A NEW OBJECT BASED QUALITY METRIC BASED ON SIFT AND SSIM

Marc Décombas[ab], *Frédéric Dufaux*[b], *Erwann Renan*[a], *Béatrice Pesquet-Popescu*[b], *François Capman*[a]

[a]Thales Communications & Security, Laboratoire MMP, 92700 Colombes, France
{marc.decombas,erwann.renan,francois.capman}@thalesgroup.com
[b]Télécom ParisTech, Dept. Traitement du Signal et des Images, 75014 Paris, France
{dufaux,pesquet}@telecom-paristech.fr

ABSTRACT

We propose a full reference visual quality metric to evaluate a semantic coding system which may not preserve exactly the position and/or the shape of objects. The metric is based on Scale-Invariant Feature Transform (SIFT) points. More specifically, Structural SIMilarity (SSIM) on windows around the SIFT points measures the compression artifacts (SSIM_SIFT). Conversely, the standard deviation of the matching distance between the SIFT points measures the geometric distortion (GEOMETRIC_SIFT). We validate our metric with subjective evaluation and reach a Spearman correlation of 0.86 for SSIM_SIFT and 0.74 for GEOMETRIC_SIFT.

Index Terms— Object-based metric, SSIM, SIFT, subjective evaluation

1. INTRODUCTION

We consider semantic video coding techniques aiming at preserving the visual quality of salient objects which may undergo small displacements and geometric deformations. Such schemes are especially suited for security and monitoring applications. Examples include content-based coding methods based on seam carving [1][2].

In this paper, our goal is to develop a full reference object-based visual quality metric to evaluate a semantic coding system under the assumption that the position and the shape of objects may have been considerably modified.

Traditional fidelity metrics such as Peak-Signal-to-Noise-Ratio (PSNR) or Structural SIMilarity (SSIM) [3] compare corresponding pixels or blocks in the reference and processed images. Therefore, these approaches fail whenever geometric displacements or deformations occur.

In [4], Wang and Simoncelli propose a complex wavelet domain image similarity measure that is insensitive to luminance change, contrast change and spatial translation. This metric is robust for small geometric distortions relative to the size of the wavelet filter. However, It does not handle large displacements, nor assesses geometric deformations.

Rubinstein et al. presents a subjective evaluation for image retargeting and intents to create an objective metric [5]. Specific features are identified in retargeted media that are more important for viewers. They conclude that the resizing method having the best subjective score is also the one having the worst score with their objective metric. Therefore, a reliable metric remains a challenge.

A full reference metric based on Scale-Invariant Feature Transform (SIFT) [6] and SSIM has been developed by Azuma et al. [7] in order to evaluate images resized by different retargeting algorithms. In [8], Liu et al. present an objective metric simulating the Human Vision System (HVS) based on global geometric structures and local pixel correspondence based on SIFT.

The common objective of [7][8] is to evaluate resized (smaller) images. However, both methods are not designed to measure compression artifacts and do not take into account geometric deformations possibly occurring in content-based coding schemes (e.g. [1][2]). Another limitation is that both metrics compare two entire images. Therefore, they fail to assess the quality of a specific (salient) object. Finally, the metric in [7] has not been validated by subjective tests.

In this paper, we introduce an object-based visual quality metric by selecting and matching SIFT points in the object to evaluate. SIFT points allow to put in correspondence the same object in the reference and processed images even though it has been geometrically distorted. Our proposed metric gives two scores. The first one applies SSIM in the neighborhood of matching SIFT points and is referred to as SSIM_SIFT. Thus, it measures traditional compression artifacts such as those resulting from H.264/AVC (Advanced Video Coding). The second score measures the geometric deformation of the object. It is based on the standard deviation of matching SIFT points coordinates and is referred to as GEOMETRIC_SIFT. These two measures have been validated by subjective evaluation following the Double Stimulus Impairment Scale (DSIS) protocol [9].

In summary, our main contribution is a metric that : (i) is object-based and not disturbed by distortions in the background, (ii) measures compression artifacts, (iii) measures geometric artifacts in the object, and (iv) is validated by subjective tests.

2. PROBLEM DEFINITION

For the purpose of object-based semantic coding, approaches based on seam carving have been proposed [1][2]. Seam carving is a method of resizing that suppresses or adds lines in non-salient parts of an image. Hereafter, we more specifically consider the method proposed in [1] without loss of generality. In this method, seam carving is applied as pre- and post-processing in conjunction with a conventional H.264/AVC video coding scheme, as illustrated in Fig. 1. Consequently, the method may introduce both traditional compression artifacts of H.264/AVC and geometric artifacts from the seam carving and synthesis.

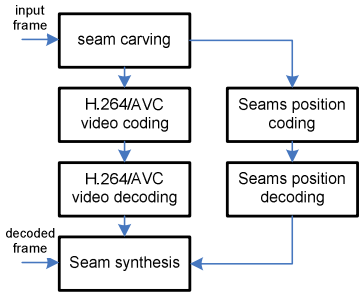


Fig. 1: Architecture of seam carving based video coding.

In this context, where salient objects are preserved but may undergo small displacements and deformations, traditional quality metrics such as PSNR or SSIM [3] fail.

3. PROPOSED OBJECT-BASED METRIC

In this paper, we propose a full reference metric to assess an object based coding system which has possibly modified the position and/or the shape of objects.

The metric relies on the combination of SIFT and SSIM to evaluate both compression artifacts and object deformations. SIFT [6] is an approach for detecting and extracting local feature descriptors invariant to different changes, in particular rotation, scaling and, in general, geometric deformations. In the proposed metric, SIFT allows to match an object from the original image with a potentially deformed object in the processed image.

Figure 2 represents the proposed full-reference metric. It takes three images as input: the original image, the processed image #1 which has been altered by seam carving but without compression, and the processed image #2 which has been modified both by seam carving and compression.

In a first step, we extract SIFT points from the original image, as well as the processed image #1. This is done in order to avoid the sensitivity of SIFT to coding artifacts. For the reference image, we only select SIFT points inside the considered object.

Next, SIFT points matching is performed from the original image towards the processed image as well as from the processed image towards the original one to increase robustness. Finally, a statistical analysis is performed on the matching distances in order to eliminate outliers. This step is useful to indentify erroneously matched SIFT pairs. Formally, a pair of points is considered an outlier if the following equation holds

$$|D(i) - \mu| > 3\sigma$$

with

$$\mu = \frac{1}{N}\sum_{i=1}^{N} D(i) \quad \text{and} \quad \sigma = \sqrt{\frac{1}{N-1}\left(\sum_{i=1}^{N}(D(i)-\mu)^2\right)}$$

where N is the number of SIFT points, i =1,.., N denotes the index of the current SIFT point, D is the distance between a pair of matching SIFT points in the original and processed images respectively, μ is the mean distance between the matching SIFT pairs, and σ is the standard deviation.

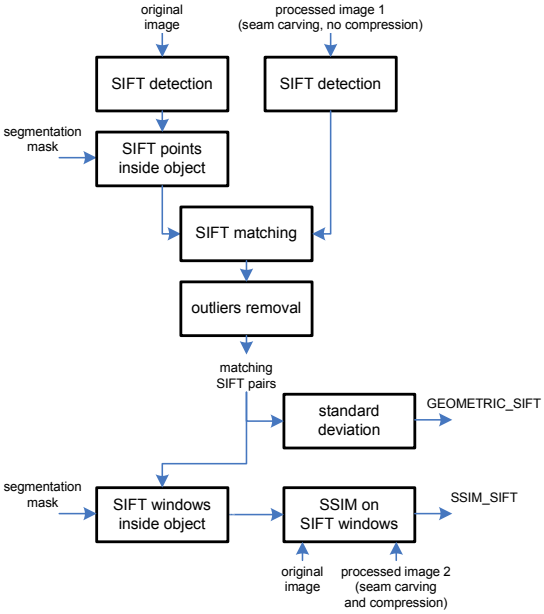


Fig. 2: Proposed object-based metric.

For GEOMETRIC_SIFT, we simply measure the standard deviation σ between matched SIFT points. This component of the metric captures the non-rigid deformation of the object.

In turn, the SSIM_SIFT component of the metric assesses the visual content of the object. First, non-overlapping W x W pixel windows are defined centered at each SIFT point and wholly contained inside the object. For this purpose, SIFT points with an associated window laying partly outside the object or with a spatial distance inferior to W pixels are discarded. The window dimension W=11 is chosen to cover enough of the surroundings.

Since SIFT points coordinates are not integer values, it can cause a mismatch of ±1 pixel, horizontally and/or vertically, when the window from the original image is compared with the window in the processed image. Thus, nine positions, representing all {-1, 0, 1} pixel shifts

horizontally and vertically, are tested and the one with the minimal Mean Square Error (MSE) is kept. Finally, SSIM is applied on all the windows defined by the above process, leading to the SSIM_SIFT measure.

4. SUBJECTIVE EVALUATION PROTOCOL

Following the ITU-R BT.500-12 recommendation [9], the protocol DSIS is chosen for subjective evaluation.

During the session, as a variation to standard DSIS, the assessor is first presented with a binary mask defining the object, then an unimpaired reference, and finally with the same picture impaired. At the beginning of each session, a training is given to the observers about the subjective assessment. In particular, assessors are specifically instructed to concentrate on the corresponding object. Afterwards, the assessor is asked to vote using the five-grade impairment scale: 5 imperceptible, 4 perceptible, but not annoying, 3 slightly annoying, 2 annoying, 1 very annoying.

Each assessor evaluates 30 images altered with different levels of artifacts spanning a large range of visual quality. The five first images are used for training and corresponding scores are discarded. Subjective scores are then processed and analyzed according to [9].

5. RESULTS

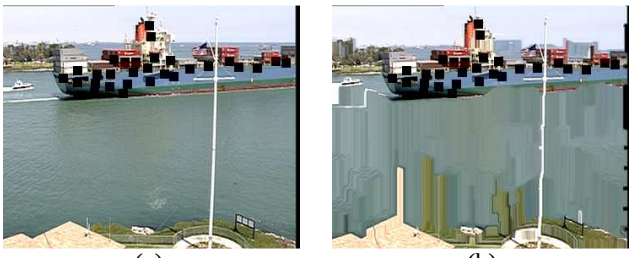
To validate our metric, we use the object-based compression method described in [1] to generate sequences presenting H.264/AVC compression artifacts, geometrical deformations and repositioning artifacts of the salient objects. Experiments are carried out using the test sequences Container and Coastguard in CIF format.

5.1. Performance of SSIM_SIFT

In a first set of experiments, we evaluate the performance of the proposed SSIM_SIFT, and in particular its ability to assess object-based visual quality in the presence of small displacements and deformations of the salient object.

Seam carving usually stops when it reaches objects defined by a saliency map. In this first experiment, seam carving parameters in [1] have been selected in order to achieve minor geometric distortions of the salient object. Nevertheless, a few seams may go through a salient object leading to artifacts. In addition, small deformations may also be introduced when reinserting seams during synthesis.

Figure 3 illustrates the proposed SSIM_SIFT metric. The Fig. 3 (a) and (b) show the SIFT windows (black squares) used to compute SSIM_SIFT in the original image and the processed image #1 (i.e. altered by seam carving but without H.264/AVC compression). In Fig. 3 (b), it can be observed that the container ship is well preserved, although the background is noticeably distorted. Moreover, the position and shape of the ship have been slightly altered.



(a) (b)

Fig. 3: SIFT windows used to compute SSIM_SIFT (black squares), (a) original image, (b) processed image #1.

As a reference, SSIM_Mask is a straightforward extension of SSIM computed on a salient object as defined by its binary mask. More precisely, SSIM is only calculated on the 11 x 11 windows which are wholly contained in the binary mask.

To validate the proposed SSIM_SIFT metric, a subjective evaluation was done with 14 non-expert assessors following the procedure described in Sec. 4. Six images quantized with QP={18, 36, 39, 42, 48}, for a total of 30 images, were shown in a random order to each assessor. We have found no outliers among assessors when following the procedure defined in ITU-R BT.500-12 [9].

Figure 4 shows the proposed SIFT_SSIM as a function of the Mean Opinion Score (MOS). The Spearman correlation is 0.86 and the Pearson correlation is 0.86 for the proposed SIFT_SSIM, showing a strong correlation. In comparison, the Spearman correlation is 0.20 and the Pearson correlation is 0.14 for SSIM_Mask. Clearly, SSIM_Mask fails as it cannot handle small geometric displacement or deformation.

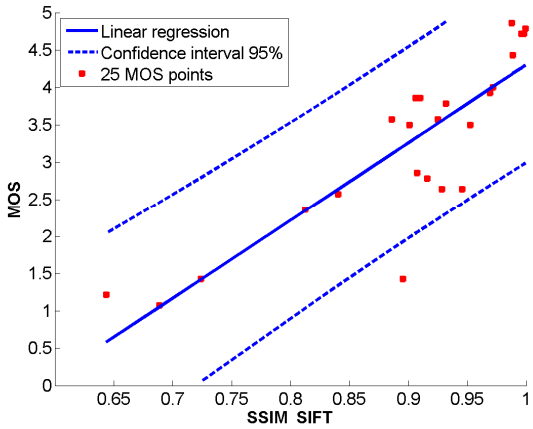


Fig. 4: SIFT_SSIM as a function of MOS.

5.2. Performance of GEOMETRIC_SIFT

We now evaluate the performance of the proposed GEOMETRIC_SIFT to measure object deformation. For this purpose, images with different levels of geometric deformation resulting from seam carving, but without

compression artifacts (QP=0), are considered. A new evaluation with 11 assessors has been performed. During this evaluation, no assessor has been detected as an outlier.

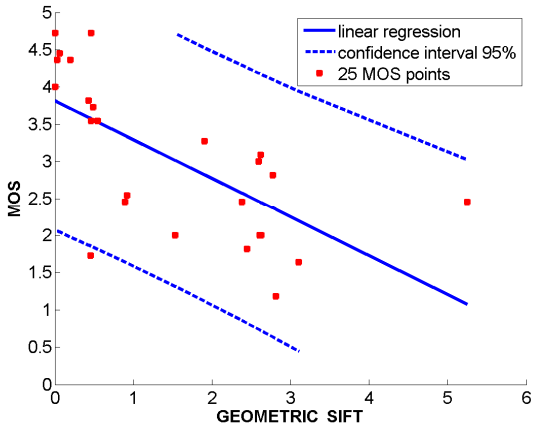
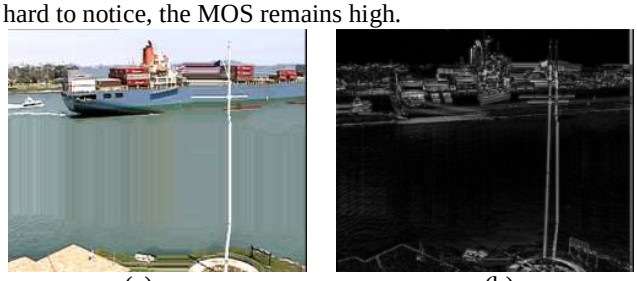


Fig. 5: GEOMETRIC_SIFT as a function of MOS.

The result of the experiment is given in Fig. 5. The Spearman correlation is -0.74 and the Pearson correlation is -0.67.

Correlations are lowered due two images with poor performances. The image corresponding to GEOMETRIC_SIFT=5.24 (right most point in Fig. 5) is shown in Fig. 6 (a). GEOMETRIC_SIFT is high, as the ship is elongated and has slightly moved as shown in Fig. 6 (b). However, as the artifact of translation and deformation is hard to notice, the MOS remains high.



(a) (b)

Fig. 6: (a) Container, frame 18, container object, (b) difference between the original and processed images.

Fig. 7: Container, frame 23, container object.

The image corresponding to GEOMETRIC_SIFT=0.46 and MOS=1.72 (lower left in Fig. 5) is shown in Fig. 7. The object of interest (the container ship) itself is well-preserved, however the borders of the object have been strongly distorted. In such a case, the assessor may have evaluated the border region instead of the ship alone. This underlines one of the limitations of this evaluation. The assessor can be influenced by the background.

6. CONCLUSION

In this paper we present an object-based full reference visual quality metric based on SIFT and SSIM. It can be used for images where the objects have their position and/or shape modified. The two proposed components have been validated by a subjective evaluation following DSIS. SSIM_SIFT gives a Spearman and a Pearson correlation of 0.86. Evaluation of deformation artifacts with GEOMETRIC_SIFT gives a Spearman correlation of -0.74 and a Pearson Correlation of -0.67.

In future work, we will aim at including additional attributes and combining different components into a single overall quality score.

7. REFERENCES

[1] M. Décombas, F. Capman, E. Renan, F. Dufaux and B. Pesquet-Popescu, "Seam carving for semantic video coding", *SPIE Proc. Application of Digital Image Processing*, San Diego, CA, 2011.

[2] Y. Tanaka, M. Hasegawa and S. Kato, "Improved image concentration for artifact-free image dilution and its application to image coding," *IEEE Proc. Int. Conf. On Image Processing*, Hong Kong, 2010.

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

[4] Z. Wang and E.P. Simoncelli, "Translation Insensitive Image Similarity in Complex Wavelet Domain" IEEE Trans. *Acoustics, Speech, and Signal Processing*, vol. 2, pp. 573-576, 2005.

[5] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A Comparative Study of Image Retargeting," *ACM Trans. on Graphics*, vol. 29, no. 5, pp. 1-10, 2010.

[6] D. Lowe, "Distinctive image features from scale invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[7] 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.

[8] Y.J. Liu, X. Luo, Y.M. Xuan, W.F. Chen, X.L. Fu, "Image Retargeting Quality Assessment," *Eurographics*, vol. 30, no. 2, 2011.

[9] Recommendation ITU-R BT.500-12, Methodology for the subjective of the quality of television pictures, 2009.