Lessons Learned from Online Classification of Photo-Realistic Computer Graphics and Photographs

Conference Paper · May 2007

Source: IEEE Xplore

3 authors, including:

S. Chang
National Tsing Hua University

Mao-Pei Tsui
University of Toledo

Lessons Learned from Online Classification of Photo-Realistic Computer Graphics and Photographs

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

Mao-Pei Tsui
Department of Mathematics
University of Toledo
Toledo, OH 43606
Email: mao-pei.tsui@utoledo.edu

***Abstract*— We presented a set of physics motivated features for classifying photographic and computer graphic images in our previous work [1]. We also deployed an online demo system for distinguishing photographic and computer graphic images in October 2005 [2], which features our geometry classifier, together with the wavelet classifier, and the cartoon classifier. On the first anniversary of its launch, we have received 1582 submitted images, through which we perform an analysis on the user behavior, the image set characteristics, and the classifier performance. We observe that online users do not provide clear judgments for about 80% of the submitted images, confirming the challenge in distinguishing photo-realistic computer graphics images from natural photographs. We also found the accuracy of our classifiers over the online submission set is consistent with that computed over an offline data set. Finally, in order to improve the online computational speed of our classifier, we perform feature selection and reduction, cutting the response time from 152 seconds to 24 seconds per image, while keeping the accuracy almost unchanged.**

I. Introduction

As model-based computer graphic rendering technology is making great strides, distinguishing photographic image from photorealistic computer graphics is getting increasingly challenging. To convince yourselves of this trend, readers can now take part in the quizzes [1] hosted by the 3D graphic company Autodesk Alias (`www.autodesk.com/alias`). The quizzes are designed for challenging users' visual judgement on photorealistic images as to be photographs or computer graphics. Furthermore, there are also experimental evidences in [3] that, to human, computer graphic images of certain scenes are visually indistinguishable from photographic images.

With high photorealism, computer graphic images naturally qualifies themselves as potential suspects for forgery images. Forgery images can be used for fraud, make-belief, and dishonest setup, while the goal of image forensics is to detect these forgery images. Under the setting of image forensics, we have proposed a set of physics-motivated features for distinguishing photographic images and photorealistic computer graphic images [1]. The physics-motivated features are obtained by studying the respective physical image generative process for photographic images and photorealistic computer graphics. The features consist of three subsets, i.e., the differential geometry features, the local patch statistics features, and the local fractal dimension features, with a total of 192 dimensions. Furthermore, since October 2005 we deploy an online demo system (accessible from `www.ee.columbia.edu/trustfoto`) for distinguishing photographic images and the general computer graphic images (i.e., including non-photorealistic computer graphic images) [2]. The online demo system features three types of classifiers: the geometry classifier, the wavelet classifier (based on the method in [4]), and the cartoon features (based on the method in [5]. The geometry classifier is based on the physics-motivated features, together with a few strategies, such as image size reduction and classifier fusion, for the purpose of efficient computation while maintaining an effective classification rate.

At the first anniversary of the launch of our online demo system, it is timely for us to review and analyze the data collected by the system. At this point of time, the system has received 1528 image submissions, which allows for a statistically meaningful analysis. We will analyze on the shortcoming of the system design, the system user behaviors, the characteristics of the images, and the performance of the online classifiers. A few important observations are that the users are not naturally keen on providing meta-information of the images, some of the submitted image are interestingly ambiguous, most of the submitted images are photographic and from the Internet, and the conclusions on the online classifier performance comparison matches those reported in our prior work [1], [2]. The lack of subjective judgments from online users also partly confirms the challenge of the task of distinguishing photo-realistic computer graphics from natural photographs.

Of the three online classifiers, our geometry classifier is the most computational expensive one. In order to improve its computation efficiency, we provide a further analysis on the physics-motivated feature performance, where we evaluate the performance of its five sub-groups of features and their combinations. The rationale is that we may drop the least contributing feature sub-group for computation cost saving. These five sub-groups of features respectively correspond to five different physical motivations and they are the surface

[1]The quizzes can be accessed from `http://www.autodesk.com/eng/etc/fake_or_foto/index.html` and `http://www.autodesk.com/eng/etc/fake_or_foto/v1/index.html`

gradient features, the second fundamental form features, the beltrami flow features, the local patch statistics features, the local fractal dimension features. Based on our analysis, by removing the least-contributing feature sub-group (local fractal dimension distribution), we are able to speed up the classifier from 152 second/image to 24 second/image, while keeping the accuracy at 83.3% (compared to the original accuracy of 83.5%).

In Sec. II, we will review the physics-motivated features described in [1]. Then, in Sec. III, we will describe the online classification system for photographs and computer graphics. In Sec. IV, we will analyze the images submitted to the online system, and the performance of the online classifiers. In Sec. V, we will consider the feature selection and reduction issue for the geometry classifier, and finally conclude with sec. VI.

II. Physics-motivated Features

In [1], we analyzed the physical differences between the image generation pipeline of photographic images and photorealistic computer computer. We thereby identified three types of differences in terms of the object surface reflectance property, the object surface geometry, and the image acquisition process. The reflectance property of real-world objects is often complex as evident in the subtle sub-surface scattering of human skin and the fluorescence phenomena where the incident light of one frequency range is absorbed by surface for the emission of light with a lower frequency range [6]. However, the computer graphics rendering often makes simplifying assumptions to decouple the interdependency between the different color channels of an image for efficient computation. In addition, the surface geometry of real-world objects is often complex, being a result of the underlying biological and physical processes. For example, the texture and wrinkles of human skin is related to a biological process, the rough surface on the wall is a result of the perennial air erosion, and the rusty surface of metal embodies a natural oxidization process. However, computer graphics rendering often model objects with polygonal surfaces with a limited resolution and hence is incapable of capturing all the subtleties of the real-world surface geometry. Finally, the photographic images are captured by a camera which often has a convex-shaped camera response function. This type of response may not be incorporated in computer graphics rendering.

Such analysis leads us to proposing a set of features derived from the differential geometry quantities (i.e., beltrami flow vectors, second fundamental form, and surface gradient), the local fractal dimension of an image, and the statistics of the local 3×3-pixel color and grayscale patches. The relationship between the features and the above-mentioned three types of physical differences between the image generation pipeline of photographic images and photorealistic computer computer are given below. More details about the features can be found in [1].

1) **Reflectance property difference**: The beltrami flow [7], a differential geometry quantity, is used to capture the above-mentioned surface reflectance property difference. The beltrami flow vector is obtained by projecting the mean curvature vector of the graph of an RGB image onto the RGB color space. Intuitively, the beltrami flow vector has an effect of capturing the local interdependency between the RGB color channels in an image.
2) **Surface geometry difference**: The surface geometry difference between the photographic image and computer graphics is captured by the second fundamental form of the graph of an image [8] (also a differential geometry quantity), the local fractal dimension, and the local patch statistics. We represent the second fundamental form by its two principle curvatures, which capture the maximal and the minimal normal curvatures of the local surface. The curvatures are capable of distinguishing sharp edges, blurred edges, sharp corner and other local 2D surface profile. Apart from the second fundamental form, the typical complexity of a real-world local surface can be aptly measured by the local fractal dimension of its photographic images [9]. In addition, the appearance of an object in a photographic image and a computer graphic image can be considered as having different styles. The appearance style has been shown to be effectively captured by the local patches of an image [10].
3) **Acquisition difference**: the acquisition difference of the photographic images and computer graphics can be captured by the surface gradient of the graph of an image. The surface gradient was shown capable of capturing the convexity property of a camera response function [1]. This observation can be further extended for estimating camera response function from a single image [11].

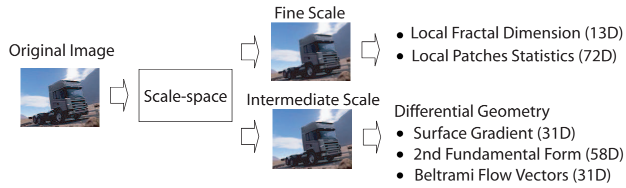
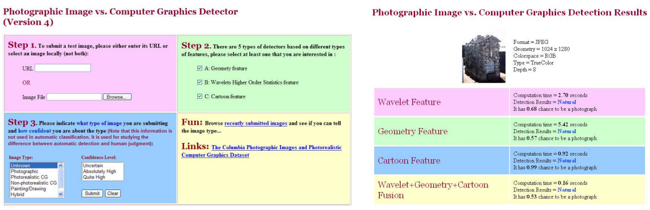


Fig. 1. The framework for feature extraction in an image scale space.

The differential geometry features, the local patch features, and the local fractal features are computed from an image at the different scales in an image scale space [12] (see Fig. 1). This represents another dimension of differences between the features. The local patch statistics features, and the local fractal dimension features are extracted from the fine scale of an image, while the differential geometry features are computed from its intermediate scale (hence oblivious to the fine-scale details in an image) as shown in Fig. 1.

III. The Online Demo System

We deployed an implementation of the physics-motivated features as an online demo system (accessible from `www.ee.columbia.edu/trustfoto`) since October 2005 [2]. The



(a) Web demo input page (b) Web demo result page

Fig. 2. The screen capture of the web demo user interface.

(a) Photograph (Photo) (b) Photorealistic CG (PRCG) (c) Non-photorealistic CG (NPRCG)

(d) Painting or Drawing (Painting) (e) Hybrid (Hybrid) (f) Others (Others)

Fig. 3. The image types to be selected by users. The keyword for the image type is given in the bracket.

goals of the online demo system are to give the web users a more concrete idea of photographic images and computer graphics classification problem, and invite the web users to help testing the three classifiers that we deployed, i.e., the geometry classifier (based on the physics-motivated features), the wavelet classifier (based on the method in [4]), and the cartoon classifier (based on the method in [5]). The open access testing scenario also allow us to evaluate the capability of the classifier in generalizing to unseen data.

Fig. 2 shows the input interface and the result page of the online system. On the input page, web users are required to input the URL of an online image or upload an image from the local computer, provide their judgement or knowledge on the image type together with the judgement confidence level, and also select the type of classifier output they wish to see. For the image type, the web users can select unknown, photographic, photorealistic computer graphics, non-photorealistic graphics, painting, hybrid, or others (see Fig. 3). Note that such user input is collected for comparing with the automatic classification results. It is not used in the automatic classification process.

IV. Analysis on the Set of Image Submission and Performance of Classifiers

We conducted a review of our online system at the first anniversary of its launch. In one year, the system received a total of 1528 image submissions. The breakdown of the user-specified image type is shown in Fig. 4 (a). The image type label can be referred to Fig. 3. Note that almost 80% of the images are labeled with the 'unknown' image type. As the 'unknown' label is the default selection, this observation

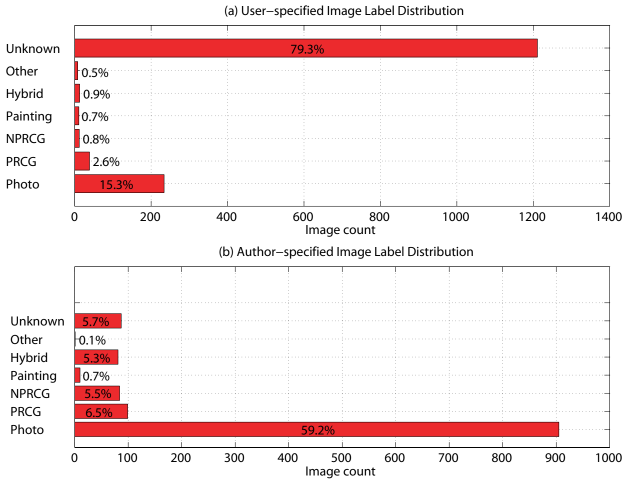


Fig. 4. (a) User-specified image label distribution, (b) Image label distribution after relabeling by first author of the paper. The image type label can be referred to Fig. 3

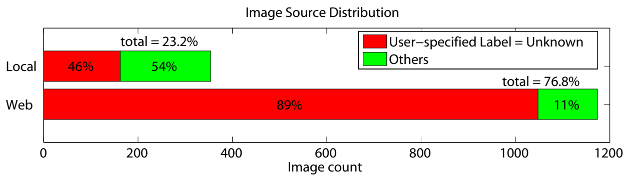


Fig. 5. Image source distribution.

indicates that for almost 80% of the times, users are either not keen on providing an image label or are truly have no idea of the image type. This is a shortcoming of our system that these two cases could not be resolved and for which we need to rectify. Fig. 5 shows the distribution of the image source, which indicates that almost 76% of the images are from the web, while the rest came from user's local machine. From Fig. 5, it is also observed that for images submitted from the local machine, users are more willing to assign an image label other than the 'unknown' label, as compared to the online images. This observation is reasonable as images from the local machine are most likely to be ones that the users know its origins, e.g., the images are captured by the users themselves.

As 90% of the web images are labeled with an 'unknown' image type and the web images make up 76.8% of the all the submitted images, we can advance our analysis by relabeling the images ourselves. The rationale for the relabeling is that for the web images which are the majority, the users' assessment of their image type is just as good as ours because we have an equal access to the meta-information concerning the images by visiting the original webpages that contain the images (we recorded the URL of the submitted images). After relabeling of the images, the distribution of the author-specified labels is shown in Fig. 4 (b). One commonality between the distribution of the user-specified labels and that of the author-specified labels is that the ratio of images with a 'photo' label to those with other labels (excluding 'duplicate' and 'unknown')

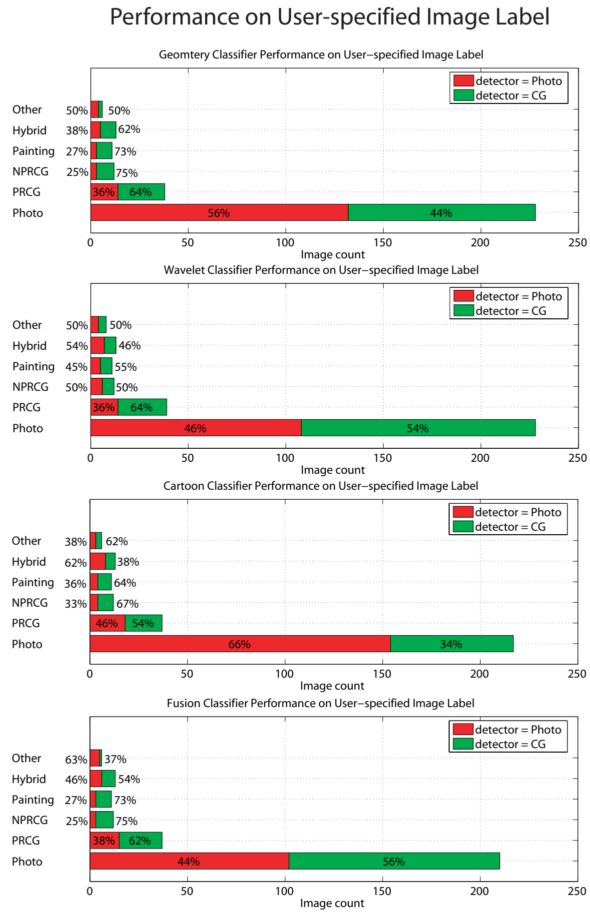


Fig. 6. Classification outcomes for the geometry, wavelet, cartoon, and fusion classifier based on the **user-specified image labels**.

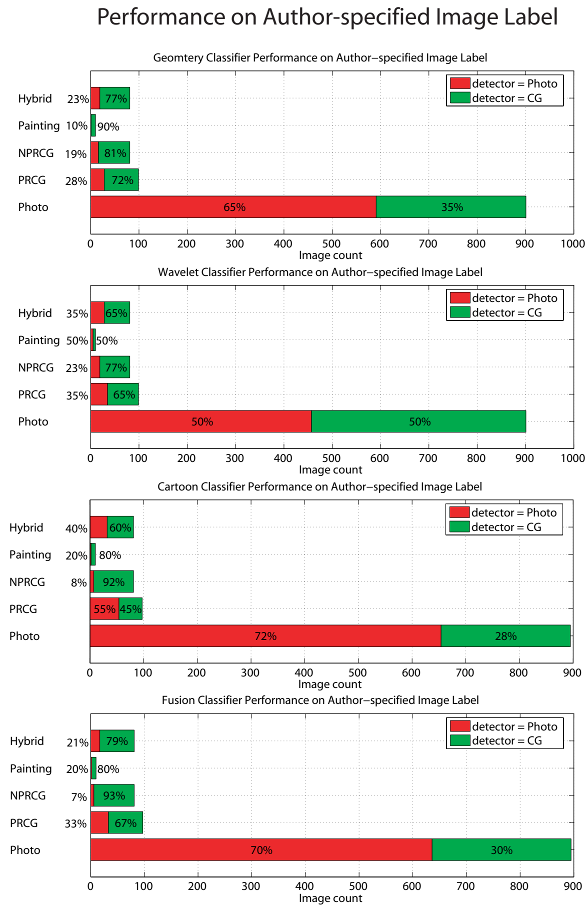


Fig. 7. Classification outcomes for the geometry, wavelet, cartoon, and fusion classifier based on the **author-specified image labels**.

is about 2.5:1. This indicates that users are more prone to submit a photographic image, which probably due to the fact that photographic images are more prevalent in the Internet, compared to other types of images. The relabeling gives us about 17% of the duplicate images (images with duplicate content). It is interesting to see such a high volume of duplicate images submitted to the classification system. One conjecture is that users may want to repeat testing of the same image to confirm the consistency of the classifier behavior.

With the user-specified image labels and the author-specified image labels, we can evaluate the accuracy of the three classifiers featured on our online system. Fig. 6 shows the classifiers' performance based on the user-specified labels and Fig. 7 shows the performance based on the author-specified labels. Note that among the three classifiers, the geometry classifier has the more balanced and satisfactory performance, as it performs equally well on the 'photo', 'PRCG', and 'NPRCG' images. The cartoon classifier performs well on the 'photo' and 'NPRCG' images, but does poorly on the 'PRCG' images. This result is understandable as the cartoon classifier is originally designed for distinguishing photographic images and non-photorealistic computer graphic images. On the other hand, the wavelet classifier does well on the 'PRCG' and 'NPRCG' images, but it performs poorly on the photographic images. We observed that many of the submitted images are highly compressed and the poor performance of the wavelet classifier could be due to its sensitivity to image compression.

The fusion classifier is obtained by combining the decisions from the three classifiers [1], and it performs very well on author-specified image labels. In this case, the fusion classifier seems to have combined the strength of all the three classifiers. However, it performs slightly poorer than the geometry classifier on the user-specified image labels.

The web users are creative in their image submission. For instance, some of the submitted images are of visually confusing image types, as shown in Fig. 8. They are photographs with a computer-graphic-like neon lighting background, and those with real paintings in the image. In this case, the photographic images can be potentially detected as computer graphics or painting, instead of being detected as photographic images. In addition, web users also submitted some hybrid images, mainly created by photomontage techniques, as shown in Fig. 9.

G=cg, W=cg, C=cg, F=cg
http://www.latimes.com/media/alternatethumbnails/photo/2006-06/24010006.jpg

G=photo, W=cg, C=cg, F=photo
http://www.iht.com/images/2006/06/25/web.0626city9ss4.jpg

G=photo, W=photo, C=photo, F=photo
http://www.spiegel.de/img/0,1020,681928,00.jpg

G=photo, W=photo, C=photo, F=photo
http://www.spiegel.de/img/0,1020,681938,00.jpg

Fig. 8. Some image with visually confusing image types. They are all judged by the author as photographic images, but they can potentially confuse the classifiers to detect them as computer graphics or painting. Shown below each image is the detection results for the geometry classifier (G), the wavelet classifier (W), the cartoon classifier (C), and the fusion classifier (F), besides the image source.

G=cg, W=cg, C=cg, F=cg
http://www.aromatherapycandlesandgifts.com/gift-baskets/gift-basket-illume-pillar-400.jpg

G=photo, W=photo, C=photo, F=photo
submitted from a local machine

G=cg, W=cg, C=cg, F=cg
http://news.tom.com/img/assets/200311/031111135215gg111120.jpg

G=photo, W=photo, C=photo, F=photo
submitted from a local machine

Fig. 9. Some hybrid images created by photomontage techniques. Shown below each image is the detection results for the geometry classifier (G), the wavelet classifier (W), the cartoon classifier (C), and the fusion classifier (F), besides the image source.

V. Classification Performance for Various Feature Sub-group Combinations

The mean and the standard deviation of the online computational speed for the geometry, the wavelets, and the cartoon classifiers are shown in Table I. The geometry classifier is the most computational intensive. In this section, we will analyze the contribution of sub-group features to the geometry classifier accuracy. The sub-group features with the least impact on the accuracy may be dropped for computational cost saving.

The set of physics-motivated features can be decomposed into five sub-groups of features according to different physical motivations. In this section, we evaluate the classification performance of each feature sub-group and their combinations.

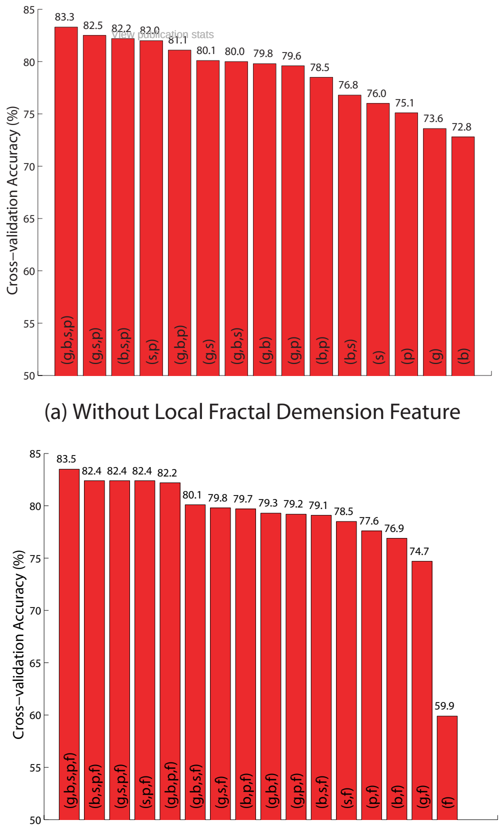
TABLE I
Statistics of the Online Classifier Speed (second per image)

Statistics	Geometry	Wavelets	Cartoon
mean	5.00	1.77	0.60
std. dev.	0.85	0.35	0.26

The main purpose is to find out the usefulness of the feature sub-group or their combinations, in terms of their classification performance.

The experiment is for classifying photographic images and photorealistic computer graphic (PRCG) images from the *Columbia Photographic Images and Photorealistic Computer Graphic Dataset* [13]. The dataset has two sets of photographic images and one set of PRCG images. The two sets of photographic images are the Personal set with 800 images acquired by the authors, and the Google set with 800 images downloaded from the Google Image Search. The PRCG set contains 800 PRCG images, downloaded from the 3D artist websites. The classification experiment is conducted using Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel. We use the LIBSVM implementation [14] for SVM and perform a grid search for the SVM kernel parameter γ and the regularization parameter C. The parameter selection is conducted with a five-fold cross-validation procedure using validation subsets partition from the training set. The classification accuracy, which is the averaged accuracy for the binary classes, is evaluated on the test set. The splitting of the entire dataset into the training set and the test set is again conducted according to a five-fold cross-validation procedure. The final classification accuracy is the average of the five-fold test accuracy. The set of geometry features can be decomposed into five sub-groups of features according to different physical motivations. These five sub-groups of features are the surface gradient features (g), the second fundamental form features (s), the beltrami flow features (b), the local patch statistic features (p), and the local fractal dimension features (f). We would like to evaluate the classification performance of each feature sub-group and their combinations.

Fig. 10 (a) shows the classification accuracy of the feature sub-group combinations excluding the local fractal dimension feature sub-group, where the results are ordered according their classification accuracy from high to low, whereas Fig. 10 (b) shows those with the local fractal dimension feature sub-group included. Note that the local fractal dimension feature sub-group has the lowest classification accuracy at 59.9%, as shown in Fig. 10 (b). Furthermore, its role is most insignificant in the complete combination of all feature sub-groups, as can be seen from the small classification accuracy difference between 83.5% for the (g,b,s,p,f) combination and 83.3% of the (g,b,s,p) combination, as compared to other feature sub-groups. The weaknesses of the local fractal dimension features are probably due to the fact we do not segment the image into smooth regions and textured regions before computing the statistics of the local fractal dimension. Such pre-segmentation may help improving the features as the fractal dimension of the



(b) With Local Fractal Demension Feature

Fig. 10. Classification performance of feature combinations (a) without and (b) with local fractal dimension features. Legend: g = surface gradient features, b = beltrami flow features, s = second fundamental form features, p = local patch statistics features.

smooth regions are mainly not interesting. The experimental verification of this suggestion will be considered in the future work. From Fig. 10, it seems that good performance (82%) can be achieved even if we use two features only (s and p). g and b does not add much on top of these two.

We analyze the computational cost of the sub-group feature extraction in Matlab 7.0 [2]. The per-image feature extraction time is as follows: 5.2s (surface gradient features), 8.7s (second fundamental form features), 6.5s (beltrami flow features), 3.0s (local patch statistics features), and 128.1s (local fractal dimension features). Note that, the local fractal dimension feature sub-group happens to be most computationally expensive. Therefore, when considering both the classification accuracy and the computational cost, the local fractal dimension features can be dropped without much impact on classification accuracy while doing so improves the computational efficiency from 152 seconds to 24 seconds per image (in Matlab 7.0).

VI. Conclusions

In this paper, we provided a further analysis on the the online system for distinguishing photographic images and computer graphics. For the online system, we analyze the 1528 image submission, received in the past one year. We found that most of the images are from the web and the 'unknown' label is the most common label specified by the users. We relabeled the images for a further analysis of the images. Based on both the user-specified and the author-specified image labels, the geometry classifier is found to have the best classification accuracy when compared to the wavelet classifier and the cartoon classifier. We also found that fusing the decision of the three classifiers is a good idea as decision fusion combines the strength of the individual classifiers. In order to improve the computational cost for the online geometry classifier, we evaluate the contribution of its five feature sub-groups to its classification accuracy. We found that the local fractal feature has an insignificant contribution to the entire set of features, apart from being most computationally expensive. By keeping the most significant feature subsets, we are able to speed up the classifier by about 6 times with very little accuracy reduction.

Acknowledgment

This work has been supported in part by NSF Cyber Trust program under award IIS-04-30258. Authors would like to thank Yu-Feng Hsu, Lexing Xie, and Martin Pepeljugoski for sharing their cameras and photographing equipments.

References

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

[2] T.-T. Ng and S.-F. Chang, "An online system for classifying computer graphics images from natural photographs," in *SPIE Electronic Imaging*, San Jose, CA, January 2006.

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

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

[5] T. Ianeva, A. de Vries, and H. Rohrig, "Detecting cartoons: A case study in automatic video-genre classification," vol. 1, 2003, pp. 449–452.

[6] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. Ginsberg, and T. Limperis, "Geometric considerations and nomenclature for reflectance," National Bureau of Standards (US), Monograph 160, October 1977.

[7] N. Sochen, R. Kimmel, and R. Malladi, "A general framework for low level vision," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 310–318, 1998.

[8] M. P. D. Carmo, *Differential Geometry of Curves and Surfaces*. Prentice Hall, 1976.

[9] A. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 661–674, November 1994.

[10] R. Rosales, K. Achan, and B. Frey, "Unsupervised image translation," 2003, pp. 472–478.

[11] T.-T. Ng, S.-F. Chang, and M.-P. Tsui, "Camera response function estimation from a single-channel image using differential invariants," Columbia University, Tech. Rep., March 2006.

[12] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA, USA: Kluwer Academic Publishers, 1994.

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

[14] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," July 2003.

[2]The computational speed of the online classifiers presented in Sec. V is one based on a C language implementation.