Minimum image quality assessment based on saliency maps: a human visual approach

João Barreira[ab], Maximino Bessa[ab], Luís Magalhães[ab]
[a]Universidade de Trás-os-Montes e Alto Douro, Apartado 1013, 5001-801 Vila Real, Portugal;
[b]INESC TEC (formerly INESC Porto), Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200 - 465 Porto, Portugal

ABSTRACT

Image quality assessment as perceived by humans is of crucial importance in numerous fields of image processing. Transmission and storage of digital media require efficient methods to reduce the large number of bits to store an image, while maintaining sufficiently high quality compared to the original image. Since subjective evaluations cannot be performed in various scenarios, it is necessary to have objective metrics that predict image quality consistent with human perception. However, objective metrics that considers high levels of the human visual system are still limited. In this paper, we investigate the possibility of automatically predict, based on saliency maps, the minimum image quality threshold from which humans can perceive the elements on a compressed image. We conducted a series of experimental subjective tests where human observers have been exposed to compressed images with decreasing compression rates. To measure the difference between the saliency maps of the compressed and the original image it was used the normalized absolute error metric. Our results indicate that the elements on the image are only perceived by most of the human subjects not at a specific compressed image quality level, but depending on a saliency map difference threshold.

Keywords: Image quality assessment, saliency map, full-reference metric, human visual system.

1. INTRODUCTION

The perceptual quality of an image plays a crucial role in most image processing applications. Image quality assessment (IQA) methods can be used to: monitor the image quality of a video acquisition system; benchmark image processing algorithms; and optimize the performance and the parameter settings in image processing systems [1]. E.g. IQA can be used to systematically evaluate the performance of different image compression algorithms that attempt to minimize the number of bits required to store an image, while maintaining sufficiently high image quality [2].

In the past few years, many objective and subjective IQA methods have been proposed. Objective methods [3] aim to automatically measure the quality of an image based on quantitative parameters typically obtained from either reference or distorted images characteristics. On the other hand, subjective methods result are directly given by humans through the mean opinion score (MOS) resulting from subjective tests [4]. As objective IQA methods do not reflect how a human may see the real image, subjective methods have been used as the most reliable way to evaluate the quality of an image for applications in which human observers are the ultimate receivers. However, subjective evaluations are typically very time-consuming and expensive. In addition, they cannot be easily and routinely performed in various scenarios (e.g. real-time systems). Therefore, there has not been presented so far any precise model that accurately matches the subjective quality and can be easily implemented into various IQ systems.

The aim of the work presented in this paper is to investigate the possibility of developing, based on saliency maps, quantitative measures that can be used to automatically predict the minimum quality of an image to humans perceive the elements on it. In order to achieve this goal, human observers were exposed to images with different compression rates and asked to identify the item presented on the image. The differences between each of the saliency maps of the compressed images and the saliency map of the reference one were computed by using a normalized absolute error (NAE) metric. The results indicate, therefore, that it is possible to provide a saliency map difference threshold after which human subjects can identify the elements on the image. The paper is organized as follows: Firstly, section 2 reviews the background and presents related work on IQA. Then, section 3 describes the procedure and results of experimental subjective tests performed with human observers. Finally, section 4 concludes the paper.

Image Quality and System Performance X, edited by Peter D. Burns, Sophie Triantaphillidou,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8653, 86530T • © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 • doi: 10.1117/12.2005380

2. BACKGROUND AND RELATED WORK

Objective IQA methods aims to provide a computational model that can automatically predict the difference between images [3]. These methods can be classified as full-reference (FR) or no-reference (NR). In full-reference methods the quality of a distorted image is evaluated by comparing it with a reference (original) image that is assumed to have perfect quality, while in no-reference methods the original image is not required to assess the image. Evaluation metrics that requires both information about the distorted image and partial information about the original are designated as reduced-reference (RR) [4]. As NR image quality assessment is a very difficult task, FR metrics that make use of distorted and original images characteristics are the main used techniques in the literature. Among them, traditional IQA methods, such as *Mean Square Error* (MSE) and *Peak Signal to Noise Ratio* (PSNR), have been a simple and popular used metric to evaluate image-processing algorithms. However, these metrics reveal weak performance, as they do not consider subjective evaluations ratings [5].

Since humans are in most cases the ultimate receivers of the image, recent IQA methods based on some characteristics of the human visual system (HVS) have been proposed. The goal of these methods is to evaluate the quality of an image as perceived by humans. HVS-based IQA methods recently developed include bottom-up models (JND [6], VDP [7]), spatial colour (S-CIELAB [8]), structural (SSIM [3], MSSIM [9]), mathematical (fuzzy [10]), and information-theoretic (VIF [11], IFC [12]). The prominent among them are SSIM (*Structural Similarity*) and VIF (*Visual Information Fidelity*). SSIM takes the idea that the HVS is highly sensitive to structural distortion of nature images, so the quality of an image is based on the degradation as perceived change in structural information. On the other hand, VIF considers IQA as an information-fidelity problem rather than a signal fidelity problem and quantifies how much information in the reference image can be extracted, to the human observer, from the distorted image. To evaluate the image quality, VIF uses *Natural Scene Statistics* (NSS) and model natural images in the wavelet domain [13]. These methods successfully model into IQA metrics low aspects of the HVS, such as luminance masking, contrast and orientation sensitivity, frequency selectivity, and texture masking. However, none of them consider an important feature of human vision, which is visual attention. This is due to the lack of fast and precise methods to model visual attention in real-time, and also due to the fact that visual attention mechanism is not yet fully understood for IQA.

Few recent works that include visual attention mechanism into objective metrics are based on the assumption that a distortion occurring in certain regions on an image may be visually more important than in others. Thus, they attempt to weight image quality evaluations according to saliency importance. In [14]-[16] different visual attention models were implemented into IQA methods. The performance gain resulting from these works are however diverging; in consequence there are still numerous apprehensions when including visual attention into objective metrics. The existing computational attention models such as presented in [17] and [18] are too complex to be implemented in IQA. Furthermore, these models are not specially designed for a specific metric, and therefore, not necessarily generally applicable. In addition, works combining saliency and minimum image quality in a perceptually meaningful way are still limited, and hardly discuss a generalized model for combining visual distortion assessment and saliency. This implies that before including visual attention mechanisms into image processing applications, it is necessary to exactly know whether and what to extent to improve the quality score of an objective metric.

3. EXPERIMENTAL TESTS

3.1 Approach

Visual attention is the cognitive process that corresponds to the ability to select and process only the most relevant regions of a visual scene, while the remainder are left relatively unprocessed [17]. A common method to identify and visualize these regions is the saliency map proposed by Koch and Ullman [17], which is a topographically map that represents visual saliency of a corresponding visual scene. The areas containing pixels with high intensity on the saliency map represent regions that can attract more attention.

The approach that is proposed on this paper aims at investigating the minimum image quality threshold from which humans can perceive the elements on an image. The minimum image quality threshold is defined based on saliency maps differences. Figure 1 presents the overview of our approach. Given an original (reference) image and a set of compressed images, in the first step are computed the saliency maps of these images. The second step calculates the normalized absolute error (NAE) between each saliency maps of the compressed images in relation to the reference. In our approach, this metric is used as a measure of quality.

During the experiments 7 photos were used. All of them were uniformly degraded using JPEG compression algorithm. The quality of the compression was intentionally varied for values of: 1, 3, 5, 8, 12, 36, and 100 %, where 100% means no compression (reference image) (e.g. 496Kb), and 1% means that the image contains 99% less information than the uncompressed reference image (e.g. 4.93Kb). The compressed quality of the images was varied to produce images at a wide range of quality, from imperceptible levels to high levels of impairment.

Then, from all the compressed images we have computed their saliency map. Among the various computational models available in the literature [18]-[20] we chose the standard Itti-Koch saliency map [18]. After that, we have calculated the normalized absolute error between the saliency map of the reference image and each of the saliency maps of the compressed, using equation (1). Figure 2 shows an example of an image used on this study with increasing image quality rates. Figure 3 shows the saliency maps of figure 2 images. It appears that with increasing image quality rates the pixels with high values on the saliency map also increase.

$$NAE = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N}\left|x_{j,k} - x'_{j,k}\right|}{\sum_{j=1}^{M}\sum_{k=1}^{N}\left|x_{j,k}\right|} \qquad (1)$$

3.2 Procedure

The 7 original input images used on the experiments consist of: 5 photos of day-to-day objects on a white background and 2 photos of day-to-day objects in an outdoor environment. All the images were captured with a camera at a resolution of 640x480 pixels and saved as PNG files. The images on a white background include: a key, two red markers, a staple, a pen, and an image with multiple objects: a pen, a key and a mobile phone. The images in an outdoor scene include: a garbage bin and a car.

The experiments with human observers occur individually and as follow. The subject sat at approximately 60 cm in front of a calibrated monitor. The images were shown to each participant in a random order in a closed room with normal indoor illumination levels. The image with the worst compressed image quality (i.e. 1%) was the first to be presented. The subject was asked to identify the item presented on the image. If s/he could not identify it, the same image was displayed but with a continuous increase on the image quality (i.e. 3, 5, ..., 100%). As soon as the subject could correctly identify the item on the image the experiment stopped. The subjects had a limited time to identify the element on each compressed image (approximately 15 seconds) and were asked to only identify the item on the image as soon as they were sure about the answer. A total of 50 non expert subjects took part in the subjective tests and all had normal or corrected to normal vision. Furthermore, previously to the experimental tests, each of them was verbally introduced to the experiment.

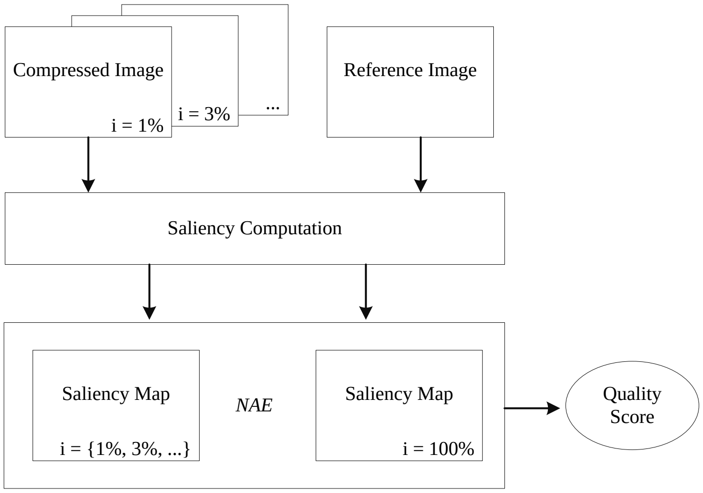

Figure 1. Overview of our approach. The normalized absolute error (NAE) of the saliency map of the compressed image in relation to the reference is the measure of quality.

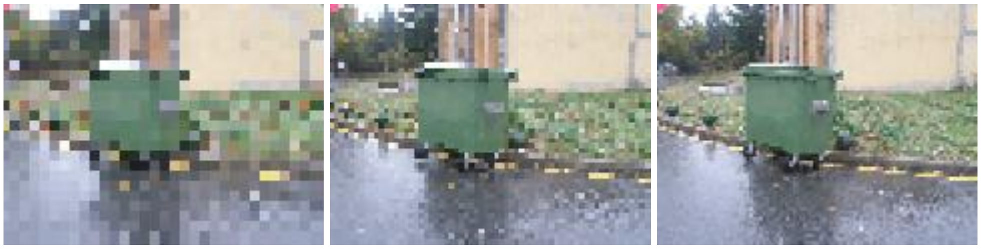

Figure 2. Compressed images used on the study with increasing quality rates (1%, 3%, and 5%).

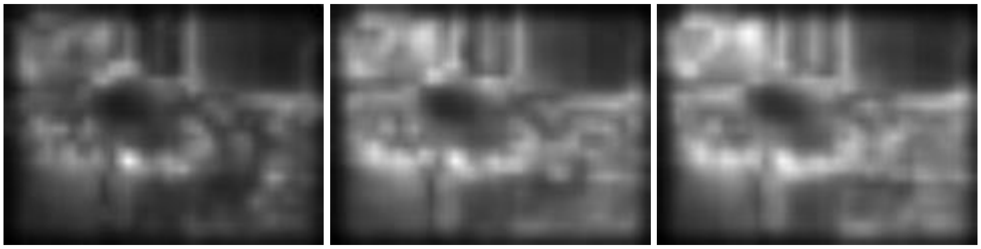

Figure 3. Saliency maps of figure 2 images.

3.3 Results and discussion

Table 1 show results from the experiments conducted with human observers. In this table is summarized the number of subjects who identify the items on the compressed images at the various quality rates. Table 2 presents the normalized absolute error between the saliency maps of the original (ideal) image and each of the saliency maps of the compressed ones. E.g. a result of 0.08 in the table 2 means that the saliency map at this quality rate (column 1) is similar to the original saliency map of that image in 92%.

The main conclusions were derived from the two tables. It appears that with an increasing on the image quality rates the salient regions within the image are also more visible. Consequently, it decreases the difference among the saliency maps. In table 2 the results with * refers to the worst image quality values whereof the human subjects participating on the study start to indentify the item on the compressed image. Comparing table 2 results with the subjective experimental tests (table 1) we can notice a threshold of 0.15 in the NAE after which the subjects could identify the elements on the image. We observe that the subjects identified the element on image 1 and 4 after the minimum threshold of 0.15. Concerning the element on image 2 it was identified by the majority of the subjects at the quality rate of 3% (NAE = 0.1370). However, one of the 50 subjects was able to identify the item on the image at the compressed image rate of 1%. According to our results, this may happen because at this quality rate the salient threshold is 0.1473, which corresponds to the borderline of the minimum saliency threshold to perceive the elements on the compressed image. This suggests that at the saliency threshold limit of 0.15 is possible to understand the elements on the image, however closer to this value more difficult it is. In image 3 the numbers of subjects who identify the items on the compressed images were split among the various quality rates. A possible explanation for these results is the fact that for the quality rates of {*i*= 3%, 5%, 8%} the saliency maps are very similar among them, which make that there is not a clearly NAE value that goes beyond the minimum saliency threshold of 0.15. At last, images 5, 6, and 7 present outstanding results in which more than 40 subjects start to identify the elements on the compressed images according to the quality rate after the saliency threshold is exceed. These results indicate, therefore, that the elements on the image only start to be perceived by human subjects not at a specific compressed image quality, but depending on the saliency threshold. That is, when:

$$NAE\_min\ (salmap(iComp),\ salmap(iRef)) \leq 15\%\ . \qquad (2)$$

We noticed that some subjects identify the elements on the image as soon as this threshold value is exceeded, while others only identify the elements on the image at higher quality rates. This may be due to the subject familiarity with the object on the image. The effects of familiarity on the recognition of objects in the brain are yet very difficult to determine [21] and are not within the scope of this paper. Nevertheless, observing the two tables the results also suggest that with NAE values inferior to 0.10, the majority of the subjects (more than 50%) identify the elements on all the compressed images. Accordingly, these results indicate that when the NAE between the saliency map of the compressed image and the reference one is equal or less than 10%, the elements on the image can be easily identified by most of human observers (3).

$$NAE\_max\ (salmap(iComp), salmap(iRef)) \leq 10\% \,. \qquad (3)$$

Table 1. Number of subjects who identified the item(s) presented on the image at different image quality rates.

Quality rates	Image1	Image2	Image3	Image4	Image5	Image6	Image7
1%	-	1	-	-	-	48	-
3%	14	22	20	12	40	2	43
5%	34	11	15	20	5	-	6
8%	2	10	7	15	3	-	1
12%	-	5	5	2	1	-	-
36%	-	1	3	1	1	-	-
100%	-	-	-	-	-	-	-

Table 2. Normalized absolute error between the saliency map of the reference image and each of the saliency maps of the compressed images computed from equation (1). The results with * refers to the worst image quality values when the subjects start to identify the item on the compressed.

Quality rates	Image1	Image2	Image3	Image4	Image5	Image6	Image7
1%	0.1752	0.1473*	0.1851	0.1715	0.1739	0.1275*	0.3469
3%	0.1370*	0.1096	0.1323*	0.1137*	0.0951*	0.0817	0.1075*
5%	0.1233	0.1084	0.1440	0.1073	0.1100	0.0705	0.0621
8%	0.1086	0.0906	0.1401	0.0839	0.0928	0.0675	0.0530
12%	0.0810	0.0723	0.1115	0.0739	0.0866	0.0584	0.0377
36%	0.0214	0.0297	0.0730	0.0234	0.0600	0.0162	0.0162
100%	0	0	0	0	0	0	0

4. CONCLUSION

In this paper we have presented experimental subjective tests where 50 human observers have been exposed to compressed images with decreasing compression rates. The results suggest that the elements on the image are only perceived by human observers not at a specific compressed image quality, but depending on a saliency map difference threshold. Accordingly, saliency not only represents more attractive regions, but can also have influence on the perceptual quality of an image. Our subjective evaluations investigate different image quality rates to quantify this impact.

Based on these results new approaches can be defined. The results can be incorporated into a full-reference objective IQ metric or embedded into rendering algorithms to minimize the computational effort, while maintaining sufficiently high image quality so that humans perceive the elements on the image. There are numerous fields of image processing where this can be important, ranging from image acquisition to compression and communication applications. In some cases, files or data streams contain more information than is actually required for a particular purpose. E.g. a picture may have more detail than the eye can reproduce and therefore does not need to be processed with a lot of fine detail. Accordingly, since IQA is in close touch with practical application, it is sensible to start from concrete conditions to design IQ methods that target at a certain application environment, while preserving the perceptual quality of the image.

Although the study presented on this paper shows clear results, there are still some issues that should be more investigated in the future. The value of the saliency map difference threshold is based on saliency maps obtained from compressed images using an existing visual attention models. However, the accuracy of these models hasn´t been always completely proved yet, which may limit the overall reliability of our conclusions. Therefore, more evaluations with other visual attention models rather than Itty-Koch algorithm can be performed. Finally, the compressed images used on this study were degraded using JPEG compression. As a future work we also intend to investigate how other types of distortions (White Noise, Gaussian Blur, etc.), and not uniformly distributed distortion over the image, may affect minimum image quality as perceived by human subjects. With these results, a saliency-based compression algorithm that accurately matches the subjective quality may be developed.

ACKNOWLEDGMENT

This work was supported in part by the Fundação para a Ciência e Tecnologia – FCT (Portuguese Foundation for Science and Technology), under the grant number SFRH/BD/73129/2010.

REFERENCES

[1] Lin, W. S., "Gauging image and video quality in industrial applications," in Advances of Computational Intelligence in Industrial Systems, 116, 117-137 (2008).

[2] Seshadrinathan, K. and Bovik, C. A., “Unifying analysis of full reference image quality assessment”, in IEEE Intl. Conf. on Image Proc., (2008).

[3] Wang, Z. and Bovik, C. A., “Modern Image Quality Assessment”, New York: Morgan and Claypool Publishing Company (2006).

[4] Wang, Z., Bovik, C. A., Seikh, H.R., Simoncelli, E.P., “Image quality assessment: From error visibility to structural similarity”, IEEE Transactions on Image Processing, 13, 600-612 (2004).

[5] Wang, Z. and Bovik, C, A., “Mean squared error: Love it or leave it? A new look at signal fidelity measures,” IEEE Signal Process. Mag., vol. 26, no. 1, pp. 98–117 (2009).

[6] Sheikh, H.R., Bovik C.A., “Information Theoretic Approaces to Image Quality Assessment”, in Handbook of Image and Video Processing, Elsevier (2005).

[7] Daly, S., "The visible difference predictor: an algorithm for the assessment of image fidelity," in A. B. Watson, ed., Digital Images and Human Vision, Cambridge, MA: The MIT Press, 179-206 (1993).

[8] Zhang, X., Wandell, B.A., “A spatial extension of CIELAB for digital color-image reproduction”, Journal of the Society for Information Display, 5(1), 61-63 (1997).

[9] Wang, Z., Simoncelli, P. E., and Bovik, C. A., “Multi-scale structural similarity for image quality assessment,” in Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers (2003).

[10] Weken, V. D., Nachtegael, M., and Kerre, E. E., "Using similarity measures and homogeneity for the comparison of images,” Image and Vision Computing, 22, 695-702 (2004).

[11] Sheikh, R. H., Bovik, C. A., and Veciana, D. G., "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Processing, 14(12), 2117-2128 (2005).

[12] Sheikh, H. R., Bovik, A.C., “Image information and visual quality”, IEEE Transaction on Image Processing, 15(2), 430–444 (2006).

[13] Gao, X., Lu, W., Tao, D., and Li, X., “Image quality assessment and human visual system,” in Proceedings of SPIE, Video Communications and Image Processing Conference, Huangshan, China, vol. 7744 (2010).

[14] Ma, Q. and Zhang, L. “Image quality assessment with visual attention,” in Proc. ICPR, pp. 1–4 (2008).

[15] Moorthy, K. A., and Bovik, C. A., “Visual importance pooling for image quality assessment,” IEEE J. Select. Topics Signal Process (Special Issue on Visual Media Quality Assessment), vol. 3, no. 2, pp. 193–201 (2009).

[16] Zhou, W., Gangyi, J., and Me, Y., “New visual perceptual pooling strategy for image quality assessment”, Journal of Electronics (China), V29 (3/4): 254-261 (2012).

[17] Koch, C., and Ullman, S., “Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology”, 4:219-227 (1985).

[18] Itti, L., Koch, C., and Niebur, E. “A model of saliency-based visual attention for rapid scene analysis,” in IEEE Trans. Patt. Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259 (1998).

[19] Frintrop S., Klodt, M., Rome, R., “A real-time visual attention system using integral images”, International Conference on Computer Vision Systems (ICVS’07), (2007).

[20] Achanta, R., Estrada, F., Wils, P., Süsstrunk, S., “Salient Region Detection and Segmentation”, International Conference on Computer Vision Systems (ICVS '08), LNCS, 5008, 66-75 (2008).

[21] Bulthoff, I., Newell, F.N., “The role of familiarity in the recognition of static and dynamic objects”, Prog Brain Res. 154, 315-25 (2006).