# ON DISCRIMINATION BETWEEN PHOTOREALISTIC AND PHOTOGRAPHIC IMAGES

Ying Wang and Pierre Moulin

Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign Urbana, IL 61801 Email: {ywang, moulin}@ifp.uiuc.edu

### ABSTRACT

This paper presents a classifier built for differentiating digital photorealistic images from digital photographs. Results show that our 144-dimensional (144-D) feature vector extracted from characteristic functions of wavelet histograms is more efficient than Lyu and Farid's 216-D feature vector [1], and Ng *et al.*'s 192-D feature vector [2]. Our classifier outperforms Lyu and Farid's state-of-art method while only requiring half of their feature extraction and testing time.

## 1. INTRODUCTION





Advances in computer graphics, such as recursive ray tracing [3] and subsurface scattering [4], have made it easier to generate photorealistic images, destined for entertainment and advertisement applications. Indeed, many computer-generated photorealistic images pass the visual test very satisfactorily. Fig. 1 shows two example images: even a careful visual examination does not reveal which one is a photograph and which one is a photorealistic image.<sup>1</sup> However, perceptual indistinguishability between photographs and photorealistic images does not imply complete indistinguishability. Even the most sophisticated rendering algorithms are based on rough models of the real world where photographs are produced through complex interactions among objects, photons, and imaging devices, let alone simplistic ones that only aim at perceptual realism. The deviation of the computer simulating process from the real-world physics results in statistical differences between photographs and photorealistic images. If we find appropriate image features that capture the statistical differences, we can design a two-class classifier, using standard methodology from computer vision and pattern recognition.

Differentiating between photographic and photorealistic images serves two purposes. On the one hand, it helps computer graphics researchers identify potential areas of improvement in their algorithms to mimic the true physical world. On the other hand, the level of photorealism in computer graphics is so convincing that we need to reestablish the trustworthiness and integrity of photographs in applications such as criminal investigation and journalistic reporting. Although currently people are mainly concerned with the integrity of digital photography [5], and reality-forging incidents happen mostly with manipulation of digital photographs [6], it is reasonable to believe that in the near future, the same incidents can happen with photorealistic images.

Two groups of researchers have focused on the differentiation between photographic and photorealistic images. Lyu and Farid [1] use an image statistical model built on a waveletlike decomposition, and take the first four order statistics (mean, variance, skewness, and kurtosis) of the subband coefficients and the inter-subband prediction errors as features with a total number of 216. Ng *et al.* [2] take a geometry-based approach to model physical differences in the generation of photographic and photorealistic images, where 192 features are extracted from analysis on local patch statistics, local fractal dimension, surface gradient, quadratic geometry, and Beltrami flow.

In this paper, we use a wavelet-based statistical model to extract features from the characteristic functions of wavelet-coefficient histograms. Our 144-D feature set takes *half* of the extraction time of Lyu-Farid's 216-D feature set, and achieves better classification results. Since we use global image statistics, we do not pinpoint specific geometrical differences in photorealistic images as Ng *et al.* may be capable of doing. However, results on the same image datasets show that our method can achieve a detection rate of 100% at a false alarm rate of 0.1% with a feature extraction and testing time about

This work was supported by NSF under grants CCR 00-81268 and CCR 02-08809.

<sup>&</sup>lt;sup>1</sup>In Fig. 1, (a) is a photorealistic image and (b) is a real photograph.

1/33 of that reported by Ng *et al.* in [2].

## 2. FEATURE EXTRACTION

The statistical features we propose for classification between photographic and photorealistic images are extracted from the characteristic functions of wavelet-coefficient histograms. For a RGB color image, we apply a three-level discrete wavelet transform to each color channel. At each level  $l \in \{1, 2, 3\}$ of color channel  $c \in \{r, g, b\}$ , there are smooth, horizontal, vertical, and diagonal subbands, denoted as  $S_{l}^{c}$ ,  $H_{l}^{c}$ ,  $V_{l}^{c}$ , and  $D_1^c$ . Hence, there are 36 (= 3 × 3 × 4) resulting subbands. To further exploit the statistics of the coefficients in subband  $D_1^c$ , we decompose  $D_1^c$  into four second-level subbands too, denoted as  $S_{2'}^c$ ,  $H_{2'}^c$ ,  $V_{2'}^c$ , and  $D_{2'}^c$ . Hence, we have a total of 48 subbands. For each subband, we partition the range of coefficients x into bins, i.e., disjoint intervals of equal lengths, calculate the number of data samples in each bin, and obtain the normalized histogram  $\{h(m)\}$ , where  $m \in \{0, 1, \dots, M-1\}$ denotes the  $m^{th}$  bin.

If we use the histograms in 48 subbands as features, they capture global image statistics fully but at the cost of large memory requirements and computational complexity. Indeed, there are usually hundreds of bins in each histogram. So, we need to extract *low-dimensional* features that are informative for classification.



(a) not-so-photorealistic(b) photorealistic-yet-noisyFig. 2. Examples of questionable photorealistic images.

To find such features, we should consider certain attributes of computer-generated images. A typical artifact of images generated with very basic algorithms is that they have simple scenes where only a small number of objects exist without much depth. Besides, those objects often have a plastic/rubber feel because of a lack of details such as dust and small defects, subtle color variation, and radiosity caused by indirect lighting or specularity caused by light sources. Fig. 2 (a) shows one such not-so-photorealistic example. One simple way to address above problems is adding noise to simulate dust or texture on object surfaces [7]. Ray-tracing algorithms are physically-accurate better solutions. However, they are often operated at low-sampling rates due to computational limits, which result in noisy artifacts [8]. Fig. 2 (b) shows one such photorealistic-yet-noisy example. While adequate noise does increase image's visual naturalness, the resulting image possesses unnatural statistical characteristics.

Fig. 3 shows the normalized histograms of the first-level diagonal subband coefficients of Fig. 1(b) (photographic), Fig.

2(a) (not-so-photorealistic), and Fig. 2(b) (photo-realistic-yetnoisy). The wavelet coefficients of photographs is well modelled by a generalized Gaussian distribution (GGD)

$$p_{\alpha,\beta}(x) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \exp\left\{-\left(\frac{|x|}{\alpha}\right)^{\beta}\right\},\qquad(1)$$

where  $\Gamma(\cdot)$  is the Gamma function,  $\alpha > 0$  is the scale parameter, and  $\beta > 0$  is the shape parameter. Most photographs have distributions with  $\beta \leq 1$ , i.e., Laplacian-like.  $\alpha = 0.7$ and  $\beta = 0.4$  are estimated for the normalized histogram of Fig. 1(b). Not-so-photorealistic images have much more zerovalued wavelet coefficients and are best modelled as a mixture of Dirac delta function  $\delta(x)$  and a GGD. In contrast, photorealistic-yet-noisy images have different characteristics: the distribution is smoother in the vicinity of x = 0 and is well modelled by a Cauchy distribution with scale parameter b

$$p_b(x) = \frac{b}{\pi [x^2 + b^2]}.$$
 (2)

b = 1.8 is estimated for the normalized histogram of Fig. 2(b).



**Fig. 3**. Normalized histograms of the first-level diagonal subbands of three intensity images. Their color versions are shown in Fig. 1(b), Fig. 2(a), and Fig. 2(b), respectively. The Haar wavelet transform is used.

The classification problem studied in this paper bears similarities to steganalysis. A basic steganalysis problem is to classify between real photographs and steganographic images generated by imposing information-bearing noise to original photographs. Similar to photorealistic-yet-noisy images, the distribution of the wavelet coefficients of steganographic images is smooth around x = 0 and Cauchy-like. In [9], we proposed a set of features and demonstrated their efficiency in the steganalysis problem. It turns out that the same set of features work well in differentiating photorealistic images and photographs. The significant difference in this paper is that we need to distinguish not-so-photorealistic images from photographs as well, i.e., histograms with extremely strong peak from those with moderate peak.

The set of features we used in [9] are calculated from the characteristic functions of normalized histograms. A *K*-point

discrete Fourier transform (DFT) is performed on the normalized histogram  $\{h(m)\}$  to obtain its characteristic function

$$H(k) = \sum_{m=0}^{M-1} h(m) \exp\{-i2\pi mk/K\}, \ 0 \le k \le K.$$
 (3)

Fig. 4 shows the corresponding DFT amplitude plots for the three normalized histograms in Fig. 3. Not-so-photorealistic images have both significant high-frequency and low-frequency components because of the exceedingly strong peak at value 0; photographic images also have significant yet weaker high-frequency components due to the moderate peak in Laplacian-like distributions; photorealistic-yet-noisy images have fast decreasing high-frequency components while packing most of their energy in low-frequency components because of the smoother curvature at value 0 in their Cauchy-like distributions. Hence, we calculate three features from each histogram: two that convey information about the high-frequency components and one that convey information about the low- to mid-frequency components. The three features are defined as

$$f_1 = \sum_{k=0}^{K/2} |H(k)| \sin(\frac{\pi k}{K}), \tag{4}$$

$$f_2 = \sum_{k=0}^{K/2} |H(k)| \sin^2(\frac{\pi k}{K}), \tag{5}$$

and

$$f_3 = \sum_{k=0}^{K/4} |H(k)| \sin(\frac{4\pi k}{K}).$$
(6)

Clearly,  $f_1$  and  $f_2$  are obtained by high-pass filtering the normalized histogram and  $f_3$  by band-pass filtering.  $f_1$  and  $f_2$ are related to upper bounds on the first and second discrete derivatives of the normalized histogram, respectively. A variant of  $f_1$  and  $f_2$  was proposed by Xuan *et al.* in steganalysis [10].  $f_3$  is crucial to improve classification performance and—to the best of our knowledge—a new feature that no one used before. From Fig. 4, we expect that for not-sophotorealistic images, they have the largest  $f_1$ ,  $f_2$  and  $f_3$ ; for photorealistic-yet-noisy images, they have the smallest  $f_1$  and  $f_2$  and the second largest  $f_3$ ; while for photographic images, they have the second largest  $f_1$  and  $f_2$  and the smallest  $f_3$ . Hence, the three features distinctively reflect the class that images belong to. There are a total of 144 features available (3 for each 48 subbands).

## 3. CLASSIFICATION RESULTS

We use a simple Fisher linear discriminant (FLD) classifier for the automatic classification. We omit the details of the FLD analysis, which can be found in [11]. Although we may be able to achieve better classification results by using more complex classifiers such as support vector machines (SVM),



**Fig. 4**. Corresponding DFT amplitude plots of histograms in Fig. 3 and filter amplitude plots used by Equations (4)—(6).

our results show that our feature selection is very effective even with the FLD classifier. We normalize the features so that they have comparable dynamic ranges.

We built our photographic (PG) dataset by downloading 4546 photographic images from www.bigfoto.com and www.freefoto.com, and our photorealistic (PR) dataset by downloading 3844 photorealistic images from www.raph.com and www.irtc.org. All photographs and photorealistic images are RGB color images, JPEG compressed with quality above 80%, and  $600 \times 800$  pixels in typical size. All the image datasets consist of a wide range of outdoor and indoor scenes, including nature (e.g., trees, animals), portraits, man-made objects (e.g., signs, architectures, cars), etc. We also downloaded the photographic (PIM) and photorealistic (PCCG) datasets collected and tested on by Ng *et al.* in [2].

We use 2273 photographic images from the PG dataset and 1922 photorealistic images from the PR dataset to train our classifier, then apply the classifier to the remaining 2273 photographic images in the PG dataset, the remaining 1922 photorealistic images in the PR dataset, the 800 photographic images in the PIM dataset, and the 606 photorealistic images in the PCCG dataset.<sup>2</sup> We randomly split the PG and PR datasets to select training/testing images so that we avoid flukes for specific splits. The reported results are averaged over 1000 random splits. Although our classifier is trained only on the PG and PR datasets, it has the same good results on the PIM and PCCG datasets too, which demonstrates our classifier's strong generality.

Fig. 5 compares the test false alarm (FA) probabilities (percentage of photographic images that are misclassified as photorealistic) based on our 144-D feature vector and on Lyu-Farid's 216-D feature vector. The classifier with our feature selection has better test false alarm probability performance, especially at low operating FA rate below 1%. It is remarkable that our feature selection has 100% detection rate (ability to correctly classify photorealistic images, Fig. 6) on both PR

 $<sup>^{2}</sup>$ Due to some broken world wide web links, we could only find 606 out of 800 images in the PCCG dataset.

and PCCG datasets, for every single one of the 1000 random training/testing splits. Lyu-Farid's feature selection has 100% detection rate on the PCCG dataset but consistently errs on some photorealistic images in the PR dataset. The trouble-some images include both complex and simple scenes. Fig. 7 shows two examples misclassified by Lyu-Farid's feature selection. Moreover, the extraction and testing time for our classifier is 2.8 seconds for a  $384 \times 384$  image while that for Lyu-Farid's is 5.6 seconds when implemented in Matlab 6.1 running on the Windows XP Professional platform and a Pentium 4 1.8GHz CPU. Both wavelet-based feature selections take much less time compared to ~ 92 seconds<sup>3</sup> taken by Ng *et al.*'s geometry-based feature selection [2].



**Fig. 5**. Test false alarm probabilities on the PG and PIM dataset by our 144-D feature vector and Lyu-Farid's 216-D feature vector vs. training false alarm probability. Results are averaged over 1000 training/testing splits.



**Fig. 6.** Test detection probabilities on the PR and PCCG dataset by our 144-D feature vector and Lyu-Farid's 216-D feature vector vs. training false alarm probability. Results are averaged over 1000 training/testing splits.

### 4. REFERENCES

[1] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Trans. on Signal Processing*, vol. 53, no. 2,



**Fig. 7**. Two examples from the PR dataset that Lyu-Farid's 216-D feature set misclassifies.

pp. 845-850, Feb. 2005.

- [2] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M. P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," *Proc. of ACM Multimedia*, pp. 239–248, Singapore, Nov. 2005.
- [3] T. Whitted, "An improved illumination model for shaded display," *Communications of the ACM*, vol. 23, no. 6, pp. 343–349, 1980.
- [4] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan, "A practical model for subsurface light transport," *Proc. of the ACM Special Interest Group on Computer Graphics and Interactive Techniques*, pp. 511– 518, ACM Press Addison-Wesley, New York, NY, 2001.
- [5] National Press Photographers Association, "Ethics in the age of digital photography," www.nppa.org/ professional\_development/self-training\_resources/eadp \_report/index.html.
- [6] S. L. Leach, "Seeing is no longer believing," www.csmonitor.com/2005/0202/p15s02-lire.html.
- [7] B. Fleming, 3D Photorealism Toolkit. Wiley, New York, 1998.
- [8] I. Wald, T. Kollig, C. Benthin, A. Keller, and P. Slusallek, "Interactive global illumination using fast ray tracing," *Proc. of Thirteenth Eurographics Workshop on Rendering*, P. Debevec and S. Gibson, Eds., Pisa, Italy, June 2002.
- [9] Y. Wang and P. Moulin, "Statistical modelling and steganalysis of DFT-based image steganography," *Proc. of SPIE Electronic Imaging*, San Jose, CA, Jan. 2006.
- [10] G. Xuan, Y. Q. Shi, J. Gao, D. Zou, C. Yang, Z. Zhang, P. Chai, C. Chen, and W. Chen, "Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions," *Proc. of Information Hiding Workshop (IHW05)*, Barcelona, Spain, June 2005.
- [11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classi-fication, 2nd ed.* Wiley, New York, NY, 2001.

<sup>&</sup>lt;sup>3</sup>This number is calculated indirectly according to the results reported in [2], where when it is 9.3 seconds for Lyu-Farid's feature extraction, it is 151.5 seconds for Ng *et al.*'s feature extraction.