivity and reliability of the forensics outcome is the primary requirement. The reliability and robustness of the decision also needs to be validated with rigorous and comprehensive test procedure to demonstrate its stability in the presence of noise. For admissibility to the court, the system should minimize the chance of falsely incriminating an innocent person by lowering the false position rate for instance. As the detection result will be debated in the court, any physical intuition on the detection result will make it more accessible and convincing to the legal professionals who may lack the technical background to grasp highly abstract technical details.

Other Practical Issues. A simple classification model and fast processing are also important. A simple classification model with small number of features will simplify the classification training procedure and requires a smaller number of training images. Fast processing is important especially for recaptured image detection so that a face authenticator secure against image or video replay attack can operate at an interactive rate.

3.1 Evaluation Metric

The current works mainly consider the problem of distinguishing photorealistic computer graphics or recaptured images from photographic images as a two-class classification problem. The performance of a binary classifier can be measured by a classification confusion matrix at an operating point of the classifier:

$$\begin{bmatrix} p(\mathrm{C} = \mathrm{positive} \mid \mathrm{L} = \mathrm{positive}) & p(\mathrm{C} = \mathrm{negative} \mid \mathrm{L} = \mathrm{positive}) \\ p(\mathrm{C} = \mathrm{positive} \mid \mathrm{L} = \mathrm{negative}) & p(\mathrm{C} = \mathrm{negative} \mid \mathrm{L} = \mathrm{negative}) \end{bmatrix}, \quad (1)$$

where the two classes are respectively identified as positive and negative while C and L respectively represent the assigned label (by the classifier) and the true label. The probability $p(\mathrm{C} = \mathrm{positive} \mid \mathrm{L} = \mathrm{negative})$ is known as false positive rate, and $p(\mathrm{C} = \mathrm{negative} \quad \mid \mathrm{L} = \mathrm{positive})$ false negative rate. The averaged classification accuracy can be computed as

$$\frac{p(\mathrm{C} = \mathrm{positive} \mid \mathrm{L} = \mathrm{positive}) + p(\mathrm{C} = \mathrm{negative} \mid \mathrm{L} = \mathrm{negative})}{2}. \quad (2)$$

The operating point of a classifier can be adjusted by shifting the decision boundary or threshold values which in turn adjust the balance of the false positive and false negative rates. *Equal error rate* is often used for evaluating a biometric system. Equal error rate refers to the false positive rate or the false negative rate of a classifier when it functions at an operating point where the two rates are equal.

4 Approaches for Photorealistic Computer Graphics Detection

Computer graphics detection has been a problem of interest since the early days of content-based image retrieval [62]. The early work focuses on non-photorealistic computer graphics [1, 36, 72]. Only recently, digital image forensics [51] provides a strong motivation to study the problem of identifying photorealistic computer graphics.

Non-photorealistic computer graphics images such as cartoons, cliparts, logos and line drawings are abundant in the Internet. There are commercial incentives to separate such graphics from photographs. For web search companies, being able to identify non-photorealistic graphics offers a value-added service to web users who wish to find cliparts to enhance their presentation slides. On the other hand, this capability could improve the precision of image search by filtering out graphics images when users is interested in photographs. Some companies may be interested in scanning the logo images in the Internet to detect trademark infringement. In all the above-mentioned applications, computation speed is crucial to ensure interactive user experience.

In the Internet, graphics images are mainly kept in common image formats as photographs and metadata often offers no clue for the image type. Therefore, image content features are used for identifying computer graphics.

4.1 Methods using Visual Descriptors

Visual descriptors refer to features motivated by visual appearance such as color, texture, edge properties, and surface smoothness. Ideally, visual descriptors should not be effective in identifying photorealistic computer graphics aiming at simulating the appearance of photograph. However, if we look at photographic and photorealistic computer graphics images on the Internet, the *distribution of their visual properties* may be different. For instance, the level of difficulty in rendering a scene increases with its geometric and photometric complexity, hence computer graphics of lower complexity may be more common than the more complex ones on the Internet. However, there is no evidence that the photographic images on the Internet have such distributional property. Computer graphics of lower complexity here refers to images of simpler scene, less color variation, or simpler textures.

Color and Edges. Simple visual descriptors have been proposed to identify non-photorealistic computer graphics [1, 36, 72, 26, 6]. For example, Ianeva *et al.* [26] observed that the cartoonist graphics has characteristics of saturated and uniform colors, strong and distinct lines, and limited number of colors. They devised a computer graphics detection method that used image features such as the average color saturation, the ratio of image pixels with brightness greater than a threshold, the Hue-Saturation-Value (HSV) color histogram, the edge orientation and strength histogram, the compression ratio, and the distribution of image region sizes. Their work is motivated with the goal for improving the accuracy of video key-frame retrieval through graphics image prefiltering. As computational efficiency is crucial for graphics detectors, Chen *et al.* [6] focused on this computational aspect of Web graphics and photographs classification.

Color and Texture. Wu *et al.* [76, 77] used visual clues such as the number of unique colors, local spatial variation of color, ratio of saturated pixels, and ratio of intensity edges to classify computer graphics and photographs. With these features, on their undisclosed data set, a k-nearest neighbor (k-nn) classifier was able to achieve an average accuracy of 76%. They also considered Gabor texture descriptor which enabled a k-nn classifier to achieve an average classification accuracy of 95% while a *support vector machine* (SVM) classifier only achieved 75% with the same set of features.

Fractal properties. Pan *et al.* [56] considered that photorealistic computer graphics on the Internet is more surreal in color and smoother in texture as compared to photographic images. Fractal dimension, a self-similarity measure, was proposed to describe the mentioned characteristics. From a scalar image $I(x,y) \in [0,1]$, a set of N binary images are computed through thresholding:

$$I_k(x,y) = \mathbf{1}\left(\frac{k-1}{N} \leq I(x,y) \leq \frac{k}{N}\right), \tag{3}$$

using an indicator function $\mathbf{1}(\cdot)$ that equals to 1 when the input argument is true, 0 otherwise. The fractal dimension of a binary image can be estimated with a simple method such as box counting [42]. They computed the simple fractal dimensions on the hue and saturation components of an image in the HSV color space. They also extracted a generalized fractal dimension [22] from an image to offer more detailed local information. On their undisclosed data set, the SVM classifier achieved an average test accuracy of 91.2%, while the method by Lyu and Farid [38] achieved 92.7% on the same data set.

4.2 Methods from Image Formation Process

Incorporating knowledge from the problem domain can potentially lead to a simpler pattern model which requires less data for training. The problem of identifying photorealistic computer graphics and photographic images is essentially a problem of identifying the different image formation processes. Therefore, a detailed understanding on the image formation processes offers an inroad into the problem. As estimating the parameters of the generative models from a single image is mostly ill-posed, the existing methods mainly extract distinguishing features motivated by the generative models indirectly.

4.2.1 Formation of Photograph

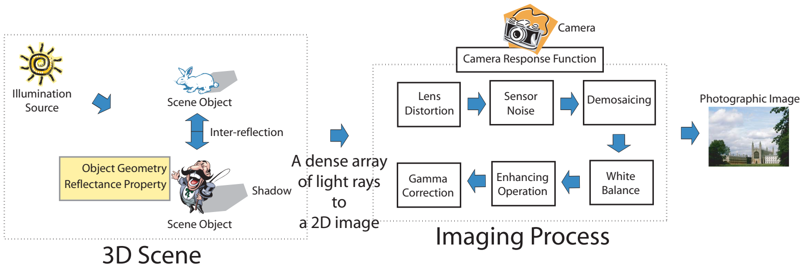


Fig. 7: Photographic image formation process.

Photographic images are in general snapshots of natural scenes with a camera. Technically, a camera samples the light radiance reflected from the scene with its optical sensor followed by a series of in-camera processing, as shown in Figure 7. The scene radiance varies with the light sources, geometry, and reflectance properties of the scene. Scene radiance could also be altered by the optical property of the participating medium such as fog and haze through which the light travels. Hence, the properties of real-world scenes define the characteristic photographic images.

Common cameras are based on the pinhole camera model, where scene radiance is mapped perspectively onto the light sensors behind the pinhole. Modern digital cameras focus scene radiance for a better optical efficiency using a lens system and digitally processes the sensor measurements to produce visually pleasuring and storage-efficient images. As photographic images are produced by cameras, certain characteristics of the camera design and the in-camera processing are present in the images. For example, *vignetting*, a visual artifact of radial brightness falloff that arises naturally from an uncompensated lens system, can sometimes be observed on photographs. *Chromatic aberration*, which manifests as color fringes at occlusion edges in an image due to lens with varying refractive indexes for different wavelengths, is also not uncommon. Apart from the mentioned *photometric distortions*, *geometric distortions* such as pincushion distortion can be visible in an image captured with a poorly designed lens.

Other imprints of camera on a photograph are related to the optical sensor and in-camera processing. Most image sensors used today including the charge-coupled device (CCD) and the complementary metal-oxide-semiconductor (CMOS) sensor are pixilated metal oxide semiconductors, which suffer from several forms of noise such as the pattern noise, dark current noise, shot noise, and thermal noise [25]. Although such camera noise is at a small degree, it can be estimated to certain extent [37].

Photographic images generally undergo *color filter array demosaicing* which is a form of image interpolation within and across color channels. Such interpolation is needed as most commercial cameras today sample the red, green, and blue color components of the scene radiance with a single sensor array, instead of three separate ones. Hence, each sensor can only measure one of the color components at a snapshot and interpolation is employed to fill in the missing measurement. Other in-camera operations include white balancing that offsets the shade of the illumination color, edge sharpening, intensity contrast enhancement, and gamma correction for dynamic range compression. The overall effect of all these operations can be modeled by a camera response function with a typical concave shape [23].

4.2.2 Formation of Photorealistic Computer Graphics

Physics-based graphics rendering described by the Kajiya's rendering equation [27] is the basis of photorealistic rendering. To achieve photorealism, graphics rendering needs to produce complex visual effects such as color blending from light interreflections among surfaces, complex outdoor illumination, and the appearance of some subtle reflectance properties of real-world objects.

Scene modeling refers to modeling of illumination, surface reflectance, and object geometry. Although computer generated images of simple diffuse scenes can be visually indistinguishable from photographs [43, 44], modeling and rendering complex scenes remain challenging. Image-based approach is an answer to this challenge. Image-based modeling incorporates photographs of real scene illumination and object appearance into the graphics pipeline and hence blurs the distinction between photographic and computer graphics images. For example, complex real-scene illumination can be modeled by an environment map derived from the photograph of a mirror sphere [46]. Similarly, spatially varying surface reflectance can be measured from multiple-view photographs [8].

An image-based model with high fidelity calls for elaborate measurement of the real illumination or objects. Hence, image-based measurement may require special devices or need to capture a large number of images which can seriously strain the storage and rendering efficiency in the graphics pipeline. To achieve efficiency, various forms of simplification are introduced in both graphics modeling and rendering. Representing color in three separate channels in the early stage of scene modeling and independent rendering of the color components are among the examples [59]. This creates differences between computer graphics and photographs.

At final stage, a synthesized image may be further touched up or color-adjusted using image editing software such as Adobe Photoshop. The post-processing step can be distinctively different from the in-camera processing in camera.

4.2.3 Prior Work Survey

Below, we give detailed description of the methods and features motivated by the image formation processes.

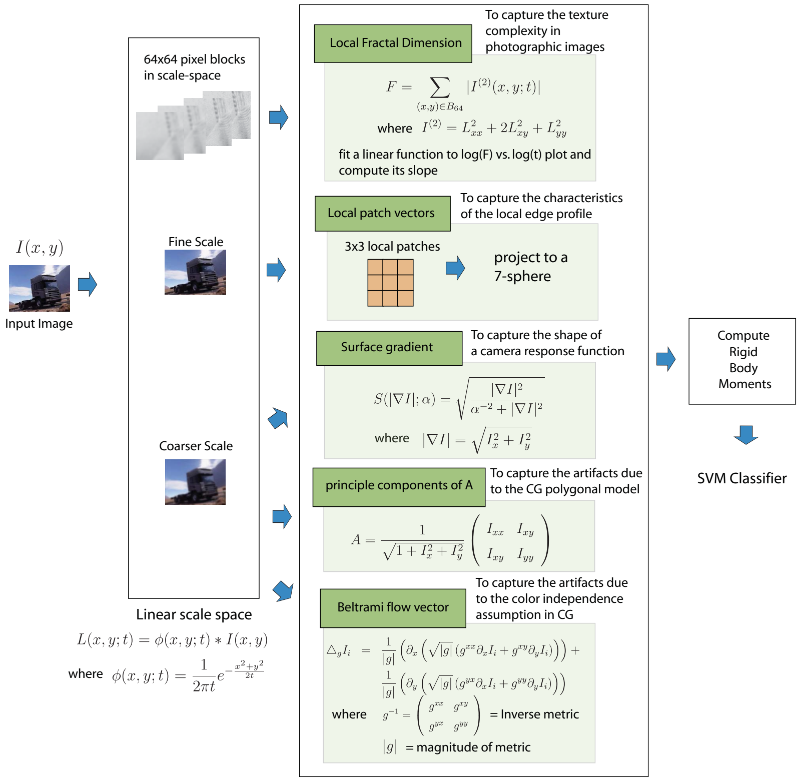

$I(x,y)$
Input Image
64x64 pixel blocks in scale-space
Fine Scale
Coarser Scale
Linear scale space
$L(x,y;t) = \phi(x,y;t) * I(x,y)$
where $\phi(x,y;t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}}$
Local Fractal Dimension
To capture the texture complexity in photographic images
$F = \sum_{(x,y)\in B_{64}} |I^{(2)}(x,y;t)|$
where $I^{(2)} = L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2$
fit a linear function to log(F) vs. log(t) plot and compute its slope
Local patch vectors
To capture the characteristics of the local edge profile
3x3 local patches
project to a 7-sphere
Surface gradient
To capture the shape of a camera response function
$S(|\nabla I|;\alpha) = \sqrt{\frac{|\nabla I|^2}{\alpha^{-2} + |\nabla I|^2}}$
where $|\nabla I| = \sqrt{I_x^2 + I_y^2}$
principle components of A
To capture the artifacts due to the CG polygonal model
$A = \frac{1}{\sqrt{1+I_x^2+I_y^2}} \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix}$
Beltrami flow vector
To capture the artifacts due to the color independence assumption in CG
$\triangle_g I_i = \frac{1}{|g|}\left(\partial_x\left(\sqrt{|g|}\,(g^{xx}\partial_x I_i + g^{xy}\partial_y I_i)\right)\right) + \frac{1}{|g|}\left(\partial_y\left(\sqrt{|g|}\,(g^{yx}\partial_x I_i + g^{yy}\partial_y I_i)\right)\right)$
where $g^{-1} = \begin{pmatrix} g^{xx} & g^{xy} \\ g^{yx} & g^{yy} \end{pmatrix}$ = Inverse metric
$|g|$ = magnitude of metric
Compute Rigid Body Moments
SVM Classifier

Fig. 8: Illustration of the feature extraction process in the work by Ng *et al.* [49].

Contrasting Photographic and Computer Graphics Formation. By contrasting the photographic and computer graphics formation processes, Ng *et al.* [49] identified three differences between them. First, photographic images are subject to the typical concave response function of cameras while computer graphics rendering pipeline may not have a standardized post-processing procedure that mimics the camera processing. Second, graphic objects may be modeled with simple and coarse polygon meshes. The coarseness of the polygons can give rise to the unnatural sharp edges and polygon-shaped silhouettes in computer graphics images. Third, the three color channels of graphics images are often rendered independently as graphics models are often in such color representation instead of the continuous color spectrum representation as in the real scenes.

The computational steps of the method in [49] are shown in Figure 8. The three mentioned differences are described using image gradient, principal curvatures, and Beltrami flow vectors. They also computed the local block-based fractal dimension and the local patch vectors. The local fractal dimension was meant to capture the texture complexity and self-similarity in photographs and the local patch vectors to model the local edge profile. An SVM classifier is trained with the features on the Columbia open data set (more details in Section 7), they attained an average classification accuracy of 83.5%.

Employing Device Noise Properties. Dehnie *et al.* [9] demonstrated that noise pattern of photographic images, extracted by a wavelet denoising filter, is different from that of computer graphics images. Hence, with their respective reference noise patterns, a test image can be classified based on its correlation to the reference noise patterns. On their undisclosed data set, the method achieved an averaged classification accuracy of about 72%.

Inspired by the directional noise in scanners, Khanna *et al.* [29] employs the features for the device-dependent residual pattern noise computed row-wise and column-wise in an image to distinguish photographic, computer graphics and scanned images. On their undisclosed data set, the method achieved an average accuracy of 85.9%.

Employing Camera Demosaicing and Chromatic Abberation. Dirik *et al.* [10] observed that an image from a camera with a Bayer color filter array experiences a smaller change if it is reinterpolated again according to the Bayer pattern as compared to other patterns. They also measured the misalignment among the color channels due to chromatic aberration where the camera lens diverges the incoming light of different wavelengths. With the two physical characteristics unique to photographs but often not present in computer graphics, the method achieved an average classification accuracy of about 90% on their undisclosed data set.

Gallagher and Chen [19] showed that the Bayer-pattern demosaicing in original-size camera images can be detected. The computational steps for their method are shown in Figure 9. Their method is based on two main observations; First, the demosaiced pixels always have a smaller variance as the original pixels, and high-pass filtering can make the demosaicing property more prominent. Second, in the green-color component of a Bayer-pattern image, the interpolated and the original pixels respectively occupy the alternate diagonal lines. Hence, the variance plot from the diagonal scan lines would display a regular pattern with a frequency of two units. Their method achieved an average classification accuracy of 98.4% on the Columbia open data set. Their method may be sensitive to image post-processing operations such as image resampling or resizing that may destroy the interpolation structure specific to the Bayer pattern.

4.3 Methods from Natural Image Statistics

The research for natural image statistics is motivated by efforts to observe, isolate, and explain the regularities inherent to natural images [66]. Due the high dimensionality of the image space, building a probability model directly on the space is

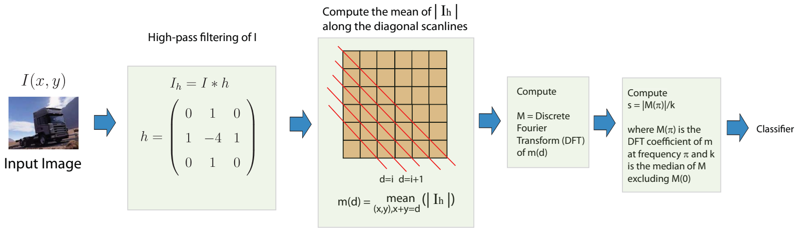

$I(x,y)$
Input Image
High-pass filtering of I
$I_h = I * h$
$h = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
Compute the mean of | Ih |
along the diagonal scanlines
d=i d=i+1
m(d) = mean (x,y),x+y=d (| Ih |)
Compute
M = Discrete Fourier Transform (DFT) of m(d)
Compute
s = |M(π)|/k
where M(π) is the DFT coefficient of m at frequency π and k is the median of M excluding M(0)
Classifier

Fig. 9: Illustration of the feature extraction process in the work by Gallagher and Chen [19].

intractable. Hence, statistics are instead derived from the lower dimensional subspaces in some transform domains such as the wavelet or Fourier domain. Some of the statistics that are motivated to explain the scale invariance properties in natural images which is in general only meaningful for image ensembles, while those motivated by applications such as image compression would be useful descriptors for single images. For example, the power law of the power spectra is statistically stable for an image ensemble while the sparse distribution of the marginal wavelet coefficients is stable for single images, as we have seen in Figure 5. The methods motivated by natural image statistics for distinguishing computer graphics are based on the belief that computer graphics images are statistically different from photographic ones.

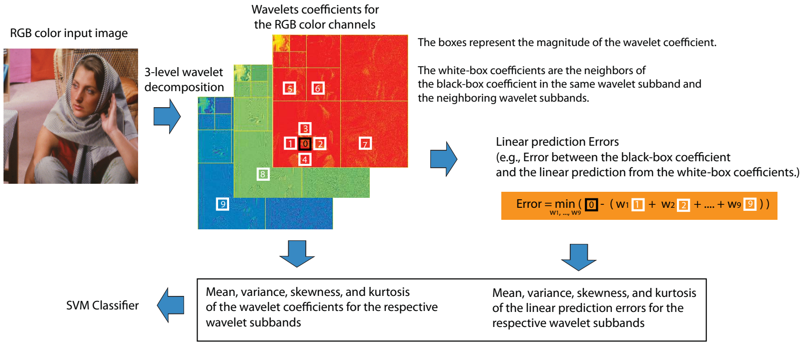

RGB color input image
3-level wavelet decomposition
Wavelets coefficients for the RGB color channels
0
1
2
3
4
5
6
7
8
9
The boxes represent the magnitude of the wavelet coefficient.
The white-box coefficients are the neighbors of the black-box coefficient in the same wavelet subband and the neighboring wavelet subbands.
Linear prediction Errors
(e.g., Error between the black-box coefficient and the linear prediction from the white-box coefficients.)
Error = min w1, ..., w9 (0 - (w1 1 + w2 2 + ... + w9 9))
Mean, variance, skewness, and kurtosis of the wavelet coefficients for the respective wavelet subbands
Mean, variance, skewness, and kurtosis of the linear prediction errors for the respective wavelet subbands
SVM Classifier

Fig. 10: Illustration of the feature extraction process in the work by Lyu and Farid [38].

Wavelet Statistics. Lyu and Farid considered that photographic images have different statistical characteristics in the wavelet domain as compared with photorealistic computer graphics images [15, 38]. The computation steps for their method are illustrated in Figure 10. An RGB input image is first decomposed into three levels of wavelet subbands. For natural-scene images, the wavelet coefficients in a subband are modeled well with a generalized Laplacian distribution and correlation exists between wavelet coefficients of adjacent subbands [66]. Lyu and Farid modeled the former distribution with statistical moments of the wavelet coefficients within a subband, and the latter using the linear prediction error of the coefficients. Figure 10 illustrates how the prediction error is computed using an example where the black-box coefficient is linearly predicted using the neighboring white-box coefficients from subbands within a neighborhoud. Four moments (mean, variance, skewness, and kurtosis) of the wavelet coefficient distribution and the linear prediction error distribution are then computed for each subband as features. On an undisclosed Internet image set, a SVM classifier achieved a classification rate of 66.8% on the photographic images, with a false-negative rate of 1.2%.

Wang and Moulin [74] observed that the characteristic function of the coefficient histogram of a wavelet subband is different for photographic, photorealistic computer graphics, and not-so-photorealistic computer graphics images. Such differences are distinct at the low and middle frequency regions. Hence, for each subband, they computed three simple features through low-passing and band-passing the characteristic functions. With the simple features, the computation speed is about four times faster than that of Lyu and Farid [38]. On their undisclosed data set, the method has comparable performance as that of Lyu and Farid [38] on a simple Fisher linear discriminant classifier. They also tested their classifier (trained using their data set) with the Columbia open data set and an abnormally high false alarm was observed. This implied the statistical discrepancy between data sets. More comments on data set differences are given in Subsection 4.6.

Power Law for Fourier Power Spectrum and Local Patch Statistics. Ng *et al.* [47] explored the usefulness of various natural image statistics for distinguishing photographic images from computer graphics. The study was based on the features related to the power law of the power spectrum of images, the wavelet statistics, and the local patch statistics. The study showed that the local patch statistics performed best in the classification, while the power law statistics performed the worst. This indicates a relationship between the classification performance and the spatial locality of these statistics. The power law statistics is computed on Fourier domain, hence it does not retain any spatial information about the image. The wavelet statistics is partially localized, while the local patch statistics with the smallest spatial support is computed on the high-contrast local patches in an image. This result indicates the importance of local features in distinguishing computer graphics images.

More recently, Zhang *et al.* [79] modeled local patch statistics as visual words and took an object recognition approach for recognizing photorealistic computer graphics.

Color Compatibility. Color composition of natural images is not random and some composition is more likely than the others. Lalonde and Efros [34] showed that the color compatibility between the foreground object and the background scene in a natural-scene image provides a statistical prior for identifying composite images. As computer graphics images may deviate from this statistical regularity, color compatibility can potentially be used to distinguish computer graphics from photographic images.

Photorealism Measures. Natural image statistics can essentially serve as a measure for photorealism. Wang and Doube [73] attempted to measure visual realism empirically. The measure consists of three visually perceivable characteristics of natural images which are surface roughness, shadow softness, and color variance. As a valid photorealism measure should track the degree of photorealism in an image, this work brought up an interesting idea of using computer game images produced in different years for photorealism evaluation, with the assumption that newer computer games are more photorealistic. Such measure is still considered weak and hence a better measure is needed for real applications.

4.4 Methods from Steganalysis

Steganography embeds confidential messages imperceptibly in a carrier (e.g., an image) in order to hide both the message and the act of message hiding. Whereas steganalysis aims at revealing the act of message hiding blindly without the help of the reference image. Some steganalysis methods detect the specific abnormal statistics in an image resulted from steganography. For example, the Chi-square test statistics on *pairs of values* that differ in the least significant bits (LSB) is good at detecting information hiding by EzStego [75]. Such technique is steganography-method-specific. It is believed that there exists *universal steganalysis methods* [30] which can detect steganography regardless of its technique. These universal methods aim at extracting discriminative statistics or features which are highly sensitive to information hiding in general.

It is believed that the procedure for extracting the distinguishing features for hidden data can be applied for capturing the statistical characteristics of photorealistic computer graphics. For example, the wavelet statistics method by Lyu and Farid [15] was originally applied for steganalysis, although it is motivated by natural image statistics. Below, we describe in detail several methods that are inspired by steganalysis methods.

Moments of Characteristic Functions. The method based on moments of characteristic functions on wavelet subbands and prediction error image originates from steganalysis [64]. Chen *et al.* [5] applied this method in hue, saturation, and value (HSV) color space for distinguishing photorealistic computer graphics. The predication error image I_e is the difference between an image I and its predicted version $\hat{I}$, $I_e = |I - \hat{I}|$. The prediction $\hat{I}(x,y)$ with a threshold value c can be computed as

$$\hat{I}(x,y)=\begin{cases}\max[I(x+1,y),I(x,y+1)] & c\leq\min[I(x+1,y),I(x,y+1)]\\ \min[I(x+1,y),I(x,y+1)] & c\geq\max[I(x+1,y),I(x,y+1)]\\ I(x+1,y)+I(x,y+1)-I(x+1,y+1) & \text{otherwise}\end{cases}. \tag{4}$$

Both the original and prediction images can be decomposed into wavelet and approximation subbands. The characteristic function $H(\omega)$ of a subband is the discrete Fourier transform (DFT) of its coefficient histogram and the n-th order statistical moment of the characteristic function is given by

$$m_n=\frac{\sum_{\omega>0}\omega^n|H(\omega)|}{\sum_{\omega>0}|H(\omega)|}, \tag{5}$$

where only the positive half the characteristic function is considered in the computation. Chen *et al.* [5] computated three levels of wavelet decomposition on the original and the prediction images for each of the HSV color channels. The first three statistical moments were computed for each subband which gave 234 features in total. On a data set expanded from the Columbia open data set, they were able to achieve a classification accuracy of 82.1%.

Sutthiwan *et al.* [68] extended the work by Chen *et al.* [5] to include moments of 2D characteristic functions for the Y and Cb components of an image in YCbCr color space. A 2D characteristic function is the DFT of a 2D histogram. They computed the features on the original image, its JPEG coefficient magnitude image and their respective prediction error images, where wavelet decomposition were performed. With a total of 780 features, a SVM classifier was trained and tested on their undisclosed image data set with an averaged test accuracy of 87.6%. With feature selection on a Adaboost classifier, the number of features was reduced to 450 while the averaged accuracy was improved to 92.7%.

In another work, Sutthiwan *et al.* [67] considered the JPEG horizontal and vertical difference images as first-order 2D markov processes and used transition probability matrices to model their statistical properties. A total of 324 features were extracted from Y and Cb channels. On their undisclosed data set, the method achieved a classification accuracy of 94.0% with a SVM classifier. With the same feature reduction method with a Adaboost classifier, a classification accuracy of 94.2% can be achieved with 150 features.

The ratio of Chi-squared Test Statistics. Rocha and Goldenstein [61] considered the statistical response of an image to pixel perturbation as a property for distinguishing photographic and computer graphics images. Pixel perturbation is performed by replacing the least significant bits (LSB) of a randomly selected set of pixels with a random binary sequence generated from a uniform distribution of the binary symbols. This pixel perturbation process is similar to the bit embedding function of the EzStego steganography method [75].

Statistical test can be performed to measure the statistical similarity between the resulting distribution with the uniform distribution as reference. The chi-squared statistics χ^2 and the Ueli Maurer Universal statistics U_T of a LSB perturbed image I_p were computed and the deviations from that of the original image I were measured by

$$r_{\chi^2} = \frac{\chi^2(I_p)}{\chi^2(I)}, \quad r_{U_T} = \frac{U_T(I_p)}{U_T(I)}. \tag{6}$$

Six versions of perturbed images I_p were generated for an image through perturbing a fixed percentage of randomly selected pixels, with the percentage corresponds to 1%, 5%, 10%, 25%, 50% and 75%. The authors validated their approach on an undisclosed image data set with 12,000 photographs and 7,500 photorealistic computer graphics images. With a SVM classifier, the method achieved an averaged accuracy of 97.2% as opposed to 82.2% by the method of wavelet high order statistics [38].

4.5 Methods from Combining Features

The different types of features are meant to capture different characteristics of an image and they have different strengths and weaknesses. A set of features can be combined to improve performance. Sankar *et al.* [63] combined the general graphics features from Ianeva *et al.* [26] , the moments of characteristic function features from Chen *et al.* [5], the local patch statistics from Ng *et al.* [49], and the image resampling features from Popescu and Farid [58]. On the Columbia open data set, a classifier with the aggregated set of features achieved an average classification accuracy of 90%.

4.6 Data Set and Performance Evaluation

Table 1 lists the performance of the various proposed methods. A direct comparison of their classification performances is not meaningful as some of the experiments were conducted on different data sets. The discrepancy between different data sets can be significant as observed by Wang and Moulin [74]. However, the methods evaluated on the Columbia open data set may be compared with a caveat that the quoted classification performance merely corresponds to a single operating point on the performance curve. More comments on performance evaluation are given below.

Properties of Internet Image Sets. Fundamentally, our aim is to build a system for recognizing photorealism and the inherent properties of camera, which are independent of image scenes. Hence, an ideal data set would be one composed of pairs of photographic and computer graphic images with identical image scenes. Such image pairs have been used for subjective experiments [43, 44, 59]. However, how to synthesize such a data set efficiently remains an open problem.

The evaluation of current works are mainly based on data sets of Internet images, hoping that the image content is diverse enough to qualify as random samples

Table 1: Tabulation of the classification accuracy for various methods on distinguishing photographic and photorealistic computer graphics images

Approach	Work	Feature dimension	Data set	Highest classification accuracy
Image formation	Ng *et al.* [49]	192	Columbia open data set	83.5%
	Dehnie *et al.* [9]	1	Internet images	72%
	Khanna *et al.* [29]	15	Internet images	85.9%
	Dirik *et al.* [10]	77	Internet images	90%
	Gallagher and Chen [19]	1	Columbia open data set	98.4%
Natural image statistics	Lyu and Farid [38]	216	Internet images	66.8% true-photo, 1.2% false-photo
	Wang and Moulin [74]	144	Internet images	comparable to Lyu and Farid [38]
	Ng *et al.* [47]	24	Internet images	83%
Steganalysis	Chen *et al.* [5]	234	Columbia open data set	82.1%
	Sutthiwan et al [68]	450	Internet images	92.7%
	Sutthiwan et al [67]	150	Internet images	94.2%
	Rocha and Goldenstein [61]	96	Internet images	97.2%
Visual cues	Wu *et al.* [76]	38	Internet images	95% with *k*-NN, 75% with SVM
	Pan *et al.* [56]	30	Internet images	91.2%
Combining features	Sankar *et al.* [63]	557	Columbia open data set	90%

in the image space or at least the two sets of images have similar distribution in the image space. Unfortunately, the photographic images and the photorealistic computer graphics on the Internet may form different distributions in the image space. For example, computer graphics that are harder to render can be less common on the Internet than the easier ones, but such statistics may not apply to photographic images on the Internet. Although filtering has been in place for the Columbia open data set to reduce the number of simplistic computer graphics, the data set is still considered far from the ideal data set. Therefore, the classification performance in Table 1 can only serve as a proxy for the capability in recognizing photorealism or the properties of camera.

Usefulness for Real Applications. The classification accuracy based on the Columbia open data set ranges from 82.1% of Chen *et al.* [5] to 98.4% of Gallagher and Chen [19]. The quoted performance may not be a sufficient indicator for the usefulness of the methods when it comes to real applications. For example, the method by Gallagher and Chen exploits the interpolation clues related to the Bayer color filter array and works well on recognizing the original-size photographic images in the data set. However, the images may be resized in real applications. The Columbia group has built a website for detecting computer graphics images submitted by Web users [48]. This exercise offers a realistic evaluation scenario for the detectors with adversarial users. It was observed that the true performance of the detectors under such scenario is in general lower than the performance numbers given in Table 1.

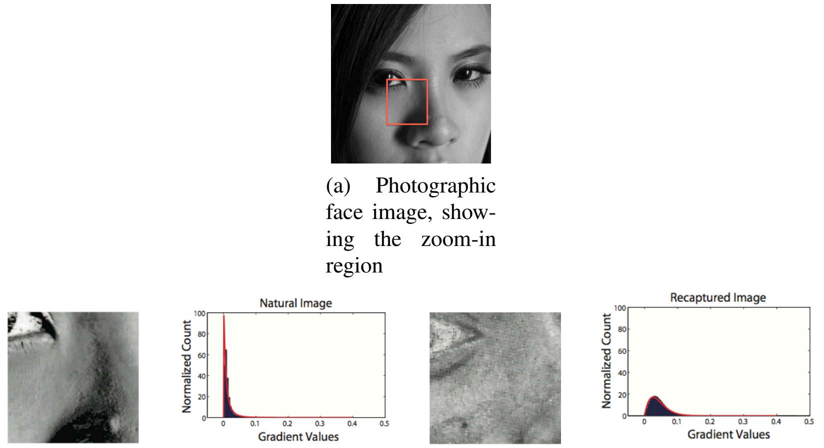
5 Approaches for Recaptured Image Detection

Technically, a recaptured image is a photograph of an image reproduction medium. There are many ways to recreate an image on a physical surface. For example, an image on paper reproduced by an inkjet printer is represented as ink dots modulated by half-toning to recreate the appearance of the image. For color laser print, the color toner particles, in a combination of cyan, magenta, yellow, and black color, are deposited and fused on a paper through heat treatment in multiple scans. Photo print recreates an image on a photo paper with finite grain size which is generally about 300 dots per inch (DPI). When an image is displayed on an LCD screen, each pixel is represented by the emitted light modulated by a liquid crystal of finite size arranged in a 2D array. Each of these physical methods has different color reproduction capability and prefered viewing conditions. Hence, the reproduced color and image appearance depends on the color reproduction technique, the physical medium, and the ambient light. Below, we describe in detail the proposed recaptured image methods.

A recaptured image detection system can be used as a countermeasure for image replay attack on a face authentication system. Liveness detection [7, 35, 57, 32] generally identifies image replay through the specific motion of the human subject that is captured on video. Unlike the image recaptured detection approach, none of the liveness detection methods can resist video play attack, e.g., playing back a face video clip on a tablet computer. Furthermore, unlike image recaptured detection, liveness detection would not be effective in detecting image replay for static objects.

Reproduction Medium Property. Yu *et al.* [78] studied the ambient light reflected off the paper as recaptured by a high resolution camera. Part of the light is reflected as specularity. Such reflectance carries a spatial pattern similar to the fine texture of a paper which is more pronounced in the specular component of the recaptured image. There are various methods to decompose an image I into its diffuse D and specular S components, where $I = D + S$ [69, 71, 41]. Figure 11b and 11d shows the normalized specular component $\hat{S} = S/(S+D)$ for a cropped out image of a real face as shown in Figure 11a and its corresponding recaptured image. The fine texture is a characteristic of recaptured images from paper printouts.

Bai *et al.* [2] modeled this texture pattern with a histogram of gradient magnitude on the specular component. The texture results in a heavy tail for the histogram as shown for the recaptured face in Figure 11e. The shape of the histogram after being normalized to a unity area is modeled by a generalized Rayleigh distribution [28]



(a) Photographic face image, showing the zoom-in region

(b) Specular ratio image (zoom-in) computed from the photographic image

(c) Specular Gradient Histogram for the photographic image

(d) Specular ratio image (zoom-in) computed from the corresponding recaptured image

(e) Specular Gradient Histogram for the recaptured image

Fig. 11: The normalized specular components of a 3D face image and its recaptured image, shown together with their corresponding histograms of the gradient of the specular images.

$$f(x) = kxe^{(\frac{x}{\alpha})^{\beta}} \qquad (7)$$

parameterized by two parameters α and β. The parameters can be used as features for distinguishing recaptured images from non-recaptured 3D scene images. A linear SVM classifier is trained with a set of 45 real face images and 45 recaptured face images. The classifier achieved 2.2% false acceptance rate and 13% false rejection rate with 6.7% equal error rate.

Color Reproduction and Recapture Scene Properties. To resolve the fine texture of a reproduction medium such as paper or computer screen requires high-resolution camera for image recapturing. Such limitation will preclude recaptured image detection on mobile cameras which have relatively lower pixel resolution. To enable recaptured image detection on mobile devices, Gao *et al.* [20] proposed a set of distinguishing features related to the photometric and scene-related properties due to the image recapturing process, which is an extension of the common photographing process as shown in Figure 12. The extension is related to the color reproduction response function f_m, the reflected radiance from the recaptured scene R and the photometric property of the second camera f_2.

A color reproduction technique may involve specific color profile, which could be a smaller portion of the full color space or has limited color resolution. This can result in specific color shade on the reproduced image such as the blue tint on some LCD display. Therefore, the covariance of the color distribution of an image can be a distinguishing feature.

The photometric response for the reproduction process and the second recapture are in general nonlinear. Hence, the cascaded photometric response of f_1, f_m, and f_2 in the image recapturing pipeline could be different from f_1 of common photographs. The difference in photometric characteristics can be captured by image gradient [49].

Apart from the reflected specularity, the light may transmit through a reproduction medium which is not entirely opaque such as a paper. Similar to specularity, the light transmitted from the back can significantly reduce the contrast and saturation of a recaptured image. Therefore, image contrast and the histogram of the chromatic components for an image can be computed to capture these phenomena.

One may place a photograph or a printout at a specific distance from the camera to control its size when appearing in the recaptured image. Doing so can potentially place the photograph outside the camera depth of field and result in image blur. Therefore, image blur can be a tell-tale sign of recapture attack. Finally, it is possible that the natural-scene background is visible in the recaptured image, besides the image reproduction medium.

Using the above-mentioned features, Gao *et al.* [20] devised a recaptured image detector for mobile devices using an SVM classifier trained on a recaptured image data set [21]. They compared the performance of the proposed features with that of the wavelet features by Lyu and Farid [38]. When the natural-scene background is visible, the detector achieved an averaged detection accuracy of 93% as compared to 86% for the wavelet features. Without the background information, the detector achieved 78% detection accuracy as compared to 68% for the wavelet features.

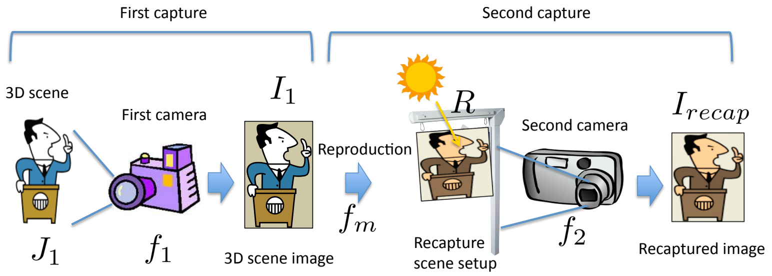


Fig. 12: An image recapturing pipeline where image recapturing (second capture) as an extension of the common photographing pipeline (first capture).

Statistical Property for LCD Screen Recaptured Images. Cao and Kot [3] studied the issue of image recapturing from LCD screens. They performed a subjective study which showed that recaptured images from LCD screen are largely perceptually indistinguishable to humans. However, there are fine differences between LCD screen recaptured images and non-recaptured ones. For instance, there is a fine grid pattern on LCD screen recaptured images which can be described statistically with multi-scale local binary pattern features [55]. They also considered the loss of details due to recapturing as a distinguishing feature and model it with the statistics of wavelet coefficients. To capture the chromatic property of LCD scene recaptured images, color features in both RGB and HSV color spaces were computed. The features were evaluated on an image data set with 2,700 LCD scene recaptured images and 2,000 non-recaptured images with an SVM classifier that achieved an equal error rate of 0.5% as compared to that of 3.4% for the wavelet features by Lyu and Farid [38].

6 Possible Attacks and Counter Attacks

In digital image forensics setting, a computer graphics detector can potentially face attacks as described below. Although attacks on recaptured image detection have not been studied, we can imagine that similar attack strategies are equally applicable to recaptured images.

1. **Recapture attack:** Ng *et al.* [49] showed that recapturing computer graphics is a convenient way to turn a computer graphics into a photograph with little perceptible changes on the image content. To reduce the risk of such attack, the computer graphics image set used for classifier training can be expanded to include the recaptured images.

2. **Histogram manipulation attack:** Sankar *et al.* [63] showed that a computer graphics detector with color histogram features is vulnerable as the histogram of a computer graphics image can be warped to mimic that of a photograph. Such attack can be prevented by detecting histogram manipulation with a local pixel correlation measure, as histogram manipulation naturally alters the correlation of local pixels.

3. **Hybrid image attack:** Sankar *et al.* [63] showed that the performance of a computer graphics detector drops significantly for hybrid images, a composite of photographic and computer graphics image regions. They found that local patch statistics [49] are effective in distinguishing hybrid images from non-hybrid ones.

Attacks on a detector succeed as the attack images deviate from the pattern model of the detector. Therefore, an approach with a system of classifiers that builds in a mechanism for anticipating and handling attacks can be helpful. Figure 13 shows an example of such system proposed by Sankar *et al.* [63]. The system begins with a classifier detecting the anticipated types of attack or forgery images. Although such an open-ended system is more secure than a single non-adaptive classifier, it is not capable of handling unseen types of attacks. An ideal system would adapt to the inputs and minimize misclassification through online learning with some forms of supervision.

The design of a digital image forensics system secure against attacks is still an open problem. For learning-based systems, a potential issue would be how to gather enough training data for constructing attack models. Furthermore, if the attackers have easy and unlimited access to the forensics system, an exhaustive search can be performed to arrive at the best attack parameters. Even when system access is limited to each user, multiple attackers can still collude to search for the best attack parameters.

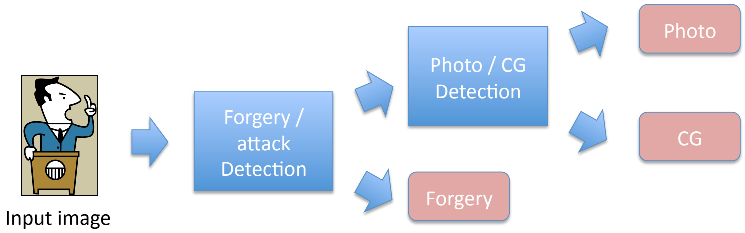


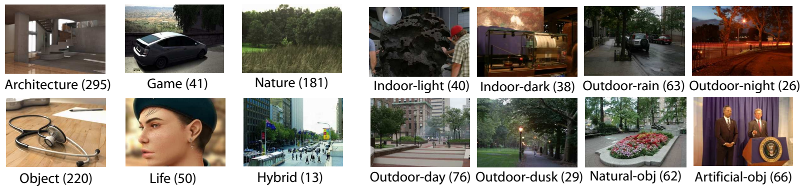
Fig. 13: A system design for computer graphics detector with a forgery or attack detection feature.

7 Resources

With photorealistic computer graphics and recaptured image detection formulated as a pattern recognition problem, data set becomes an essential component of the research. An open benchmark data set does not only provide data for experiments, it also serves as a basis for comparing various detection methods. For classification of photographic and photorealistic computer graphics images, Ng *et al.* [50] constructed the Columbia open data set. Whereas the classification of recaptured and non-recaptured images, Gao *et al.* [21] constructed the I^{2}R open data set.

These data sets are limited in diversity and could not cover all types of images one may encounter in real applications. For example, when Ng and Chang [48] deployed their computer graphics detector online, images with different characteristics from those in the open Columbia data set were observed. These images include composite images, recaptured images, and images with graphics posters. Hence, the online classifier presents a useful case study for a real-world application [52] and helps to address the limitation of the data sets. Samples from such online studies may be used to grow the data sets, whenever their ground truth labels are reliable.

7.1 Benchmark Data Sets



(a) Photorealistic Computer Graphics (b) Natural-scene Photograph

Fig. 14: The image categories in the Columbia open data set.

Columbia Open Data Set. The Columbia benchmark data set was constructed and made accessible to the research community [50]. The data set consists of 800 personal photographic images, 800 photographic images obtained through Google Image Search, 800 photorealistic computer graphics images from 3D artist websites, and 800 recaptured computer graphics images. To ensure diversity in the image content and lighting, there are various subcategories in the data set, as shown in Figure 14. As the goal for the data set was to support studies on photorealism instead of just as a sample set of Internet computer graphics per se, the downloaded computer graphics are perceptually filtered with majority votes from three human observers by assessing the photorealism level of the images through visual inspection.

I^{2}R Open Data Set for Smart Phone Recaptured Images. Gao *et al.* [21] constructed a smart phone recaptured and non-recaptured image data set, which considered the variations in the image recapturing pipeline, as shown in Figure 15. There is a variety in the first camera, the reproduction device, and the second camera. The first cameras are mainly high-quality single-lens reflex (SLR) cameras. The second cameras are smart-phone cameras with lower sensor quality, which include the front and back cameras of a smart phone. The back camera which is meant for phototaking in general has a higher resolution than the front camera which is meant for video conferencing or facetime applications. Various forms of reproduction medium were considered including laser print, ink print, photo print, computer LCD screen, and smart-phone LCD screen. An important characteristic of this data set is the matched-content pairs for the recaptured and non-recaptured images as shown in Figure 16a. Each pair of images can be geometric aligned as shown in Figure 16b to facilitate an unbiased study on their photometric difference, independent of the im-

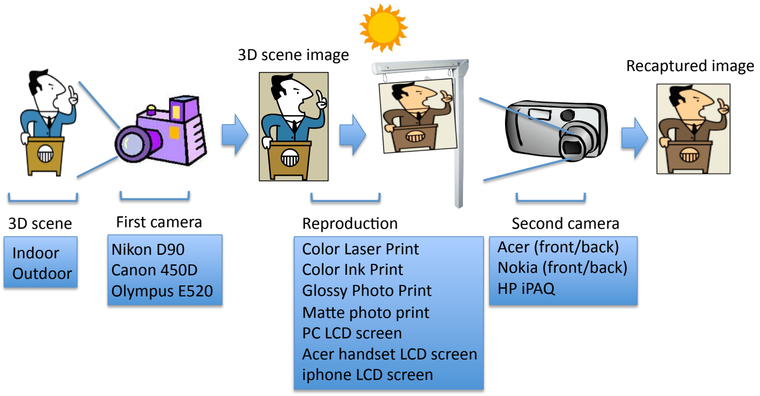


Fig. 15: The diversity in image recapturing pipeline considered for constructing the I^{2}R open dataset.

age content. In fact, the matched-content property is equally desirable for the open Columbia data set as it would make the data set ideal for studying photorealism in a content-independent manner.

7.2 Online Evaluation System

The Columbia online demo system [48] offers a platform for users to try out the fully automatic computer graphics detection function on any test images of their choices. Users are free to try out any attack strategy on the online system. In [52], an evaluation on the Columbia online system was done considering its performance and the specifics of the images submitted by users. The system also allows comparison of different detection algorithms and features (geometry, wavelet, and cartoon features). Among the images submitted by users, there are unconventional images with vague classes. These images include photographs with graphic content embedded in the 3D real-world scenes, and the images composed of both photograph and computer graphics as shown in Figure 17[4].

[4] Figure 17a was from `http://www.latimes.com/media/alternatethumbnails/photo/2006-06/24010006.jpg`, Figure 17b `http://www.iht.com/images/2006/06/25/web.0626city9ss4.jpg`, Figure 17c `http://www.spiegel.de/img/0,1020,681928,00.jpg` and Figure 17d `http://www.spiegel.de/img/0,1020,681938,00.jpg`.

(a) Examples of recaptured images. The top row shows the examples of recaptured images with the real environment as the background. The bottom row shows the corresponding cropped images of the first row.

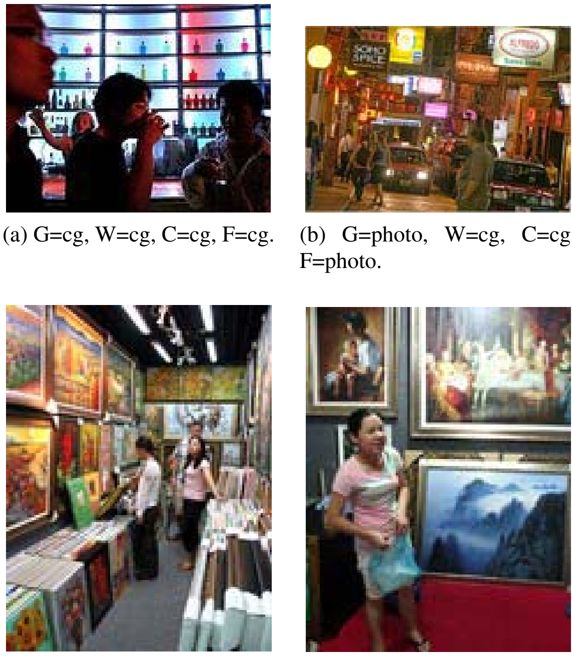
(b) Example geometrically aligned images in the I^{2}R open data set. The top row of images are the real-scene images, while the bottom row of images are the corresponding recaptured ones.

Fig. 16: Example images in the I^{2}R open data set.

8 Open Issues and Future Research Directions

Computational measure for photorealism or photograph-ness remains an open problem despite advances in multimedia forensics and perceptual studies for computer graphics. While perception of photorealism can be studied through subjective experiments, the computational model of photorealism requires detailed modeling of the photographic image pipeline and the non-photographic ones such as computer graphics rendering and image recapturing. Hence, estimating the related distinguishing features from the image models is a research challenge for computer vision and graphics.

The definition of photograph evolves with time, so is it for computer graphics and recaptured images. The current camera model is largely based on the pinhole model. The camera model will eventually evolve for enabling new functionalities for cameras, as has been actively pursued by researchers in computational photography [60]. While disruptive changes may not be imminent, the current camera imaging pipeline has always been changing in a gradual manner with small advances in hardware. For example, Foveon introduced a new sensor called the Foveon X3 sensor (a CMOS sensor) which can mimic the color negative film by stacking the RGB color sensitive elements on top of each other, in layers, at each pixel site. As a result, it can measure three RGB colors at each site and hence demosaicing is no longer needed. The



(a) G=cg, W=cg, C=cg, F=cg.

(b) G=photo, W=cg, C=cg, F=photo.

(c) G=photo, W=photo, C=photo, F=photo.

(d) G=photo, W=photo, C=photo, F=photo.

Fig. 17: The images with vague classes submitted to the Columbia online computer graphics detector. The classification results for four types of classifier are shown under each image. The shorthands G, W, C and F respectively represent the geometry classifier [49], wavelet classifier [38], cartoon classifier [26] and the fusion of the former three classifiers.

dynamic definition of photography poses challenges to machine learning. Adaptive and online learning methods [4, 31] are required to keep track of the changes in the photography model and update its pattern model without relearning from scratch.

A good data set is important for a pattern recognition problem. As pointed out in Section 4.6, the current photographic and computer graphics data set can be bettered by having content matching pairs, a feature similar to that of the I^{2}R open data set for smart phone recaptured images [21]. The photographic and computer graphics pairs with matched content enables a better investigation of computational photorealism in a content-independent manner. Establishing this data set would require significant efforts in computer graphics modeling and rendering.

The current approach for distinguishing computer graphics and recaptured images from photographs has been mainly through discriminative learning. This approach lacks flexibility in adding new classes of images. Generative model for various types of images is probably more efficient for future-proof system designs.

Eventually, computer graphics or recaptured image detectors are meant for real-world scenarios. One of the applications is in the court of law. As described in Section 3, such detector needs to be robust, rigorously evaluated, interpretable in physical terms, and capable of handling attacks. The current works have been lacking in addressing these practical issues and more efforts are needed to bring the research closer to real-world applications.

9 Conclusions

Distinguishing computer graphics and recaptured images from photographic images is important for image classification and countering security risks. Despite a number of works in these areas, there are still many issues that remain open. The security risk due to not being able to distinguish photographic and non-photographic images is real. This risk greatly reduces the value of photographs and their applications in computer vision. The alternative image synthesis will become more sophisticated with advances in computer graphics and computer vision. Therefore, the security risk will certainly become more prominent in the coming years. The body of work presented in this chapter represents an initial step in containing the mentioned security risks and more exciting results are anticipated in this area of research as we move forward into the future.

Appendix: True Image Labels for Images in Figure 2 and 4

Figure 2a and 2c are computer generated, while Figure 2b and 2d are photographic images. Figure 4a and 4a are photographic images of 3D face, while Figure 4b and 4d are recaptured images.

References

1. V. Athitsos, M. J. Swain, and C. Frankel. Distinguishing photographs and graphics on the world wide web. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 10–17, 1997.
2. J. Bai, T.-T. Ng, X. Gao, and Y.-Q. Shi. Is physics-based liveness detection truly possible with a single image? In *Proc. of IEEE International Symposium on Circuits and Systems*, 2010.
3. H. Cao and A. C. Kot. Identification of recaptured photographs on LCD screens. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010.
4. R. Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.
5. W. Chen, Y.-Q. Shi, and G. Xuan. Identifying computer graphics using HSV color model and statistical moments of characteristic functions. In *Proc. of IEEE International Conference on Multimedia and Expo*, pages 1123–1126, 2007.
6. Y. Chen, Z. Li, M. Li, and W. Y. Ma. Automatic classification of photographs and graphics. In *Proc. of IEEE International Conference on Multimedia and Expo*, pages 973–976, 2006.

7. T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland. Multimodal person recognition using unconstrained audio and video. In *International Conference on Audio- and Video-Based Person Authentication*, pages 176–181, 1999.
8. K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):34, 1999.
9. S. Dehnie, T. Sencar, and N. Memon. Digital image forensics for identifying computer generated and digital camera images. In *Proc. of IEEE International Conference on Image Processing*, pages 2313–2316, 2006.
10. A. E. Dirik, S. Bayram, H. T. Sencar, and N. Memon. New features to identify computer generated images. In *Proc. of IEEE International Conference on Image Processing*, volume 4, pages 433–436, 2007.
11. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley-Interscience, 2001.
12. M. Everingham, A. Zisserman, C. Williams, L. Van Gool, M. Allan, C. Bishop, O. Chapelle, N. Dalal, T. Deselaers, and G. Dorko. The 2005 PASCAL visual object classes challenge. *Machine Learning Challenges*, pages 117–176, 2006.
13. H. Farid. Creating and detecting doctored and virtual images: Implications to the child pornography prevention act. Technical report, TR2004-518, Dartmouth College, Computer Science, 2004.
14. H. Farid and M.J. Bravo. Perceptual discrimination of computer generated and photographic faces. *Digital Investigation*, 2011.
15. H. Farid and S. Lyu. Higher-order wavelet statistics and their application to digital forensics. In *IEEE Workshop on Statistical Analysis in Computer Vision*, 2003.
16. L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In *Workshop on Generative-Model Based Vision*, 2004.
17. R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of Computer Vision and Pattern Recognition*, 2003.
18. J. A. Ferwerda. Three varieties of realism in computer graphics. In *Proc. SPIE Human Vision and Electronic Imaging*, volume 3, 2003.
19. A. C. Gallagher and T. Chen. Image authentication by detecting traces of demosaicing. In *Proc. of Computer Vision and Pattern Recognition*, 2008.
20. X. Gao, T.-T. Ng, B. Qiu, and S.-F. Chang. Single-view recaptured image detection based on physics-based features. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2010.
21. X. Gao, B. Qiu, J. Shen, T.-T. Ng, and Y.-Q. Shi. A smart phone image database for single image recapture detection. In *Proc. of International Workshop on Digital Watermarking*, 2010.
22. P. Grassberger. Generalized dimensions of strange attractors. *Physics Letters A*, 97(6):227–230, 1983.
23. M. Grossberg and S. K. Nayar. What is the space of camera response functions? In *Proc. of Computer Vision and Pattern Recognition*, 2003.
24. R. Hartley and A. Zisserman. *Multiple view geometry*, chapter 6, page 164. Cambridge university press, 2000.
25. G. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.
26. T. I. Ianeva, A. P. de Vries, and H. Rohrig. Detecting cartoons: A case study in automatic video-genre classification. In *Proc. of IEEE International Conference on Multimedia and Expo*, volume 1, 2003.
27. J. T. Kajiya. The rendering equation. In *Proc. of ACM SIGGRAPH*, pages 143–150, 1986.
28. A. H. Kam, T.-T. Ng, N. G. Kingsbury, and W. J. Fitzgerald. Content based image retrieval through object extraction and querying. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, 2000.
29. N. Khanna, G. T. C. Chiu, J. P. Allebach, and E. J. Delp. Forensic techniques for classifying scanner, computer generated and digital camera images. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1653–1656, 2008.

30. M. Kharrazi, H. T. Sencar, and N. Memon. Benchmarking steganographic and steganalysis techniques. *Proceedings of SPIE Electronic Imaging*, 2005.
31. J. Kivinen, A. J. Smola, and R. C. Williamson. Online learning with kernels. *IEEE Transactions on Signal Processing*, 52(8):2165–2176, 2004.
32. K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. *Image Vision Computing*, 27(3):233–244, 2009.
33. S. Kovach. This guy just exposed a major security flaw in ice cream sandwich. `http://www.businessinsider.com/ice-cream-sandwich-face-unlock-2011-11`, 2011.
34. J.F. Lalonde and A.A. Efros. Using color compatibility for assessing image realism. In *Proc. of Int'l Conf. on Computer Vision*, 2007.
35. J. Li, Y. Wang, T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra. In *Proc. of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5404, pages 296–303, Aug 2004.
36. R. Lienhart and A. Hartmann. Classifying images on the web automatically. *Journal of Electronic Imaging*, 11(4):445–454, 2002.
37. J. Lukáš, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Security and Forensics*, 16(3):205–214, 2006.
38. S. Lyu and H. Farid. How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 53(2):845–850, 2005.
39. D. R. Magee and R. D. Boyle. Detecting lameness using re-sampling condensation and multi-stream cyclic hidden markov models. *Image and Vision Computing*, 20(8):581–594, 2002.
40. S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.
41. S. Mallick, T. Zickler, P. Belhumeur, and D. Kriegman. Specularity removal in images and videos: A pde approach. *Proc. of European Conf. on Computer Vision*, pages 550–563, 2006.
42. B. B. Mandelbrot. *The fractal geometry of nature*. W. H. Freeman, 1982.
43. G. W. Mayer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance. An experimental evaluation of computer graphics imagery. In *Proc. of ACM SIGGRAPH*, pages 30–50, 1986.
44. A. McNamara. Exploring perceptual equivalence between real and simulated imagery. In *Proc. of the ACM symposium on Applied perception in graphics and visualization*, page 128, 2005.
45. J. Michels, A. Saxena, and A. Y. Ng. High speed obstacle avoidance using monocular vision and reinforcement learning. In *Proc. of Int'l Conf. on Machine Learning*, pages 593–600, 2005.
46. G. Miller and C. R. Hoffman. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH 84 Advanced Computer Graphics Animation seminar notes*, volume 190, 1984.
47. T.-T. Ng and S.-F. Chang. Classifying photographic and photorealistic computer graphic images using natural image statistics. Technical report, ADVENT Technical Report, 220-2006-6, Columbia University, Oct 2004.
48. T.-T. Ng and S.-F. Chang. An online system for classifying computer graphics images from natural photographs. In *Proc. of SPIE*, volume 6072, pages 397–405, 2006. `http://apollo.ee.columbia.edu/trustfoto/trustfoto/natcgV4.html`.
49. 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 *Proc. of ACM International Conference on Multimedia*, pages 239–248, 2005.
50. T.-T. Ng, S.-F. Chang, Y.-F. Hsu, and M. Pepeljugoski. Columbia photographic images and photorealistic computer graphics dataset. Technical report, ADVENT Technical Report, 203-2004-3, Columbia University, Feb 2005.
51. T.-T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun. Passive-blind image forensics. In *Multimedia Security Technologies for Digital Rights*, pages 111–137. Elsvier, 2006.

52. T.-T. Ng, S.-F. Chang, and M.-P. Tsui. Lessons learned from online classification of photorealistic computer graphics and photographs. In *IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, 2007.
53. T.-T. Ng and M.-P. Tsui. Camera response function signature for digital forensics – part i: Theory and data selection. In *IEEE Workshop on Information Forensics and Security (WIFS)*, Dec 2009.
54. D. Ngo. Vietnamese security firm: Your face is easy to fake. `http://news.cnet.com/8301-17938_105-10110987-1.html`, 2008.
55. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
56. F. Pan, J. B. Chen, and J. W. Huang. Discriminating between photorealistic computer graphics and natural images using fractal geometry. *Science in China Series F: Information Sciences*, 52(2):329–337, 2009.
57. G. Pan, L. Sun, Z. H. Wu, and S. H. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Proc. of Int'l Conf. on Computer Vision*, pages 1–8, 2007.
58. A. C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of resampling. *IEEE Transactions on Signal Processing*, 53(2):758–767, 2005.
59. P. Rademacher, J. Lengyel, E. Cutrell, and T. Whitted. Measuring the perception of visual realism in images. In *Proc. of the Eurographics Workshop on Rendering Techniques*, pages 235–248, 2001.
60. R. Raskar, J. Tumblin, A. Mohan, A. Agrawal, and Y. Li. Computational photography. *Proc. of Eurographics State of the Art Report*, 2006.
61. A. Rocha and S. Goldenstein. Is it fake or real? In *XIX Brazilian Symposium on Computer Graphics and Image Processing*, pages 1–2, 2006.
62. Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, 10(1):39–62, 1999.
63. G. Sankar, V. Zhao, and Y. H. Yang. Feature based classification of computer graphics and real images. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1513–1516, 2009.
64. Y. Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, and C. Chen. Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2005.
65. E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.
66. 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*, 18(1):17–33, 2003.
67. P. Sutthiwan, X. Cai, Y.-Q. Shi, and H. Zhang. Computer graphics classification based on markov process model and boosting feature selection technique. In *Proc. of IEEE International Conference on Image Processing*, pages 2877–2880, 2009.
68. P. Sutthiwan, J. Ye, and Y.-Q. Shi. An enhanced statistical approach to identifying photorealistic images. *Digital Watermarking*, pages 323–335, 2009.
69. R. T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2):178–193, 2005.
70. A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 854–869, 2007.
71. S. Umeyama and G. Godin. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(5):639–647, 2004.
72. F. Wang and M. Y. Kan. NPIC: Hierarchical synthetic image classification using image search and generic features. *Image and Video Retrieval*, pages 473–482, 2006.

73. N. Wang and W. Doube. How real is really? a perceptually motivated system for quantifying visual realism in digital images. In *Proc. of IEEE International Conference on Multimedia and Signal Processing*, volume 2, pages 141–149, 2011.
74. Y. Wang and P. Moulin. On discrimination between photorealistic and photographic images. In *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 161–164, 2006.
75. A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In *Information Hiding*, pages 61–76. Springer, 2000.
76. J. Wu, M. V. Kamath, and S. Poehlman. Color texture analysis in distinguishing photos with computer generated images. In *Proc. of the UW and IEEE Kitchener-Waterloo Section Joint Workshop on Knowledge and Data Mining*, 2006.
77. J. Wu, M. V. Kamath, and S. Poehlman. Detecting differences between photographs and computer generated images. In *Proc. of the 24th IASTED international conference on Signal processing, pattern recognition, and applications*, pages 268–273, 2006.
78. H. Yu, T.-T. Ng, and Q. Sun. Recaptured photo detection using specularity distribution. In *Proc. of IEEE International Conference on Image Processing*, pages 3140–3143, 2008.
79. R. Zhang, R. Wang, and T.-T. Ng. Distinguishing photographic images and photorealistic computer graphics using visual vocabulary on local image edges. In *Proc. of International Workshop on Digital-forensics and Watermarking*, 2011.