Tian-Tsong Ng and Shih-Fu Chang

Identifying and Prefiltering Images

Distinguishing between natural photography and photorealistic computer graphics

© BRAND X PICTURES

Given the ability of photorealistic computer graphics (photorealistic CG) to emulate photographic images, as seen in movies and the print media today, there is little doubt that uninformed viewers can easily mistake photorealistic computer-generated graphics for photographic images. In fact, there was already evidence 20 years ago [1] that to the naked eye certain computer graphics were visually indistinguishable from photographic images. Such convincing photorealism qualifies computer graphics as a form of image forgery that can be unscrupulously exploited. Some popular Web sites [2] even highlight examples of computer-generated photorealism that human eyes find indistinguishable from photographic images.

While human experts can be called upon to examine and detect these deceptive computer images, this will eventually become infeasible given the enormous number of images created and circulated every day. Therefore, automated and effective computer techniques that can be used to distinguish between natural photographic images and photorealistic CG are needed to identify or prefilter these fabricated images. In criminal investigation (e.g., child pornography), the

Digital Object Identifier 10.1109/MSP.2008.931077

verification of images captured from real-world scenes (instead of synthesized by CG tools) is important from the viewpoint of law interpretation and enforcement in some countries (such as the United States). For example, an owner of child pornographic images may try to avert legal prosecution by declaring the images in question to be computer-generated, rather than images of real minors.

Apart from digital forensics, methods for distinguishing between photographic and photorealistic CG images are useful for automatic image-type classification, indexing, and retrieval. In [3], the performance of video key-frame retrieval showed improvement when the computer graphics key frames that included two-dimensional (2-D) graphics were detected and prefiltered prior to retrieval. Nowadays, it is common to see graphic content in our everyday environment, such as posters on notice boards, movie ads at bus stops, traffic signposts, and so on (see Figure 8 for examples). As these types of images are often printed on flat surfaces, they provide a strong clue about the scene geometry. Therefore, methods for distinguishing between photographic and photorealistic CG images that can be used to detect such graphic content are also potentially useful for image/scene understanding.

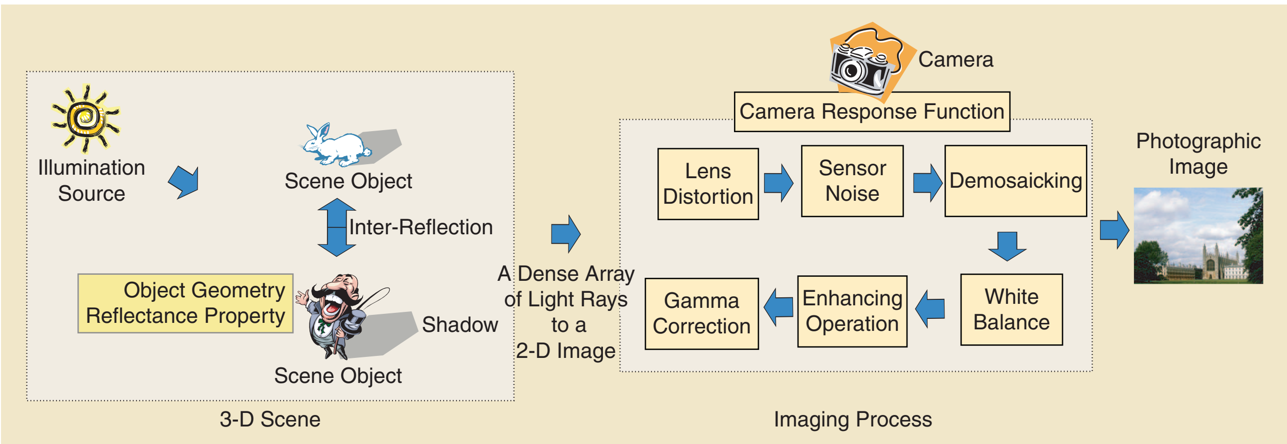
In the following section, we summarize the general processing pipeline used in typical imaging devices and the general approaches used in computer graphics to emulate the realism of natural photographs. We will highlight the commonalities and differences between these two processes in order to select effective cues that may be used to design automatic techniques for detecting computer graphics.

FORMATION OF PHOTOGRAPHIC IMAGES

A typical real-world scene consists of light sources and three-dimensional (3-D) objects, where the light sources illuminate the scene while the objects reflect, refract, transmit, or diffuse the light impinging on their surfaces. This process of light interaction fills the 3-D space with a dense array of light rays of various intensities, which can be visualized as a set of panoramic images or light rays at different 3-D locations. This set of light rays can be represented as the seven-dimensional (7-D) plenoptic function [4] of the space (three dimensions), orientation (two dimensions), time (one dimension) and wavelength of the light (one dimension) or equivalently as the four-dimensional (4-D) light field, defined as static radiance as a function of position and direction in free space [5]. Through the light rays a 3-D scene communicates its visual information to a human observer and a camera. When a camera takes a snapshot of the scene, the camera effectively samples an instantaneous cone of light rays at a position in the 3-D space. The sampled cone of radiance is recorded on the film in a film camera or by a sensor array in a digital camera as a 2-D image. This process of photographic image formation is illustrated in Figure 1.

Despite the variety in consumer camera brands, the physical structure and in-camera operation pipeline of most cameras are quite similar. In Figure 1, a typical in-camera operation pipeline is shown. The radiance from a scene point is focused on an optical sensor through a lens or a lens system. Due to the foreshortening of the views between the sensors and the scene as well as the aperture, the transmittance of most lenses falls off as it occurs further away from the optical center. This effect is called vignetting and sometimes results in a noticeable radial brightness falloff on an image. Furthermore, when the refractive index of a lens is different for different wavelengths of light, a light ray could diverge over the wavelength as it passes through the lens (a phenomenon called chromatic aberration), resulting in color fringes on an image. Besides that, an imperfect lens can also result in geometric distortion of an image at various degrees.

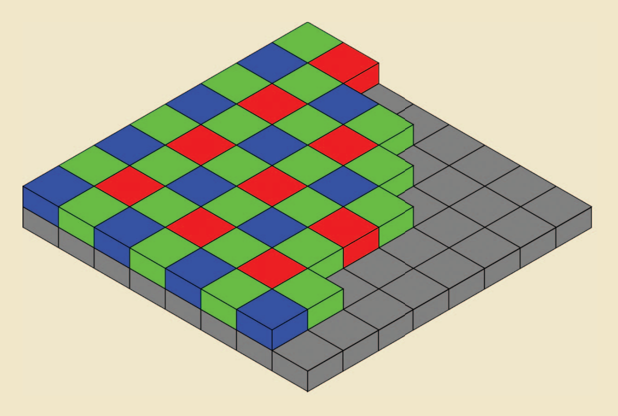
The role of the optical sensor is to convert the light energy into electric voltage levels, namely radiance measurements. Most image sensors used today, including the charge-coupled device (CCD) and the complementary metal-oxide-semiconductor (CMOS) sensor, are pixilated metal-oxide semiconductors. Therefore, they all suffer from similar forms of noise, such as pattern noise, dark current noise, shot noise, and thermal noise, albeit to a different extent [6]. (For further details about camera noise, see [7].)



[FIG1] The photographic image-formation process.

The actual color of a light is determined by its spectral power distribution. As its perceptual effect can be approximated by a red, green, and blue (RGB) trichromatic model, most cameras measure RGB colors. For most consumer cameras with a single sensor array, a mosaicked color filter array (CFA) is applied on the sensor, with the Bayer pattern being the most common type [8]. For mosaicked sensors, only one out of the three colors is measured at each pixel site; hence, two other colors are missing for each site, as shown in Figure 2. The process of interpolating the missing measurements is called demosaicking. Demosaicking may result in physically incorrect colors as well as undesirable artifacts. Recently, Foveon introduced a new sensor called the Foveon X3 sensor (a CMOS sensor) that can mimic color negative film by stacking the RGB color-sensitive elements on top of each other, in layers, at each pixel site [9]. As a result, it can measure three RGB colors at each site, and demosaicking is no longer needed. Apart from the Foveon sensor, demosaicking is also not needed for the more expensive three-CCD cameras [10], where the incoming light is physically split by a trichroic prism into the three RGB components, measured by three separate CCD sensors.

A typical camera also performs white balancing to offset the tinge of the illumination color, enhances sharpness and contrast, and performs gamma correction for dynamic range compression. Finally, the overall effect of the above-mentioned operations from the radiance measurement at the sensors to the final intensity output by the camera can be modeled by a camera response function with a typical concave shape [11].



[FIG2] Bayer filter mosaic for image sensor array. CG images may not involve such filters during the synthesis process and thus can be distinguished from natural photographic images.

FORMATION OF PHOTOREALISTIC COMPUTER GRAPHICS

The motivation of synthesizing realistic image has been a driving force behind the various technical breakthroughs in the physics-based graphics rendering that attempts to emulate the photographic image formation process described in the previous section. Realistic image synthesis is important for applications such as simulation, design, education, and advertising. Despite the general notion of a realistic image, its exact definition remains a subject of scholarly debate. In [12], three varieties of realism were defined for computer graphics: physical realism, which provides the same visual stimulation as the scene; photorealism, which produces the same visual response as the scene; and functional realism, which provides the same visual information as the scene (such as an object's shape and depth). Of these three varieties, the image forensics community is mainly concerned with photorealistic CG. With today's CG techniques, despite being extremely computationally expensive, photorealistic CG is definitely achievable.

The verisimilitude of photorealism is often the collective result of various visual effects contained within a 3-D scene. There are various levels of complexity in the scene and object geometry, illumination, and object reflectance of a scene, as illustrated in Figure 3. Complex scene geometry gives rise to the color-bleeding effect, where the color of an object may spill over to its surrounding due to the interreflection between scene objects. Complex illumination consists of multiple types of light sources both near and far and coming from all directions. Finally, complex object reflectance gives rise to the appearance of transparency, translucency, reflection, diffuse shading, and specularity. All these effects define photorealism.

(a) (b) (c)

[FIG3] Complex components in real-world image formation lead to the many complex and subtle perceptual effects required in photorealism: (a) a complex environment light is reflected on a mirror sphere; (b) subsurface scattering reflectance of a grassy surface results in translucency; (c) the fur of a lion causes unique perceptual responses due to its complex geometry and reflectance properties.

A 3-D graphics system takes a scene made up a set of light sources and geometric primitives with material properties as input and renders a CG image as output. Therefore, the two important components of photorealistic graphics synthesis are 1) scene modeling, which includes the modeling of the illumination, object reflectance, and object geometry in a scene; and 2) scene rendering. With realistic scene modeling and correct light-transport simulation, photorealistic CG can be generated.

OTHER TECHNICAL ISSUES CONCERNING DISTINGUISHING PHOTOGRAPHIC AND COMPUTER GRAPHICS IMAGES INCLUDE BENCHMARK DATASETS FOR PERFORMANCE EVALUATION AND EFFICIENT SYSTEM IMPLEMENTATION.

MODELING OF ILLUMINATION, OBJECT REFLECTANCE, AND OBJECT GEOMETRY

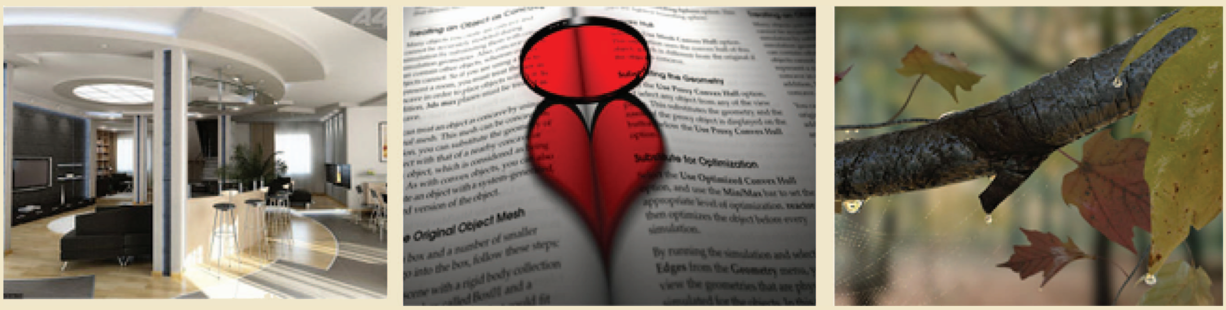
In recent years, scene modeling has undergone a transition from simple parametric modeling to image-based modeling. Image-based models are now able to capture the complexity of a real-world scene for which the corresponding parametric models only offer a crude approximation. For scene illumination, synthetic light sources (e.g., point/area light) as used in CG are far too simple as compared with the illumination present in the real scene, which can be measured as an environment map using a mirror sphere [13] or with other similar methods. The environment map can then be used to model a complex light source in CG rendering. However, if the dynamic range of the environment map is insufficient for the real-scene dynamic range, the rendered image may still be different from the photographic one.

Simple parametric surface reflectance models (such as the Phong model) are incapable of producing certain visual effects, e.g., translucency. To model the complex reflectance of real-world surfaces, measuring the reflectance from real surface samples is more effective. For instance, spatially varying surface reflectance (texture) can be measured from multiple-view photographs [14]. However, despite the more accurate image-based reflectance model, the color of a reflectance model is often represented in an RGB trichromatic format, and the rendering is often done independently for each of the color channels. The result may fall short of the actual interaction of light between objects.

In the CG pipeline, the geometry of objects is often represented as a polygonal mesh. The geometry of the real-world objects in the form of a polygonal mesh can be obtained via range scanning [15]. Although increasing the mesh resolution can improve the accuracy of the geometric representation, it leads to a higher computational load; hence a compromise is needed and is often adopted in practice. Despite the various shading-interpolation methods (e.g., Phong shading and Gourand shading) for removing the faceted appearance of a polygonal object during rendering, obvious artifacts such as the polygonal silhouette edge can still arise from a coarse-resolution mesh.

CG RENDERING

CG rendering simulates the light transport between the illumination sources and the object surfaces. The light transport can be very complicated, as it may involve multiple bounds of light-from one location of the scene to the others that give rise to visual effects such as soft shadow, color bleeding, caustics, and so on. This complex process can be described by an integral rendering equation [16]. Current photorealistic CG rendering methods such as ray tracing and radiosity effectively amount to solving the rendering equation in approximation. Solving the rendering equation exactly is very challenging, as no closed-form solution exists for the general case. Various forms of assumption result in various levels of photorealism in rendering. At the simplistic end, only light originating directly from light sources is considered in rendering a local surface, and the interreflection between surfaces is ignored. With ray tracing and radiosity, on the other hand, multiple light bounds between surfaces are simulated to produce the global illumination effects often featured nowadays in 3-D rendering software such as Autodesk 3D Max Studio (see Figure 4). However, the



[FIG4] Examples of photorealistic computer graphics rendered using Autodesk 3-D Max Studio. Note the effects of global illumination rendering, such as glass refraction.

approximation adopted by such rendering techniques will only result in simulating a simplified light transport in the scene, and the resulting rendered image may still be different from the photographic one. For example, the ray-tracing technique, which discretizes the view plane, may need to generate a very large number of rays in order to render a diffuse scene, while the radiosity technique, which discretizes the environment, may need a very large amount of storage to render objects with sharp specularity [17]. Finally, after an image is rendered, it may often be processed via a simplified camera model (e.g., only with gamma correction) in order to produce a photographic appearance, while most of the in-camera operations (such as demosaicking) are not applied.

[SOME POPULAR WEB SITES EVEN HIGHLIGHT EXAMPLES OF COMPUTER-GENERATED PHOTOREALISM THAT HUMAN EYES FIND INDISTINGUISHABLE FROM PHOTOGRAPHIC IMAGES.]

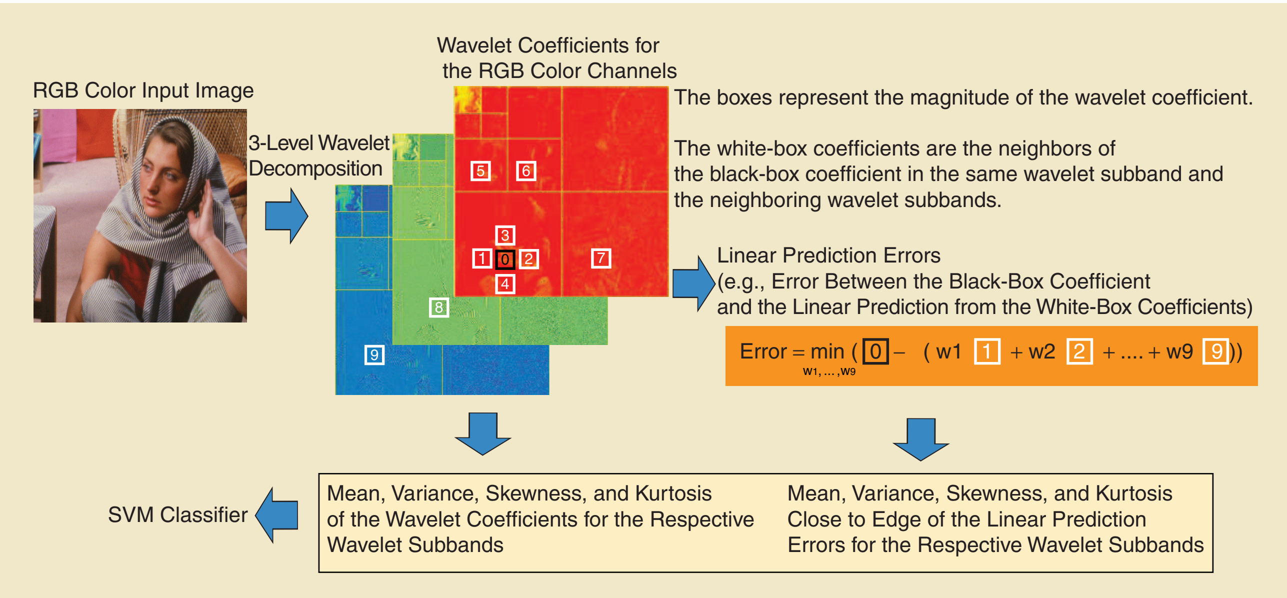
In this section, we have highlighted some differences between photographic image formation and CG image formation, when practical approximations enter into scene modeling and rendering. Some of these differences, such as the absence of the camera operations and the color-independence assumption used in CG image production, have been exploited to a certain extent for distinguishing the two groups of images (see next section), but the artifacts due to insufficiently complex lighting models and simplified light-transport simulation in CG have yet to be explored in this regard.

STATE OF THE ART IN DISTINGUISHING PHOTOGRAPHIC AND PHOTOREALISTIC IMAGES

After describing the image formation process for photographic and CG images, we now review the state of the art in automatic classification of images from these two different sources.

In an early work [3], classification of photographic and general CG video key-frame images was proposed as a prefiltering step for a more effective video key-frame retrieval system. The CG images being considered were mainly of the cartoon type, with saturated, uniform colors, strong, distinct lines, and so on. The proposed CG detection method uses image features such as the average color saturation, the ratio of image pixels with brightness greater than a threshold to total pixels, the hue-saturation-value (HSV) color histogram, the edge orientation and strength histogram, the compression ratio, and the pattern spectrum. In contrast, the CG images being studied in digital forensics are ones with higher photorealism that that seen in the simple CG images discussed in [3].

In prior work in digital forensics, one branch of development considers statistical models of images. Lyu and Farid considered that photographic images have different statistical characteristics in the wavelet transform domain as compared with photorealistic CG images [18]. The computational steps for their method are illustrated in Figure 5. An RGB input image is first decomposed into three levels of wavelet subbands. In wavelet decomposition, a natural-scene image is bandpassed and disintegrated into subbands of wavelet coefficients for different scales and orientations, as shown in Figure 5. Note that the wavelet coefficients corresponding to an edge respond strongly in a subband when the edge matches the scale and orientation of the subband. In the study of nature image statistics [19], it is well known that, for a natural-scene image, the wavelet coefficients in a subband are Laplacian distributed, and correlation exists



[FIG5] Computational steps for the statistical method in [18].

between wavelet coefficients of the adjacent subbands. Lyu and Farid modeled the former statistics using the moments of the wavelet coefficients of a subband and the latter using the linear prediction error of the coefficients. Figure 5 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 both the same subband and the adjacent subbands. The prediction error is an indication of the correlation strength between the coefficients. Four moments (mean, variance, skewness, and kurtosis) of the wavelet coefficient distribution and the linear prediction error distribution are then computed from each subband and used as input features for an SVM classifier. Lyu and Farid performed a classification experiment on a dataset with 40,000 photographic images and 6,000 photorealistic CG images in which the SVM classifier achieved a classification rate of 66.8% on the photographic images, with a false negative rate of 1.2%.

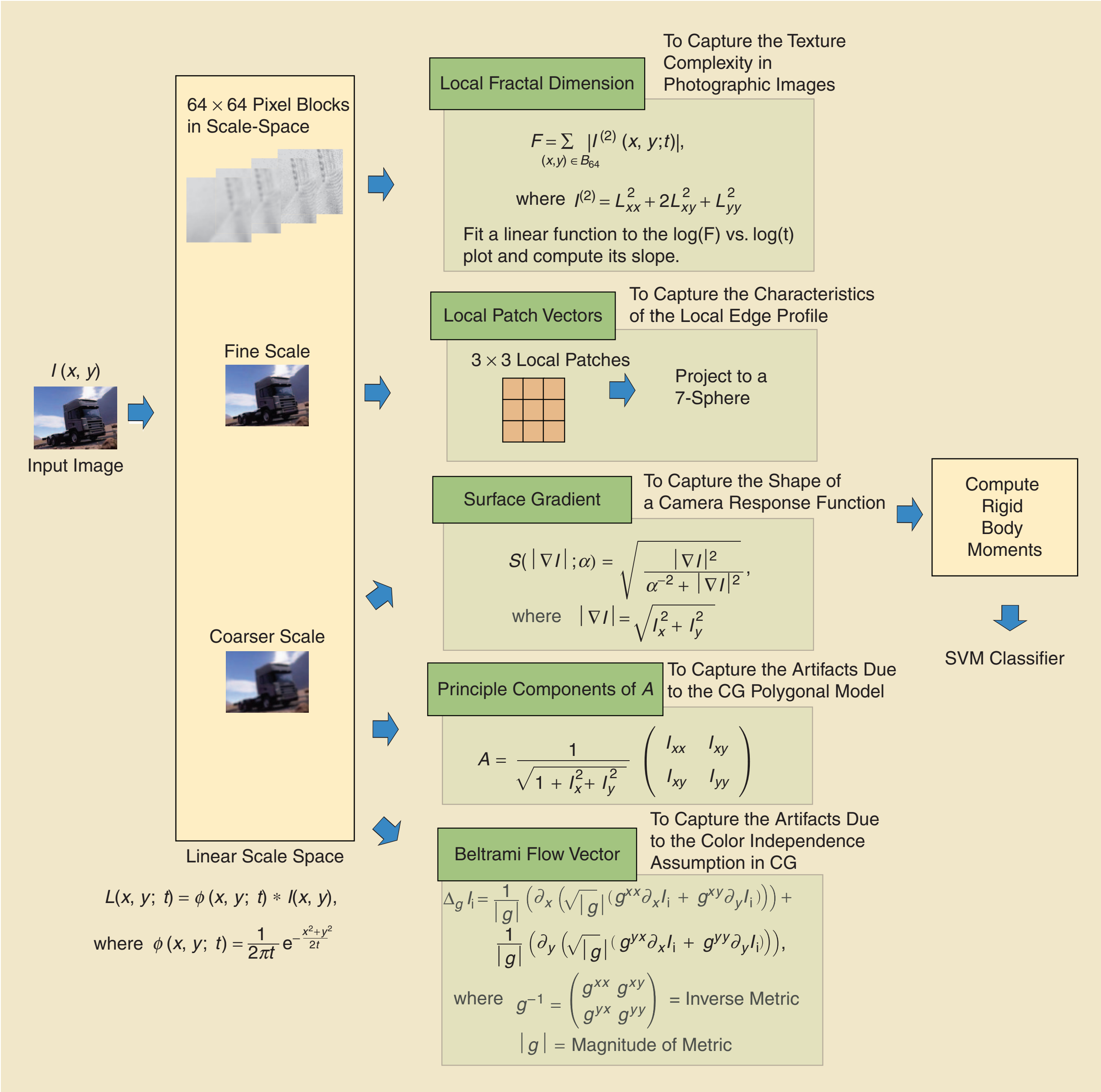
In the same vein, [20] proposed a simplified way to compute the statistical features in the wavelet transform domain and is able to achieve a classification performance comparable with that of [18] but with a four-fold reduction in computational time. Chen et al. [21] proposed that the HSV color space is a more effective color space than the RGB color space for computing the wavelet statistics. In their work, instead of computing the four moments (mean, variance, skewness, and kurtosis) of the wavelet coefficient distribution and the linear prediction error distribution, as in [18], the moments of their characteristic function were computed. On their dataset, they were able to achieve a classification accuracy of 82.1%.

THE MOTIVATION OF SYNTHESIZING REALISTIC IMAGE HAS BEEN A DRIVING FORCE BEHIND THE VARIOUS TECHNICAL BREAKTHROUGHS IN THE PHYSICS-BASED GRAPHICS RENDERING THAT ATTEMPTS TO EMULATE THE PHOTOGRAPHIC IMAGE FORMATION PROCESS.

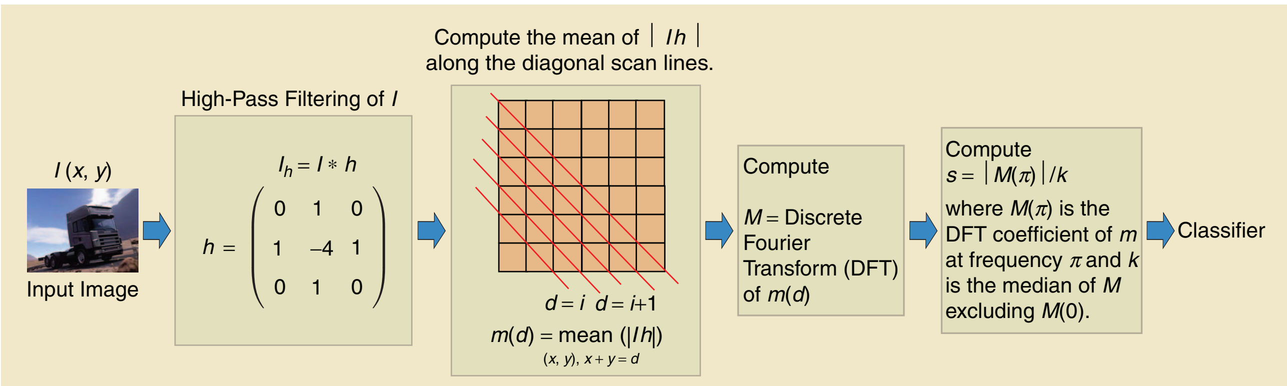
Another branch of development considers the physical models of images. Ng et al. [22] studied the photographic and CG image formation processes as shown in Figure 1 and identified three aspects wherein the two processes have discrepancies. First, photographic images are subject to the typical concave response function of cameras, while the CG rendering pipeline may not have a standardized postprocessing procedure that mimics camera processing. It was shown that the effect of the camera response function manifests on the image gradients. Second, the polygonal representation of the CG object geometry is often too coarse for capturing real-world object geometry with many fine details. The coarseness of the polygon can give rise to sharp edges and polygon-shaped silhouettes. These sharp features were measured using the principal curvatures of the image function. Third, as mentioned earlier, the three color channels of computer graphics images are often rendered independently to reduce computational load. The independence assumption is in general not true in real scenes, where the incident light of a certain spectral band on a surface may induce reflected light of a different spectral band. To capture this discrepancy, independence between the color channels is measured using the Beltrami flow vectors on an RGB image function. The computational steps of the method used in [22] are shown in Figure 6. Apart from the above-mentioned image gradient, principal curvatures, and Beltrami flow vectors, the authors of [22] also computed the local block-based fractal dimension and the local patch vectors. The local fractal dimension was meant to capture the texture complexity in photographic images, and the local patch vectors were meant to capture the characteristics of the local edge profile. The five features, taken together, produce a vector field on the image domain. The statistics of the vector fields in the form of rigid body moments were computed and used as the input features for an SVM classifier. On the authors' open dataset (to be discussed later in this article), an average classification accuracy of 83.5% was attained.

Additional camera-related characteristics associated with the in-camera operation pipeline as shown in Figure 1 can also be used to separate the photographic and CG images. The authors of [23] considered camera noise characteristics, and the authors of [24] considered both demosaicking and the chromatic aberration of cameras. In [23], the authors demonstrated that the noise pattern of photographic images, extracted via a wavelet denoising filter, is distinguishable from that of CG images. Hence, the two groups of images can be represented by their respective reference noise patterns, and a test image can be classified based on its correlation with the reference noise patterns. On their own dataset, the method achieved an averaged classification accuracy of about 72%. In [24], the authors observed that an image originally generated from the Bayer pattern will experience a smaller change if it is reinterpolated again based on the Bayer pattern as compared with other patterns. This observation enabled detection of camera images generated from the Bayer pattern. They also measured the misalignment among the color channels due to the chromatic aberration that operates on the incoming light of different wavelengths in a camera. Using these two physical characteristics unique to natural photos, the method used achieved an average classification accuracy of about 90% on the authors' own dataset.

In [25], the presence of the Bayer-pattern type of demosaicking in original-size camera images was detected and used to distinguish images captured by cameras from CG images, which normally do not bear such cues. The computational steps for their method are shown in Figure 7. Their method is based on two main observations: 1) the interpolated coefficients always have a smaller variance than the



[FIG6] The computational steps of the method used in [22].



[FIG7] Computational steps for the method used in [25].

original coefficients, and high-pass filtering can make the difference more distinct; and 2) in the green-color-channel image with the Bayer pattern, the interpolated and original coefficients are respectively grouped together in alternate diagonal scan lines. Therefore, the variance of the diagonal scan lines should have a frequency of two. Their method achieved an average classification accuracy of 98.4% on the open Columbia dataset [26]. However, their method could be sensitive to image resampling or resizing, where the restrictive interpolation structure of the Bayer pattern may be destroyed.

THE APPLICATIONS ARISING FROM TACKLING THIS GENERAL PROBLEM INCLUDE DIGITAL FORENSICS, PREVENTING RECAPTURING ATTACKS, IMAGE INDEXING, AND IMAGE/SCENE UNDERSTANDING.

In a digital forensics setting, there may be attackers who want to beat the photographic and photorealistic CG classifier. As one form of attack, one could transform a CG image into a photographic one by recapturing it using a camera. The rephotographed image is confusing to the detector, as it bears the content of the original CG image while also carrying photographic characteristics (such as the camera response function and the mosaic filter). Ng et al. demonstrated that this issue can be partially addressed by including the rephotographed images in the dataset for classifier training [22]. More recently, a physics-based method based on the specularity distribution of a printed surface was proposed to detect rephotographed images [27]. In [27], it was observed that microstructures on the surface of printing paper, appearing as a high-frequency random noise pattern, can be found in the specularity component of a high-resolution image, even when the paper is printed with image contents. The random noise–like specularity pattern of photographed CG images can be very different from the normal specularity pattern of objects in natural-scene images. The method is particularly effective for distinguishing the human face (structured image content) when printed on paper and hence is useful as a countermeasure for face authentication spoofing.

BENCHMARKS AND EVALUATION

Other technical issues concerning distinguishing photographic and computer graphics images include benchmark datasets for performance evaluation and efficient system implementation. The Columbia benchmark dataset was constructed and made accessible to the research community [26], and an online public classification system was deployed [28]. The Columbia dataset consists of 800 personal photographic images, 800 photographic images obtained through Google Image Search, 800 photorealistic CG images from 3-D artist Web sites, and the recapture of the 800 CG images. The classification accuracies reported using this dataset are 83% [22], 82% [21], and 98% (for nonresized images and nonrecaptured images only) [25]. The online classification system in [28] offers a way for users to try out the fully automatic detection functions on test images they choose themselves, allowing for an interactive diagnostic setting in which to investigate successes and failures of the existing solutions. In [29], an evaluation on the Columbia online system was done in terms of its performance with regard to images submitted by users. While the classifiers in general perform well on images belonging to a clear-cut category, there is a group of images that confuse the classifiers. This group includes both photographs with graphic content embedded in 3-D real-world scenes and images composed using both photographic and CG elements, as shown in Figure 8.

http://www.iht.com/images/2006/06/25/web.0626city9ss4.jpg
(a)

http://www.spiegel.de/img/0,1020,681938,00.jpg
(b)

http://news.tom.com/img/assets/200311/031111135215gg111120.jpg
(c)

[FIG8] Confusing images for the Columbia online classification system. These images include photographs with graphic content in the (a) and (b) 3-D real-world scenes and (c) images composed using both photographic and CG elements.

[TABLE 1] EXTENSION OF COMPUTER GRAPHICS CONTENT TO MORE GENERAL TYPES OF IMAGES. CLASSIFICATION OF THE EXTENDED TYPES WILL LEAD TO APPLICATIONS BROADER THAN IMAGE FORENSICS.

GENERATION PROCESSES	SOURCE TYPES		
	3-D GRAPHICS	2-D GRAPHICS	REAL SCENES
COMPUTER RENDERING	PHOTOREALISTIC COMPUTER GRAPHICS	CARTOONS, MECHANICAL DRAWINGS, ETC.	IMAGE-BASED RENDERED IMAGES AND NONPHOTOREALISTIC RENDERINGS OF PHOTOGRAPHS
PHYSICAL PHOTOGRAPHING	RECAPTURED COMPUTER GRAPHICS, PRINTED IMAGES, BILLBOARD IMAGES, TRAFFIC SIGNPOSTS, ETC.		PHOTOGRAPHIC IMAGES

OPEN PROBLEMS AND FUTURE DIRECTION

One way to assess the effectiveness of a statistical classifier is through extensive testing, and this requires a large image dataset. The Columbia benchmark dataset described above was constructed in 2005 and consisted of 3,200 images of various content types [26]. While the number of images in the dataset can be increased, images rendered using more recent techniques also need to be included to keep up with advances in CG technology. Another way to reliably assess and boost the classifier performance is through the physical modeling of scenes and cameras. More detailed physical modeling, which includes lighting modeling, global illumination effect modeling, and modeling of various rendering techniques, could be pursued to supplement the physical characteristics that have been employed in current methods [22]–[25].

Photorealistic CG is a subset of general CG, which also includes line drawings, cartoons, nonphotorealistic renderings of photographs, and drawings incorporating geometric design. Furthermore, a computer-generated graphic may not always be the direct output of a computer: it may be rephotographed after having been printed on a surface or displayed on a computer screen. In the more general setting, if the image forensics problem is extended to include the above-mentioned nonphotorealistic CG and the graphic content in a photograph, it will open doors to a wider range of important applications. After all, the respective techniques for distinguishing the natural photographic images from other groups of images in Table 1 share certain commonalities. An immediate advantage of considering the more general problem is that we will improve our ability to handle the confusing cases shown in Figure 8, which are essentially photographs with graphic content.

ONE WAY TO ASSESS THE EFFECTIVENESS OF A STATISTICAL CLASSIFIER IS THROUGH EXTENSIVE TESTING, AND THIS REQUIRES A LARGE IMAGE DATASET.

In the general problem for future consideration, we may want to examine the five distinct classes of images shown in Table 1, which are grouped according to content source (either CG or real scenes) and the final generation processes (either rendered by computer tools or photographed by imaging devices). In addition to the case of graphic content in a photograph, Table 1 also shows a class of computer-generated images (in the top-right box) synthesized from photographs of a real 3-D scene, using both image-based rendering (e.g., interpolation of photographs taken from different viewpoints—see [30]) and nonphotorealistic rendering of photographs (e.g., generating a cartoon-style image from a photograph—see [31]).

The applications arising from tackling this general problem include digital forensics, preventing recapturing attacks, image indexing, and image/scene understanding. While the works reviewed in this article provide a starting point for addressing this classification problem, further techniques for recognizing computer graphics content without exclusively relying on the photographic properties of the imaging pipeline are required. Such content-based recognition is challenging. For example, in the case of rephotographing, the computer graphic is displayed or printed on a real physical surface and lit by a real illumination [27]. Techniques exploiting camera cues alone will no longer prevail; instead, content-based analyses (like those revealing natural scene properties and statistics) will become more important. In fact, the rephotographing of images is an art form by itself; representative works include those by Prince, who rephotographed magazine advertisements in the 1970s [32]. These works, interestingly, created an intense debate about photographic copyright issues and the authenticity of photographic images.

Emerging works using computational photography [33] also blur the boundaries among the categories in Table 1, and arriving at clear distinctions becomes even more difficult. Under such a paradigm, imaging devices will no longer be simple acquisition systems that faithfully record scene radiance. Instead, sophisticated manipulation techniques may be embedded in the imaging pipeline inside a "camera" to modify and enhance the images such that the signatures associated with the conventional concept of photorealism may no longer be completely valid and need to be redefined. However, the knowledge and tools discussed in this article about the characterization of CG content will still remain valuable.

CONCLUSIONS

In this article, we described the problem of distinguishing between photographic and photorealistic CG images in digital forensics and other related applications, surveyed the related work, and introduced some benchmark resources and online demonstrations that are publicly available. We also described the possible extension of the problem to include general CG content. This extension leads to a wider range of important applications beyond digital forensics, such as image indexing and image/scene understanding. Finally, we highlight some future directions for a larger and more diverse CG image dataset, a more detailed modeling of the physical characteristics, and the more general problem of content-based computer graphics recognition.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their very valuable comments. Our work has been supported in part by the Singapore A*STAR Graduate Academy and the U.S. National Science Foundation (awards IIS-04-30258 and CNS-0716203).

AUTHORS

Tian-Tsong Ng (ttng@i2r.a-star.edu.sg) received his Ph.D. in electrical engineering from Columbia University, his M.Phil. in information engineering from Cambridge University 2001, and his B.Eng. in electrical engineering from Malaya University. He was a Singapore A*STAR scholar. He joined the Institute for Infocomm Research in 2007, as a researcher. He works on image forensics and computational photography.

Shih-Fu Chang (sfchang@ee.columbia.edu) directs the Digital Video and Multimedia Lab and currently chairs the Department of Electrical Engineering of Columbia University. He has made significant contributions in multimedia search, media forensics, adaptive video transmission, and international standards for multimedia indexing. He has been recognized with the IEEE Kiyo Tomiyasu Award, Navy ONR Young Investigator Award, IBM Faculty Development Award, ACM Recognition of Service Award, and NSF CAREER Award. He and his students have received two Best Paper and six Best Student Paper Awards from IEEE, ACM, and SPIE. He has worked in different advising/consulting capacities for IBM, Microsoft, Kodak, PictureTel, and several other research institutions. He is an IEEE Fellow and past editor-in-chief for *IEEE Signal Processing Magazine*.

REFERENCES

[1] G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, "An experimental evaluation of computer graphics imagery," *ACM Trans. Graph.*, vol. 5, no. 1, pp. 30-50, Jan. 1986.

[2] *Fake or Foto* [Online]. Available: http://www.autodesk.com/eng/etc/fakeorfoto/quiz.html

[3] T. Ianeva, A. de Vries, and H. Rohrig, "Detecting cartoons: A case study in automatic video-genre classification," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, Baltimore, MD, 2003, vol. 1, pp. 449–452.

[4] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computation Models of Visual Processing*, M. Landy and J. A. Movshon, Eds. Cambridge, MA: MIT Press, 1991, pp. 3-20.

[5] M. Levoy and P. Hanrahan, "Light field rendering," *ACM SIGGRAPH*, New Orleans, LA, 1996, pp. 31–42.

[6] J. Luká , J. Fridrich, and M. Goljan, "Digital camera identification from sensor noise," *IEEE Trans. Inform. Sec. Forensics*, vol. 1, no. 2, pp. 205–214, June 2006.

[7] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Recognit. Machine Intell.*, vol. 16, no. 3, pp. 267–276, Mar. 1994.

[8] B. E. Bayer, "*Color imaging array*," U.S. Patent 3 971 065, 1976.

[9] *Foveon Company* [Online]. Available: http://www.foveon.com

[10] *Three-CCD* [Online]. Available: http://en.wikipedia.org/wiki/Three-CCD

[11] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?" in *Proc. IEEE Computer Vision and Pattern Recognition*, Madison, WI, 2003, vol. 2, pp. 602–609.

[12] J. A. Ferwerda, "Three varieties of realism in computer graphics," in *Proc. SPIE Human Vision and Electronic Imaging*, San Jose, CA, 2003, pp. 290–297.

[13] G. S. Miller and C. R. Hoffman, "Illumination and reflection maps: simulated objects in simulated and real environments," in *Course Notes for Advanced Computer Graphics Animation*. Chicago, IL: ACM SIGGRAPH, 1984, pp. 290–297.

[14] K. J. Dana, B. Van-Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.

[15] *Cyberware* [Online]. Available: http://www.cyberware.com

[16] J. T. Kajiya, "The rendering equation," *ACM SIGGRAPH*, Dallas, TX, 1986, vol. 20, no. 4, pp. 143-150.

[17] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips, *Introduction to Computer Graphics.* Reading, MA: Addison-Wesley, 1994.

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

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

[20] Y. Wang and P. Moulin, "On discrimination between photorealistic and photographic images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, 2006.

[21] W. Chen, Y. Q. Shi, and G. Xuan, "Identifying computer graphics using HSV color model and statistical moments of characteristic functions," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, Beijing, China, 2007, pp. 1123–1126.

[22] T.-T. Ng, S.-F. Chang, Y.-F. Hsu, L. Xie, and M.-P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *Proc. ACM Multimedia*, Singapore, 2005, pp. 239–248.

[23] S. Dehnie and T. Sencar, and N. Memon, "Digital image forensics for identifying computer generated and digital camera images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Atlanta, GA, 2006, pp. 2313-2316.

[24] A. E. Dirik, S. Bayram, H. T. Sencar, and N. Memon, "New features to identify computer generated images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, San Antonio, TX, 2007, vol. 4, pp. 1522-4880.

[25] A. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition, Workitorial on Vision of the Unseen*, Anchorage, AK, 2008, pp. 1–8.

[26] T.-T. Ng, S.-F. Chang, Y.-F. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," Columbia Univ. ADVENT Tech. Rep. 205-2004-5, 2005. [Online]. Available: http://www.ee.columbia.edu/ln/dvmm/downloads/PIM_PRCG_dataset/

[27] H. Yu, T.-T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, San Diego, CA, 2008, pp. 3140–3143.

[28] T.-T. Ng and S.-F. Chang, "An online system for classifying computer graphics images from natural photographs," in *Proc. SPIE Electronic Imaging*, San Jose, CA, 2006. [Online]. Available: http://apollo.ee.columbia.edu/trustfoto/trustfoto/natcgV4.html

[29] T.-T. Ng, S.-F. Chang, and M.-P. Tsui, "Lessons learned from online classification of photo-realistic computer graphics and photographs," in *Proc. IEEE Workshop on Signal Processing Applications for Public Security and Forensics (SAFE)*, Washington, D.C., Apr. 2007, pp. 1–6.

[30] S. M. Seitz and C. R. Dyer, "ViewMorphing: Synthesizing 3D metamorphoses using image transforms," *ACM SIGGRAPH*, Aug. 1996, pp. 21–30.

[31] J. Lansdown and S. Schofield, "Expressive rendering: A review of non-photorealistic techniques," *IEEE Comput. Graph. Applicat.*, vol. 15, no. 3, pp. 29–37, May 1995.

[32] Richard Prince Art. [Online]. Available: http://www.richardprinceart.com/

[33] R. Raskar, J. Tumblin, A. Mohan, A. Agrawal, and Y. Li, "Computational photography," in *Proc. Eurographics Tutorials*, Prague, Czech Republic, 2007.

SP