1D-05

What is Wrong with the SSIM Quality Metric

Zheng Liu

Toyota Technological Institute

Abstract: Structural similarity (SSIM) measurement is well accepted by the community of image quality assessment due to its simplicity and efficiency and has been widely applied in different applications. However, images are subject to varied degradation and a universal index always has limitation to quantify every kind of degradation. This paper discusses the relevant issues and proposes an image contour based approach for image quality assessment.

Introduction: Objective quality assessment for video and image plays an important role for modern multimedia communications and applications. Numerous image quality indices have been proposed so far and integrated for improved performance[1]. Basically, the quality indices can be categorized into reference-based and non-reference based approaches. The reference-based approaches may employ a full reference model or reduced reference model for quality assessment. A leading quality index is the structural similarity (SSIM) measure proposed by Wang et al., which takes the image local luminance, contrast, and structure into account[2]. This method has demonstrated its effectiveness and efficiency for varied applications. Based on this idea, a number of quality indices were proposed [1]. However, images are subject to varied degradation and a universal index always has its limitation to quantify every kind of degradation.

In this paper, we propose to use the image segmentation to quantify the accessibility of the informative features in an image subject to certain degradation. With the assumption that the full reference (ground truth image) is available, both the input and reference images are segmented and the derived contour maps are compared in the framework of contour mapping measure (CMM)[3]. The new quality index S is defined with the F measure by CMM precision and recall.

A Segmentation based Index: In this study, a new image quality index is defined under the framework of the gPb-ucm (globalized probability of boundary - ultrametric contour map) algorithm, which is top-ranked at Berkeley segmentation benchmark study, for image contour detection[4].

The contour maps of the reference and input images are then compared to identify the quality of the input image. For an algorithm boundary A and a ground-truth boundary B, precision (P) and recall (R) are respectively defined as:

$$\begin{aligned} P &= \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} = \frac{CM(A,B)}{|A|} \\ R &= \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} = \frac{CM(A,B)}{|B|} \end{aligned} \quad (1)$$

where $CM(\cdot)$ refers to the method to match the points on the two boundaries and $|\cdot|$ is the size of the boundary. In this study, a contour mapping measure defined as following is adopted.

$$CM(A,B) = \frac{1}{N}\delta([A],[B]) \quad (2)$$

where N is the number of mapped point pairs and $[\cdot]$ means the cyclic-shifted version of A and B. The distance $\delta(\cdot)$ is obtained by an improved dynamic programming approach. As CM is a relative value, it is not appropriate to define an index. Thus, the quality index S can be simply defined as the F-measure, once P and R values are obtained:

$$S = \frac{PR}{\alpha R + (1-\alpha)P} \quad (3)$$

Letting $\alpha = 0.5$, the S index is expressed as:

$$S = \frac{2PR}{P+R} \quad (4)$$

Experimental Results: In the experiments, the Lena image was degraded by: impulsive salt and pepper noise, additive Gaussian noise, multiplicative speckle noise, mean shift, contrast stretching, blurring, and JPEG compression. The degraded images except mean-shifted image were prepared to have a similar SSIM value. The shift of the mean is still in the range $[0,255]$. Thus, SSIM index identifies a similar quality as original Lena image. Following the procedure described in previous section, the contour mapping measure and S index are calculated and listed in Table 1.

Table 1. Quality assessment of image "Lena".

Image	Distortion Type	SSIM	CMM	S
Lena1	Impulsive salt & pepper	0.6395	10.61	0.6764
Lena2	Additive Gaussian	0.6392	19.46	0.6532
Lena3	Speckle noise	0.6400	10.14	0.7144
Lena4	Mean shift	0.9547	8.04	0.7629
Lena5	Contrast stretch	0.6398	18.86	0.4561
Lena6	Blurring	0.6395	31.21	0.1532
Lena7	JPEG compression	0.6395	26.57	0.3279

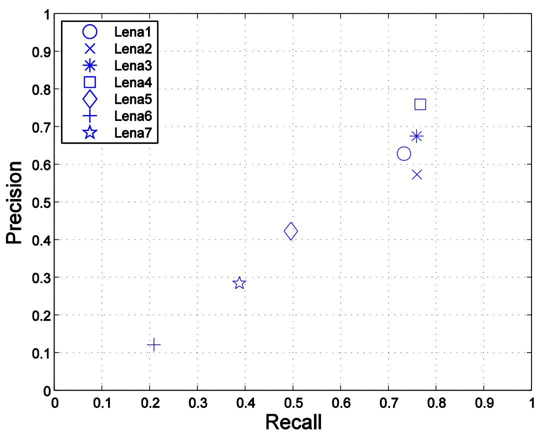

Figure 1. Precision and recall plot for Lena images.

For the precision-recall plot in Fig.1, high precision corresponds to a low false positive rate while high recall indicates a low false negative rate. For the speckle, salt & pepper, and additive Gaussian noises, the detected contours have comparable recall values, but with variations on precision, i.e. false positive, which leads to the final difference in S (F) value. The plot shows the clear differences between the Lena images undergoing the rest three operations.

For the mean-shifted image, the mean value is 99 while the mean of original Lena image is 122. To see the impact of mean value, the SSIM index is calculated between Lena and an image with uniform intensity value, i.e., a "flat" image without any contents, from 1 to 255. The SSIM curve is shown in Fig.2. The plot shows the maximum SSIM value 0.5450 is reached when the mean value is 127. The corresponding SSIM value of mean value 122 is 0.5447, which is near the maximum. However, these values are meaningless for image quality. The SSIM index does not work, but we can see how the mean value affects the SSIM index.

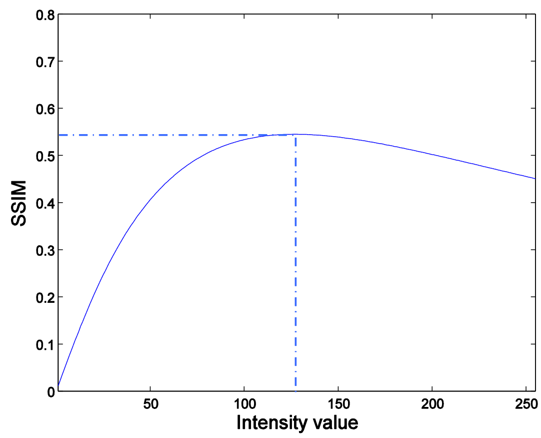

Figure 2. The impact of image mean value on SSIM index.

Conclusion: In this paper, an image quality index based on image segmentation and contour mapping measure is proposed. It provides a different perspective for the image quality assessment in comparison with the well-known SSIM index.

References:

[1] T.-J. Liu, W. Lin, and C.-C. Kuo, "Image quality assessment using multi-method fusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1793-1807, 2013.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: from Error Measurement to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 1, 2004.

[3] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Proceedings of 7th IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*, 2010, pp. 1-8.

[4] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898-916, 2011.