An Online System for Classifying Computer Graphics Images from Natural Photographs

Tian-Tsong Ng and Shih-Fu Chang

{ttng,sfchang}@ee.columbia.edu

Department of Electrical Engineering
Columbia University
New York, USA

ABSTRACT

We describe an online system for classifying computer generated images and camera-captured photographic images, as part of our effort in building a complete passive-blind system for image tampering detection (project website at `http://www.ee.columbia.edu/trustfoto`). Users are able to submit any image from a local or an online source to the system and get classification results with confidence scores. Our system has implemented three different algorithms from the state of the art based on the geometry, the wavelet, and the cartoon features. We describe the important algorithmic issues involved for achieving satisfactory performances in both speed and accuracy as well as the capability to handle diverse types of input images. We studied the effects of image size reduction on classification accuracy and speed, and found different size reduction methods worked best for different classification methods. In addition, we incorporated machine learning techniques, such as fusion and subclass-based bagging, in order to counter the effect of performance degradation caused by image size reduction. With all these improvements, we are able to speed up the classification speed by more than two times while keeping the classification accuracy almost intact at about 82%.

Keywords: Computer graphics, photograph, classification, online system, geometry features, classifier fusion

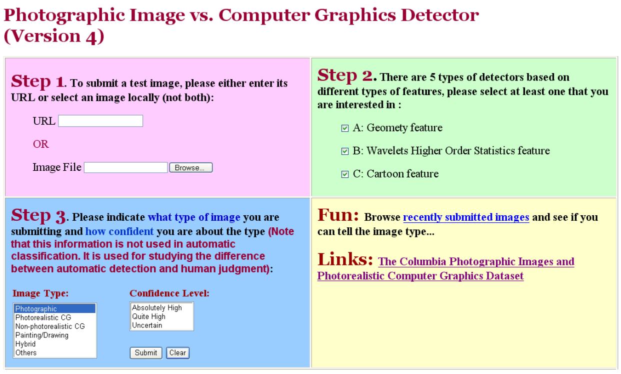
1. INTRODUCTION

The level of photorealism capable by today's computer graphics makes distinguishing photographic and computer graphic images difficult. Alias, a 3D computer graphics company, runs an online challenge (`http://www.fakeorfoto.com`) for people to distinguish five photographic and five computer graphic images. In fact, there are experimental evidences that, for certain sorts of scenes, computer graphic images are visually indistinguishable from photographic images by human.[1]

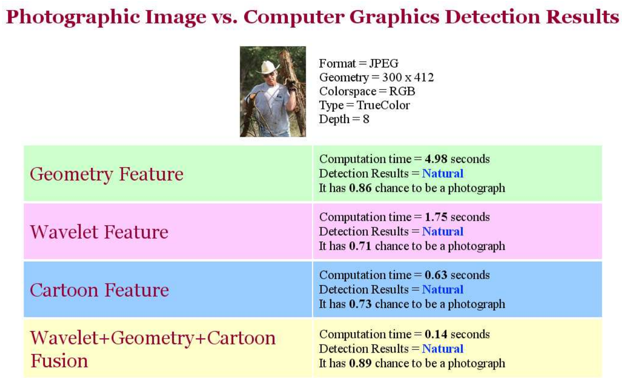
In our prior work,[2] we have addressed the problem of distinguishing photorealistic computer graphics (PRCG) and photographic images (PIM), in the context of passive-blind image forensics. Due to the high photorealism achievable for PRCG, it can be considered a form of image forgery and the goal of passive-blind image forensics is to find out the actual production source of an image without any prior information. We approach the problem by analyzing the differences of PIM an PRCG, in terms of their image generative process. To capture these physical differences, we derive a set of geometry features from fractal geometry, differential geometry and local patch statistics. By the geometry features, we are not only able to explain the physical differences between PIM and PRCG, the resulting classifier also outperforms those reported in the prior work.[3,4]

As part of the research effort, we have developed an online system for classifying computer graphics (CG) and PIM (at `http://www.ee.columbia.edu/trustfoto/demo-photovscg.htm`). Our goal for the online system is as follows:

1. We would like to further evaluate our technique in an open and realistic environment. In the Internet, the system would encounter a wide variety of images, some of which may have not been seen in the laboratory experiment.
2. We would like to offer an opportunity to compare the various proposed techniques for classifying CG/PRCG and PIM. This effort provides a motivation for further innovation.
3. The system may be used as an educational tool for promoting the awareness on the credibility of the online images as we see today. Such awareness is important when the falsehood in images is getting more prevalent.



(a) Image submission page



(b) Classification result page

Figure 1. (a) The web interface for the online CG and PIM classification system, where users can submit images. (b) A classification result page for the online system. The page lists information about the submitted image, and the classification decision as well as the confidence score for three types of classifiers, i.e., the geometry, the wavelet and the cartoon classifier.

The goal of this paper is to describe our online CG and PIM classification system and present the system design challenges as well as our strategy in resolving the challenges. In Section 2, we describe the online system and the corresponding design challenges. Then, we describe the features used by the classifiers in Section 3 and the dataset used for training the classifiers in Section 4. Finally before the conclusions, we present our strategy in overcoming the challenges and also the experimental results which vouch for our strategy in Section 5.

2. ONLINE SYSTEM DESCRIPTION

The online CG and PIM classification system provides a web interface for users to submit images, either by providing the URL of the image or uploading an image from a local computer, as shown in Figure 1(a). Then, the user selects the types of classifier to be used. We currently implemented three types of classifier, which are the geometry classifier,[2] the wavelet classifier[3] and the cartoon classifier.[4] Before submitting the image, the user provides some information about the image for survey purposes, according to their knowledge or judgment. The survey information includes the type of the image and the user's confidence on the image type (computer graphics or natural photos). There are six image types that the users can select as an input to the survey, as shown in Figure 2(b).

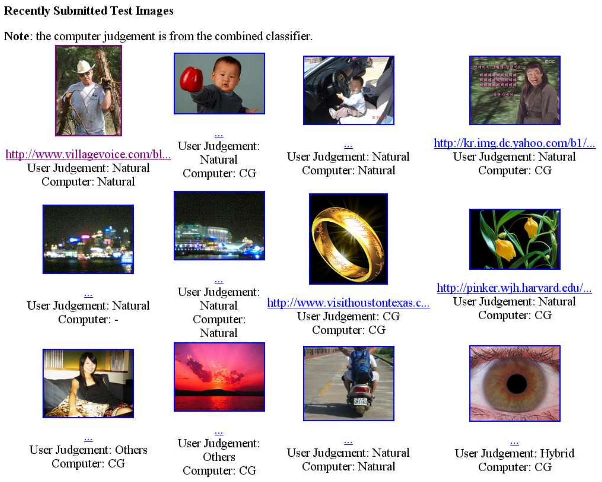
The submitted image will be processed and classified. The meta-information of the submitted image and the classification decisions including the confidence score for the image as being a photograph (a real number between 0 and 1) will be shown on a result page, as shown in Figure 1(b). The classification decision for each of the selected classifiers will be shown in a same table and may be useful for comparison. Also shown is the results for the fusion classifier (obtained by fusing the decision of the geometry, the wavelet and the cartoon classifier). The users can also browse the most recently submitted images and see how the fusion classifier do in terms of the consistency between the classifier decision and the human judgment (provided as part of the survey information), as shown in Figure 2(a).

2.1. System Design Challenges

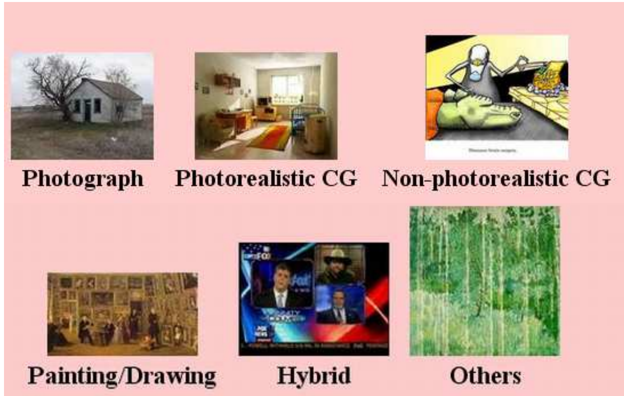
In designing the online classification system, we face the following challenges:

1. Processing speed: For user-friendliness, the system should not take too long (e.g., a few minutes) for processing a submitted image. The submitted images can be of various size and the processing time in general depends on the image size. Therefore, reducing the processed image size (e.g., by resizing or central-region cropping) is a strategy for improving the processing speed.

2. Classification accuracy: For the usefulness of the system, the classification should have a reasonably good classification accuracy as compared to random guessing. However, as reduction in the processed image size may degrade the classification accuracy, the criteria of processing speed and classification accuracy are in a tradeoff relationship. Therefore, we need to find a strategy to improve the classification accuracy for the case of reduced-size images.



(a) Submission history page



(b) Classification of image type

Figure 2. (a) A web page that shows the most recently submitted twelve images. Also recorded are the image type according to the user's judgment and the classifier decision. (b) The six types of images that users can specify as the survey information.

3. The diversity of the input images: It is not too surprising that an online system would encounter input images of diverse types, which includes PRCG, non-photorealistic CG, CG-and-PIM-hybrid images, painting or drawing and so on. However, most of the experiments presented in the related paper[2–4] do not consider such a wide spectrum of images. In order to obtain a well-performing classifier, we need to include images of wide diversity into the training dataset.

Before describing how we overcome the above challenges, we will first describe the features used for the geometry classifier, the wavelet classifier and the cartoon classifier (in Section 3), and the dataset used for training of the online classifiers (in Section 4).

3. CLASSIFICATION FEATURES

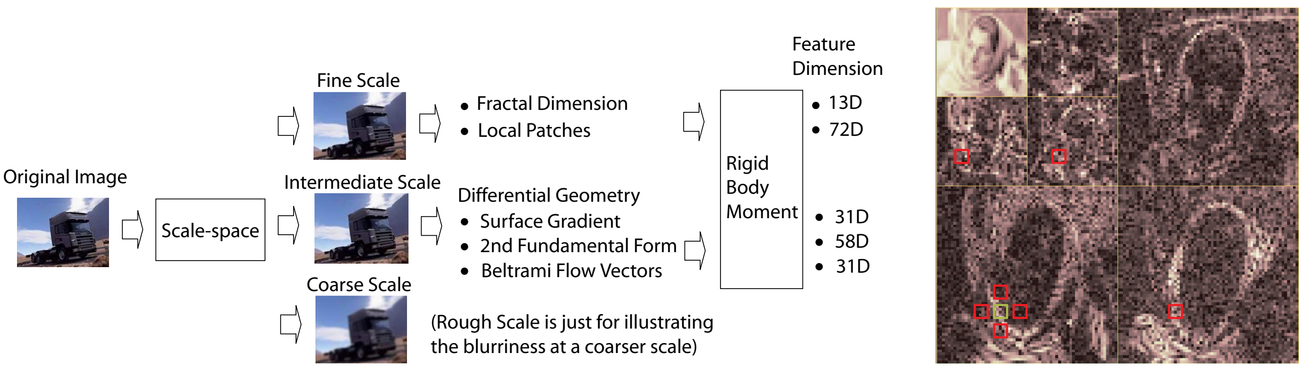
Our online classification system consists of three types of classifier, respectively based on the geometry features, the wavelet features and the cartoon features, which are described in the following subsections.

3.1. Geometry Features

In our prior work,[2] the problem of photographic images (PIM) and photorealistic computer graphics (PRCG) classification is approached by analyzing the physical differences between the image generative process for PIM and PRCG. This approach provides a physical explanation for the actual differences between PIM and PRCG, while the geometry features from this approach outperforms the features in the prior work.

Specifically, the difference between PIM and PRCG can be briefly summarized below.

1. **Object Model Difference**: The surface of real-world objects, except for man-made objects, are rarely smooth or of simple geometry. Mandelbrot[5] has showed the abundance of fractals in nature and also related the formation of fractal surfaces to basic physical processes such as erosion, aggregation and fluid turbulence. Furthermore, surface such as human skin is full of subtleties and a result of the natural biological process. However, the computer graphics 3D objects are often represented by the polygonal models. Although the polygonal models can be arbitararily fine-grained, it comes with a higher cost of memory and computational load. Furthermore, such a polygonal model is not a natural representation for fractal surfaces.[6] A coarse-grained polygonal model may be used at the perceptually insignificant area for saving computational resources.



(a) Geometry feature framework

(b) Wavelet feature extraction

Figure 3. (a) The geometry-based image description framework. (b) An illustration on how the linear prediction error of a wavelet coefficient (at the green box location) is computed using the wavelet coefficients (at the red box locations) in the neighboring spatial location, orientation and scale, for a grayscale image case. For a color image, the wavelet coefficients of other color channels can also be used for computing the prediction error.

2. **Light Transport Difference**[7]: The physical light field captured by a camera is a result of the physical light transport from the illumination source, reflected to the image acquisition device by an object. The precise modeling of this physical light transport involves an 8D function of the object's reflectance property, hence its simulation requires substantial computational resources. Therefore, a simplified model based on the assumption of isotropy, spectral independence and parametric representation is often used.

3. **Acquisition Difference**: PIM carry the characteristics of the imaging process, while PRCG may undergo different types of post-processing after the rasterizer stage. There is no standard set of post-processing techniques, but a few possible ones are the simulation of the camera effect, such as the depth of field, gamma correction, addition of noise, and retouching.

The above differences are captured using the geometry features derived from differential geometry, fractal geometry and local patch statistics. Specifically, we have developed a two-scale image description framework, as shown in Figure 3(a). At the finest scale of the linear Gaussian scale-space, the geometry can be characterized by the local fractal dimension and also by the "non-parametric" local patches.[8] At an intermediate scale, when the fine-grained details give way to a smoother and differentiable structure, the geometry can be best described in the language of differential geometry, where the surface gradient, the second fundamental form and the Beltrami flow vectors are computed. The total dimension of the geometry features is 192.

3.2. Wavelet Features

A prior work[3] uses the wavelet-based natural image statistics for the PIM and PRCG classification. The method extracts the first four order statistics (mean, variance, skewness and kurtosis) of the in-subband wavelet coefficients and also computes the first four order statistics of the linear prediction error for the wavelet coefficients using the coefficients from the neighboring spatial location, scale, orientation and the other color channels, as illustrated in Figure 3(b). The total dimension of the wavelet features is 216.

3.3. Cartoon Features

In another prior work,[4] the problem of PIM and CG classification is approached by modeling the characteristics of the general (i.e., including both photorealistic and non-photorealistic) computer graphics. We call these features the cartoon features. The cartoon features include the average color saturation, the ratio of image pixels with brightness greater than a threshold, the Hue-Saturation-Value (HSV) color histogram, the edge orientation and strength histogram, the compression ratio and the pattern spectrum. The total dimension of the cartoon features is 108.

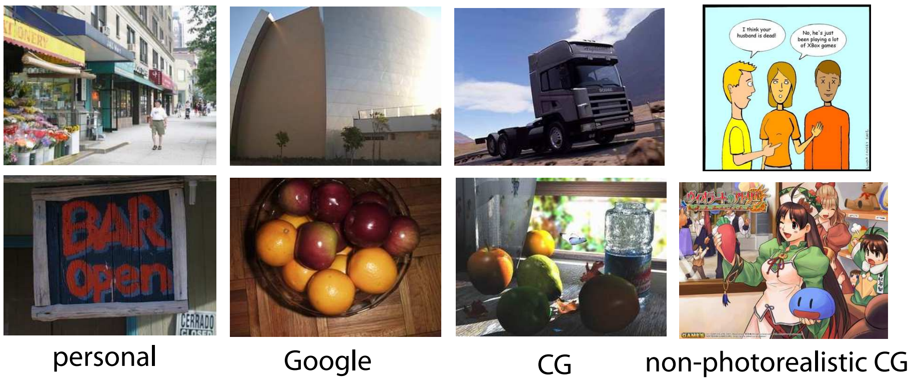


Figure 4. Example images from the four categories of the online classifier training dataset.

4. DATASET

In our prior work,[2] the experiments are performed on the *Columbia Photographic Images and Photorealistic Computer Graphics Dataset*[9] which includes only photorealistic computer graphics. One of the challenges in system design is the potential wide diversity of the input images and non-photorealistic CG images are actually very common on the Internet. In order to improve the classification performance, we have to include another set of 800 non-photorealistic computer graphics (NPRCG) into the online classifier training dataset. As a result, our training image set consists of four categories as shown in Figure 4. Note that in training the classifiers used for this online system, the recaptured CG category in the Columbia Photographic Images and Photorealistic Computer Graphics Dataset is excluded as the recaptured CG category mainly caters for the recapturing attack experiment. Below, we list the details of the four image categories in the online classifier training dataset.

1. 800 PRCG (*PRCG*): These images are categorized by content into architecture, game, nature, object and life, see Figure 5(a). The PRCG are mainly collected from various 3D artists (more than 100) and about 40 3D-graphics websites, such as `www.softimage.com`, `www.3ddart.org`, `www.3d-ring.com` and so on. The rendering software used are such as 3ds MAX, softimage-xsi, Maya, Terragen and so on. The geometry modelling tools used include AutoCAD, Rhinoceros, softimage-3D and so on. High-end rendering techniques used include global illumination with ray tracing or radiosity, simulation of the camera depth-of-field effect, soft-shadow, caustics effect and so on.

2. 800 PIM images (*Personal*): 400 of them are from the personal collection of Philip Greenspun, they are mainly travel images with content such as indoor, outdoor, people, objects, building and so on. The other 400 are acquired by the authors using the professional single-len-reflex (SLR) Canon 10D and Nikon D70. It has content diversity in terms of indoor or outdoor scenes, natural or artificial objects, and lighting conditions of day time, dusk or night time. See Figure 5(b).

3. 800 PIM from Google Image Search (*Google*): These image are found by searching Google Image Search with the following keywords: architecture, people, scenery, forest, statue, and so on. We choose only the images that are judged to be natural photographs. The judgements are made by the authors subjectively, thus there is potential risk of human misclassification. We intentionally add this category to ensure that the CG and photograph categories have similar content such as architecture, people, etc.

4. 800 NPRCG (*NPRCG*): The NPRCG includes cartoon, drawing, 2D graphics, images of presentation slide, logo images and so on. These images are collected from the Google Image Search.

5. STRATEGIES FOR OPTIMIZING THE SPEED-ACCURACY TRADEOFF

5.1. Image Downsizing

For addressing the system design challenge of processing speed, we downsize the input images such that the longer dimension of the downsized image is of 360 pixels (the entire content of an image is retained, so is the aspect ratio). As the

Figure 5. (a) Subcategories within *CG* and (b) Subcategories within *personal* image set, the number is the image count.

Table 1. SVM Classification Accuracy for Different Image Size Reduction Strategies

Classifier	Original size	Downsizing	Central Cropping
Geometry	83.8%	78.2%	79.9%
Wavelets	81.2%	77.3%	72.8%
Cartoon	76.1%	73.1%	75.9%

algorithm for extracting the features is proportional to the image size, the reduction of computational load due to image downsizing is substantial (about more than two times in average) as the average size of the web images is around 700×500 pixels.

In a prior work,[3] computational efficiency is obtained by cropping out the smaller-size central area from the images. In order to evaluate the two strategies of reducing image size, we train support vector machines (SVM) of the LIBSVM[10] implementation using the dataset mentioned in Section 4 for separate cases where the images are downsized, central-cropped and with the original size. In this case, the *Personal* and *Google* categories form a class, while the *PRCG* and *NPRCG* categories form the opposite class. We use the radial basis function (RBF) kernel for the SVM and model selection (for the regularization and the kernel parameters) is done by a grid search[10] in the joint parameter space. The classification performance we report hereupon is based on a five-fold cross-validation procedure.

Table 1 shows the SVM classification accuracy corresponding to the two different image size reduction strategies, i.e., downsizing and central cropping, while comparing to the SVM performance when there is no reduction of image size. Note that central cropping results in a slightly better classification accuracy for the geometry ($+1.7\%$) and the cartoon ($+2.8\%$) classifier, but causes a serious drop of performance for the wavelet (-4.5%) classifier. In this case, it is inconclusive on which image reduction strategy is better, but the degradation of the classifier performance due to image downsizing is more uniform over all the classifiers.

5.2. Classifier Fusion

The image downsizing results in an average of 4.2% decrease (over the three classifiers) in the classification accuracy when comparing to the case of no image size reduction. To counter the performance decrease, we consider fusing the SVM classification outputs from the geometry, the wavelet and the cartoon classifiers. The effort of fusing a set of base classifiers is only justified when the fusion classifier outperforms the best among the base classifiers. According to an analysis,[11] this only happens when the individual base classifiers are reasonably accurate, individual performances are comparable, and their errors are uncorrelated. Although these conditions may be not satisfied in practice, fusion of classifiers often leads to a better classification performance and three fundamental reasons for the good performance are proposed.[12] Firstly, in the case of a small dataset, different classifiers may provide different but equally good hypothesis. Fusion of these hypothesis reduces the risk of choosing the wrong hypothesis. Secondly, for learning algorithms which could only achieve local optima, fusion of hypothesis corresponding to the multiple local optima may be a better approximation to the target function. Thirdly, for the case where the target function is not in the hypothesis space of the individual base classifiers, the fused classifier may have an enhanced hypothesis space which includes the target function.

We evaluate three fusion strategies, the normalized ensemble fusion,[13] the score concatenation fusion, the majority vote fusion. The normalized ensemble fusion procedure consists of three steps:

Table 2. Downsized Image Classification Accuracy for Different Fusion Strategies

Normalized ensemble	Score concatenation	Majority vote	Best base classifier (geometry)
79.7%	80.0%	77.4%	78.2%

1. **Normalization of the distance of the data points to the SVM decision boundary**: The paper suggests three types of normalization schemes, i.e., rank normalization, range normalization and Gaussian normalization, for producing normalized scores.

2. **Combination of the normalized scores**: The paper suggests functions such as minimum, maximum, average, product, inverse entropy and inverse variance for combining the normalized scores.

3. **Finding the optimal fusion strategy**: The optimal fusion strategy is obtained by searching over all the normalization schemes and the combination functions.

The features obtained from the normalized ensemble fusion are then used for training the final fusion SVM.

Before performing the score concatenation fusion and the majority vote fusion, the binary SVM output is mapped into posterior probability by fitting the empirical posterior histogram of the distance of the data points to the decision boundary using a parametric sigmoid function[14]:

$$P(y = 1|f) = \frac{1}{1 + exp(Af + B)} \tag{1}$$

where f is the distance from the decision boundary and (A, B) are the sigmoid model parameters. The score concatenation fusion concatenates the posterior probability of all the base classifiers and the concatenated features are used for training the final SVM. The majority vote fusion selects the final decision by a majority vote from the base classifiers. In our evaluation, we compute the classification accuracy for the three above-mentioned fusion strategies, using the dataset of downsized images. The classification accuracy is as shown in Table 2. Note the accuracy reported here is for classifying downsized images. From the results, we see the use of classifier fusion indeed improves the accuracy in classifying downsized images by about 1.8%, as compared to the best performing base (geometry) classifier.

Note that the normalized ensemble fusion and the score concatenation fusion perform equally well on the dataset, while the the majority vote fusion result seems to be just about the average of the classification accuracies of the three base classifiers. Due to the simplicity of the score concatenation fusion as compared to the normalized ensemble fusion, we choose to use the score concatenation fusion method.

5.3. Exploiting Dataset Heterogeneity

The increase of the classification accuracy for the score concatenation fusion is only 1.8%, which is quite minor. To further improve the classification accuracy for downsized images, we aim to find a better way to generate a set of base classifiers by exploiting the diversity and heterogeneity within the dataset. As we can see that while the *PRCG* and the *NPRCG* image categories form a single class (CG), they are actually two very different types of images. Similarly, the *Personal* and the *Google* image categories are different in the sense that the *Personal* category consists of the typical camera images from a few high-end cameras with diverse content while the *Google* category contains images from diverse models of camera potentially having undergone various types of additional post-processing.

From the dataset, we can generate nine (i.e., 3^2) sets of two-class data by exhaustively combining the elements of the power set of the two classes, as shown in Figure 6. We train a SVM base classifier for each of the two-class data subset combinations. The binary outputs of the SVM base classifiers are mapped to posterior probability by fitting a sigmoid function, as given in Equation 1. Then, the posterior probability for all base classifiers are combined through the score concatenation fusion for training the final fusion SVM classifier.

The accuracy of the fusion classifiers for the dataset of original size images and downsized images are shown in Table 3. The above approach is conceptually similar to a common machine learning technique, called bagging, in which different subsets of data are used to train multiple classifiers to be combined later. Each individual classifier has the potential to

Table 3. Classification Accuracy After Considering Dataset Heterogeneity

Original image	Downsized images
84.0%	82.0%

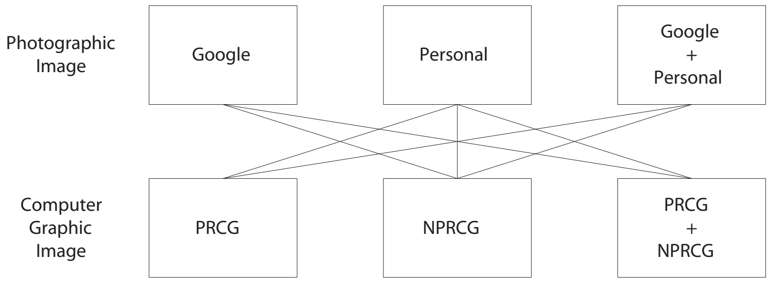


Figure 6. Nine combinations of the data subsets.

capture the differences between classes contained in each unique subset of data. The difference of our implementation from the conventional bagging is that in our case subsets of data are obtained according to the data subtypes, rather than sampling of the whole set. Note that for the downsized images, the gain in classification accuracy after considering dataset heterogeneity is 2%, while there is very little gain for the original size images (84.0% compared to best performing geometry classifier with a classification accuracy of 83.8%). To explain the observation, we conjecture that image downsizing has made the base classifiers less stable and decreased the correlation of error between the base classifiers. Therefore, according the analysis mentioned before, there is a greater chance for classifier fusion to be effective in the case of downsized images.

In summary, we address the system design challenges by reducing the processed image size, including an additional set of non-photorealistic computer graphic images into the training dataset, and exploiting the heterogeneity within the dataset. By adopting this strategy, the per-image processing time is reduced by more than two times, the classification accuracy degrades only by 2.0% as compared to the case with no image downsizing, and the system can now handle non-photorealistic computer graphic images.

6. CONCLUSIONS

In this paper, we have described our online photographic images and computer graphic images classification system and presented three system design challenges, i.e., the high precessing speed, the high classification accuracy and the capability of dealing with diverse image types. We have also shown our strategy in addressing the challenges by reducing the processed image size, choosing the right methods for reducing image size, including an additional set of non-photorealistic computer graphic images into the training dataset, and exploiting the heterogeneity within the dataset.

ACKNOWLEDGMENTS

This project is supported in part by the NSF CyberTrust program (IIS-04-30258) while the first author is supported by the Singapore A*STAR scholarship.

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 SIGGRAPH* **5**(1), pp. 30–50, 1986.
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,” in *ACM Multimedia*, (Singapore), November 2005.
3. S. Lyu and H. Farid, “How realistic is photorealistic?,” *IEEE Transactions on Signal Processing* **53**, pp. 845–850, February 2005.
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*, **1**, pp. 449–452, 2003.

5. B. B. Mandelbrot, *The fractal geometry of nature*, San Francisco: W.H. Freeman, 1983.
6. A. Pentland, "On describing complex surface shapes," *Image and Vision Computing* **3**, pp. 153–162, November 1985.
7. T. Akenine-Moller, T. Moller, and E. Haines, *Real-Time Rendering*, A. K. Peters, Ltd., MA, 2002.
8. A. B. Lee, K. S. Pedersen, and D. Mumford, "The nonlinear statistics of high-contrast patches in natural images," *International Journal of Computer Vision* **54**(1), pp. 83–103, 2003.
9. T.-T. Ng, S.-F. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," ADVENT Technical Report 205-2004-5, Columbia University, Feb 2005.
10. C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," July 2003.
11. L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis Machince Intelligent* **12**(10), pp. 993–1001, 1990.
12. T. G. Dietterich, "Ensemble methods in machine learning," *Lecture Notes in Computer Science* **1857**, pp. 1–15, 2000.
13. B. L. Tseng, C.-Y. Lin, M. Naphade, A. Natsev, and J. R. Smith, "Normalized classifier fusion for semantic visual concept detection," in *IEEE International Conference in Image Processing*, pp. 535–538, 2003.
14. J. C. Platt, *Advances in Large Margin Classifiers*, ch. Probabilities for Support Vector Machines, pp. 61–74. MIT Press, 1999.