VISUAL SALIENCY-BASED CONFIDENTIALITY METRIC FOR SELECTIVE CRYPTO-COMPRESSED JPEG IMAGES

Noé Le Philippe, Vincent Itier, William Puech

LIRMM, UMR 5506 CNRS, University of Montpellier
860 rue de St Priest, Bat. 5, 34095 Montpellier, FRANCE

ABSTRACT

For security reasons, more and more digital data are transferred or stored in encrypted domains. In particular for images, selective format-compliant JPEG encryption methods have been proposed for the last ten years. Since encryption is selective, in order to reduce the processing time and to be format-compliant, it is now necessary to evaluate the confidentiality of these selective crypto-compressed JPEG images. It is known that image quality metrics, such as PSNR or SSIM, give a very low correlation with a mean opinion score (MOS) for low quality images. In this paper, we propose an efficient confidentiality metric based on the visual saliency diffusion. We show experimentally that this metric is well correlated with a MOS and efficient to evaluate the confidentiality of selective crypto-compressed JPEG images.

Index Terms— JPEG, confidentiality metric, visual saliency, encryption

1. INTRODUCTION

With the increasing use of social networks, more and more images are diffused, shared and exchanged on the internet. Image quality assessment is divided into two main fields, no reference image quality assessment (NR-IQA), which refers to situations 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 images are available.

The PSNR is the most well known metric, but it has been shown to not interact very well with the human visual system (HVS), especially with low quality images. Although the SSIM [1] interacts better with the HVS, it is not consistent across the whole range of image quality. Similar metrics [2–5] exhibit the same deficiencies, either on low or high quality images, as shown in [6]. There has not yet been devised a confidentiality 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 useful to do so: decide whether or not an image is confidential. Confidentiality means that a HVS cannot understand the image content.

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. Selective encryption methods using sign encryption have been shown insecure by Said [8]. Selective encryption can be applied on specific areas such as faces in order to protect privacy [9]. This method which relies on XOR operation with the AES algorithm, performs compression and encryption in the same process. Selective encryption is sufficient to hide sensitive information, such as text [10]. Moreover, it has the advantage that it does 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, authors also encrypt the quantization table. Block and coefficient scrambling is used in [12–15]. Simple scrambling methods tend to increase the size of the file if there is no verification of the run-length for example. Inter-block shuffle and non-zeros AC scramble methods have been shown insecure to sketch attack by Li and Yan [16].

This paper presents a full reference perceptual metric based on visual saliency which assesses image confidentiality after encryption or alteration for example.

Section 2 presents the dataset we have created. Then, in Section 3, we discuss the evaluation and rating of the images, by human (MOS) as well as by standard image quality metrics. In Section 4, we present the proposed confidentiality metric based on visual saliency. Finally, we conclude and open a few perspectives in Section 5.

2. DATASET CREATION AND UTILIZATION

The crypto-compression method we used is targeted towards JPEG images. We have six parameters that we can decide to use or not to generate selective crypto-compressed images. *FIBS* and *SJCC* are the two used encryption methods. *AC* and *DC* control which part of the DCT coefficients is encrypted and two additional parameters, *chrominance* and *luminance* decide which of the luminance or chrominance (or both) DCT coefficients are encrypted. There must be at least one encryp-

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

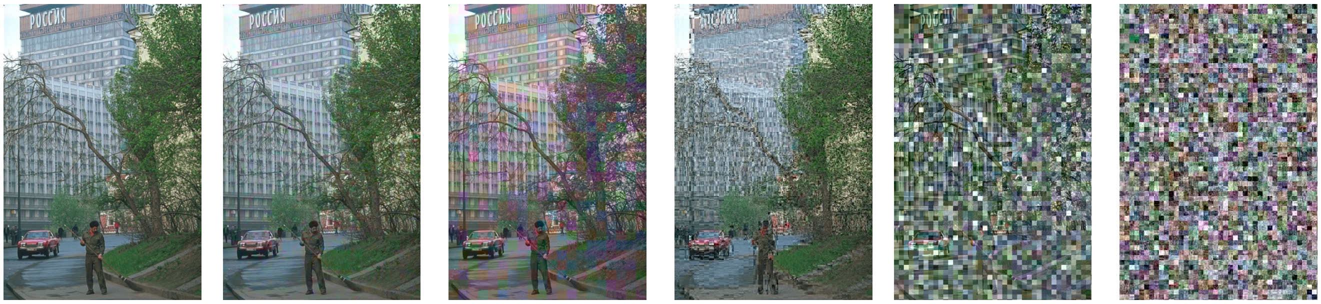

(a) Original image (b) MOS $\simeq 5$ (c) MOS $\simeq 4$ (d) MOS $\simeq 3$ (e) MOS $\simeq 2$ (f) MOS $\simeq 1$

Fig. 1: An example of images using different selective crypto-compression methods with their corresponding MOS.

tion method, at least one type of coefficient, and chrominance or luminance selected for the image to be affected. This results in 27 relevant combinations obtained by combining these parameters. The distortions range from completely indecipherable images to almost invisible transformations, as shown in Fig.1. From this, we can generate distortions for different situations as well as a wide range of distortions to show that different transformations result in the same level of degradation.

The *SJCC* corresponds to the selective JPEG crypto-compression method proposed by Puech *et al.* [17]. This can be useful for low resolution visualization 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 substituted to the amplitudes in the original bitstream.

The *FIBS* corresponds to a full inter-block shuffle (FIBS), proposed by Li and Yuan [16]. This method scrambles 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 200 images from the BSDS500 [18] dataset as our input images for a total of $27 \times 200 = 5400$ crypto-compressed images.[1]

3. IMAGE EVALUATION

A mean opinion score is an arithmetic mean of ratings given by humans for a particular stimulus. It is a single number, generally from 1 to 5, used to describe the quality of the current stimulus, where 5 is the best score and 1 is the worst. We conducted our evaluation on 41 different people, male and female from 17 to 53 years old, using the following scale:

1: The distortion is unbearable, nothing is visible
2: The distortion is very annoying, I can barely make-out 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

A score of 1 corresponds to a fully confidential image, where no information about the content is available while a score of 5 corresponds to an image with no apparent distortion.

An example of the five values for this MOS is illustrated in Fig. 1. The participants had to rate 81 images picked randomly from the proposed database, three for each distortion. The sessions were 10 to 15 minutes long, depending on the person. The distortion order was shuffled differently for each evaluation.

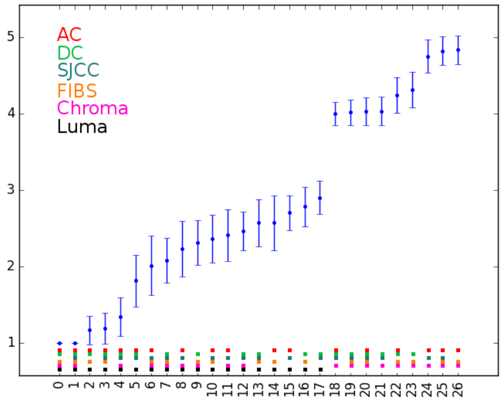


Fig. 2: MOS for the 27 distortions.

The images were evaluated in a dim room, on a 3840 × 2160 pixel, 75 inch Sony monitor, the person evaluating the images was about 2.5 meters away and the height of the monitor was around eye level. The user could only see one image

[1] www.lirmm.fr/∼wpuech/dataset

at a time, a new image was shown once the previous had been rated. The MOS obtained during the evaluation are given in Fig. 2 for each distortion. We can see that after distortion $\#17$ there is a large gap in the MOS. This is due to the absence of the parameter $luminance$, the $FIBS$ and $SJCC$ 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 does not interact well 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, a contrast 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]: 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 have been changed.

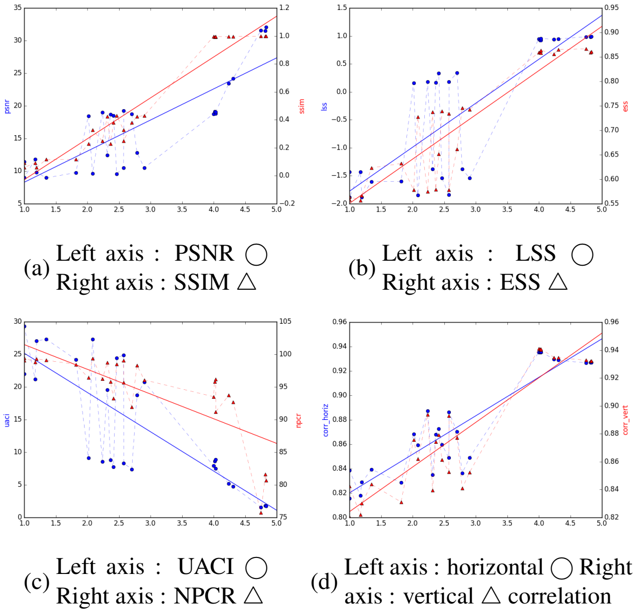
UACI [20, 21]: Is the unified averaged changed intensity. It is the average intensity difference between two images. Its range is [0;100], where a fully encrypted image has a value close to 33.

Our goal is to predict which rating a human would give to a selective crypto-compressed 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 however is not the case, as shown in Fig. 3.

As we can see from Fig. 3, 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. VISUAL SALIENCY-BASED CONFIDENTIALITY METRIC

In this section, we present the new metric we designed, the results we obtained and our analysis of them. Our metric is based on visual saliency in images and a saliency map. This idea is defined as image utility in [22] for JPEG compression improvement. 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 understandable. According to [23], important information is located in salient areas. Our reasoning is twofold: if salient areas are consistent in both the original and the processed image, the same amount of information is present in the image and the content is readily available. On the contrary, if no salient areas can be found, then the content is hidden. We try to compute to which extent the visual saliency of two images are similar to extract a score.



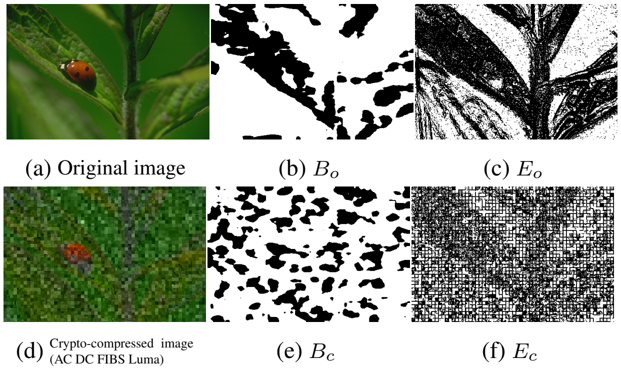
(a) Left axis : PSNR ◯ Right axis : SSIM △

(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, a: PSNR and SSIM, b: LSS and ESS, c: UACI and NPCR, d: horizontal and vertical correlation.


(a) Original image

(b) B_o

(c) E_o

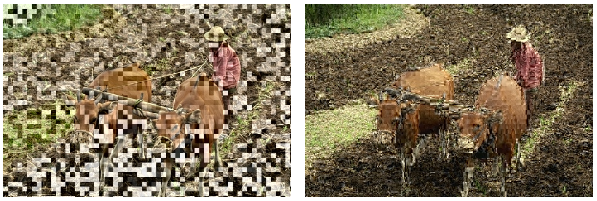
(d) Crypto-compressed image (AC DC FIBS Luma)

(e) B_c

(f) E_c

Fig. 4: Maps obtained for an original image a), a crypto-compressed image d), b) and e) saliency maps, c) and f) edge maps.

Let M_o be the saliency map of the original image and M_c be the saliency map of the crypto-compressed image. A threshold is applied to M_o and M_c to only keep the most salient areas of each image, the best threshold has been experimentally found to be 15% more salient areas, it is the point with the highest correlation with the MOS. Two binary im-



(a) $SJCC$ on the luma DC coefficients

(b) $SJCC$ on the luma AC coefficients

Fig. 5: Global noise caused by a $SJCC$ on DCT coefficients.

ages are thus created, B_o from M_o (Fig. 4b) and B_c from M_c (Fig. 4e). The score based on the visual saliency is then:

$$v_{saliency} = \frac{\sum_{i=0}^{width} \sum_{j=0}^{height} B_o(i,j) \times B_c(i,j)}{\sum_{i=0}^{width} \sum_{j=0}^{height} B_o(i,j)}, \qquad (1)$$

where $B_o(i,j) \times B_c(i,j)$ is equal to 1 when both pixels are equal to 1 and $\sum_{i=0}^{width} \sum_{j=0}^{height} B_o(i,j)$ is the number of pixels of value 1 in the original image.

The proposed metric gives good results for either high or low quality images, where it is already able to predict the MOS. The results however are not as good for mid quality images, when the MOS is around 2 and 3. Because of this, we can only tell that an image is either fully confidential or not confidential at all, but not to which degree it is confidential. This is because we are not able to compute a meaningful saliency map on images with a global, patterned noise, such as a $SJCC$ on the DC coefficients of the luminance channel for example (Fig. 5a). Another problem we encountered is that visual saliency performs too well on distorted images where the global structure is intact but the fine grained details are not available, typically when an $SJCC$ is applied to the AC coefficients of the luminance channel (Fig. 5b).

Because of these two types of distortions, and their variations, using only visual saliency is not a realistic approach for an image confidentiality metric. We introduce a second score, v_{edges}, based on the Sobel operator [24] in an attempt to stabilize our first score $v_{saliency}$.

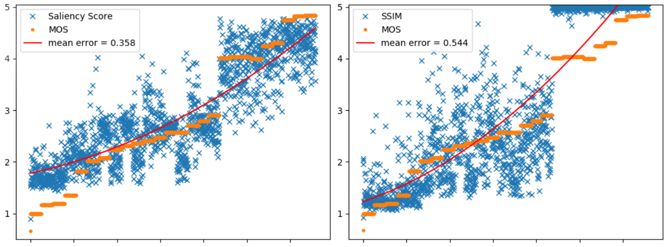
Distortions such as Fig. 5a do not hinder the edge detection, making it a good candidate to balance visual saliency defects. Our second score is computed the same way as our first one: two maps S_o and S_c are computed using the Sobel operator for the original and the crypto-compressed image. A threshold is then applied to S_o and S_c to only keep the strongest edges, thus creating two other bitmaps, E_o and E_c (Fig. 4c and Fig. 4f), the threshold has been experimentally found to be 10% of the edges. Our second score is then obtained using Eqn. (1) with the images obtained from the Sobel operator. The final score is:

$$v = \alpha * v_{saliency} + (1 - \alpha) * v_{edges},$$

where the weight α was experimentally determined to be 0.6.

The results we obtained after running our metric on our dataset are available in Fig. 6a, with Fig. 6b, the SSIM, for comparison since the SSIM presents better results comparing to other metrics. It shows that our metric takes into account the gradation of the distortions and produces no large gaps between values. The euclidean distance of our metric to the MOS of every distortion is 0.4323 for our metric and 0.5095 for SSIM and the euclidean distance of our metric to individual image rating is 0.6674 and 0.6699 for SSIM. With our metric, it is then possible to classify images as confidential or not as a function of the estimated MOS.

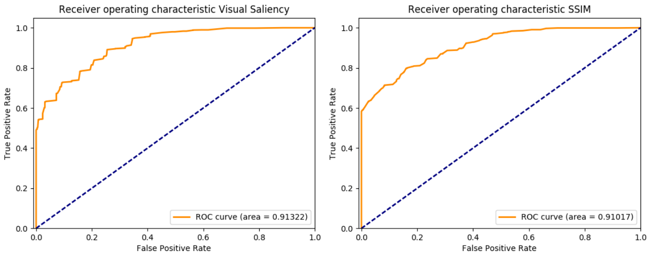
The threshold for a confidential image would be between a MOS of 2 and 3. We used our metric as a classifier to try and decide whether images are confidential, the results are available on the ROC curve illustrated in Fig. 7.



(a) Scores obtained with our metric, sorted on the image distortion MOS

(b) Scores obtained with SSIM, sorted on the image distortion MOS

Fig. 6: Comparison of results for our metric and SSIM.



(a)

(b)

Fig. 7: ROC curves of: a) Our metric, b) SSIM.

5. CONCLUSION

In this paper we proposed a dataset composed of selective crypto-compressed images for image quality assessment. The images were rated by human observers to obtain a mean opinion score. We also introduced a new confidentiality metric based on visual saliency. Our dataset was used as a benchmark to evaluate our metric and we noticed that we obtained better results compared to quality metrics such as SSIM.

Future work would include a more in depth analysis of several different datasets and each of its 27 distortions, as well as a more refined metric based on visual saliency, which shows great potential. We think it might be possible to generalize this metric to other applications, not necessarily only for selectively encrypted JPEG images.

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] Edward T Scott and Sheila S Hemami, “Image utility estimation using difference-of-gaussian scale space,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 101–105.

[23] 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.

[24] Irvin Sobel, “An isotropic 3× 3 image gradient operator,” *Machine Vision for three-demensional Sciences*, 1990.