

Computer- Generated Image Identification Based on Image Gradient

Haiyu Li, Yanjun Sun

Department of Basic, Aviation University of Air force, Changchun, 130021, China

haiyu9871@tom.com

Keywords: computer- generated image;natural image; image identification

Abstract. In consideration about human eyes' incapability of computer- generated image identification, this thesis proposes an identification algorithm that uses image gradient in difference analysis between texture of computer generated images and natural images; image gradient is an efficient image detail presentation, which is feature that the algorithm takes advantage of for characteristic value collection and further categorization with SVM classifier. Test results indicate that the algorithm proposed in this thesis is capable of efficient identification on computer-generated image and real images.

Introduction

With rapid progress in development of digital cameras and smart phones, study on application and operation of digital image processing technology is showing more and more significance. Meanwhile, promotion in digital image alternation technology, as a result of development in this in this information era, creates different kinds of image processing software that is capable of remarkably vivid scene image generation, which causes trouble in human eyes' identifying factuality of such images and unprecedented challenge in image identification.

So far, image alternation has become a popular phenomenon. However, alarming events are happening while people benefiting from digital images. For example, Zhou Zhenglong claimed that he had found extinct south China tiger in 2007. The photo Zhou proposed caused a global shock. However, this south China tiger photo 'took' by Zhou was lately authenticated as a computer-generated photo created with a tiger on new year print and a forest scene image, as shown in Fig.1 and Fig.2. Fake photo events in international society are also familiar. In 2003, an experienced journalist with Los Angeles Times sent back a group of photos about a British soldier raising his hands to stop an Iraqi man who was holding a baby from going through the crowd. The photo was deemed as a classical news photo for its precise composition and vivid presentation of the tension, but was also lately authenticated as a computer- generated photo.

Rising amount of fake events is increasing people's doubt about factuality of digital photos, and methods for preserving factuality and completion of digital images have thus become an issue that worth a lot to academics, while evidence collection on factuality and completion of digital photos has also become a hot issue to the public.



Fig.1. Photo of the 'South China Tiger' Fig.2. New Year Print of South China Tiger

At present, both domestic and foreign scholars have conducted many researches on digital image evidence collection. For example, Fraid and his colleagues propose methods like high order statistic characteristic model, analogue sampling effect quantization of DCT- factor histogram, linear combination rate of neighboring pixel, and have realized splendid accomplishment in computer-generated image authentication, such as the geometrical feature based identification on computer- generated image and real images mentioned in bibliography [3]. Though commencement

of domestic researches on digital image evidence collection is a bit later than foreign studies, this issue is also causing more and more attention in domestic academic world since Shanghai Jiao Tong University, Dalian University of Technology and several other universities.

Starting from differences between computer-generated images and real images, this thesis uses calculated characteristic values of images to identify computer-generated and real images by classification. It is proved in experiments that the method proposed in this thesis is capable of efficient identification on computer-generated images and real images.

Differences between Computer-Generated Images and Real Images

Theory of digital image is as shown in Fig. 3, i.e. digital camera uses CFA interpolation for image generation while obtaining pixel value of three channels of R, G and B from natural light that goes through CCD sensor of the camera. Due to complexity of real images, the process of interpolation and hardware conditions, noise point is inevitable in real images; however, image generation on computer software are about images obtained from different kinds of image edition tools being consummated, sharpened, smoothened or even re-obtained by repeating interpolation and other processes with computer software, which causes different noise point mode of cameras took by digital photos and noise point mode in images generated by computer software. It is thus reasonable to conduct characteristic value collection on the basis of such difference between two kinds of images, and perform texture feature identification on natural images and computer-generated images.

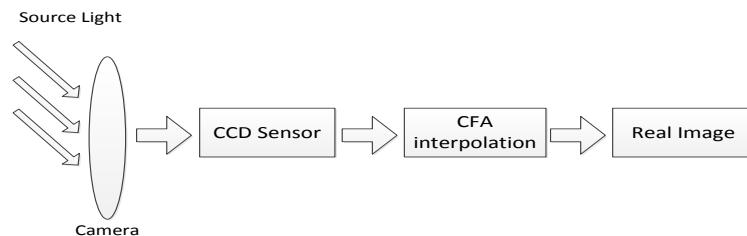


Fig.3. Theory of Digital Image

Algorithm of Image Characteristic Value Collection

This thesis uses image gradient in characteristic value collection. It is available for continuous function to calculate its change rate by calculating its first order partial derivation. Led by the same thought, it is also available to apply this first order partial derivation calculation on images, as shown in Formula 1.

$$\begin{cases} R_x(x, y) = f_x(x) \\ R_y(x, y) = f_y(x) \end{cases} \quad (1)$$

For images, first order partial derivation is an indicator of pixel value change on that direction. Relatively smooth areas of an image usually appear smaller derivation value, while marginal areas appear larger values. Larger derivation value also appears at areas with higher noise frequency. Such derivation calculation may appropriately point out marginal areas of an image.

In real situations, derivation of image and other free data may be indicated in gradient, as shown in Formula 2.

$$\begin{cases} R_x(x, y) = f(x+1, y) - f(x, y) \\ R_y(x, y) = f(x, y+1) - f(x, y) \end{cases} \quad (2)$$

Filtering models on X and Y, i.e. $f_x = (-1, 1)$ and $f_y = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ may be figured out with Formula 2.

In order to further emphasize edge information, Formula 2 may be transferred in gradient calculation, as shown in Formula 3.

$$\begin{cases} R_x(x, y) = f(x+1, y) - f(x-1, y) \\ R_y(x, y) = f(x, y+1) - f(x, y-1) \end{cases} \quad (3)$$

Formula 3 is actually a conversion that transfers filtering models on X and Y into $f_x = (-1, 0, 1)$

and $f_y = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$. Result of the conversion indicates that such method has very limited effect on

smooth areas in an image, while edge information emphasize effect of the method is very obvious. Since detail image information, as indicated in Section 1, is the most essential difference between computer-generated and real images, the method that Formula 3 implies may further strengthen such difference.

Due to possibility of minus derivation value in computer gradient calculation, dimension of the final characteristic vector may increase to a level that time cost of the algorithm becomes uncontrollable if all gradient values are involved in data collection. Moreover, gradient value calculation on X and Y axle is also necessary, which is unfavorable to final characteristic value calculation. As a settlement to this issue, a further improvement is thus made to the algorithm proposed in this thesis, namely calculating gradient amplitude with existing gradient on X and Y axle, as shown in Formula 4. Gradient amplitude is an effective solution to minus derivation value, and is thus involved in characteristic image calculation. As shown in Fig.4, (a) is a common gray level image, (b) is an image processed with gradient calculation, namely characteristic image.

$$grad = \sqrt{R_x^2 + R_y^2} \quad (4)$$

It is implied by Fig. 4 that gradient amplitude is an effective protection for edge information of an image, which makes gradient amplitude a perfect presentation for edge texture features. The step following gradient amplitude image, i.e. characteristic image, calculation is generation of the histogram about the calculated characteristic image and calculation about the final characteristic vector, as shown in Fig.5. It is indicated in the figure that amplitude of gradient in smoother areas are more intensive, while pixel value on edge areas is smaller.



(a) gray level image



(b) characteristic image

Fig.4. Comparison on Gradient Amplitude

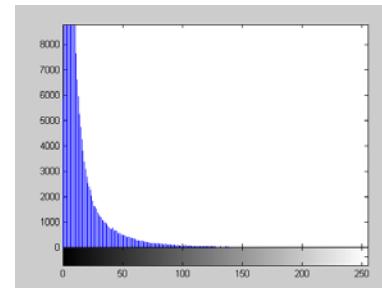


Fig.5. Gradient Histogram

Algorithm Processes

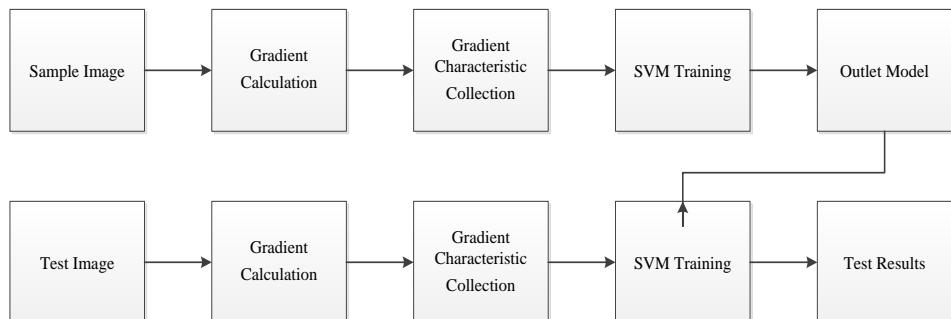


Fig.6. Processes of Algorithm

Essence of this algorithm is to conduct computer-generated image authentication with gradient

characteristics of an image. Detailed steps are as follows: divide all test samples into two groups of training and testing, then collect gradient characteristics of sample images, and finally use SVM classifier to test images after training samples with SVM classifier.

Experiment Results and Analysis

The experiment was performed on a computer with AMD Athlon(tm)7750, 2.7GHz CPU, 1GB RAM, programming environment of Microsoft Visual 2010, programming language of C Language, and classifier of SVM. 800 natural images are from digital image library of Columbia University, 800 computer-generated images are from several foreign professional 3D image websites.

Abundant experiments that had been performed for proving validity of this algorithm finally appear expected output, as shown in Fig. 1. All 400 training samples and 200 testing samples are collected by random selection, rate between training images and testing images is limited to 2:1.

Table 1 Experiment Result Comparison

Algorithm	Accuracy
Proposed Algorithm	82.50%
Bibliography 4	72.25%

Characteristic vector of this algorithm is directly extracted from statistical histogram, which indicates that dimension of the vector is 256d. It is reflected by experiment results that accuracy of this algorithm fits real conditions. This algorithm is of higher accuracy on images with more complex texture information, but is of insufficient accuracy on images with fewer texture images.

Meanwhile, accuracy of this algorithm on image identification may also be affected by effect of CCD sensor and CFA interpolation to imaging process of real images (especially those scenes with strong vibration). However, experiment results indicate that the algorithm is of good accuracy on computer-generated image identification.

Summary and Expectations

Due to different imaging process, texture of computer-generated images and real images is of obvious difference. In consideration of unique advantage that image gradient has in image texture complexity description, this thesis focuses on value collection towards characteristics of computer-generated and real images, uses gradient amplitude to further figure out characteristic vector from characteristic image histogram, and finally conduct classification with SVM classifier. Experiment results prove that algorithm proposed in this thesis is of effectiveness in computer-generated image identification and good potential in real applications.

References

- [1] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. Pattern Analysis and Machine Intelligence, 2002,24(7) 971-987.
- [2] Farid H,LyuS.Higher-order wavelet statistics and their application to digital forensics.IEEE Workshop on Statistical Analysis in Computer Vision.
- [3] Tian-Tsong Ng,Shih-Fu Chang. Using geometry invariants for camera response function estimation[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.
- [4] Danhong Yao,Buo Su,Shenghong Li. Computer-generated image detection methods based on fractal dimension [J]. communication technology.2008,41 (12)256-258.
- [5] X. Gao1, B. Qiu, J. J. Shen, T. -T. Ng and Y. Q. Shi.Asmart phone image database for single image recapture detection[C]//the 9th International Conference on Digital Watermarking, 2010, 90-104.