2012 International Conference on Advanced Computer Science Applications and Technologies

Detecting Computer Generated Images for Image Spam Filtering

Zubaidah Muataz Hazza, Normaziah Abdul Aziz
Department of Computer Science
Kulliyyah of Information and Communication Technology
International Islamic University Malaysia, Malaysia.
zubyda_moutaz@yahoo.com, naa@iium.edu.my

Abstract— **Image spam continues to be one of cyber security problem today. Spammers used image spam as a technique to by-pass conventional email filters. Anti-Spammers used image classification as a method to detect images spam by extracting different features of the image. One of the important features used is color features. Several works used different color analysis to differentiate image spam, most of these works used supervised methods trying to differentiate computer generated images which is mostly like to be a spam and natural images. Supervised methods have its weaknesses, such as high cost in computation, requires training data, and rapid changes in spammers behaviors. This paper develops an unsupervised method using HSL geometric model (Hue, Saturation, and Luminance) to distinguish computer generated (CG) and natural images. Rules and Heuristics are defined by using HSL variables. The proposed method mainly depends on Saturation and Lightness values and their histograms. Experiment results shows that the combination of these variables can give high classification accuracy results.**

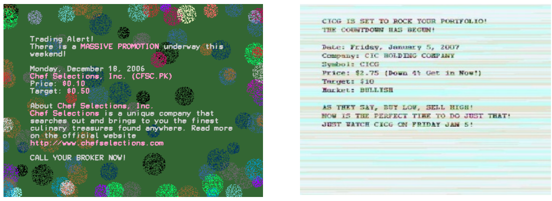
Keywords-HSL; image spam; lightness; saturation; normalized histogram;

I. INTRODUCTION

Image based spam is one of techniques used by spammers to by-pass the conventional email filters. It represents images that contain text messages of the spammers. Image based spam has then been further developed to embed clickable areas to redirect users to malicious websites.

The use of image spam by spammers has many advantageous to them; image spam has broken all text based filters, and they can send their messages in many designs to attract the users. The trend of image spam has developed from the first wave with simple text converted into image as shown in Fig.1 (a). The next trend is obfuscating the text in the image, to make it even harder for Optical Character Recognition (OCR) based filters to recognize the text, and today's trend is with design to make it appear legitimate as shown in Fig.1 (b).

On the other hand, the Anti-spammers started using the high level features such as email size, header analysis, image width and height, image types and so on, and low level features such as color and texture analysis for the image itself, to identify image spam.



Trading Alert!
There is a MASSIVE PROMOTION underway this weekend!

Monday, December 18, 2006
Chef Selections, Inc. (CFSC.PK)
Price: $0.10
Target: $0.50

About Chef Selections, Inc.
Chef Selections is a unique company that searches out and brings to you the finest culinary treasures found anywhere. Read more on the official website
http://www.chefselections.com

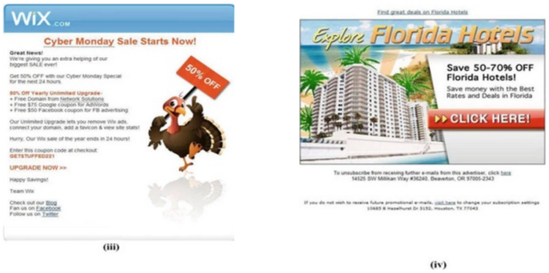
CALL YOUR BROKER NOW!

CICG IS SET TO ROCK YOUR PORTFOLIO!
THE COUNTDOWN HAS BEGUN!

Date: Friday, January 5, 2007
Company: CIC HOLDING COMPANY
Symbol: CICG
Price: $2.75 (Down 4% Get in Now!)
Target: $10
Market: BULLISH

AS THEY SAY, BUY LOW, SELL HIGH!
NOW IS THE PERFECT TIME TO DO JUST THAT!
JUST WATCH CICG ON FRIDAY JAN 5!

Figure 1(a). Early versions of image Spam



Cyber Monday Sale Starts Now!

Great News!
We're giving you an extra helping of our biggest SALE ever!

Get 50% OFF with our Cyber Monday Special for the next 24 hours.

50% Off Yearly Unlimited Upgrade-
- Free Domain from Network Solutions
- Free $75 Google coupon for AdWords
- Free $50 Facebook coupon for FB advertising

Our Unlimited Upgrade lets you remove Wix ads, connect your domain, add a favicon & view site stats!

Hurry, Our Wix sale of the year ends in 24 hours!

Enter this coupon code at checkout:
GETSTUFFED221

UPGRADE NOW >>

Happy Savings!

Team Wix

Check out our Blog
Fan us on Facebook
Follow us on Twitter

Find great deals on Florida Hotels

Save 50-70% OFF Florida Hotels!
Save money with the Best Rates and Deals in Florida

To unsubscribe from receiving further e-mails from this advertiser, click here
14525 SW Millikan Way #36240, Beaverton, OR 97005-2343

If you do not wish to receive future promotional e-mails, visit here to change your subscription settings
10685 B Hazelhurst Dr 3152, Houston, TX 77043

Figure 1(b). Later versions of image Spam

Color analysis is a main method used for image spam classification. Color analysis uses methods such as color histogram, color saturation, color coverage, and the number of colors. Previous works used common assumption that the image spam often are computer-generated graphics with specific properties [1] [2] [3]. Most of the previous works used supervised methods which depend on dataset in order to train the machine learning methods to get reasonable accuracy. Spammers nowadays generate new templates and keep producing new types or version of spam mails. Due to that, using unsupervised methods is required. In this paper the use of HSL geometric model is studied to analyze and differentiate computer generated from natural images. Then a model is developed that can detect computer generated images in high accuracy and efficient time processing. The results show that the method can be efficiently used and combined with other features to classify the current trend of image spam.

978-0-7695-4959-0/13 $25.00 © 2013 IEEE
DOI 10.1109/ACSAT.2012.38

The outline of the paper is as follows: section II will discuss the related works, section III on HSL model; section IV will discuss the proposed method; in section V the results and its discussion are presented; finally the conclusion and future work will be discussed in section VI.

II. Related Works

Different color analysis methods have been used and selected, such as in [1] the graphics are detected based on the assumption that computer-generated graphics usually contain homogeneous background and very little texture; wavelets was used to analysis the texture. In [2] color features are used such as colour saturation and heterogeneity, and then SVM was used for classification. In [3], the authors used colour variance, prevalent colour coverage, and the number of colours as well as colour saturation features.

In [4], the authors proposed four properties to be specific of image spam: color moment, color heterogeneity, conspicuousness, and self-similarity.

The features used in [5] were related to spatial information (pixel coordinates), colour and texture.

Image histograms were one of the main methods used to extract the color features, such as in [6], The Authors Proposed supervised detection method builds its training dataset based on two image features; colour and gradient orientation histograms, then probabilistic boosting tree (PBT) is used to distinguish spam images from ham images. In [7], the color properties selected were color saturation and color histogram, the authors relied on the assumption that legitimate images typically convey a much larger number of colors than spam images. In [8], the authors extracted gradient histogram as a key feature of spam images and applied BP neural network as classifier, while in [9] gray histograms are used to extract color features and SVM to classify, and in [10], the authors used the color histogram and the color moment.

III. HSL Model

HSL (Hue, Saturation, and Lightness) color model is based on intuitive color parameters, being derived from the RGB color cube. It is represented by a double hexagonal pyramid as can be seeing in Fig. (2) (Hearn et all, 1994). Hue (H) specifies an angle about the vertical axis of the pyramid, varying from 0_o, that corresponds to the red, to 360_o.

The parameter H possesses indefinite value for the gray scale which varies from black to white. Saturation (S) is measured along the horizontal radius of the pyramid and specifies the relative purity of the color. This parameter varies from 0 (gray scale) to 1 (pure colors). Lightness (L), measured along the vertical axis, possesses value 0 for black and 1 for white. It specifies the amount of light in the color.

Lightness is the variable has been used in our approach, under assumption that the computer generated objects will have higher lightness due to the sharpness of the objects.

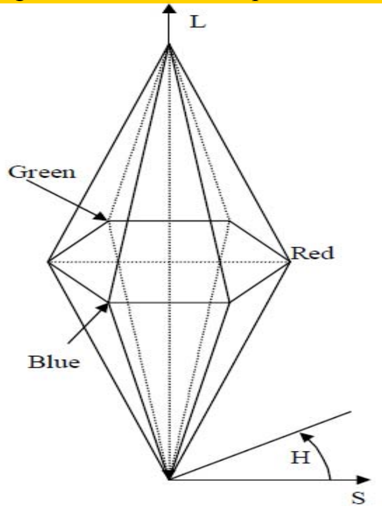


Figure 2. Representation of HSL Model

IV. Methodology

There is an argument that image spam are often artificial, and contain clearer and sharper objects than legitimate images; thus, their colour distribution should be less smooth [5]. Under assumption that in computer generated images have high lightness, especially the one used for image spam, due to the existence of light background to make it easy to read the message. From an initial analysis, it was found that in image spam, the mean value of lightness is greater than 0.5 to make the text in the image readable, and for natural images the mean value of the lightness is between 0.2 and 0.6. as shown in Fig.3 due to the smooth distribution of the natural colors.

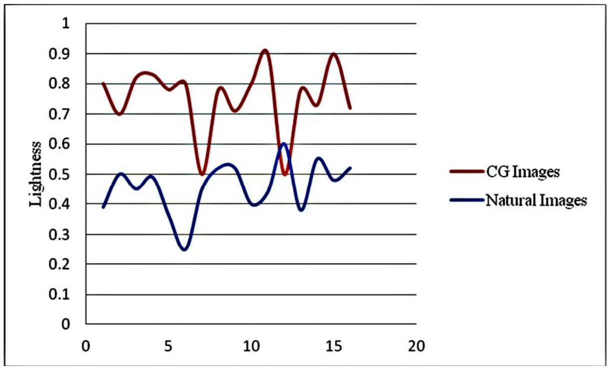


Figure 3. Example of Lightness values distribution

To classify the images, a general rule is defined which classified 83% of the dataset, including general computer generated images and spam images and define special rule to classify special cases in image spam.

Due to the sharpness of computer generated components and the lack of natural gradient as in the natural images, it was found that the normalized lightness histogram shows peak for the color of the background, while for natural images, the distribution of colors have close values, as in Fig. 4, based on that a threshold value is defined as 0.05, to differentiate and compare the components of the histograms.

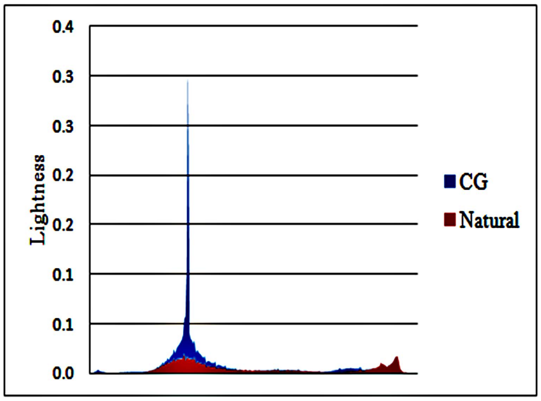
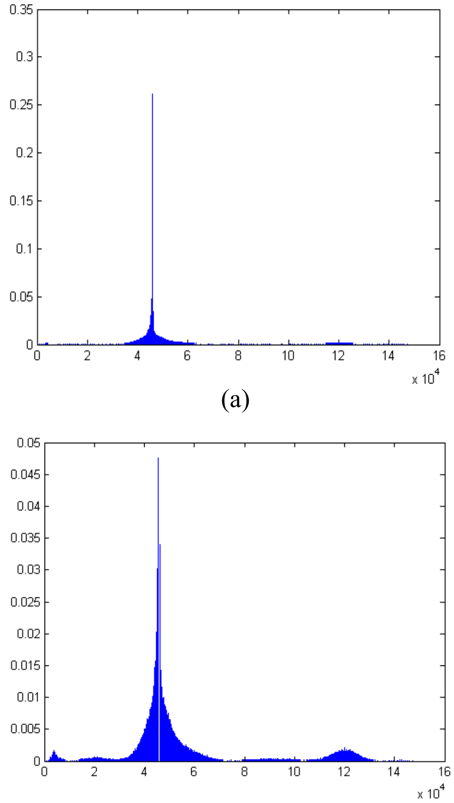


Figure 4. Lightness histogram for CG and natural image



(a)

(b)

Figure 5. Lightness histogram of CG image, a) before process, b) after process

The lightness histogram is processed, by removing any lightness value greater than 0.05. Fig. 5 shows the histogram of a computer generated image before and after the process.

The rate of change is computed by taking the maximum values before and after the process as follows:

$$\text{rate of change} = |\,(\max_a - \max_b)\,/\,\max_b\,| \times 100 \qquad (1)$$

Where $\max_a$ is the maximum value in the histogram after process and $\max_b$ is the maximum value in the histogram before the process. If the rate of change is greater than 10%, lead to a conclusion that there is a peak value.

In combination with the rate of change, another variable is used $\text{Final}_{\text{mean}}$, is the mean value of the following variables:

L_{mean}: The mean value of the lightness array
L_{max}: The maximum value of the normalized histogram of lightness
S_{mean}: The mean value of the Saturation array
S_{max}: The maximum value of the normalized histogram of saturation

These variables are selected by studying their response for CG and natural images. It was found that for CG images, the lightness and saturation has high mean values compared to natural images, the same is for the lightness histogram; the reason behind that is the existence of the sharp objects and the lack of the natural gradient that exist in natural images.

From Fig. 6, it is clear that CG images have higher $\text{Final}_{\text{mean}}$ than natural and that the range of natural images is between 0.05- 0.25.

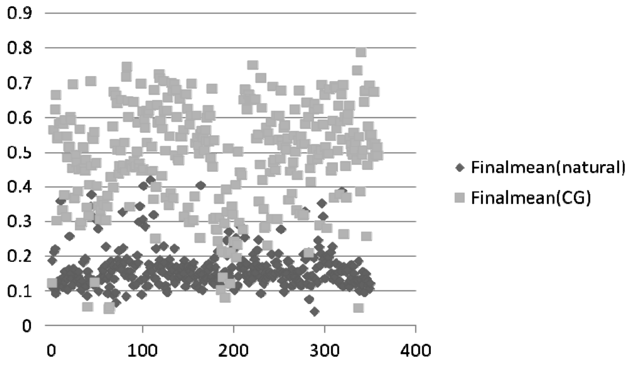


Figure 6. Distribution of $\text{final}_{\text{mean}}$

Based on the above, a general rule is defined as follows:

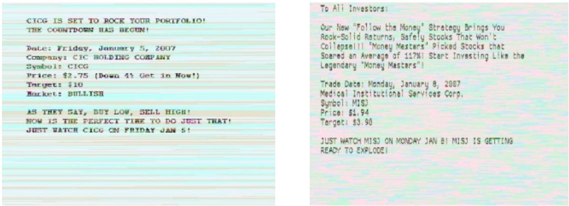
$$\text{General Rule} = \left\{ \begin{array}{l} \text{CG if Rate of change>10 \& Final}_{mean} > 0.25 \\ \text{Natural otherwise} \end{array} \right\} \qquad (2)$$

In the general rule, the image is classified as computer generated if satisfy the rate of change is greater than 10 and final$_{mean}$ is greater than 0.25. The next section will discuss the results and will present a special rule defined to handle special cases of computer generated images.

V. Results and Discussion

MATLAB 2009a was used for developing the method, the data set consists of (500) natural images which was taken from ICDAR 2003 and other resources from the Internet, and (350) computer generated images consists of image spam samples and general computer generated images.

For some image spam types with colorful gradually-changing background, Fig. 7, where it was misclassified and classified as natural in [6], our method had successfully classified them..



CICG IS SET TO ROCK YOUR PORTFOLIO!
THE COUNTDOWN HAS BEGUN!

Date: Friday, January 5, 2007
Company: CIC HOLDING COMPANY
Symbol: CICG
Price: $2.75 (Down 4% Get in Now!)
Target: $10
Market: BULLISH

AS THEY SAY, BUY LOW, SELL HIGH!
NOW IS THE PERFECT TIME TO DO JUST THAT!
JUST WATCH CICG ON FRIDAY JAN 5!

To All Investors:

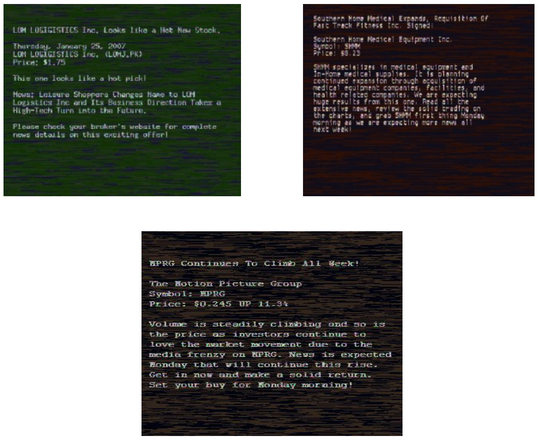
Our New "Follow the Money" Strategy Brings You
Rock-Solid Returns, Safely Stocks That Won't
Collapse!!! "Money Masters" Picked Stocks that
Soared an Average of 117%! Start Investing Like the
Legendary "Money Masters"!

Trade Date: Monday, January 8, 2007
Medical Institutional Services Corp.
Symbol: MISJ
Price: $1.94
Target: $3.90

JUST WATCH MISJ ON MONDAY JAN 8! MISJ IS GETTING
READY TO EXPLODE!

Figure 7. challenged CG images successfully classified

The general rule classified 83% of the computer generated images, and fails to classify old shapes of image spam, Fig.8; where all the HSL values are low and there is no peak due to darkness value in the image is high, and thus the rate of change will be zero.



LOM LOGIGISTICS Inc. Looks like a Hot New Stock.

Thursday, January 25, 2007
LOM LOGIGISTICS Inc. (LOMJ.PK)
Price: $1.75

This one looks like a hot pick!

News: Leisure Shoppers Changes Name to LOM
Logistics Inc and Its Business Direction Takes a
High-Tech Turn into the Future.

Please check your broker's website for complete
news details on this exciting offer!

Southern Home Medical Expands, Acquisition Of
Fast Track Fitness Inc. Signed!

Southern Home Medical Equipment Inc.
Symbol: SHMM
Price: $0.23

SHMM specializes in medical equipment and
In-Home medical supplies. It is planning
continued expansion through acquisition of
medical equipment companies, facilities, and
health related companies. We are expecting
huge results from this one. Read all the
extensive news, review the solid trading on
the charts, and grab SHMM first thing Monday
morning as we are expecting more news all
next week!

MPRG Continues To Climb All Week!

The Motion Picture Group
Symbol: MPRG
Price: $0.245 UP 11.3%

Volume is steadily climbing and so is
the price as investors continue to
love the market movement due to the
media frenzy on MPRG. News is expected
Monday that will continue this rise.
Get in now and make a solid return.
Set your buy for Monday morning!

Figure 8. Example of images successfully classified by special rule

To solve that another heuristic is defined, trying to filter out such type of images, by examining the values of natural images that have closed values with this type of image spam, it was found that the mean value of these images is within the natural images area, but these values can be separated by raisin the mean of the natural images based on S$_{mean}$ and L$_{mean.}$ A new variable SpecialScore is calculated based on these variables, and it was found that for these types of image spam, it will have SpecialScore value less than 0.2, as shown in Fig.9, the special score of natural images have risen leaving this type of image spam with low score, The special rule is defined as follows:

$$\text{Special Rule} = \left\{ \begin{array}{l} \text{CG if SpecialScore<0.2} \\ \text{Natural otherwise} \end{array} \right\} \qquad (3)$$

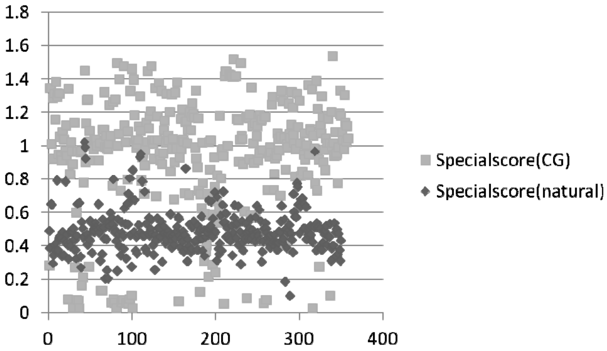


Figure 9. Distribution of SpecialScore

Two measurements were used to evaluate the performance of our method; True Positive Rate (TPR) and False Positive Rate (FPR) which are computed as follows:

$$\text{True Positive Rate} = \frac{\text{Correctly detected}}{\text{Total number of images}} \qquad (4)$$

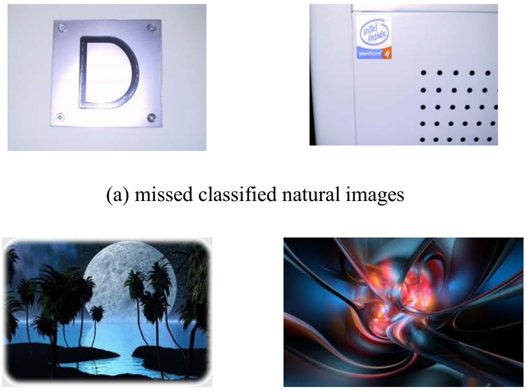
$$\text{False Positive Rate} = \frac{\text{Incorrectly detected}}{\text{Total number of images}} \qquad (5)$$

And average processing time was computed as 0.67 second/image; the results of the experiment are presented in Table 1.

TABLE I. PERFORMANCE EVALUATION

	Classified	**TPR**	**FPR**
CG images (350)	332/350	0.9486	0.0514
Natural images (500)	465/500	0.93	0.0700

In Fig. 10, Examples of incorrect classified images are shown.



(a) missed classified natural images

(b) missed classified CG Images

Figure 10. Examples of Incorrect classification

VI. CONCLUSION AND FUTURE WORK

Differentiating computer generated images from natural images is discussed in this paper; unsupervised algorithm is defined based on HSL geometric model. The algorithm mainly defined to help in image spam classification, where no training data is required and it processes the image in efficient time.

The results show that the algorithm performs well and give high classification rate even against challenged image spam.

For future work, further investigation will be done on the performance of the algorithm especially for the images failed to classify, and investigating possibilities to combine HSL with other geometric models to enhance the performance. For image spam filtering, the proposed method will be used in combination with additional features such as detecting the text features, and analyzing the header of the image spam,

REFERENCES

[1] C.-T. Wu, K.-T. Cheng, Q. Zhu, Y.-L. Wu, 2005. Using visual features for anti-spam filtering. In: Proc. IEEE Int. Conf. on Image Processing, Vol. III. pp. 501–504.

[2] H. Aradhye, G. Myers, J. A. Herson, 2005. Image analysis for e_cient categorization of image-based spam e-mail. In: Proc. Int. Conf. on Document Analysis and Recognition, pp. 914–918.

[3] Q. Liu, Z. Qin, H. Cheng, and M. Wan, Efficient Modeling of Spam Images. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, 2010, pp. 663-666.

[4] B. Byun, C. Lee, S. Webb, C. Pu, A discriminative classifier learning approach to image modeling and spam image identification, in: Proc. CEAS, 2007.

[5] B. Mehta, S. Nangia, M. Gupta, W. Nejdl, 2008. Detecting image spam using visual features and near duplicate detection. In: Proc. 17th Int. Conf. on World Wide Web. ACM, pp. 497–506.

[6] Y. Gao, M. Yang, X. Zhao, B. Pardo, Y.Wu, T. Pappas, and A. Choudhary, “Image spam hunter,” in Proc. 33th IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Las Vegas, NV, Apr. 2008.

[7] C. Xu, K. Chiew, Y. Chen, and J. Liu, 2012, Fusion of Text and Image Features: A New Approach to Image Spam Filtering, Practical Applications of Intelligent Systems, AISC 124, pp.129–140, Springer Berlin Heidelberg, 2012.

[8] M. Soranamageswair, C. Meena, A novel approach towards image spam classification, Int. J. of Computer Theory and Engineering, 2011: 3 (1), pp. 84 – 88.

[9] P. Li, H. Yan, G. Cui , Y. Du, Integration of Local and Global Features for Image Spam Filtering, Journal of Computational Information Systems 8: 2 (2012) 779–789.

[10] T. Liu, W. Tsao, C. Lee, 2010, A High Performance Image-Spam Filtering System, Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, 445-449, 2010

[11] C. Wang, F. Zhang, F. Li, Q. Liu, Image spam classification based on low-level image features, in: Proc. ICCCAS, 2010, pp. 290 – 293.