A REDUCED-REFERENCE METRIC BASED ON THE INTEREST POINTS IN COLOR IMAGES

M. Nauge, M.-C. Larabi and C. Fernandez

XLIM-SIC Lab., University of Poitiers, France

ABSTRACT

In the last decade, an important research effort has been dedicated to quality assessment from subjective and objective points of view. The focus was mainly on Full Reference (FR) metrics because of the ability to compare to an original. Only few works were oriented to Reduced Reference (RR) or No Reference (NR) metrics, very useful for applications where the original image is not available such as transmission or monitoring. In this work, we propose a RR metric based on two concepts, the interest points of the image and the objects saliency on color images. This metric needs a very low amount of data (lower than 8 bytes) to be able to compute the quality scores. The results show a high correlation between the metric scores and the human judgement and a better quality range than well-known metrics like PSNR or SSIM. Finally, interest points have shown that they can predict the quality of compressed color images .

1. INTRODUCTION

Quality assessment is becoming an important issue in the framework of image and video processing. This need is expressed by the fact that the quality threshold of end-users has been shifted up because of the large availability of high fidelity sensors at very affordable prices. This observation has been made for different application domains such as printing, compression, transmission, and so on. Starting from this, it becomes very important to manufacturers and producers to provide products of high quality to attract the consumer. This high interest for quality means that tools to measure it have to be available.

Hence, two types of measurements are possible: Subjective assessment that takes the end-observer in the loop of assessment and objective assessment based on the use of mathematical tools that can use properties of the Human Visual System (HVS). In the recent literature [1] hundreds of papers have proposed objective quality metrics dedicated to a large panel of image and video applications.

Basically, objective evaluation metrics can be categorized into three groups: full-reference (FR), no-reference (NR), and reduced-reference (RR) metrics. FR metrics often seen as fidelity metrics, need full information of the original images and demand ideal images as references which can be hardly achieved in practice. The traditional methods of FR (such as peak signal-to-noise-ratio PSNR) are based on pixel-wise error and have not always been in agreement with perceived quality. Recently, some FR metrics based on HVS have been proposed like weighted signal-to-noise-ratio (WSNR) metric [2] by using the Contrast Sensitivity Function (CSF), to mimic the human early vision. Other metrics introduce more HVS properties, such as perceptual difference (Pdiff) metric [3] with a complex modeling of the HVS. So, in this metric the images are transformed from RGB to CIELAB. Then Barten's CSF and Daly's visual masking are applied. The pixel-wise difference between the original and the impaired image is then computed, and the obtained values are considered as perceptible if they are above achromatic and chromatic predefined thresholds. Another approach consists in the use of natural scene statistics in the wavelet domain, such as Information Fidelity Criterion (IFC) metric [4]. This model captures two important, and complementary, distortion types: blur and additive noise. The IFC is the mutual information between the source and the distorted images. The IFC is not a distortion metric, but a fidelity criterion. It theoretically ranges from zero (no fidelity) to infinity (perfect fidelity). Wang et al. introduced in [5] new framework for quality assessment based on the degradation of structural information. They developed a Structural SIMilarity (SSIM) and demonstrated its performances through a complete set of natural images.

At the opposite of FR metrics, NR metrics aim to evaluate distorted images without any cue from the source. However, most of the proposed NR quality metrics [6] are designed for one or a set of specific distortion and are unlikely to be generalized for evaluating images degraded with other types of distortions. While RR [7] metrics are between FR and NR, they make use of a part of the information from original images in order to evaluate the visual quality of the distorted ones. They extracted information from the reference represents small ratio with the regards of the original size. This allows RR metrics to be used in very specific fields (transmission or monitoring) where the original data cannot be availiable or does not exist.

The objective of this paper is to build a new RR metric able to predict image quality that correlate with human judgment while having same or better performance than recent FR metrics. One important thing is related to the complexity for especiailly for embedded applications. The main difficulty for designing RR metric is to extract pertinent information in order to have comprehensive data for a robust and accurate prediction. Our focus is on the importance of interest points such as Harris, SIFT,... for the prediction of the perceived quality with a link to the image content in terms of active/flat regions. For this study, the compression is the targeted application and all the experiments are performed in such a direction.

The remainder of this paper is organized as follows: The next section is dedicated to the description of the metric flowchart, followed by the experimental section. This paper ends with some conclusions and gives some future directions.

2. PROPOSED METRIC

In order to build a new metric with a high correlation with human judgment, it is helpful to understand the human response for different kind of image distorsions. With blurred images, like those obtained using JPEG 2000 compression, a human observer can have difficulties to perceive contours or interest regions. In this case, the HVS needs to exploit the scene semantic to reconstruct and under-

stand the content of the image. With blocking effect, like JPEG compression, there are many horizontal and vertical contours generated because of the 8×8 scheme. This creates a conflict between inherent contours and the artificial ones. In this case the HVS needs and adaptation effort to understand and recognize the objects in the scene. So, one can conclude that the perceived quality is directly related to the effort made by the HVS in the pattern recognition task.

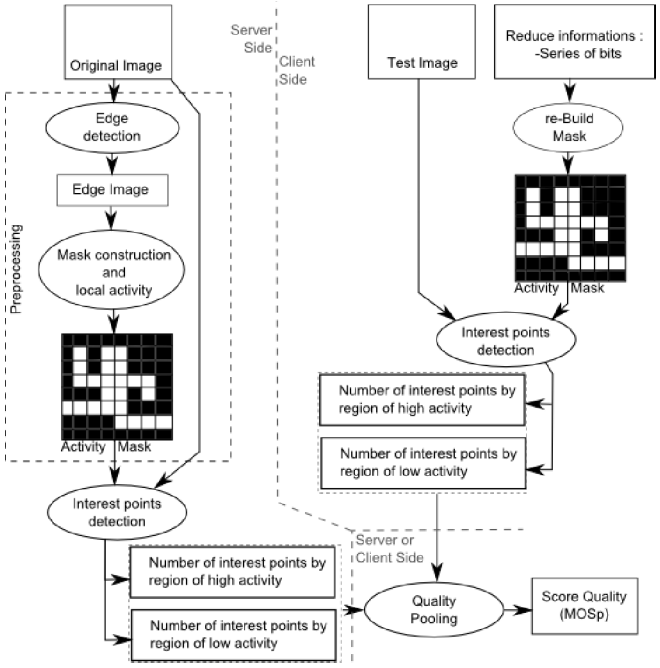


Fig. 1. Flowchart of the proposed metric

On the one hand, images are not uniform and so they can contain regions that attract the human gaze more than others.This behavior is confirmed by eye-tracking experiments. On the other hand, to date many algorithms have been developed to mimic the eye-tracking process in order either to estimate the salient regions or to detect the interest points of an image. These latter have attracted an important research effort. Hence, in the literature, one can find algorithms to predict interest point, like Harris corner detector [8], SIFT detector [9] or SURF detector [10]. Interest points are very important for the characterization of an object or a texture; they are involved in the detection of contours and textures, and allow the discrimination of shapes and objects. These algorithms are classicaly used in motion detection and object recognition. Some works like the one contained in [11] studied the performance of these detectors for differents kinds of image distortions such as translation, rotation, compression,.... they noticed that these tools are invariant to scale and rotation, but less effective with JPEG and JPEG 2000 compression. From the studies, the number and position of interest points seem to be related to the bitrate which make them appropriate for predicting visual quality/fidelity. From the last observation, we have used interest point detectors for the prediction of perceived quality of an image having undergone a process such as compression, transmission.... This represents the originality of this paper.So, the proposed metric can be summarized by the flowchart given in figure 1. The main idea is to extract from the source image a reduced reference information for the quality assessment process . The reference is based on the image partition, the activity of the partitions and their associated interest points. For the evaluation of the target image, the same algorithm is applied by using the same partition than the source image. In this section we will describe the complete scheme of the metric.

2.1. Interest points detection

In this work, The method used to detect the interest points is the one developed by Harris [8]. The Harris corner detector is based on the local auto-correlation function of a signal. This latter measures the local changes of the signal with patches shifted by a small amount in different directions. A corner can be defined as the intersection of two edges. A corner can also be defined as a point for which there are two dominant and different edge directions in its local neighborhood. In natural images, these corners are present along the edges and very dense in the texture. The number and the position of detected corners can be guided by different factors, such as neighborhood size, aperture size...For our metric, the Harris parameters were chosen to maximize the number of interest points. So, their number is not limited they can be very close to each other. For instance, interest points can be dense for textures where distortions can be the most visible.

Figure 2 gives some results of color points of interest extraction for an image compressed using JPEG at 2 different bitrates. It is easy to see that the number of interest points decreases when the compression ratio increases. This is valid for all compression schemes such as JPEG 2000 and JPEG XR.

Thus, the blocking effect introduced by JPEG creates a loss of interest points in some regions and generates new ones at the corners of each block. The blurring introduced by JPEG 2000 also spreads the interest points on textures and contours. It is obvious that after a compression, there is a loss of interest points in regions of high texture and contour activity. However, in the flat area like the sky, we notice an increase of the interest points because of the impairments. Hence, it will be useful to patition the image in two kind of region to make the metric sensitive to the decrease of interest points in high activity region and the increase of interest points in low activity region.

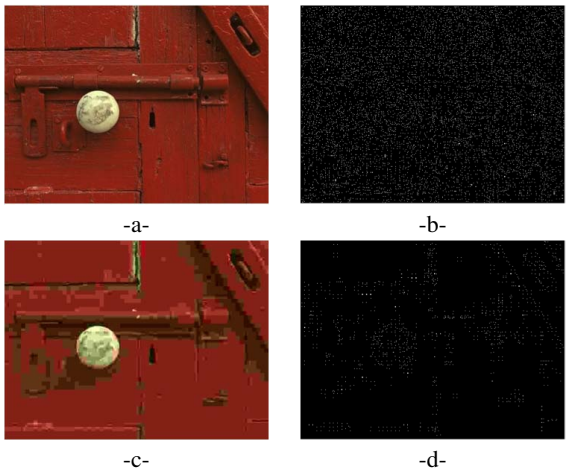


Fig. 2. Extraction of color points of interest for a JPEG compressed image at 2 different bitrates

2.2. Construction of the activity mask

To ensure the robustness of the information coming from the extraction of interest points, it is essential to decompose the image into two regions of different texture activity. And also in order to approach the issues related to the perception of salient areas. This would limit the size of information to embed with the image. The idea is to partition the image into blocks of large size and to give a label to each block to indicate that it should be treated as an area of low activity or high activity. Then measure the local activity (low or high activity) of each partition. It is possible to perform this activity measurement in the frequency domain, or with the help of wavelet, but in this version of the metric we used a gradient operator in the spatial domain. Each block having an average activity higher than a threshold T is considered of high activity and otherwise of low activity. The threshold has been fixed at 15% after several experiments on a large database containing different several types of images.

For the simulation part of this paper, we fixed the number of image partitions to 64 (8×8) in order to limit the size of data to be embedded. This means that the partition is sent as succession of bits of 1 if the block is of high activity and 0 otherwise. At this stage the amount of data needed to be in the RR mode is 8 bytes. This number will increase depending on the adopted number of partitions.

3. SCORES PREDICTION

The quality factor is obtained by using four pieces of information. The first values (resp. original and impaired) are the number of points of interest in areas of high activity (HA), the other two (resp. original and impaired) are the number of points of interest in areas of low activity (LA). The metric score is given by the following equation:

$$MOS_p = w_{HA} \times S_{HA} + w_{LA} \times S_{LA} \quad (1)$$

where w_{HA} is the proportion of partitions of high activity and $w_{LA} = 1 - w_{HA}$ is the proportion of partitions of low activity. S_{HA} and S_{LA} are respectively the scores for HA and LA partitions. They are computed as follows :

$$S_{**} = \begin{cases} 1 - VDiff/maxP & if maxP > 0 \\ 1 & else \end{cases} \quad (2)$$

where $VDiff$ is the absolute difference of the number of interest points between the original and impaired images and $maxP$ is the maximal number of points.

4. EXPERIMENTAL RESULTS

This section is dedicated to the experimentation of the proposed metric by using a standard framework as defined by the Video Quality Expert Group [12]. For this, we used 3 available images databases: Live [13], Toyama [14] and TID [15]. These databases are widely used in the literature because they provide comprehensive subjective scores for each image, at each compression ratio and for each coder.

Table 1 gives some examples of scores obtained using the proposed metric. The results show the subjective score (MOS: Mean Opinion Score), the predicted scores (MOSp) for HA and LA partitions and their combination. We can notice that mixing the scores of HA and LA partitions allows to have scores close to those given by human observers (MOSp Mix). This means that both partitions (to a certain extent) are used by a human observer to judge the quality of an image. However, for the rest of the experimentation two versions named QIP1 (using only the HA partition) and QIP2 (mixed partitions) will be used.

Table 1. Some examples of quality scores (MOSp) obtained with the proposed metric in comparison to subjective scores (MOS).

Bdd JPEG	ImgO	ImgD	MOS Sub	MOSp Mix	MOSp HA	MOSp LA
LIVE1	ocean	img8	0.700	**0.800**	0.936	0.608
LIVE1	caps	img24	0.709	**0.777**	0.914	0.737
Tid2008	i03	i03-10-3	0.351	**0.255**	0.473	0.151
Tid2008	i09	i09-10-2	0.571	**0.538**	0.818	0.343

In order to study the performance of the proposed metric, we compared to several state-of-the-art metrics such as PSNR, SSIM [5], Pdiff [3], IFC [4] ... The comparison is made using the prediction accuracy (RMSE) and the prediction monotonicity (Pearson correlation) as defined by the VQEG. A benchmark of recent state-of-the-art metrics is available in [16].

From tables 2, 3, we can notice that the RR metrics QIP1 and QIP2 have correlation results that are comparable or better than results of FR metrics from the literature. Note also that QIP2 is particularly effective on JPEG, and its short version i.e. QIP1 seems to be sufficient for jp2k. The efficiency of QIP2 for JPEG compression can be explained by the usage of Harris corner detector. It s designed to catch corners and the blocking effect of JPEG introduces four corners by DCT block.

Table 2. Pearson correlation for JPEG images

DB\Mc	IFC	Pdiff	SSIM	P-L	WSNR	QIP1	QIP2
Toyama	0,975	0,982	0,971	0,973	0,960	0,975	**0,985**
LIVE1	0,951	0,974	0,944	0,986	**0,989**	0,966	0,984
Tid2008	0,957	0,967	0,984	0,985	0,986	0,981	**0,988**
All	0,962	0,973	0,967	0,961	0,963	0,975	**0,987**

Table 3. Pearson correlation for JPEG 2000 images

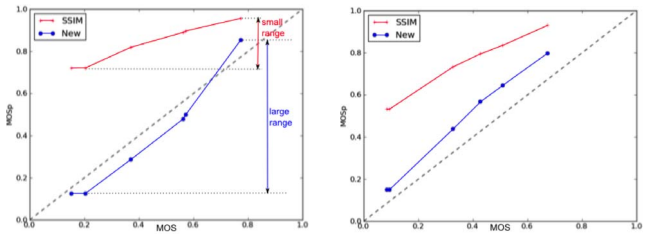
DB \Mc	IFC	Pdiff	SSIM	P-L	WSNR	QIP1	QIP2
Toyama	0,959	0,982	0,966	0,970	**0,958**	0,971	0,948
LIVE1	0,961	**0,983**	0,944	0,978	0,981	**0,983**	0,974
Tid2008	0,957	**0,986**	0,978	0,984	0,979	0,973	0,949
All	0,960	**0,983**	0,964	0,978	0,972	0,977	0,962

It is really difficult to judge the contribution of a metric in comparison to others. The real advantage of the one proposed in this paper lies in the relevance of the values that it provides. Where metric scores are difficult to interpret, requiring a run on a set of several magnitudes of distortions to get an idea of how to interpret the score, this metric gives values directly understandable, spread over a larger range. Table 4 makes a comparison using two images having the same PSNR but visually very different. While metrics like PSNR-HVS [17] and SSIM give scores in a very small range, our metric transcript the perceptual impression.

To prove the relevance of the predictions and not to introduce controversial factors, we compared the results of predictions without normalization or fitting.

The comparison given in Figure 3 is made with SSIM on images from the LIVE database on which SSIM is the most efficient. We use only a subset, but they represent a variety of content, close-up view, landscape, animal, face.

Note that SSIM never give predictions below 0.5; it never informs that an image is not "watchable" or "usable". It does not exploit the full range values available, as shown on -Parrots- image (Figure 3). The proposed metric has a much larger dynamic.

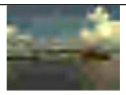


-Parrots- -Sailing-

Fig. 3. Results of our metric vs. SSIM for several images

So each prediction is comprehensible. In addition, this metric is a reduced reference one. This means that very little amount of information is used for quality prediction (64 bits for the mask and the 2 integers containing the number of HA and LA points). Unlike modified versions of PSNR or SSIM which require the reference image. Our metric can be used in contexts of digital communication and monitoring for judging the quality of color images transmitted through different channels.

Table 4. Comparison of ranges of quality scores between the proposed metric and state-of-the-art metrics.

Images \Metrics	PSNR	PSNR-H	SSIM	QIP2
	24	24.45	0.78	**0.99**
	24	19.8	0.58	**0.09**

5. CONCLUSION

In this paper, we proposed a new reduced reference metric based on interest points extracted from high and low activity partitions of the image. The major innovation is the use of interest points for the prediction of percevied image quality. This metric is fast and not complex for a real time usage. Moreover the range of score prediction is large. So each score is explicit and understandable with regards to human judgement. The obtained results are very encouraging and proved a good correlation with human judgment. The advantages of this metric are the reduced reference, high speed, robustness and accuracy. One important conclusion of this work is that interest points detector are suitable for the prediction of visual quality on compressed images. This study opens plenty of future directions such as studying the performance of Harris detector with regards to SIFT or SURF detector for quality assessment. It will be also interesting to replace the activity part by a saliency predictor that integrates suitable HVS models. An important extension can be the integration of temporal information for video quality assessment (videoconferencing, VOD ...).

6. ACKNOWLEDGEMENTS

This work is supported by the ANR project VERSO - CAIMAN. We also thank the anonymous reviewers for their valuable remarks and suggestions.

7. REFERENCES

[1] Marius Pedersen and Jon Yngve Hardeberg, "Survey of full-reference image quality metrics," in *GCIS*, Gjovik, Norway, 2009.

[2] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," in *IEEE transactions on image processing*, 2000, pp. 636–650.

[3] H. Yee, "A perceptual metric for production testing," *Journal of Graphics Tool*, pp. 33–40, 2004.

[4] H. R. Sheikh, *Image Quality Assessment Using Natural Scene Statistics*, PhD thesis, University of Texas at Austin, 2004.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[6] D. Hands, D. Bayart, A. Davis, and A. Bourret, "No reference perceptual quality metrics: approaches and limitations," in *HVEI XIV*, feb 2009, vol. 7240.

[7] U. Engelke and H.-J. Zepernick, "Perceptual-based quality metrics for image and video services: A survey," in *3rd EuroNGI Conference on Next Generation Internet Networks*, Trondheim, Norway, 2007.

[8] C. Harris and M. Stephens, "A combined corner and edge detector," in *4th Alvey Vision Conference*, Manchester, 1988.

[9] David G. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, 2004, vol. 60, pp. 91–110.

[10] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[11] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[12] VQEG, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessement," Tech. Rep. PHASE I 2008, VQEG.

[13] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 1," http://live.ece.utexas.edu/research/quality.

[14] Y. Horita, Y. Kawayoke, and Z. M. Parvez Sazzad, "Image quality evaluation database," http://mict.eng.u-toyama.ac.jp/mict/index2.html.

[15] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "Color image database for evaluation of image quality metrics," 2008, http://www.ponomarenko.info/tid2008.htm.

[16] M. Nauge, C. Larabi, and C. Fernandez, "Image quality metric benchmarking on compressed image databases," in *IS&T/SPIE Electronic Imaging*, California, United States, 2010.

[17] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli., "Two new full-reference quality metrics based on hvs," in *Second International Workshop on Video Processing and Quality Metrics*, Scottsdale USA, 2006.