

# Classifying Photographic and Photorealistic Computer Graphic Images using Natural Image Statistics

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

ADVENT Technical Report #220-2006-6  
Oct 2004

## Abstract

As computer graphics (CG) is getting more photorealistic, for the purpose of image authentication, it becomes increasingly important to construct a detector for classifying photographic images (PIM) and photorealistic computer graphics (PRCG). To this end, we propose that photographic images contain natural-imaging quality (NIQ) and natural-scene quality (NSQ). NIQ is due to the imaging process, while NSQ is due to the subtle physical light transport in a real-world scene. We explicitly model NSQ of photographic images using natural image statistics (NIS). NIS has been used as an image prior in applications such as image compression and denoising. However, NIS has not been comprehensively and systematically employed for classifying PIM and PRCG. In this work, we study three types of NIS with different statistical order, i.e., NIS derived from the power spectrum, wavelet transform and local patch of images. The experiment shows that the classification is in line with the statistical order of the NIS. The local patch NIS achieves a classification accuracy of 83% which outperforms the features derived from modeling the characteristics of computer graphics.

## 1 Introduction

Traditional photographs were considered faithful records of the state of a real-world event, as its manipulation is not only technically challenging (e.g. requires



Figure 1: Examples of authentic images (top row) and photorealistic CG (bottom row).

contriving multiple exposures of a film in a darkroom), traces of manipulations are normally revealing. Unfortunately, today's digital images, being just an array of numbers, are susceptible to tampering. Even back in 1989, ten percent of all color photographs published in the United States were digitally retouched or altered, according to the Wall Street Journal estimate [1].

While compositing camera images is a popular means of creating image forgery, a more versatile way is however through computer graphics (CG) technique, where images of arbitrary scene composition and arbitrary viewpoints can be generated as long as the 3D model of the scene and objects are available. Photorealism (a visual fidelity close to that of real-world photographic images) has long been the Holy Grail of computer graphics research and leads to CG techniques such as the *physics-based rendering*, which simulates the physical light transport process, and *image-based rendering*, which synthesizes images of novel viewpoints from a set of images taken from other viewpoints. To feature the photorealism of the current CG technology, Alias, one of the major 3D CG company, challenges viewers to distinguish CG from photograph (<http://www.fakeorfoto.com>).

The main contributions of this work is that we develop an effective means for distinguishing PRCG from PIM through a PIM model based on NIS. Specifically, we study the NIS derived from the image power spectrum, the wavelet transform, and the local image patches. The power spectrum and the local patch NIS have not been used before for classifying PIM and PRCG. The experiment is conducted using a dataset of images, where examples of them are shown in Fig. 1.

In this work, we consider PRCG detection as an important problem of the passive-blind image authentication where an image is authenticated without using any prior information of the image. Being passive and blind, there is no need

for pre-extracting a digital signature from an image, nor pre-inserting a watermark into an image.

In the following section we define the characteristics of authentic images and explain why PRCG falls short of being authentic. In Sec. 3, we will describe the prior work for PIM and CG classification and then provide a short survey on NIS. In Sec. 4, we will detail on the NIS features being employed in this work. In Sec. 5, we will describe our experimental dataset, followed by the classification results in Sec. 6. In Sec. 7, we will discuss several interesting aspects of this work before coming to the conclusions.

## 2 Definitions

### 2.1 Image Authenticity

A good definition of image authenticity should be conducive for deciding whether an image is authentic. The definition may be different dependent on the availability and the reliability of certain prior information of an image. In the extreme case where the provenance information of an image (e.g., captured by what camera, by who and through what process an image is produced) is known, image authenticity should be evaluated based on the provenance information. When there is no prior information is available, image authenticity should be evaluated based on the intrinsic quality of authentic images.

We identify two intrinsic qualities of authentic images, which we call natural-imaging quality (NIQ) and natural-scene quality (NSQ). NIQ captures the characteristics of images due to the imaging acquisition process. For the case of a digital camera, the image acquisition process consists of low-pass filtering, lens-distortion, color filter array interpolation, white-balancing, quantization, and non-linear transformation [2]. A PRCG may be highly photorealistic but it lacks NIQ as it has not undergone a physical acquisition process. On the other hand, NSQ captured the image characteristics due to the subtle physical interaction of the illumination and objects in a real-world scene. NSQ includes the correct shadows, shading, surface foreshortening, and inter-reflection, as well as realistic object texture. A manipulated image such as photomontage may have a reduced NSQ as it may have a misplaced shadow. Although re-photographing could restore the NIQ of the manipulated image, it cannot undo the lack of NSQ.

A PRCG may not have a perfect NSQ, due to the various simplification in a CG rendering process. The elements of a high-quality PRCG are the soft shadows, complex lighting, global illumination, realistic reflectance model, and realistic geometric model. The computational complexity and the technical challenges make it difficult for a PRCG to have all the above-mentioned elements. The disparity of

NSQ between PIM and PRCG is the main theme of this work, where we characterize the NSQ of PIM using NIS.

## **2.2 Natural/Authentic/Photographic Images**

In the NIS literature [3], natural images are generally defined as the photographic images of scenes which human visual system is commonly exposed to (as opposed to satellite, or microscopic images). In this work, we consider PIM to be of the natural scene and hence PIM is equivalent to natural images. As a PIM satisfies NIQ and NSQ, it be an authentic image. As we do not consider photomontage in this work, the term "natural image", "authentic image" and "photographic image" are deemed interchangeable.

# **3 Prior Work**

## **3.1 Photographic Images vs. Computer Graphics (CG) Classification**

CG is generally defined as any imagery generated by a computer, which includes PRCG, 2D drawing, and cartoon. In [4], the problem of classifying the CG and the photographic video key frames is considered, for the purpose of improving the video key retrieval performance. In this case, the CG video key frames include also those of cartoon and 2D drawing. The authors identified the CG main characteristics as having few and simple color, patches of uniform color, strong black edges and containing text. Features inspired by these CG characteristics are used for the classification task and achieved CG detection rate of 82% and 94% respectively on the TREC-2002 video corpus and the Internet images.

Farid and Lyu [5] has briefly described an experiment on classifying PIM and PRCG using higher-order statistics wavelet features (originally employed for steganographic message detection) which achieved a detection rate of 98.7% and 35.4% respectively for PIM and PRCG<sup>1</sup>. The higher-order statistics (HOS) wavelet features are in fact a form of wavelet NIS.

## **3.2 Natural Image Statistics (NIS)**

The main goal of the NIS studies is to observe, discover and explain the statistical regularities in natural images [7]. The study of natural images through a statistical approach, instead of a deterministic mathematical model, gains ground due to the complexity of natural images. NIS, being a form of natural image model, has

---

<sup>1</sup>Our work is done before the publication of the further work by Lyu and Farid on PIM and PRCG classification [6]

found application in texture synthesis, image compression, image classification and image denoising.

In late 80's, Field [8] discovered the power law for the power spectrum of natural images,  $S(f_r)$ . The power law can be expressed as Equ. 1 and when taking the natural logarithm of Equ. 1 we obtain Equ. 2.

$$S(f_r) = \frac{A}{(f_r)^\alpha} \quad (1)$$

$$\log S(f_r) = \log A - \alpha \log f_r \quad (2)$$

where  $(f_r)^2 = (f_x)^2 + (f_y)^2$  is the radial spatial frequency of the 2D image power spectrum  $S_{2D}(f_x, f_y)$ . Power law implies the scale-invariant/fractal/self-similar property of natural images (which implies the non-existence of an absolute scale for natural images) because a power spectrum which is scale-invariant satisfies Equ. 3 and the only continuous solution to Equ. 3 is in the form of a power law function as in Equ. 1.

$$S\left(\frac{f_r}{\gamma}\right) = K(\gamma)S(f_r) \quad (3)$$

The exponent of the power law function for natural images,  $\alpha$ , is *empirically* found to be about the value of two. This empirical result implies that the power of natural images is constant over the octave bands.

The power law of natural images has been widely accepted for an ensemble of images. For a single image, the power law is also empirically found to be valid but with a larger deviation from the ideal power law function as in Equ. 1 [3], although some argues otherwise [9]. Besides that, the power law exponent  $\alpha$  was found to be different for different image types, such as images of a forest scene and that of the man-made objects. In a recent study [10], the power law of CG images is shown to be insensitive to image processing operations such as gamma correction and lossy compression, as well as to the particularity of rendering, such as with/without diffuse inter-reflection and hard/soft shadow. The insensitivity to the image processing operations is a good news, but being unable to discriminate the advanced photorealism effects is a bad news. At the same time, the authors [10] also found that the power law is closely related to the geometric aspect of an image (e.g., the distribution of edges). Although PRCG seems to follow the same power law, in this work we are interested to find out how the power spectrum of PIM and PRCG are different through a detailed model which will be described in Sec. 4.1.2.

Another major discovery about natural image statistics is non-gaussianity of natural images (i.e., there exists higher-order correlations of image pixels). A



Figure 2: An example of image style translation; input natural image (left) output van Gogh style image (right). Source: [15]

study [11] shows that there are interesting patterns in the kurtosis and the trispectrum (the fourth order moment spectra) of an ensemble of whitened images (i.e., second-order de-correlated), which contain only image phase information. Furthermore, **higher-order correlations** between wavelet coefficients are also found among adjacent scales, orientations and locations [12].

Some well-known works in NIS explore the joint probabilistic distributions of pixels in local image patches. In [13], the scale-invariant statistics of  $3 \times 3$  image patches were studied by categorizing image patches according to a set of prototypical patterns with different complexity. Whereas in [14], the empirical distribution of  $3 \times 3$  normalized high-contrast image patches was studied in a eight-dimensional Euclidean space and the probability mass was found concentrating around a two-dimensional manifold. It is interesting to note that the empirical distribution captures the differences between camera (optical) images and range images, where the differences is related to the image formation and sensor model. The reason for studying only the high-contrast image patches is because the interesting image features are richer in the high-contrast image regions. The choice of the local patch NIS in our study is inspired by the recent successes of the patch-based image model in various image processing task, including image style translation [15], image segmentation [16] and image scene synthesis. In particular, the work of image style translation [15] has demonstrated the effectiveness of the local image patch in capturing the style of an image category, as shown in Fig. 2. In our work, we can consider PIM and PRCG as two image categories with different styles.

## 4 NIS and CG Features

In this paper, we study the NIS from the natural image power spectrum (a second order statistics) and the high dimensional probability distribution of local image

patches. We then compare the performance of the features extracted from these NIS to those of the wavelet NIS [5] and the CG features [4]. Note that the power spectrum NIS, the wavelet NIS, and the local image patch NIS have a different statistical order. On the other hand, an image forgery detection technique should be robust to the common image-processing operations such as scaling, compression and so on, we therefore will also study the scale and the rotation invariant property of the NIS features.

## 4.1 Power Law of Natural Image Power Spectrum

In [3], a detailed modeling of power spectrum is applied to luminance channel image. We apply the modeling technique separately to each of the individual RGB color channel, for which the power law also holds [17]. To compute the power spectrum features on images of the same size, we first downsize all images such that the smaller dimension of an image becomes 350 pixels and then estimate the features using the central 350×350-pixel portion of the downsized images.

### 4.1.1 Estimation of Power Spectrum

To reduce frequency leakage, a single channel image  $I(x, y)$  of size  $N \times N$  pixels is windowed by a circular Kaiser-Bessel function  $w(x, y)$  and mean-subtracted before computing its Discrete Fourier Transform (DFT) as in Equ. 4.

$$F(f_x, f_y) = \sum_{(x,y)} \frac{I(x, y) - \mu}{\mu} w(x, y) \exp(2\pi i(x f_x + y f_y)) \quad (4)$$

where

$$i = \sqrt{-1}$$

$$w(x, y) = \frac{I_o\left(\pi\alpha\sqrt{1 - \frac{4}{N^2}(x^2 + y^2)}\right)}{I_o(\pi\alpha)}, \quad -\frac{N}{2} < x, y \leq \frac{N}{2}$$

$$\mu = \frac{I(x, y)w(x, y)}{w(x, y)}$$

and

$$\sum_{(x,y)} (w(x, y))^2 = 1$$

Then, the power spectrum of the image is given by Equ. 5

$$S(f_x, f_y) = \frac{|F(f_x, f_y)|^2}{N^2} \quad (5)$$

and the radial-frequency power spectrum is computed from Equ. 6.

$$S(f_r) = \sum_{\phi} S(f_r \cos(\phi), f_r \sin(\phi)) \quad (6)$$

#### 4.1.2 Modeling of Natural Image Power Spectrum

For a power spectrum that follows power law shown in Equ. 2, the plot of  $\log S(f_r)$  versus  $\log f_r$  would be a straight line with a slope of  $-\alpha$  and an intercept of  $\log(A)$ . To compute the slope and the intercept, we perform a least square error linear fit on the plot. To estimate the goodness of the linear fit, we compute the root-mean-square (RMS) error of the fit. Besides that, the oriented log-contrast which was shown to improve the power spectrum model [8] is also computed as in Equ. 7.

$$c_{(\phi_1, \phi_2)}^2 = \sum_{\phi_1 < \arctan \frac{f_y}{f_x} < \phi_2} \log(S(f_x, f_y)) \quad (7)$$

We compute the orientated log-contrast on the eight orientation pie-slice of  $45^\circ$ . The power spectrum feature is naturally scale-invariant because it follow the power law. However, the orientated log-contrast may not be rotation-invariant as the power spectrum of natural images is known to be anisotropic in the sense that dominant energy concentrates at the horizontal and vertical orientations.

#### 4.2 Higher-Order Correlation of Wavelet Coefficients

NIS motivated by the higher-order correlation of the cross-subband wavelet coefficients is used in [5] for classification of PIM and PRCG. The wavelet NIS consists of the mean, variance, skewness and kurtosis of the marginal wavelet coefficients in each subband and the mean, variance, skewness and kurtosis of the linear prediction error of the wavelet coefficients which captures the cross-subband correlations.

#### 4.3 Local Image Patch Distribution

The analysis of  $3 \times 3$  contrast-normalized image patch in [14] provides a mathematical framework for the high-dimensional probability mass distribution (PMF) of image patches. The paper reveals that the geometrical structure of the high-contrast image regions (e.g., the edge region) captures the difference between images with different generative process. This inspires us to expand the modeling of

local image geometrical structure by capturing the 1D geometrical structure around the high contrast region in the RGB color space, while reusing the same mathematical framework.

### 4.3.1 Analysis of $3 \times 3$ contrast-normalized patch distribution

The contrast of an image patch in a vector representation,  $\tilde{x} = [x_1, \dots, x_9]^T$ , is given by the D-norm in Equ. 8

$$\|\tilde{x}\|_D = \sqrt{\sum_{i \sim j} (x_i - x_j)^2} \quad (8)$$

where  $i \sim j$  represents the 4-connected neighborhood. Equ. 8 can be expressed in matrix form as in Equ. 9.

$$\|\tilde{x}\|_D = \sqrt{\tilde{x}^T D \tilde{x}} \quad (9)$$

where the  $D$  matrix is given by Equ. 10.

$$D = \begin{pmatrix} 2 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{pmatrix} \quad (10)$$

Before constructing the full distribution,  $3 \times 3$  image patches are mean-subtracted and contrast-normalized as in Equ. 11.

$$\tilde{y} = \frac{\tilde{x} - \frac{1}{9} \sum_{i=1}^9 x_i}{\left\| \tilde{x} - \frac{1}{9} \sum_{i=1}^9 x_i \right\|_D} \quad (11)$$

This step projects the image patches to a 7-dimensional ellipsoid embedded in a 9-dimensional Euclidean space,  $\tilde{S}^7 \subset \mathbb{R}^9$ , and the ellipsoid is represented as in Equ. 12.

$$\tilde{S}^7 = \left\{ \tilde{y} \in \mathbb{R}^9 : \sum_{i=1}^9 \tilde{y}_i = 0, \quad \tilde{y}^T D \tilde{y} = 1 \right\} \quad (12)$$

The particular form of matrix  $D$  can be diagonalized and whitened by the 2-dimensional Discrete Cosine Transform (DCT) basis. Ignoring the constant DCT basis, the DCT basis matrix can be written as  $A = [\tilde{e}_1, \dots, \tilde{e}_8]$ , where  $\tilde{e}_i, i = 1, \dots, 8$  are the DCT non-constant basis. Whitening of  $D$  matrix gives  $A^T D A = I$ , where  $I$  is the identity matrix.

Hence, the transformation of  $\tilde{y}$  to  $\tilde{v}$  by  $\tilde{y} = A\tilde{v}$  would project  $\tilde{y}$  to points onto a 7-dimensional sphere in a 8-dimensional Euclidean space,  $\tilde{S}^7 \subset \mathbb{R}^8$ , as given by Equ. 13.

$$\tilde{S}^7 = \left\{ \tilde{v} \in \mathbb{R}^8 : \sum_{i=1}^8 \tilde{v}_i = 0, \quad \tilde{v}^T \tilde{v} = 1 \right\} \quad (13)$$

The distribution of the data points on  $\tilde{S}^7$  can be approximated by the histogram binning technique. The histogram bins are the Voronoi cells with their centers form a dense set of sampling points on  $\tilde{S}^7$ . Such set of sampling points is given by the solution of the sphere-packing problem in  $\mathbb{R}^8$ , which gives us 17520 bins in total.

### 4.3.2 Modeling Local Image Geometrical Structure

While the image edge region is considered to be most informative of the difference of image generative process, the non-zero contrast patches in the non-edge region (e.g., the weak-edge patches) could be useful too. The information inherent in the geometrical structure variation in the luminance channel could be limited. This would mean that it is harder to tell PRCG from PIM perceptually for a grayscale image. Therefore, we extend the geometrical modeling to 1D geometrical structure which captures the geometric transitions in RGB color space around the edge pixels. As a result, we altogether obtain four types of sampling methods (each may have several patterns) (see Figure 3):

### 4.3.3 Model PMF

For each image, we extract 4000 patches (whenever possible) for each type of the sampling patterns. Hence, for each sampling pattern  $t, t = 1, \dots, T$ , we construct a model PMF respectively for PIM and PRCG. Let the category index be  $y \in \{0, 1\}$  with 0 and 1 representing PIM and PRCG respectively. Each image contributes  $T$  sets of patches,  $X_t$ , where  $t = 1, \dots, T$  and  $X_t = \{x_{ti} | i = 1, \dots, N_t\}$ . Let  $[x]$  denotes the bin index of a patch  $x$  with  $[x] \in \{b_1, \dots, b_{17520}\}$ . Given a training set,

$$\left\{ \left\{ (X_t^{(m)}, y_t^{(m)}); |X_t^{(m)}| = N_t^{(m)} \right\}_{t=1}^T \right\}_{m=1}^M$$

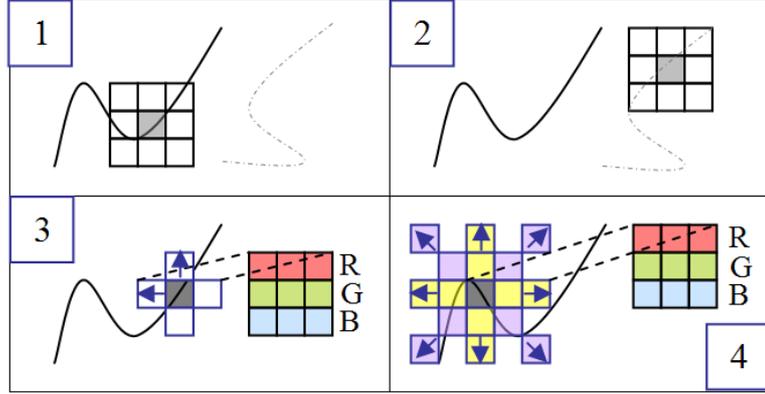


Figure 3: (1) 2D patch centered at edge points (in luminance channel) (2) 2D non-zero-contrast patch centered at non-edge points (in luminance channel) (3) 1D patch centered at edge points (use RGB channels as features/in vertical and horizontal directions) (4) 1D patch along different gradient directions from edge points (use RGB channels as features/sampled in eight directions)

with  $M$  training images (each has  $N_t^{(m)}$  patches for each sampling pattern and is assigned a corresponding label  $y_t^{(m)}$ ), the model PMF is given by Equ. 14:

$$P_{yt}^{\text{model}}(B = b_j) = \frac{\sum_{m=1}^M \sum_{i=1}^{N_t^{(m)}} \mathbf{1}([x_{ti}^{(m)}] = b_j, y_t^{(m)} = y) + 1}{\sum_{m=1}^M N_t^{(m)} \mathbf{1}(y_t^{(m)} = y) + 17520}, \quad t = 1, \dots, T \quad (14)$$

where  $\mathbf{1}(\cdot)$  is the indicator function. Note that Laplacian smoothing (add one for each bin) which is well-known in text document classification is applied in Equ. 14 to smooth out the empirical estimate of the model PMF.

Given a new image, we sample patches of different sampling patterns from the image. For each sampling pattern, we form an empirical PMF as in Equ. 15.

$$P_t^{\text{Emp}}(B = b_j) = \frac{\sum_{i=1}^{N_t^{(m)}} \mathbf{1}([x_{ti}^{(m)}] = b_j) + \frac{1}{800}}{N_t(1 + \frac{1}{800})}, \quad t = 1, \dots, T \quad (15)$$

Note that the empirical PMF are smoothed with a factor relatively consistent with the amount of smoothing in the model PMF. Then, for each sampling pattern, we compute the Kullback-Leibler (KL) distance between the empirical PMF and the model PMF, as in Equ. 16. The KL distances will be used for image classification.

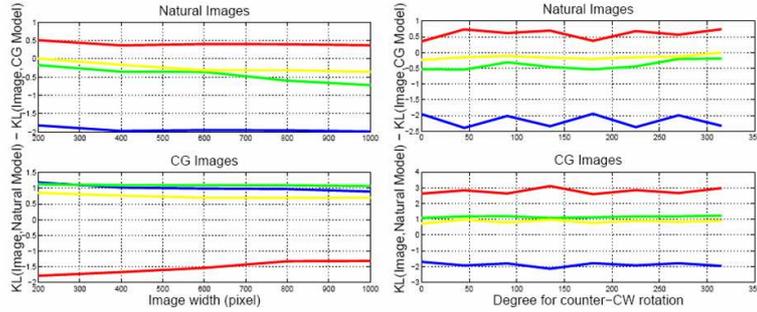


Figure 4: The variation of the average KL distance difference when some example images from the natural and CG category are scaled to different sizes (left) and rotated with different angles (right)

$$KL\left(P_t^{\text{Emp}}\|P_{yt}^{\text{model}}\right) = \sum_{j=1}^{17520} P_t^{\text{Emp}}(B = b_j) \log\left(\frac{P_t^{\text{Emp}}(B = b_j)}{P_{yt}^{\text{model}}(B = b_j)}\right) \quad (16)$$

The patch NIS feature is found to be approximate scale and rotation invariant as Fig. 4 shows how

$$\frac{1}{T} \sum_{t=1}^T \left( KL\left(P_t^{\text{Emp}}\|P_{0t}^{\text{model}}\right) - KL\left(P_t^{\text{Emp}}\|P_{1t}^{\text{model}}\right) \right)$$

(the averaged KL distance difference) of some example images vary as they are rotated to different angles and scaled to different sizes (each line in Fig. 4 is corresponding to an image).

#### 4.4 CG features

In [4], features motivated by CG characteristics is proposed for classifying the photographic and the CG video key frames. The main characteristics of CG video key frames are identified as having few and simple colors, patches of uniform color, strong black edges and containing text. By modeling the CG characteristics, features such as average color saturation, ratio of image pixels with brightness greater than 0.4, HSV color space histogram, edge orientation and strength histogram, compression ratio and pattern spectrum (i.e., the distribution of object size) were used.

## 5 Dataset

We initiated a dataset collection project for producing a dataset tailored for the passive-blind image forgery detection research. At the first stage, we collected a set of PIM and PRCG, to be used in the experiments. Examples of the images are shown in Fig. 1.

### 5.1 Authentic and Natural Images

The PIM category consists of 800 images which are authentic (directly from a camera and are not photomontage) and of scenes commonly encountered by human. Two main characteristics of this PIM set are its diversity from the point of view of image generative process and its readiness to facilitate the studies of image processing effects on a forgery detection technique. Images are generated as light rays from the illumination source are reflected off the scene objects before being captured by a camera, which is operated by a photographer. The PIM category has diversity in light sources (indoor bright/dim, outdoor daylight/dusk/night/rain), object types (natural/manmade/artificial), camera model (Canon 10D and Nikon D70, which are known to use different makes of the camera main chip) and photographers (three persons). Besides that, we recorded the images simultaneously in the high-quality JPEG format and RAW format. The RAW format images are the direct output from the imaging sensor (hence not lossy compressed and free from any image operation), therefore we can study the effect of image operations on image forgery detection algorithm using these RAW format images. The original size of the images is about  $3000 \times 2000$  pixels. In order to match the size of the PRCG images, which have an average dimension of about 630 pixels, we resize the PIM with bicubic resampling to the size of about  $730 \times 500$  pixels.

### 5.2 Photorealistic CG Images

The PRCG set consists of 800 PRCG collected from a list of reputable and trustable 3D graphics company websites and the professional 3D artist websites. Of the many subcategories of PRCG, we only selected those which are of good photorealism and with scenes that are commonly encountered by human. The subcategories in the PRCG set are ‘architecture’, ‘people and animals’, ‘objects’, ‘scenery’ and ‘games’. Subcategories such as ‘fantasy’ and ‘abstract’ are intentionally excluded.

Table 1: Feature List

Features	Dimension
Power spectrum model (PS Features)	33
Local Patch Features	24
Wavelets higher-order statistics (HOS)	72
CG Features	108

## 6 PIM vs. PRCG Classification

### 6.1 Summary of Image Features

We evaluate three types of NIS features discussed in Sec. 4 including the second-order power spectrum model features, the wavelets higher-order statistics features [5] and the local patch features. We compare these NIS features (modeling natural images) with features that model computer graphics characteristics [4]. The dimension of the features is shown in Table 1.

### 6.2 Support Vector Machine (SVM) Classification

We use SVM (from the LIBSVM implementation [18]) with radial basis function (RBF) kernel as our classifier. The best classifier parameters (soft-margin parameter,  $C$ , and the RBF kernel parameter,  $\gamma$ ) are selected using a grid search strategy, through a five-fold cross-validation on the training set. Although the features being compared are of different dimensionality, there is less concern of classifier overfitting as SVM is based on the principle of structural risk minimization (i.e., not minimizing the empirical risk but the regularized risk which bounds the true risk from above). Furthermore, the classification receiver operating characteristic (ROC) curve is estimated through five-fold cross-validation to avoid overfitting of the classifier.

### 6.3 Classification Results

Fig. 5 shows the five-fold cross-validation ROC (positive being PIM) for the features listed in Table 1, as well as certain fusions of them. Fig. 6 shows the example classification results of the local patch classifier and the CG classifier. For classifier fusion, we choose to fuse the decision value of the SVM classifier output instead of simply concatenating the input feature vectors because of the large difference between the patch NIS feature and CG feature in input space dimensionality. Such difference would result in a bias of the fused decision toward the feature with a higher dimensionality. As SVM output was shown to fit well with a distribution

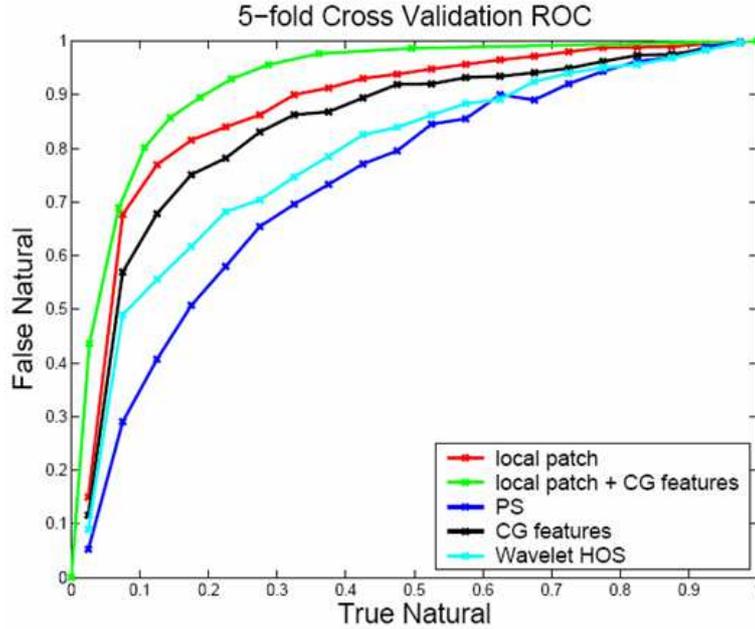


Figure 5: Classification accuracy (in ROC) for distinguishing CG from natural images

from the exponential family, hence logistic regression fusion would be an ideal option for fusing SVM decisions, assuming independent decisions. For logistic regression fusion, the posterior probability of the class label 1 is given by Equ. 17 where  $f_{svm}$  being a vector of decision value from SVM classifiers. The linear coefficients  $(a, b)$  are learnt by maximizing the class label likelihood.

$$p(y = 1|f_{svm}) = \frac{1}{1 + \exp(a^T f_{svm} + b)} \quad (17)$$

Below are the observations from the experiment:

1. The second-order NIS, i.e., PS performs worst in the classification. The wavelet HOS which can be considered a higher-order NIS is doing better, while the image patch features, which is derived from the full-distribution of local patches, performs the best. Hence, there is a trend that the higher the statistical order of the NIS, the better it captures the unique characteristics of PIM.
2. The CG-motivated features are performed surprisingly well despite the fact

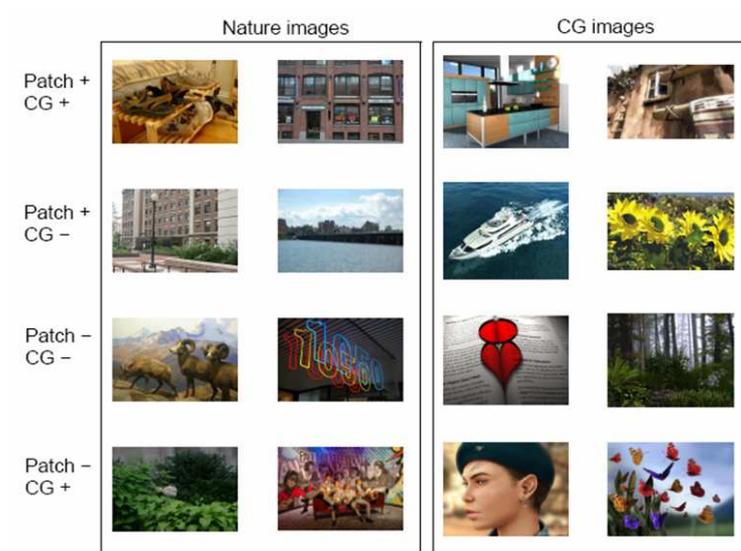


Figure 6: Example of the classification results; Row 1: images correctly classified by both the patch feature and CG feature. Row 2: patch feature correct, CG feature wrong; Row 3: both features wrong; Row 4: patch feature wrong, CG feature correct.

that the CG in our dataset are photorealistic. Of all the CG features, the contribution from the color histogram is the most significant. This observation indicates that the color of the PRCG is still quite different from that of PIM, despite not being visually obvious.

3. Since the patch features and CG features are separately modeling the characteristics of natural images and CG, the classification performance improves when combining the two features.

## 7 Discussion

### 7.1 Possible Extensions for Local Patch Distribution Modeling

Currently, we are using image patch of size  $3 \times 3$ . We believe that more interesting structure can be captured if we increase the patch size. To do this we need to overcome the difficulties of analyzing the full distribution with a quadratic dimensionality increase. There could be two potential ways to overcome such difficulty:

1. We can capture the structures inherent in a larger patch size by having a scale-space representation of image patch in a manageable dimensionality [19].
2. We may be able to learn a probabilistic generative model for a large set of images using the epitome learning framework [16]. In this case, the map from an individual image to the epitome synthesizes the image, while the reverse map from the epitome to the large set of images provide the full probability distribution of the patches. However, the learning procedure may be computationally demanding.

### 7.2 Adversarial Attack on Passive-Blind Techniques

It is natural to assume that there will be adversarial attacks on any image forgery detection techniques. Therefore, robustness against adversarial attack is critical. This work does not study such aspect comprehensively. If attackers have access to the detector or have full knowledge about the detection algorithm, they can repeatedly test and refine an image forgery (within the constraints of photorealism), until the image forgery escapes detection. Such attack is known as oracle attack in the digital watermarking literature and has been a serious threat to the public watermarking system. The proposed countermeasures to such attack in watermarking system are also applicable to our case. These techniques include converting the parametric decision boundary into a fractal one [20] and modifying the detector

temporal behavior such that time taken for returning a decision will be lengthened when the sequence of detections carries the mark of an oracle attack [21].

## 8 Conclusions

In this paper, we showed a way to distinguish PRCG from PIM by modeling PIM using NIS. Specifically, we propose novel features derived from local patch distributions and the power spectrum of images. The NIS features complement the features inspired by the CG characteristics. The patch-based NIS which recently found successes in various image processing application performs well in the classification task, when the image geometrical structure is sufficiently captured. Furthermore, the performances of the NIS features are in line with the corresponding statistical order.

## 9 Acknowledgements

This project is supported in part by NSF CyberTrust program (IIS-04-30258) and the first author is supported by Singapore A\*STAR Scholarship. The authors would like to thank Lexing Xie for the helpful discussion.

## References

- [1] C. Amsberry, "Alterations of photos raise host of legal, ethical issues," *The Wall Street Journal*, Jan 1989.
- [2] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of CCD imaging process," in *IEEE International Conference on Computer Vision*, July 2001.
- [3] A. v. d. Schaaf, "Natural image statistics and visual processing," PhD thesis, Rijksuniversiteit Groningen University, 1998.
- [4] T. Ianeva, A. de Vries, and H. Rohrig, "Detecting cartoons: A case study in automatic video-genre classification," in *IEEE International Conference on Multimedia and Expo*, vol. 1, 2003, pp. 449–452.
- [5] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Workshop on Statistical Analysis in Computer Vision*, Madison, Wisconsin, June 22 2003.
- [6] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 845–850, February 2005.

- [7] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, “On advances in statistical modeling of natural images,” *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.
- [8] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [9] M. S. Langer, “Large-scale failures of f-a: a scaling in natural image spectra,” *Journal of the Optical Society of America A*, vol. 17, pp. 28–33, 2000.
- [10] E. Reinhard, P. Shirley, M. Ashikhmin, and T. Troscianko, “Second order image statistics in computer graphics,” in *ACM Symposium on Applied perception in graphics and visualization*, Los Angeles, California, 2004, pp. 99–106.
- [11] M. G. A. Thomson, “Higher-order structure in natural scenes,” *Journal of the Optical Society of America A*, vol. 16, no. 7, pp. 1549–1553, 1999.
- [12] E. P. Simoncelli, “Modelling the joint statistics of images in the wavelet domain,” in *SPIE 44th Annual Meeting*, Denver, CO, 1999.
- [13] D. Geman and A. Koloydenko, “Invariant statistics and coding of natural microimages,” in *IEEE Workshop on Statistical and Computational Theories of Vision*, Fort Collins, CO, 1999.
- [14] A. B. Lee, K. S. Pedersen, and D. Mumford, “The nonlinear statistics of high-contrast patches in natural images,” *International Journal of Computer Vision*, vol. 54, no. 1, pp. 83–103, 2003.
- [15] R. Rosales, K. Achan, and B. Frey, “Unsupervised image translation,” in *IEEE International Conference on Computer Vision*, 2003, pp. 472–478.
- [16] N. Jojic, B. J. Frey, and A. Kannan, “Epitomic analysis of appearance and shape,” in *IEEE International Conference on Computer Vision*, Nice, France, 2003.
- [17] C. A. Parraga, G. Brelstaff, T. Troscianko, and I. R. Moorehead, “Color and luminance information in natural scenes,” *Journal of the Optical Society of America A*, vol. 15, no. 3, pp. 563–569, 1998.
- [18] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A practical guide to support vector classification,” July 2003.

- [19] K. S. Pedersen and A. B. Lee, “Toward a full probability model of edges in natural images,” in *European Conference on Computer Vision*, Copenhagen, Denmark, 2002.
- [20] A. Tewfik and M. Mansour, “Secure watermark detection with non-parametric decision boundaries,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, pp. 2089–2092.
- [21] I. Venturini, “Counteracting oracle attacks,” in *ACM multimedia and security workshop on Multimedia and security*, Magdeburg, Germany, 2004, pp. 187–192.