Full-Reference and Reduced-Reference Quality Metrics based on SIFT

Joumana Farah[(1)], Marie-Rita Hojeij[(1)], Jihad Chrabieh[(1)], Frédéric Dufaux[(2)]

[(1)] Telecommunications Department, Faculty of Engineering,
Holy-Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon

[(2)] Institut Mines-Télécom ; Télécom ParisTech ; CNRS LTCI
46 rue Barrault, F - 75634 Paris Cedex 13, FRANCE

ABSTRACT

In the last decade, an important research effort has been dedicated to implement objective image quality assessment metrics that reflect effectively human perception. Therefore, the aim of this paper is to propose new objective metrics that fulfill the demands of the image quality assessment field. For this sake, we propose two main full-reference (FR) quality metrics, and then adapt them in such a way to constitute several new reduced-reference (RR) quality metrics, for the case where the complete reference image is not available. We evaluate the influence of five types of distortion such as JPEG, JPEG2000, Gaussian Blur, AWGN, and Contrast change, on the image quality. The proposed metrics are based on the number of Scale-Invariant Feature Transform (SIFT) points, the number of SIFT matches between the unpaired and distorted images, and the Structural Similarity index (SSIM). In order to validate our proposed metrics, we compute the correlation between our metrics' scores and the subjective evaluation results. The results show a high correlation and a better quality range compared to well-known metrics, as well as a good robustness to reduced-reference situations.

Index terms – Full reference, Objective image quality assessment, Reduced reference, SIFT, SSIM, Subjective evaluation.

1. INTRODUCTION

Digital images are becoming an essential component of people's daily life. They experience various processing stages before reaching the end user. These stages can provoke significant degradation in the perceived visual quality of the images. Therefore, quality evaluation techniques become very important. For this intention, two main approaches are possible, either to ask an observer to judge the perceived image quality, leading to subjective image quality assessment, or to develop an objective measurement metric that should be highly correlated with human judgment. Since subjective evaluation is affected by physical and psychological parameters [1], measuring the quality of an image becomes a complicated task. Objective methods of evaluation which can predict the human viewer response have become very popular and new metrics are being continuously proposed [2, 3].

Basically, objective evaluation metrics can be categorized into three groups: full-reference (FR), no-reference (NR), and reduced-reference (RR) metrics.

FR metrics need to access the original reference image [4]; they can serve in studying the performance of an image codec where the encoder parameters are optimized according to the obtained performance. Such performance is assessed by directly comparing the original image to the one obtained after decoding. The Mean Squared Error (MSE) [5], the Peak Signal-to-Noise Ratio (PSNR) [2], and the Structural Similarity index (SSIM) [3] are widely used FR quality metrics.

RR metrics are useful when only partial information regarding the original image is available [6]; they can be used in the contexts of digital communications and monitoring, for judging the quality of images transmitted through different channels. For instance, the Harris corner detector metric [7] relies on the reduced-reference concept. At the opposite of FR metrics, No-Reference (NR) quality metrics [8] have access only to the distorted image, which makes the evaluation process a rather harsh task.

On the other hand, classical metrics that depend only on image intensity, such as MSE, PSNR, and SSIM, are widely criticized, because they fail whenever geometric displacements or deformations occur. For this reason, our proposed metrics will be based on local features extracted from the image, such as the Scale Invariant Feature Transform (SIFT) [9], which has the advantage of being practically invariant to changes of brightness, rotation, scaling, and small changes of viewpoint. Since the SIFT algorithm can capture the main characteristics of an image, it has often been used for image matching and object recognition [10]. In this work, we construct two main FR quality metrics that use the SIFT features, and then adapt them in order to form several RR quality metrics that meet the access limits constraints to the reference image. In order to validate our metrics, we measure the correlation between subjective experimental results, such as the Mean Opinion Score (MOS), obtained from known databases [11, 12] and the estimated objective proposed metrics' results. Correlation results show a more accurate quality prediction than several well-known metrics, a larger range of score predictions, as well as a good robustness to quantization and resizing, in the context of reduced-reference applications.

The remainder of the paper is structured as follows. The related work is first presented in Section 2. Then, the proposed full-reference metrics are described in Section 3, and their counterpart reduced-reference versions in Section 4. Experimental results are shown in Section 5 in order to evaluate and analyze the performance of the proposed approaches. Finally, conclusions and possible future extensions are drawn in Section 6.

2. MAJOR RELATED WORK

2.1 SSIM_SIFT metric

SSIM_SIFT is a full-reference visual quality metric based

on the calculation of SSIM in blocks surrounding correspondent feature points between two images. This approach was initially proposed in [13] and then adapted in [14]. However, in [13], it was only used to assess the quality of resized images, on one hand, and no comparison with subjective scores or other quality metrics was done. In [14], it was proposed to evaluate the quality of objects in an image that experienced seam carving, and no estimation of the entire image quality was done.

2.2 Reduced-Reference metric based on Harris corner detector

This metric [7] uses interest points extracted from the image, at the encoder side, using the Harris corner detectors [15]. The number of Harris corners is transmitted for high activity and low activity areas, along with the image activity mask. The paper results [7] show a high correlation between the metric scores and human judgment. However, this metric has been tested on JPEG and JPEG2000 only, and necessitates an accurate calculation of the activity mask, which yields an increase in the encoder complexity.

3. PROPOSED FULL REFERENCE METRICS

3.1 Quality metric based on the number of SIFT points and the number of matched SIFT points

In contrast with previous approaches, the proposed metric relies on both the number of SIFT points extracted from the original image and on the number of matched SIFT points between the original and distorted images.

For this purpose, in a first step, we extract the SIFT points from the original image, as well as from the distorted image. Next, we perform SIFT points matching in such a way that the best candidate match for each SIFT keypoint is found in the processed image, by identifying the keypoint with the minimum Euclidean distance for the invariant descriptor vector.

After intensive simulations, we noticed that the number of matched SIFT points is highly correlated with the subjective MOS score, for a wide range of distortion types. This observation is behind the reason of introducing it in the objective score measurement, together with the number of SIFT points extracted from the original image. A possible FR metric score that involves both parameters can be given by:

$$\mathrm{MOS}_{match} = \frac{NOM}{NOS},$$

where *NOS* is the number of SIFT points extracted from the original image (which is also the number of matched points in the absence of distortion) and *NOM* is the number of matched points between the original and the impaired images. In other words, the metric is also the ratio between the actual number of matched points and their maximal possible number. Hence, MOS_{match} lies between 0 and 1, where MOS_{match}=1 means that the two images are identical.

3.2 Quality metric based on SIFT matches, SSIM, and the distance between SIFT matches

The SSIM_SIFT metric in [14] is an object-based metric intended to assess the visual content of an object existing in an image. In our study, this metric is implemented in such a way to assess the visual quality of the entire image. Similarly to [14], rectangular windows of W x W pixels are defined centered at each matched SIFT points x_j in the original image and y_j in the impaired one. A window dimension of W=11 is chosen in order to cover enough of the surroundings. We have added zero-padding in case a part of the window centered on the SIFT point lies outside the image. Then, SSIM [3] is calculated between windows surrounding each pair of matching points x_j and y_j.

Rather than simply averaging the SSIM values between the corresponding windows to yield the final quality measure, as was done in [14], we propose to improve this metric by introducing, within the calculation, the distance between the matched SIFT points. This distance is incorporated in the global average estimation through a proper weight based on the Euclidean distance D_j between the descriptor vector of the SIFT point and its corresponding match. This way, a higher importance is given to the SSIM windows that correspond to a low distance between a SIFT point and its match. This new FR metric can be written as:

$$\mathrm{Weighted\_SSIM\_SIFT} = \sum_{j=1}^{M} W_j \times SSIM(x_j, y_j),$$

where $\sum_{j=1}^{M} W_j$ =1 and M the number of matched SIFT points (*NOM*). Several weight options were tested in this work, and the following one was retained, since it yielded the best performance:

$$W_j = \left(1/D_j\right) / \sum_{j=1}^{M} \left(1/D_j\right).$$

4. PROPOSED REDUCED-REFERENCE METRICS

4.1 Reduced-reference implementation of the quality metric using the number of SIFT points and the number of SIFT matches

In this situation, the information that is needed to allow a correct matching between SIFT points, without the need for sending the whole image, are: the number of SIFT points in the original image, and the keypoint descriptor of each SIFT point. This information is extracted from the original image and transmitted through a RR channel. The keypoint descriptor of each SIFT point is a vector of 128 normalized values. Since the descriptors are floating-point values, they should be quantized before being encoded and transmitted. Therefore, the quantization level will certainly determine the amount of information to be transmitted for each image and, at the same time, the loss of information that can influence the SIFT matching process. The amount of transmitted data for each level of quantization is equal to: *NOS* x 128 x *NOB*, where *NOS* is the average number of SIFT points extracted from the original image (measured from a set of observations), and *NOB* the number of bits used to encode a descriptor.

Once this information is available at the receiver, the quality metric can be calculated in the same way as in the

FR 'MOS_{match}'. This new RR metric will be called '$MOS_{match\_Reduced}$'.

4.2 Resized version of the RR metric based on SIFT points and SIFT matches

In order to reduce the average number of SIFT points to be transmitted for each image, we propose to resize the reference image before extracting and transmitting the partial information to the client side. At the receiver side, before computing the matching with the distorted image, the latter must undergo the same resizing factor as the original image. This RR metric will be referred to as '$MOS_{match\_Reduced\_Resized}$'.

4.3 RR implementation of the weighted metric based on SIFT matches and SSIM

Since the SSIM calculation of the correspondent SIFT points necessitates the access to the pixel values in the reference image, it appears difficult to rely on partial information that can be extracted from the image and transmitted instead of the whole image. Therefore, the only way to reduce the amount of transmitted data is to resize the reference image before transmission. At the client side, the metric calculation remains the same as in the FR version, except for the resizing of the distorted image. This RR metric will be called 'Weighted_SSIM_SIFT_Reduced'.

5. EXPERIMENTAL RESULTS

In order to assess the quality of the proposed metrics, we rely on the standard framework defined by the Video Quality Expert Group [16] for the validation of objective quality measures. For this purpose, we use the well-known image database TID2008 [12], which provides comprehensive subjective scores for each image, at each compression level and each distortion type. For instance, TID2008 contains 25 reference images, with 17 types of distortions, 4 levels of distortions for each type, and the corresponding MOS scores obtained by subjective experiments (normalized between 0 and 1). All images are saved in the database in Bitmap format (384 x 512) without any compression.

5.1 Performance of MOS_{match}

In order to analyze the performance of the proposed metric, we compare it to state-of-the-art metrics such as PSNR [2] and SSIM [3]. The simulations are carried out onto the entire database TID2008. The comparison is made using the prediction monotonicity (Pearson and spearman correlations [16]) and reported in Table 1.We can see that our FR metric MOS_{match} has correlation results that are generally better than results of FR metrics from the literature, for JPEG, JPEG2000, Gaussian Blur, and Contrast change. The PSNR metric only gives better correlations for the case of AWGN.

To prove the relevance of the predictions, we compare our predicted results without normalization or fitting. This comparison, shown in Figure 1, is made towards SSIM on 500 images with different types of distortion. We can see that SSIM give scores in a narrow range (i.e. does not exploit the full range of values), since it never gives predictions below 0.5; it never informs that an image is not "watchable" or "unusable". On the contrary, the proposed MOS_{match} metric gives values around the ideal linear case and exhibits a much larger dynamic.

Table 1. Comparative study of the proposed metric with state-of-the-art metrics

	SSIM	PSNR	MOS_{match}
JPEG Pearson	0.9300	0.8892	**0.9545**
JPEG Spearman	0.9270	0.8711	**0.9360**
JEPG2000 Pearson	0.9416	0.8657	**0.9473**
JPEG2000 Spearman	0.9623	0.8300	**0.9636**
AWGN Pearson	0.7669	**0.9327**	0.8369
AWGN Spearman	0.8310	**0.9115**	0.8559
Gaussian Blur Pearson	0.8782	0.8376	**0.9189**
Gaussian Blur Spearman	0.8996	0.8682	**0.9206**
Contrast change Pearson	0.7004	0.6043	**0.8541**
Contrast change Spearman	0.6329	0.6126	**0.8385**

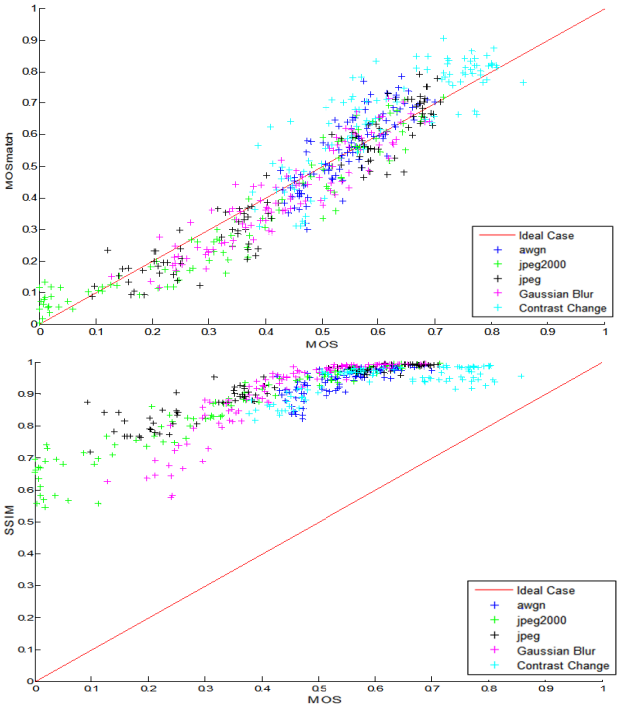

Fig. 1. Results of our metric (top) vs. SSIM (bottom) for 5 types of distortion

5.2 Performance of Weighted_SSIM_SIFT

In Table 2, we compare the performance of this metric versus other existing metrics and our proposed metric MOS_{match}. We can see that Weighted_SSIM_SIFT allows a significant improvement over the other ones for JPEG2000 and Gaussian Blur. It also allows filling a big part of the performance gap between PSNR and MOS_{match}, for the case of AWGN.

5.3 Performance of $MOS_{match\_Reduced}$ and $MOS_{match\_Reduced\_Resized}$

With the aim to explore the quality predicted by our proposed RR metrics $MOS_{match\_Reduced}$ and $MOS_{match\_Reduced\_Resized}$, we evaluate the obtained scores for

different transmission data rates (i.e. different levels of quantization) and different resizing factors, in terms of their correlation with the subjective MOS. The obtained results for $MOS_{match\_Reduced}$ (Table 3) show that, as predicted, the correlation between the proposed metric scores and the subjective MOS increases when the quantization level (i.e. the number of bits) increases. Meanwhile, the amount of transmitted data also increases. For instance, for an image size of 384 x 512, and considering an average number of SIFT points per image of 550, the amount of data to be transmitted, for 4, 5, and 6 bits, is respectively 35, 44, and 52.8 Kbytes per image. A tradeoff between the accuracy of the predicted quality and the amount of transmitted data should therefore be done, depending on the requirements of the application in use.

Table 2. Comparative results of Pearson and Spearman correlations for Weighted_SSIM_SIFT, in comparison to SSIM, PSNR, SSIM_SIFT, and MOS_{match}

	SSIM	PSNR	SSIM_ SIFT	Weighted SSIM_SIFT	MOS_{match}
JPEG Pearson	0.9300	0.8892	0.9232	0.9292	**0.9545**
JPEG Spearman	0.9270	0.9011	0.9295	0.9319	**0.9360**
JEPG2000 Pearson	0.9416	0.8657	0.9489	**0.9592**	0.9473
JPEG2000 Spearman	0.9623	0.8300	0.9473	0.9563	**0.9636**
AWGN Pearson	0.7669	**0.9327**	0.8867	0.8971	0.8369
AWGN Spearman	0.8310	**0.9115**	0.8898	0.9001	0.8559
Gaussian Blur Pearson	0.8782	0.8376	0.9282	**0.9393**	0.9189
Gaussian Blur Spearman	0.8996	0.8682	0.9351	**0.9471**	0.9206
Contrast change Pearson	0.7004	0.6043	0.6621	0.6823	**0.8541**
Contrast change Spearman	0.6329	0.6126	0.5991	0.6238	**0.8385**

Table 3. Comparative results of the $MOS_{match\_Reduced}$ and $MOS_{match\_Reduced\_Resized}$ metrics, in comparison to Harris corner metric.

	Harris corner [7]	$MOS_{match\_Reduced}$ (4 bits)	$MOS_{match\_Reduced}$ (6 bits)	$MOS_{match\_Reduced\_Resized}$ 192x256
JPEG Pearson	0.9121	0.9155	**0.9523**	0.9421
JPEG Spearman	0.9210	0.9142	**0.9398**	0.9327
JEPG2000 Pearson	0.9412	0.9214	0.9563	**0.9588**
JPEG2000 Spearman	0.9314	0.9319	**0.9682**	0.9594
AWGN Pearson	0.8018	0.8016	**0.8310**	0.8101
AWGN Spearman	0.8251	0.8219	**0.8422**	0.8019
Gaussian Blur Pearson	0.9002	0.8901	**0.9119**	0.9067
Gaussian Blur Spearman	0.9010	0.9029	**0.9191**	0.9012
Contrast change Pearson	0.7372	0.8101	**0.8637**	0.8456
Contrast change Spearman	0.7281	0.8215	0.8475	**0.8512**

When images are resized to 192x256 prior to quantization, the average number of SIFT points to be transmitted per image is reduced to 290, leading to a reduction in the data rate to 27.8 KB instead of 52.8 KB (for the case of a 6-bit representation). We have also tested this metric for images resized to 96x128; similar performance gains were obtained, but are not presented here for the sake of concision. It can be seen from Table 3 that good correlation results have been maintained in spite of the resizing, still significantly better than the RR Harris corner metric [7]. This shows that our MOS_{match} metric is very robust to different kinds of processing (i.e. resizing, quantization, etc.) applied to the original image.

As for the reduced-reference Harris corner metric [7], no image resizing is done (i.e. the algorithm is taken in its original form from [7]). Besides, the amount of data to be transmitted depends on the number of partitions considered in the calculation of the activity mask. For instance, for 64 partitions, it requires the transmission of only 8 bytes. However, it should not be forgotten that this reduction in the amount of data is at the expense of an increase in the encoder complexity to obtain an accurate segmentation mask.

5.4 Performance of Weighted_SSIM_SIFT_Reduced

Table 4 shows a comparison between the Harris corner metric and our Weighted_SSIM_SIFT_Reduced, for the case where images are resized to 192x256. Results show again a significant improvement in the correlation with our RR metric, especially for AWGN and Gaussian blur.

Table 4. Comparative results between Weighted_SSIM_SIFT_Reduced and Harris corner metric.

	Harris corner [7]	Weighted_SSIM_SIFT_Reduced 192x256
JPEG Pearson	0.9121	**0.9123**
JPEG Spearman	0.9210	**0.9256**
JEPG2000 Pearson	**0.9412**	0.9312
JPEG2000Spearman	0.9314	**0.9465**
AWGN Pearson	0.8018	**0.8721**
AWGN Spearman	0.8251	**0.8849**
Gaussian Blur Pearson	0.9002	**0.9239**
Gaussian Blur Spearman	0.9010	**0.9325**
Contrast change Pearson	**0.7372**	0.6237
Contrast change Spearman	**0.7281**	0.6239

6. CONCLUSION

This paper has proposed several new ways for automatically estimating the visual quality of an image, using objective techniques that can capture the influences of distortions similarly to the way humans perceive them. The proposed metrics lead to very encouraging results, especially when compared to previous approaches, and prove a high correlation with subjective evaluation. In addition, their range of score prediction is large, such that each score is explicit and understandable with regards to human judgment. There are several possibilities for further improvement of the proposed metrics. For instance, the objective quality measures developed here with the SIFT descriptors could also be tested with other detectors such as SURF [17]. Furthermore, in the RR version of the metric using the number of SIFT points and SIFT matches, significant reductions in the transmitted data rates could be achieved by decreasing the amount of descriptors (classically taken as 128) to be transmitted for each SIFT point, and also by controlling the number of SIFT points extracted per image. Finally, we are currently working on a combination of the two proposed metrics MOS_{match} and Weighted_SSIM_SIFT, such as to obtain a metric with an optimized performance for all types of distortion.

REFERENCES

[1] W. Osberger (1999). “Perceptual Vision Models for Picture Quality Assessment and Compression Applications”. Ph.D. thesis, Queensland, University of Technology, Brisbane, Australia.

[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] 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.

[4] Marius Pedersen and Jon Yngve Hardeberg, “Survey of full reference image quality metrics,” in *GCIS*, Gjovik, Norway, 2009.

[5] B. Girod, “What’s wrong with mean-squared error,” in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[6] 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.

[7] M. Nauge, M.-C. Larabi and C. Fernandez, "A reduced-reference metric based on the interest points in color IMAGES", PCS 2010.

[8] D. Hands, D. Bayart, A. Davis, and A. Bourret, “No reference perceptual quality metrics: approaches and limitations,” in *HVEI XIV*, Feb. 2009, vol. 7240.

[9] D. Lowe, “Distinctive image features from scale invariant keypoints,” *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2005.

[11] 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.

[12] 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

[13] D. Azuma, Y. Tanaka, M. Hasegawa, and S. Kato, “SSIM based image quality assessment applicable to resized images,” *IEICE Tech. Rep.*, vol. 110, no. 368, pp. 19-24, 2011.

[14] M. Décombas, F. Dufaux, E. Renan, B. Pesquet-Popescu, F. Capman, "A new object based quality metric based on SIFT and SSIM", ICIP 2012.

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

[16] V. Q. E. Group (2000). “Final report from the video quality experts group on the validation of objective models of video quality assessment”, march 2000. Available at: http://www.vqeg.org.

[17] 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.