AUTHOR GUIDELINES FOR ICIP 2017 PROCEEDINGS MANUSCRIPTS

Noé Le Philippe, William Puech

School A-B
Department A-B
Address A-B

Vincent Itier

School C-D
Department C-D
Address C-D

ABSTRACT

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words, and should be identical to the abstract text submitted electronically along with the paper cover sheet. All manuscripts must be in English, printed in black ink.

***Index Terms*—** One, two, three, four, five

1. INTRODUCTION

More and more content is being shared everyday on the internet. Some of it needs to be securely transfered, the problem of encryption arises. Full encryption with methods such as AES for exemple are often not needed in addition to not being possible due to computing power constraints. Instead, partial or selective encryption is used, where the goal is sufficient encryption.

When an image is intented to be consumed by a human, the most accurate measure of its confidentiality is a Mean Opinion Score (MOS), where actual people rate the image. It is however not a realistic way to rate the distortion of an image as it is way to expensive and time consuming, security and quality metrics were introduced as a means to automate the process.

Image quality assessment is divided in two main fields, no reference image quality assessment (NR-IQA), which refers to cases where only the processed image is available, with no extra information, and full reference image quality assessment (FR-IQA), where both the processed and the original image are available. In this paper, we focus on FR-IQA and we quality metrics as security metrics, since as explained in Section 3, security is achieved through low quality. The PSNR is the most well known metric, but has been shown not to be well correlated with the human visual system (HVS), especially on low quality images. The SSIM [1], even if better correletated with the HVS, is not consistent across all image qualities. Similar metrics [2–5] exhibit the same deficiencies, either on low or high quality images, as shown in [6], there is not yet a security metric that consistently rates images across all the MOS spectrum. Most quality metrics fail to predict a MOS on low quality images, precisely where it would be most important to do so: decide whether or not an image is confidential.

The most popular image compression standard is JPEG [7]. In order to exploit both efficient compression and encryption, format compliant methods are designed to produce content compatible with format specifications. There exists several format compliant JPEG encryption methods which can be used in this context. Partial encryption methods using sign encryption have been shown unsecure by Said [8]. Partial encryption can be applied selectively on automaticaly detected faces [9]. This method which relies on XOR operation with the AES algorithm, performs the compression and the encryption in the same process. Partial encryption is sufficient to hide sensitive information, such as text [10]. Moreover, it has the advantage to not change the size of the encrypted file. A reversible watermarking method in encrypted domain has been proposed by Qian *et al.* [11]. This method relies on XOR operation but for more visual masking author encrypt also quantization table. Blocks and coefficients scrambling is used in [12–15]. Simple scrambling methods tend to increase the size if there is no verification of the run-length for example. Inter-block shuffle and non-zeros AC scrambles methods have been shown insecure to sketch attack by Li and Yan [16].

In this paper, we present a new metric...

Section 2 presents the dataset we used and how it was created. Then, in Section 3 we discuss the evaluation and rating of its images, by human (MOS) as well as by security metrics. We thus introduce a new metric for image evaluation based on the visual saliency in Section 4. Finally we conclude and open a few perspectives in Section 5.

2. CREATION AND UTILIZATION OF THE DATASET

The cryptocompression method we used is targeted towards JPEG images. We have six parameters that we can enable or not to generate cryptocompressed images. *Shuffle* and *XOR*

are the parameters that decide the actual encryption method. *AC* and *DC* control which part of the DCT coefficients is encrypted and two additionnal parameters, *chrominance* and *luminance* decide which of the luminance, chrominance (or both) DCT coefficients is encrypted. As there must be at least one encryption method, at least one type of coefficient, and chrominance or luminance selected, we have selected a total of 27 distortions by combining these parameters. The distortion ranges from completely indecipherable images to almost invisible perturbations, as shown in Fig. 1. This way, we hope to have appropriate distortions for different use cases.

The *XOR* parameter corresponds to the method proposed by Puech *et al.* [17]. This method partially encrypts an image. This can be useful for partial visualisation, even if we only use it on the whole image, and it has two main strengths: it does not increase the size of the JPEG bitstream and it changes the DCT coefficients histogram. We encrypt the amplitude part of non null AC coefficients *i.e.* the concatenation of the amplitude of each coefficient of each block $[A_0^i, ..., A_k^i, ..., A_0^n, ..., A_k^n]$, where n is the number of blocks. The amplitude sequence is denoted $A = [a_0, ..., a_l]$ where l is the number of amplitude bits. A standard stream cipher function is used to generate a pseudo-random sequence $E = [e_0, ..., e_l]$ from a secret key. This sequence is XORed with the incoming plaintext to produce a ciphered sequence $\widetilde{A} = \widetilde{a}_0, ..., \widetilde{a}_l$ where $\widetilde{a}_i = a_i \oplus e_i, i \in [0, l]$. The encrypted sequence is substitued to the amplitudes in the original bitstream.

The *shuffle* parameter corresponds to a full inter-block shuffle (FIBS), proposed by Li and Yuan [16]. This method can scramble DC coefficients as well as same frequency AC coefficients. As it scrambles all coefficients, run length encoding does not perform as well and the size of the image can increase. According to the authors, the use of all AC coefficients, zero as well as non-zero, creates a more secure image, less sensitive to jigsaw puzzle attacks [].

We used the training images from the BSDS500 [18] dataset as our input images for a total of $27 \times 200 = 5400$ cryptocompressed images.

3. IMAGE EVALUATION

We conducted our evaluation on *N* different people. They had to give a score from 1 to 5 to the images, where 5 is the best score and 1 is the worst:

1 : The distortion is unbearable, nothing is visible
2 : The distortion is very annoying, I can barely guess the content
3 : The distortion is annoying, but I can see the content
4 : The distortion is slightly annoying, but the content is clear
5 : The distortion is not annoying at all

An example of the 5 MOS is illustrated Fig. 1. They had to rate 81 images, three for each distortion. The sessions were 10 to 15 minutes long, depending on the person. Each image has been seen at most once by each user, to prevent them from recognising it and give it a higher score. The distortions order was shuffled differently for each evaluation and repeated three times in the same order.

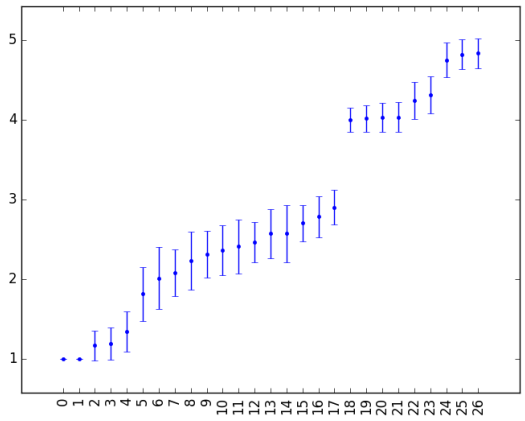


Fig. 2: MOS for the 27 distortions.

The images were evaluated in a dim room, on a *details of the screen* screen, about n meters away and around eyes level. The user could only see one image at a time, a new image shown once the previous had been rated. The MOS obtained during the evaluation are given Fig. 2. We can see that after distortions $\#17$ there is a large gap in the MOS. This is due to the absence of the parameter $luminance$, the $shuffle$ and XOR are only performed on the chrominance, hence the better ratings. We give an overview of a few selected metrics we used for image analysis. For a more in-depth review, we refer the reader to [6].

PSNR: Even though it is known that the PSNR is not well correlated with human judgment, it is still widely used due to its speed and ease of use. The range is [0; $+\infty$], where two identical images would have a PSNR of $+\infty$.

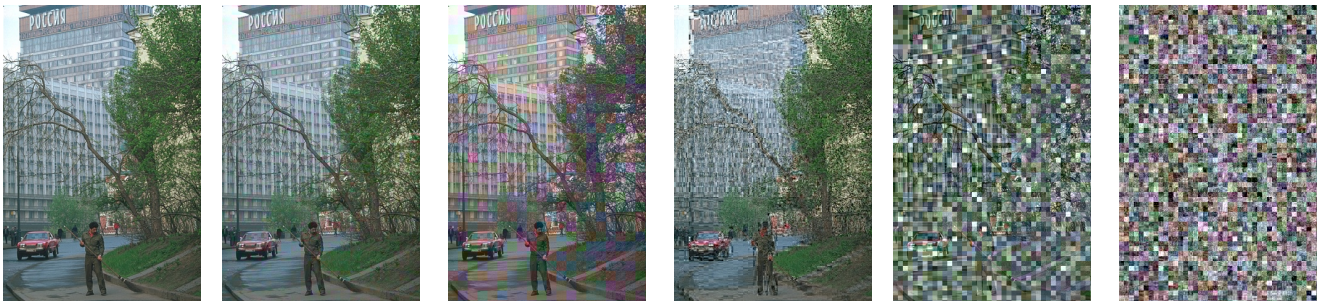
SSIM [1]: (Structural Similarity Index Measure). A luminance score, a contrast score and a structure score are combined to obtain the actual SSIM score. It has a range of [0;1] where identical images have a score of 1.

ESS [19]: (Edge Similarity Score). It uses non overlapping 8x8 block directions to compare images. With the range [0;1], a higher score reflects a less distorted image.

LSS [19]: (Luminance Similarity Score). It uses non overlapping 8x8 block average luminance to compare images. With the range [-8.5; 1] for default parameters of $\alpha = 0.1$ and $\beta = 3$, a higher score reflects a less distorted image.

NPCR [20,21]: It is the number of pixel changes between images. Its range is [0;100], where a fully encrypted image has a NPCR close to 100, where almost all the pixels changed.

UACI [20, 21]: It is the unified averaged changed intensity. It is the average intensity difference between two images.

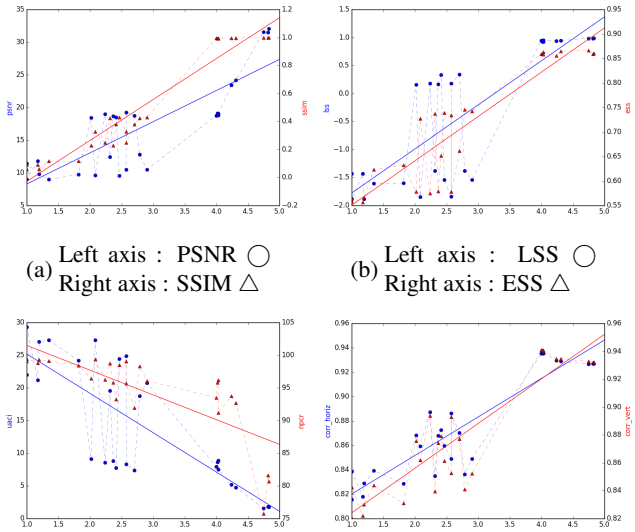


(a) Original image (b) MOS $\simeq 5$ (c) MOS $\simeq 4$ (d) MOS $\simeq 3$ (e) MOS $\simeq 2$ (f) MOS $\simeq 1$

Fig. 1: Example of images for different selective encryption methods with their corresponding MOS

Its range is [0;100], where a fully encrypted image has a value close to 33.

Our goal is to predict the rating a human would give to an image. In the best case scenario, a metric would be totally correlated with human rating and could be used to completely replace humans in image evaluation, this is however not the case, at least not for the metrics we selected, as shown in Fig. 3.



(a) Left axis : PSNR ◯ Right axis : SSIM △ (b) Left axis : LSS ◯ Right axis : ESS △

(c) Left axis : UACI ◯ Right axis : NPCR △ (d) Left axis : horizontal ◯ Right axis : vertical △ correlation

Fig. 3: Plots of different metrics with the MOS on the x-axis, 3a: PSNR and SSIM, 3b: LSS and ESS, 3c: UACI and NPCR, 3d: horizontal and vertical correlation

As we can se from these figures, most metrics actually follow a rough line, but distortions 5 to 17 are problematic and prevent us from predicting the MOS. These distortions also happen to be between a MOS of 2 and 3, where the threshold for a confidential image would be. Even the SSIM, which is the most accurate metric in our experiment, fails to predict the MOS.

4. NEW METRIC

In this section, we present the new metric we designed, the results we obtained and their analysis. Our metric is based on the visual saliency in images, and more specifically, and saliency map, a grayscale image where less salient pixels are darker than more salient pixels in the original image. The visual saliency is interesting in our case for image quality assessment because we want to know whether the meaning of the content of an image is accessible. According to [22], important information is located in salient areas. Our idea is that if salient areas are consistent in the original image and in the processed image, the content is readily available.

Let M_o be the saliency map of the original image and M_p be the saliency map of the processed image. A threshold is applied to M_o and M_p to only keep the most salient areas of each image, the best threshold has been experimentally found (Fig **??**) to be 5% more salient areas. Two binary images are thus created, B_o from M_o and B_p from M_p. A first value is computed as such:

$$v = \frac{\sum_{i=0}^{width} \sum_{j=0}^{height} f(B_o(i,j), B_p(i,j))}{sum(B_o)},$$

where $f(.,.)$ is:

$$f(x,y) = \begin{cases} 1, & \textbf{if } x = 1 \textbf{ and } y = 1 \\ 0 & Otherwise \end{cases}$$

and $sum(B_o)$ is actually the number of pixels of value 1.

The main problem with the visual saliency is that it does not perform well on images with a global noise, such as a XOR on the DC coefficients of the luminance channel for example (Fig **??**). In an attempt to address this problem, we introduce a second value, computed using the sobel edge detector [].

5. CONCLUSION

6. REFERENCES

[1] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[2] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[3] Ye Yao, Zhengquan Xu, and Jing Sun, "Visual security assessment for cipher-images based on neighborhood similarity," *Informatica*, vol. 33, no. 1, 2009.

[4] Lingling Tong, Feng Dai, Yongdong Zhang, and Jintao Li, "Visual security evaluation for video encryption," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 835–838.

[5] Jing Sun, Zhengquan Xu, Jin Liu, and Ye Yao, "An objective visual security assessment for cipher-images based on local entropy," *Multimedia Tools and Applications*, vol. 53, no. 1, pp. 75–95, 2011.

[6] Heinz Hofbauer and Andreas Uhl, "Identifying deficits of visual security metrics for images," *Signal Processing: Image Communication*, vol. 46, pp. 60–75, 2016.

[7] Gregory K Wallace, "The jpeg still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[8] Amir Said, "Measuring the strength of partial encryption schemes," in *IEEE International Conference on Image Processing 2005*. IEEE, 2005, vol. 2, pp. II–1126.

[9] José M Rodrigues, William Puech, and Adrian G Bors, "Selective encryption of human skin in jpeg images," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 1981–1984.

[10] Michael Pinto, William Puech, and Gérard Subsol, "Protection of jpeg compressed e-comics by selective encryption," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 4588–4592.

[11] Zhenxing Qian, Xinpeng Zhang, and Shuozhong Wang, "Reversible data hiding in encrypted jpeg bitstream," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1486–1491, 2014.

[12] Xiam Niu, Chongqing Zhou, Jianghua Ding, and Bian Yang, "Jpeg encryption with file size preservation," in *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on*. IEEE, 2008, pp. 308–311.

[13] Andreas Unterweger and Andreas Uhl, "Length-preserving bit-stream-based jpeg encryption," in *Proceedings of the on Multimedia and security*. ACM, 2012, pp. 85–90.

[14] Kazuki Minemura, Zahra Moayed, KokSheik Wong, Xiaojun Qi, and Kiyoshi Tanaka, "Jpeg image scrambling without expansion in bitstream size," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 261–264.

[15] SimYing Ong, KokSheik Wong, Xiaojun Qi, and Kiyoshi Tanaka, "Beyond format-compliant encryption for jpeg image," *Signal Processing: Image Communication*, vol. 31, pp. 47–60, 2015.

[16] Weihai Li and Yuan Yuan, "A leak and its remedy in jpeg image encryption," *International Journal of Computer Mathematics*, vol. 84, no. 9, pp. 1367–1378, 2007.

[17] Adrian G. Bors William Puech and José Marconi Rodrigues, *Advanced Color Image Processing and Analysis*, Springer Science & Business Media, 2013.

[18] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[19] Yinian Mao and Min Wu, "Security evaluation for communication-friendly encryption of multimedia," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*. IEEE, 2004, vol. 1, pp. 569–572.

[20] Guanrong Chen, Yaobin Mao, and Charles K Chui, "A symmetric image encryption scheme based on 3d chaotic cat maps," *Chaos, Solitons & Fractals*, vol. 21, no. 3, pp. 749–761, 2004.

[21] Yaobin Mao, Guanrong Chen, and Shiguo Lian, "A novel fast image encryption scheme based on 3d chaotic baker maps," *International Journal of Bifurcation and Chaos*, vol. 14, no. 10, pp. 3613–3624, 2004.

[22] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.