Knowles, H. D., Winne, D. A., Canagarajah, C. N., & Bull, D. R. (2003). Attack characterisation of compressed images. In IEEE International Conference on Consumer Electronics (ICCE 2003) Los Angeles, CA, USA. (pp. 66 - 67). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICCE.2003.1218810

TUPM 6.8

ATTACK CHARACTERISATION OF COMPRESSED IMAGES

H. D. Knowles, D. A. Winne, C. N. Canagarajah and D. R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol, Bristol, BS8 1UB, UK

ABSTRACT

We propose the use of robust watermarks to enable the characterisation of attacks even after lossy compression, such as JPEG and JPEG2000. A previously constructed Bayesian framework is used to allow characterisation of attacks from a predetermined library, and the double watermarking technique as earlier proposed by the authors is employed to generate the features used to drive the classifier. The results show that the developped techniques perform well for both types of compression.

INTRODUCTION

With the proliferation of powerful home computers, the use of digital media is ever increasing. In addition to the copyright issues this raises, of equal importance is that of the authentication of said media. Perhaps the most palpable use is that of evidence authentication, for example in a court of law. There are a plethora of additional uses, however. Consider the case of a photographic agency: a paper will not intentionally download and publish an image they know to have been altered, but what is to stop the photographer tampering with the image to improve its commercial potential? There are many systems that provide authentication of uncompressed images [1, 2, 3]. However, it is frequently impractical to download and store say a 5 megapixel image, and thus compression is often used. We therefore propose a system that is capable of determining which of a library of attacks has occurred, and can operate on images compressed using either JPEG or JPEG2000.

DOUBLE WATERMARKING

In order to be able to characterise the attack that has taken place, it is self-evident that some part of the watermark must remain after the attack. Indeed, in order that a wide variety of attacks with ranging severity are to be classifiable, it is desirable that the watermark will degrade relatively slowly, and have some presence even after the most severe of attacks. For this reason a robust watermark is embedded with a masking function to ensure the maximum possible watermark energy is inserted, without the watermark becoming visible.

We build on our earlier work in [4] and use the double watermarking process described therein. The Bayesian framework constructed in [5] is used to enable characterisation of the attacks to take place. The data may be modelled as either normal [6] or skew-normal (SN) [7]. There are a number of reasons for choosing a parametric model. Firstly, parametric techniques do not suffer from sparsity issues in the same way that non-parametric techniques do, nor do they require large multi-dimensional histograms to be stored, a potentially inhibitive feature for example for image authentication for police officers in the field. The final desirable property of the Gaussian distribution is that evaluation of the full density function is not necessary, thus computational requirements are reduced. We extend existing work to consider the case where the images have been compressed using lossless compression, and compare two such systems – JPEG and JPEG2000.

THE SKEW NORMAL DISTRIBUTION

The SN distribution is described in detail in [7]. The advantage of the skew-normal over the normal distribution from a classification perspective is that it has an extra degree of freedom, thus potentially enabling more accurate fitting of the data to the model. Examples of how α, the skew parameter, affects the distribution are given in Figure 1.

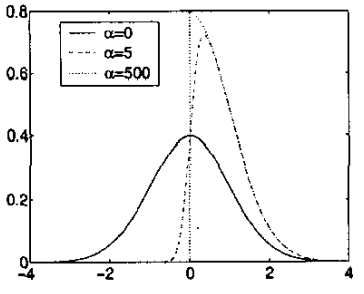

Fig. 1. Variations in Skew-Normal density with skew (α)

0-7803-7721-4/03 $17.00 © 2003 IEEE

RESULTS

The images were first watermarked, compressed and then attacked. After the training process, which in this case involves estimating the mean and the covariance matrix (& skew for the SN distribution), the performance of the classifier was evaluated. The first set of experiments used JPEG compression with a Quality Factor of 95, whilst the second used JPEG2000 compression with a compression ratio of 5:1. Both types of compression yield similar PSNR for Lena.

For the Gaussian classifier, previous experiments on uncompressed images gave an overall misclassification rate of 5.7% [6] e.g. averaged over all attacks, 5.7% of the time the classifier produces an erroneous decision . For JPEG, the error rate dropped to 5.0%, whilst for the JPEG2000 case it increased only slightly to 7.4%. For the SN classifier, the uncompressed error rate is 4.3%, which also drops for the JPEG compressed case, this time to 3.2%. It rises to 6.4% in the case where JPEG2000 compression has been applied. Therefore clearly the SN classifier outperforms the Gaussian classifier, as would be expected on account of the closer fit of the model to the data afforded by the skew term.

Results in Figure 2 give some preliminary results showing how the system is able to localise attacks. By comparing Figures 2(a) and 2(b) we can see that the averaging attack is correctly identified. The location to which an unsharp mask has been applied is also correctly identified (Figure 2(c)). However, there are some areas where false positives occur (see Figure 2(d)). By comparison with Figure 2(a), it can be seen that these errors occur in textured and edge regions, which have an increased similarity with regions which have been unsharp masked. It is hoped that future work will reduce this percentage of false positives.

Further discussion as to the cause of the change in error rates will be presented, along with consideration as to why differing regions have different probabilities of misclassification and potential solutions. We also include a study examining how the assumed Gaussianity/skew-normality differs from the true data, and discuss how this will affect the performance of the system. We also consider methods for reducing these discrepancies.

CONCLUSIONS

In the proposed paper we show that by using a double watermarking strategy, characterisation and localisation of any attack from a predetermined attack is possible even after compression with low probabilities of misclassification. The improved fit between the data and the model for the skew-normal distribution yields a lower misclassification rate than for the Gaussian case, with the trade-off of increased computational complexity.

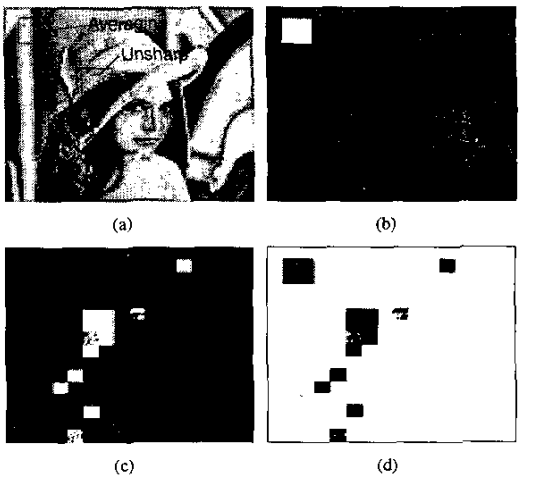

Fig. 2. (a) Lena after attack. Areas (white) for Normal classifier where (b) averaging with 5-by-5 filter, (c) unsharp mask, and (d) nothing, are the MAP estimate of the attack that has taken place. JPEG compression (QF=95) precedes the attack.

REFERENCES

[1] D. A. Winne, H. Knowles, D. R. Bull, and C. N. Canagarajah, "Digital watermarking in wavelet domain with predistortion for authenticity verification and localization," in *Security and Watermarking of Multimedia Contents IV*. SPIE, January 2002, vol. 4675.

[2] J. Fridich, "Image watermarking for tamper detection," in *Proceedings ICIP-98 (IEEE International Conference on Image Processing)*, October 1998.

[3] M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," in *Proceedings ICIP-97 (IEEE International Conference on Image Processing)*, 1997.

[4] Henry Knowles, Dominique Winne, Nishan Canagarajah, and Dave Bull, "Towards Tamper Detection and Classification with Robust Watermarks," 2003, To appear in *ISCAS'03*.

[5] Henry Knowles, Dominique Winne, Nishan Canagarajah, and Dave Bull, "A Bayesian Approach to Attack Characterisation using Robust Watermarks," 2002, To appear in *VCIP'03*.

[6] Henry Knowles, Dominique Winne, Nishan Canagarajah, and Dave Bull, "Reduced Complexity Attack Characterisation using Discriminant Functions for the Gaussian Distribution," 2002, To appear in *VIE'03*.

[7] A. Azzalini and A. Capitanio, "Statistical Applications of the Multivariate Skew-normal Distribution," *J. Roy. Statist. Soc.*, vol. 61, no. 3, pp. 579–602, 1999.