# FEATURE BASED CLASSIFICATION OF COMPUTER GRAPHICS AND REAL IMAGES

*Gopinath Sankar*[1], *Vicky Zhao*[2], *Yee-Hong Yang*[1]

[1]Department of Computing Science
[2]Department of Electrical and Computer Engineering
University of Alberta, Canada

## ABSTRACT

Photorealistic images can now be created using advanced techniques in computer graphics (CG). Synthesized elements could easily be mistaken for photographic (real) images. Therefore we need to differentiate between CG and real images. In our work, we propose and develop a new framework based on an aggregate of existing features. Our framework has a classification accuracy of 90% when tested on the de facto standard Columbia dataset, which is 4% better than the best results obtained by other prominent methods in this area. We further show that using feature selection it is possible to reduce the feature dimension of our framework from 557 to 80 without a significant loss in performance (< 1%). We also investigate different approaches that attackers can use to fool the classification system, including creation of hybrid images and histogram manipulations. We then propose and develop filters to effectively detect such attacks, thereby limiting the effect of such attacks to our classification system.

**Index Terms —** graphics, forgery, authentication, hybrid images, histogram manipulation

## 1. INTRODUCTION AND RELATED WORKS

Media that include photos, audios and videos could be forged to deceive viewers for commercial and political reasons. Forgeries vary in the scale and scope of their applications. In this paper we are interested in forgeries where photorealistic computer graphics (CG) images are used as replacement for photographic (real) images. Recent advances in computer graphics make it possible to create models that are physically accurate without visual artifacts and lighting inconsistencies. Also, with the help of advanced image editing software like Adobe Photoshop, computer generated images look very photorealistic. Therefore, to prevent image forgeries and to verify the authenticity of an image, it is of crucial importance to differentiate between computer graphics and real images. A brief survey of prior related works sheds light on different techniques which aim at differentiating CG from real images [1, 2, 3, 4]. Ianeva *et al.* [5] propose features for differentiating cartoon from real images with an accuracy of 94%. Their features are prominently based on histograms, including color and edge histograms, average saturation and threshold brightness. Their approach has been extended for classifying CG from real images with an accuracy of 72% on the Columbia dataset [3]. Tsong *et al.* [2] have developed physics motivated features for distinguishing CG and real images. They identify features based on the scene-characteristics of CG images like its smooth nature, simplified geometry and absence of natural statistics. They further include features based on capturing-device characteristics, by targeting the differences in the post-processing process of CG and real images. Their classification accuracy is 83% on the Columbia dataset. Chen *et al.* [1] develop a method using statistical moments of wavelet coefficients in the HSV color space with an accuracy of 82%. The methods described earlier deal with different properties which differentiate CG from real images. For example, cartoon features focus on properties of CG which are cartoonic, the moment-based feature focuses on the derivatives of histogram of an image, physics features focus on the physical attributes which separate real images from CG. This motivates us to build a framework by combining the prominent features from existing methods together with the texture interpolation features we propose in this paper. Our framework based on the aggregate set achieves 90% classification accuracy with 557 features. These results are 4% better than the best results from the existing methods (which are based on feature differences between CG and real images, and not based on capturing device characteristics). Furthermore, there is little prior work on studying attacks against such classification systems. This motivates us to investigate possible strategies that an attacker can use to deteriorate the performance of the classification system. We further identify features which resist such attacks.

## 2. FEATURE SELECTION

In addition to the existing features, our framework consists of a new feature based on texture interpolation for classifying CG from real images. Textures in CG undergo transformations such as scaling, rotation, affine transforms, *etc*. These transformations introduce periodic correlations in an image. In general, periodic correlations are observed in images which are resampled [6]. However, CG contains several textures which are duplicated and resampled by different amounts depending on factors including size of object, orientation, position in the scene, *etc*. These different resamplings introduce several distinctive periodic correlations. The presence of many distinct periodic correlations in an image can be used to identify an image as CG. In order to capture the different periodic correlations in an image, we segment the image to rectangular tiles of equal sizes. We assume that the textures resampled are large enough that there exists a minimum of two tiles containing textures resampled at different amounts. We use the method developed by Popescu *et al.* [6] to obtain the probability map depicting the periodic correlations belonging to each tile. We model the different periodic correlations of the whole image by local patch statistics of the probability map of an image (refer [2] for the destination of local patch statistics features). We further capture the periodic correlations caused by moments of characteristics functions of the probability map and its prediction component. In general, the prediction error component and the wavelet transforms of images capture their spatial correlations. However, in the case of the probability map the spatial neighborhood

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

of a pixel specifies the periodic correlations present in them, hence the moment-based features are capable of capturing the periodic correlations caused by texture resampling. The normalized moments of the characteristic functions used in the moment-based method [1] are defined as

$$M_n = \sum_{\omega=1}^{N/2} \omega^n \left|\varphi(\omega)\right| \Big/ \sum_{\omega=1}^{N/2} \left|\varphi(\omega)\right| \qquad (1)$$

where $\varphi(w)$ is the characteristic function value at frequency $w$, $n$ is the moment order and $N$ is the range of the histogram. The characteristic function $\varphi_x(w)$ of the probability density function $f(x)$ is nothing but the complex conjugate of its Fourier transform $F(w)$

$$\varphi_x(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)dx = \overline{\left(\int_{-\infty}^{\infty} e^{-i\omega x} f(x)dx\right)} = \overline{F(\omega)}. \qquad (2)$$

Further, the prediction error image is the difference between the original image and its predicted version (predicted based on its neighbors) [7]. The prediction algorithm used by the moment-based method is

$$\hat{x} = \begin{cases} max(a,b) & c \leq min(a,b) \\ min(a,b) & c \geq max(a,b) \\ a+b-c & otherwise \end{cases} \qquad (3)$$

where $\hat{x}$ is the predicted value of x. The neighborhood of $x$ is given by the ordering $\begin{array}{|c|c|}\hline \text{x} & \text{b} \\ \hline \text{a} & \text{c} \\ \hline \end{array}$. We achieve a classification performance of 78% using our texture interpolation method. We further evaluated the individual performance of all the features used for classifying CG from real images (please consult the original papers [1,2,5] for detailed descriptions of features). Our experimental results suggest that, only the following features give a classification accuracy above 60% on the Columbia dataset. The rest of the features proposed in the literature can achieve no more than 60% in classification accuracy on the Columbia dataset:

- color histogram feature from the cartoon set of features (69% accuracy with 45 features) [5],
- moment-based features in the YCbCr color space (86% accuracy with 234 features) [1],
- local patch statistics features from the physics set of features (69% accuracy with 96 features) [2], and
- features based on texture interpolation (78% accuracy with 182 features).

We include the above list of features in our framework for classifying CG from real images. The color histogram and local patch statistics features give a classification performance of 69% and 70%, respectively. For the moment-based features, Chen *et al.* predicted a better performance using YCbCr color space when compared to their results in the RGB and HSV color spaces included in their paper. Indeed, our experimental results indicate that the YCbCr color space performs with an accuracy of 86%, better than 84% with HSV and 81% with RGB color spaces. We further obtain variations of up to 8% using other color spaces. The aggregate set of features achieve a classification accuracy of 90% using the Columbia dataset. We apply the greedy hill-climbing algorithm for feature selection on the aggregate set to reduce its dimension. The results in figure 1 suggests that we could achieve 90% classification performance with less than 1% variation using as few as 80 features compared to the original 557 features in the aggregate set. In the reduced set containing 80 features, 44 are from the moment-based method, 24 from our texture interpolation method, and 6 each from color histogram and patch statistics. Our greedy approach is not optimal for obtain in maximum performance with a reduced set of features. However we have demonstrated that it is possible to reduce the feature dimension of the aggregate set without loss of classification accuracy.

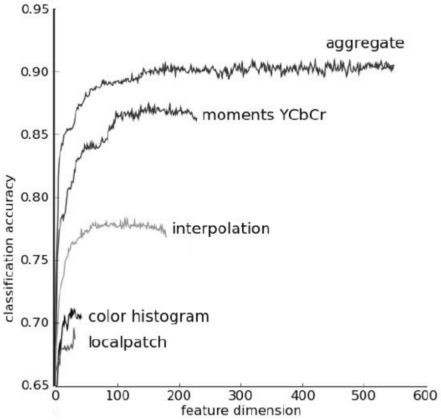

**Fig. 1**. Feature dimension vs classification accuracy

## 3. EFFECT OF MALICIOUS MANIPULATIONS ON CLASSIFICATION PERFORMANCE

In this section, we are concerned with forgery techniques wherein the contents of an image are modified purposefully in order to forge its authenticity. Especially we focus on histogram manipulations since they alter the histogram of an image which deeply affects the moment-based techniques which is the lifeblood of our framework. Further, hybrid images provide a potent threat since they combine the features of both CG and real images, thereby limiting the ability of our framework to identify them. Though we could envisage other possible fronts for forgeries, we focus our discussions on these since they have favorable properties such as trivial in implementation, visually indiscernible, and mainly because they are derived based on the discriminating features themselves and hence potentially could affect our framework the most. Moreover, unintentional manipulations like compression, filtering, changing the contrast of the image, *etc.* could affect the features extracted for classification. Since these manipulations directly affect the features which differentiate CG from real images it is beyond the scope of this paper to address their effect on image authentication. However, we acknowledge that this raises an important issue for image authentication in general and in future hope to increase the robustness of our framework.

### 3.1. Hybrid images

Hybrid images are generated by compositing CG and real images. Designers often introduce real world elements in graphics for texturing. For example, it is possible to make a CG image look more realistic by applying textures of real world objects like sky, water, grass, *etc.* Hybrid images cannot be grouped either with CG or real images since they contain features from both classes. Hence one cannot predict if a hybrid image would be classification as CG or real image. Though we do not aim at identifying hybrid images directly, we are interested in them in the context of image forgery. A malicious attacker could fool our classification system by introducing CG/real patches to a real/CG image affecting its classification result. Hence hybrid images present an obstacle to our classification system. For our experiments, we generated hybrid images by compositing CG and real images from the Columbia dataset. We produced 800 hybrid images for our experiment. 400 of them have 75% real image contents and 25% CG contents and the rest with 25% real and 75% CG contents. The hybrid images are created through image splicing, where the 25% CG/real image contents are pasted onto an image which is 100% real/CG. We train our classification on the training set consisting of the original 1600 images from the Columbia dataset. We test our classification on these 800 hybrid images. The classification results are as follows. We achieve only 77% accuracy in identifying the majority 75% contents in the hybrid images i.e., images contains 75% CG/ 75% real contents are classification as CG/real images with 77% accuracy. The classification accuracy is reduced by 12% for hybrid images compared to the Columbia dataset containing CG and real images. Please refer to figure 2 for the results. To resist such attacks and minimize the impact of such hybrid images on the performance of our systems, we propose a two-class classifier to differentiate between pure CG/real images and hybrid images. Based on the feature set that we selected in section 2, we test the performance of each set of features on detecting hybrid images and the results are shown in figure 2. Our results show that it is possible to model the traits of the hybrid images. The local patch statistics feature performs the best with 73% detection rate. The aggregate set of features performs with 75% accuracy. In our proposed framework, we use local patch statistics features for detecting hybrid images since it achieves similar performance to the aggregate feature set using a much smaller feature set (96 features in local path compared with 557 features in the aggregate set). The superior performance of local patch statistics features could be explained as follows: Local patch statistics model the joint distribution of patch structures in an image. Even though both CG and real patches are present in a hybrid image, the joint distribution of patch structures of CG and real images are different from the individual distributions of CG or real images. Hence hybrid images have properties which are unique when compared to CG or real images. This also explains the hybrid image detection rate of 73%, which is similar to the CG vs. real image classification rate of 69% for patch features. It is worthy mentioning that although the moment-based features have superior performance with 86% accuracy when differentiating CG from real images, their performance on detecting hybrid images is only 67%. This is due to the fact that the hybrid images contain high correlations due to the presence of real image patches. These correlations are measured by the prediction error component of moment based features. Therefore, for moment-based features, hybrid images share their properties with real images, resulting in poor detection performance.

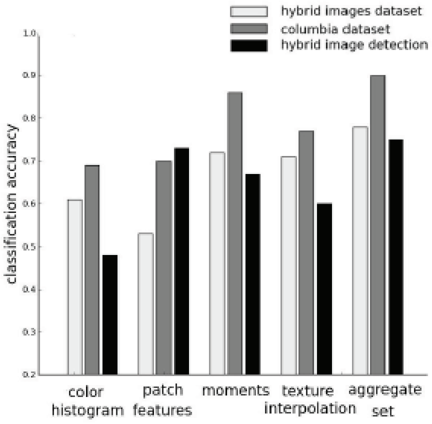

**Fig. 2**. The barchart illustrates the following results: Classification results for the hybrid image dataset with the classifier trained on the original Columbia dataset; Classification results on the original Columbia dataset; and Detection results for the hybrid images.

### 3.2. Histogram manipulation

Histogram manipulation is a commonly available technique in many image editing software to manipulate the brightness or contrast of an image. In the context of image forgery, an attacker would like to manipulate the histogram of CG/real image to falsify its authenticity as real image/CG. For example, given a CG and a real image we would like the CG image to have the histogram of that of the real image. Since our framework contains techniques like color histogram and moments of characteristic functions which are based on the histogram of the image, we expect them to be affected by this operation. We define the success of histogram manipulation attacks by the extent of decrease in the classification performance of our classifier. Also manipulations should not leave any perceptually noticeable traces of forgery. It should be noted that even though the forged image created has the histogram of the target image, wavelet transforms of level 1, 2, *etc*, of the target and forged images do not contain identical histograms. This is because the wavelets created depend on the spatial arrangement of pixels in an image. When modifying the pixel intensity values, histogram manipulation only considers the current pixel but not its neighbors. Since the visual contents are different between the source and target image pairs, the local arrangement of pixels are different. Therefore, histogram manipulations can only change the histogram of the forged image to that of the target image, but not the wavelet transforms of the forged image to the corresponding wavelet transforms of the target image. The visual quality of the histogram manipulated image depends on the target histogram chosen for manipulation. From the several choices present for the target image, we choose a target histogram which is similar to that of the source image. Earth movers distance (EMD) is used as a similarity metric. If the target histogram is dissimilar to that of the source image, the resulting histogram manipulation could possibly be identified on observation. In this paper, we use the YCbCr color space for manipulations. The Columbia dataset containing 1600 images is used for training our classification system. The testing is undertaken on 400 histogram manipulated images (200 real images with the histograms of CG, and 200 CG with the histograms of real images). The results are shown in Figure 3, and we can see that the classification performance of the aggregate set of feature drops by up to 30% under the histogram manipulation attacks. Classification accuracy for the histogram manipulated images and the original images are shown side-by-side for comparison. As expected features based on the histogram of the image such as color histogram features, the moment based features are greatly affected by this manipulation. The application of histogram manipulation leads to changes in the local pixel correlations present in an image. This property is used in identifying the histogram manipulated images. We could use the following set of features for identifying local pixel correlations in an image: prediction error, local patch statistics and interpolation features. The moment-based features use predic-

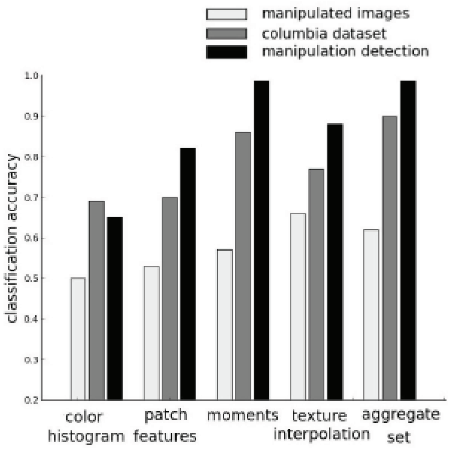

**Fig. 3**. The barchart illustrates the following results: Classification results for histogram manipulated image dataset with the classifier trained on original Columbia dataset. Classification results on the original Columbia dataset. Detection results for histogram manipulated images.

tion error to detect the high correlations present in the real images, and they are used here for identifying histogram manipulation images. Please refer to figure 3 where results of detecting histogram manipulations are shown. The moment-based features are successful in detecting histogram manipulations with an accuracy of 99%. We use the prediction error component of the moment-based features in our proposed framework for detecting histogram manipulated images.

## 4. THE PROPOSED FRAMEWORK

Based on the classification results and our knowledge of possible attacks against our classification, we propose a new framework for differentiating between CG and real images. Refer to figure 4 for an illustration of our framework. The framework consists of two parts, forgery detection system and CG vs real classification system. The purpose of the forgery detection system is to detect attempts by malicious entities, including histogram manipulation and hybrid image creation to manipulate the images used for classification. We develop this forgery detection system to detect and filter forged images created by these attempts. We use the hybrid image detector proposed in section 3.1 to identify hybrid images with an accuracy of 73%, and the histogram manipulation detection filter proposed in section 3.2 to detect histogram manipulation with an accuracy of 99%. Our framework offers the following benefits. It has a better classification accuracy than all the individual methods. It contains a forgery detection system, making it harder for an attacker to compromise. We need not have the same set of features for classification and forgery detection. Hence it is possible to optimize either system independently, without worrying about its effects on the other system.

## 5. CONCLUSION

In this paper we proposed and built a framework based on the aggregate set of features to achieve a classification rate of 90% on the Columbia dataset. We have improved on the results using the moment-based features suggested by Chen *et al.*, which provides a classification rate of 86% on the YCbCr color space. We analyze the resistance provided by our framework from possible attacks including histogram manipulation and creation of hybrid images. Both of these attacks are very successful and decrease the accuracy of our framework by around 30%. We are able to detect histogram manipulated images at 99% accuracy, and hybrid images with 73% accuracy. We use these features as filters in building a forgery detection system to prevent such manipulated images to be processed by our classification system.

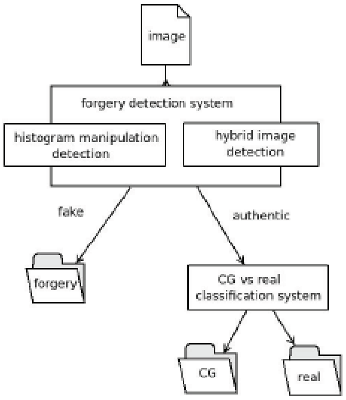

**Fig. 4**. Framework for classifying CG and real images with forgery detection filter.

## 6. REFERENCES

[1] W. Chen, Y.Q. Shi, and G. Xuan, *"Identifying Computer Graphics using HSV Color Model and Statistical Moments of Characteristic Functions"*, Multimedia and Expo, 2007 IEEE International Conference on, pp. 11231126, 2007.

[2] TT. Ng, SF. Chang, J. Hsu, L. Xie, and MP. Tsui, *"Physics motivated features for distinguishing photographic images and computer graphics"*, in MULTIMEDIA 05: Proceedings of the 13th annual ACM international conference on Multimedia, New York, NY, USA, 2005, pp. 239-248, ACM.

[3] J. Hsu T.-T Ng, S.-F. Chang and M. Pepeljugoski, *"Columbia photographic images and photorealistic computer graphics dataset"*, Tech. Rep. 205-2004-5, ADVENT, Columbia University, 2004.

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

[5] TI Ianeva, AP de Vries, and H. Rohrig, *"Detecting cartoons: a case study in automatic video-genre classification"*, Multimedia and Expo, 2003. ICME-03. Proceedings. 2003 International Conference, vol. 1, 2003.

[6] AC Popescu and H. Farid, *"Exposing digital forgeries by detecting traces of resampling"*, Signal Processing, IEEE Transactions, vol. 53, no. 2, pp. 758-767, 2005.

[7] MJ Weinberger, G. Seroussi, and G. Sapiro, *"The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS"*, Image Processing, IEEE Transactions on, vol. 9, no. 8, pp. 1309-1324, 2000.