Accepted in the 11th International Conference on Information Sciences, Signal Processing and their Applications (ISSPA 2012), 2012

AN FFT-BASED VISUAL QUALITY METRIC ROBUST TO SPATIAL SHIFT

Guangyi Chen and Stéphane Coulombe

Department of Software and IT Engineering, École de technologie supérieure, Université du Québec,
1100 Notre-Dame Street West, Montréal, Quebec, Canada H3C 1K3.
Email: {Guangyi.Chen@etsmtl.ca, Stephane.Coulombe@etsmtl.ca}

ABSTRACT

In recent years, several metrics have been developed for measuring image visual quality, including the MSSIM and the visual information fidelity (VIF). However, these metrics are not robust to spatial shifts, meaning that when the reference and distorted images are misaligned by a few pixels, these metrics will produce very low scores, which is undesirable. In this paper, we extend the SSIM metric to make it robust to spatial shifts by first pre-processing the input images with the Fast Fourier transform (FFT). We then apply the magnitude of the transformed Fourier coefficients to the existing metrics because these coefficients are shift-invariant. Our assumption is that if we shift the image by a small amount of pixels, then it will not affect the perceived quality. Experimental results show that the proposed method is attractive for measuring the visual quality of 2D images as it is far less complex than the current approach, which consists in performing global motion estimation to align the input images prior to applying the metrics, and offers better accuracy.

Keywords: *Fast Fourier transform (FFT); robustness to spatial shift; image visual quality; quality metrics.*

1. INTRODUCTION

Measuring the visual quality of an image has been a very interesting topic for many years. There are three types of image quality metrics: full-reference (FR), reduced-reference (RR), and no-reference (NR). The FR metrics, which require reference images to be available, are the most popular. RR and NR metrics require partial or no information about the reference images, and are therefore hard to develop. The most popular visual quality metrics include the following FR metrics: the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), and the visual information fidelity (VIF). The peak signal-to-noise ratio (PSNR) is a metric for measuring the quality of two images, *x* and *y,* of size *m*×*n*, and is defined as:

$$PSNR\ (x,y) = 10\log_{10}\left(\frac{255^{2}}{\frac{1}{mn}\sum_{i,j}(y(i,j)-x(i,j))^{2}}\right). \quad (1)$$

However, it is not a good image quality metric as it is not well matched to the perceived image quality. Wang et al. developed the structural similarity (SSIM) index [1] by comparing local correlations in luminance, contrast, and structure between reference and distorted images. The SSIM index is defined as:

$$SSIM(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1}\times\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2}\times\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad (2)$$

where μ_x and μ_y are sample means of images *x* and *y*, σ_x^2 and σ_y^2 are sample variances, and σ_{xy} is the sample cross-covariance between *x* and *y*. The constants C_1, C_2 and C_3 stabilize SSIM when the means and variances become small. The mean SSIM (MSSIM) over the whole image gives the final quality measure. Rezazadeh and Coulombe [2] developed a novel discrete wavelet transform framework for full reference image quality assessment, while Qian and Chen [3] proposed four reduced-reference metrics for measuring hyperspectral images after spatial resolution enhancement.

Sheikh and Bovik [4] developed a visual information fidelity (VIF) index for full-reference measurement of image visual quality. Let $\boldsymbol{e}=\boldsymbol{x}+\boldsymbol{n}$ be the reference image**,** and $\boldsymbol{n}$ the noise with zero-mean normal distribution $N(0,\sigma_n^2 I)$. Also, let $\boldsymbol{f}=\boldsymbol{y}+\boldsymbol{n}'=\boldsymbol{g}\boldsymbol{x}+\boldsymbol{v}'+\boldsymbol{n}'$ be the test image, where $\boldsymbol{g}$ represents the blur, $\boldsymbol{v}'$ the additive zero-mean Gaussian white noise with covariance $\sigma_n^2 I$, and $\boldsymbol{n}'$ the noise with zero mean normal distribution $N(0,\sigma_n^2 I)$. Then, VIF can be computed as the ratio of the mutual information between $\boldsymbol{x}$ and $\boldsymbol{f}$, $I(x,f\mid z)$, and the mutual information between $\boldsymbol{x}$ and $\boldsymbol{e}$, $I(x,e\mid z)$, for all wavelet subbands except the lowest approximation subband:

$$VIF(x,y) = \frac{\sum I(x;f\mid z)}{\sum I(x;e\mid z)} \quad (3)$$

These metrics however assume that the original and distorted images are spatially aligned and perform poorly, when that is not the case. One solution to this problem is to first align the two images, using the correlation, prior to applying the quality metrics. The problem though is that correlation pre-processing is computationally very complex: either the correlation spatial search zone is too small, and proper alignment cannot be achieved or the search zone is too large, and requires a lot of computations.

In this paper, we introduce a new pre-processing method for measuring the visual quality of 2D images. We propose to take the FFT to the two input images, and then apply standard visual quality metrics to the FFT spectra. We can thus obtain improved metric values when compared with standard metrics such as MSSIM. Our method is good at measuring the visual quality of an image that has different kinds of distortions. Experimental results show that our proposed method outperforms existing MSSIM metric in a number of test cases, even if they are preceded by a spatial alignment stage.

The rest of this paper is organized as follows. Section 2 proposes our pre-processing method by taking the Fourier spectra of the two input images. We then apply the standard image quality metrics to the generated Fourier spectra in order to obtain improved metric values. Section 3 conducts some experiments in order to show the advantages of our proposed method over standard image quality metrics. Finally, Section 4 concludes the paper.

2. PROPOSED METHOD

Existing metrics for measuring image visual quality do not perform well when the target images have certain specific distortions. For example, if the distorted image is a spatially shifted version of the original image by a few pixels, then humans will likely score this distorted image very high. However, existing metrics will give much lower scores than humans, as will be demonstrated in the experimental section. In this section, we propose a novel method for pre-processing the two input images by taking the Fourier transform and obtaining the magnitude of the Fourier coefficients (spectra), and then applying existing metrics to the Fourier spectra. Proceeding as such, we obtained improved metrics for measuring image visual quality. One important property of the Fourier spectra is that they are invariant to image spatial shifts, which is very useful in tackling the problem at hand.

It is well known that low frequency Fourier coefficients contain importation features of the input image while those with frequencies that are too high contain unstable features. We therefore chose to use the center region [m/4:3m/4, n/4:3n/4] of the Fourier spectra as the input to the existing MSSIM metric, where the image size is $m \times n$ pixels.

$$abs_X = abs_X[m/4:3m/4, n/4:3n/4] \quad (4)$$

$$abs_Y = abs_Y[m/4:3m/4, n/4:3n/4] \quad (5)$$

Since the size of the center region is one-fourth that of the original image, much less CPU time will be required for our proposed method than for the MSSIM.

The proposed method for measuring the visual quality of a pair of images, X and Y, can be summarized as follows:

1. Take the 2D FFT of both images, X and Y, and we obtain $F_X = FFT_2(X)$ and $F_Y = FFT_2(Y)$.
2. Get the magnitude of F_X and F_Y, and we have $abs_X = |F_X|$ and $abs_Y = |F_Y|$.
3. Extract the center region of abs_X and abs_Y according to equations (4) and (5).
4. Calculate MSSIM(abs_X, abs_Y) according to equation (2).

It may be suggested to first compensate the global spatial shift, and then apply standard MSSIM image visual quality metric to the compensated image. This approach should work for the global spatial motion of the test images. We selected the global motion compensation (MC) method proposed in [5], which is a simple, fast, and reliable method providing integer pixel precision. In that paper, $X(v)$ and $Y(v)$ are two images, with v being a spatial integer index vector for the underlying 2D rectangular lattice. Also,

$$F_X = FFT_2(X) \text{ and } F_Y = FFT_2(Y). \quad (6)$$

Then, they defined the cross-correlation function defined as:

$$k(v) = FFT_2^{-1}(F_X F_Y^*), \quad (7)$$

where FFT_2^{-1} is the inverse 2D Fourier transform and * means it is a complex conjugate. The estimated motion vector is given as:

$$v_{opt} = \arg\max_v k(v). \quad (8)$$

We conducted experiments for the case of MC+MSSIM, and found out that our proposed method outperforms MC+MSSIM in terms of accuracy; furthermore, our proposed method requires much less CPU time than MC+MSSIM.

The major contribution of this paper is as follows. We have generalized the MSSIM spatial domain metric to the frequency domain, and our new method outperforms the spatial domain metric significantly in the presence of a spatial shift. Our proposed method is well suited for images with many kinds of distortions. In addition, the FFT is very fast and accurate. Therefore, our proposed method is attractive for measuring image visual quality. We know that the original MSSIM is not robust to spatial shifts and we do not expect it to be. By applying MSSIM to the magnitude of the FFT coefficients of the reference and test images, we make the MSSIM robust to spatial shifts. This is an advantage of our proposed method in this paper.

3. EXPERIMENTAL RESULTS

We conducted experiments on the LIVE Image Quality Assessment Database, Release 2 [6], which consists of 779 distorted images derived from 29 original images using five types of distortion. The distortions include JPEG compression, JPEG2000 compression, Gaussian white noise, Gaussian blurring, and the Rayleigh fast fading channel model. In this experiment, we considered three performance measures, namely, the correlation coefficients (CC) between the difference mean opinion score (DMOS) and the objective model outputs after nonlinear regression, the root mean square error (RMSE), and the Spearman rank order correlation coefficient (ROCC). Table 1 shows the experimental results for this database. It can be seen that our proposed method outperforms and is much faster than the MSSIM standard metric. The execution time for the metrics with global motion compensation (MC) is the longest among the methods compared in this paper, except for VIF, which exhibits the highest accuracy but at a cost of very high computational complexity. Fig. 1 shows the scatter plots of DMOS versus MSSIM. It can be seen that our proposed method is better than the SSIM visual quality metric even when pre-processing of the images by global motion compensation (MC) is used in our experiments.

We also tested the performance of our proposed method using an augmented image database obtained by shifting every image in the LIVE database by 0, 2, 4, 6, 8, 10 pixels in both the horizontal and vertical directions. The augmented database thus contained 6 times more distorted images than the original LIVE database. In these simulations, we made the assumption that for such small spatial shifts, the human score (DMOS) will be the same as for the distorted image, but without additional spatial shift. Note that for a spatial shift of W pixels, we need to reduce the horizontal and vertical dimensions of both the reference (keep the top-left part) and the distorted images (keep the bottom-right part) by W pixels. Therefore, for large spatial shifts, the proposed methodology would not be appropriate as the contents of the reference and distorted images would be too different (due to the cropping of different regions of the image). The results are presented in Table 2. Fig. 2 shows the scatter plots of DMOS versus MSSIM, by shifting 0, 2, 4, 6, 8, 10 pixels in both the horizontal and vertical directions. The time was measured using un-optimized Matlab implementations of the metrics on a Dell PC with a 3.40GHz CPU and 12GB of memory. It can be seen that our proposed method outperforms the original MSSIM in this case, which indicates that our proposed method is valid in measuring image visual quality even if the images are not perfectly aligned.

The proposed pre-processing method does not work for VIF because the VIF is based on different principles from SSIM, and so we do not include the experimental results of the new VIF in this paper. As shown in Tables 1 and 2, while the original VIF achieves state-of-the-art performance, it does however require nearly ten times as much CPU time as MSSIM. As a consequence, there is a trade-off between accuracy and speed when choosing a metric in real-life applications.

4. CONCLUSIONS

Measuring the visual quality of an image or a video sequence is a very challenging task. The existing metric used for this purpose includes MSSIM, which does not perform well when there are spatial shifts in the images to be measured. In this paper, we have proposed a novel pre-processing method for measuring image visual quality. We take the Fourier spectra of the two images as input to existing metric MSSIM. Our proposed method can handle the situations very well when the target images are distorted. This advantage is attributable to the fact that the spectra of the Fourier transform are invariant to spatial shifts of the input images. Our experimental results show that our proposed pre-processing step improves the measuring scores when compared with existing MSSIM metric without pre-processing. In addition, our proposed method utilizes much less CPU time than the existing MSSIM metric.

Further investigation needs to be carried out in order to improve our proposed pre-processing method, by considering both rotation invariant and scale invariant metrics for measuring image visual quality. It is well known that spatial domain metrics, including the MSSIM and VIF, are very sensitive to translation, scaling, and rotation of images. In order to overcome these limitations, Sampat et al. [7] proposed a new complex wavelet image similarity index, CW-SSIM, which is robust to small rotations and translations. However, when the spatial shifts become more significant in scope, CW-SSIM does not perform well. Our proposed method is more robust because the Fourier transform is shift-invariant, and we apply it on the whole image. Based on this work, we will attempt to propose new metrics based on the dual-tree complex wavelet transform (DTCWT). The DTCWT developed by Kingsbury [8] has an approximate shift invariant property and offers better orientation selectivity. These good properties could make the DTCWT a better candidate for invariant metrics for measuring image visual quality. We will also be studying the FFT-based technique proposed by Reddy and Chatterji [9] to achieve translation, rotation, and scale invariance, and then apply standard FR metrics to the normalized images. In addition, new metrics with the affine-invariant property are currently being developed by the authors of this paper.

5. ACKNOWLEDGEMENT

This work was supported by Vantrix Corporation and by the National Sciences and Research Council of Canada (NSERC) under the collaborative research and development program (NSERC-CRD 326637-05).

REFERENCES

[1] 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, 2004.

[2] S. Rezazadeh and S. Coulombe, "Novel discrete wavelet transform framework for full reference image quality assessment," *Signal, Image and Video Processing*, pp. 1-15, September 2011.

[3] S. E. Qian and G. Y. Chen, "Four reduced-reference metrics for measuring hyperspectral images after spatial resolution enhancement," *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vienna, Austria, pp. 204-208, July 5-7, 2010.

[4] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing,* vol. 15, no. 2, pp. 430-444, 2006.

[5] G. Varghese and Z. Wang, "Video denoising based on a spatiotemporal Gaussian scale mixture model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 7, pp. 1032-1040, 2010.

[6] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE image quality assessment database release 2," http://live.ece.utexas.edu/research/quality.

[7] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385-2401, Nov. 2009.

[8] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no 3, May 2001, pp. 234-253.

[9] B. S. Reddy and B. N. Chatterji, "An FFT-Based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing,* vol. 5, no. 8, pp. 1266-1271, 1996.

Table 1. Performance comparison of image quality assessment for all distorted test images in the LIVE database. The metrics used are MSSIM and VIF. Global motion compensation (MC) as a pre-processing step is also compared in our experiments. It can be seen that our proposed pre-processing method improves the measuring scores of MSSIM. The best performing method is highlighted in bold font (not considering VIF).

Quality Index	Method	CC	RMSE	ROCC	Time in Seconds
MSSIM	Original	0.9042	11.6703	0.9104	123.52
	MC	0.9040	11.6866	0.9104	166.58
	Proposed	**0.9131**	**11.1421**	**0.9214**	**86.44**
VIF	Original	0.9595	7.6945	0.9636	792.99
	MC	0.9593	7.7127	0.9635	846.19

Table 2. Performance comparison of image quality assessment for the augmented LIVE database obtained by shifting every image in the LIVE database by 0, 2, 4, 6, 8, 10 pixels in both the horizontal and vertical directions. The metrics used are MSSIM and VIF. Global motion compensation (MC) as a pre-processing step is also compared in our experiments. It can be seen that our proposed pre-processing method improves the measuring scores of MSSIM. The best performing method is highlighted in bold font (not considering VIF).

Quality Index	Method	CC	RMSE	ROCC	Time in Seconds
MSSIM	Original	0.2377	26.54	0.2007	539.97
	MC	0.9013	11.84	0.9074	856.86
	Proposed	**0.9067**	**11.52**	**0.9143**	**310.40**
VIF	Original	0.2473	26.47	0.1367	4634.52
	MC	0.9560	8.02	0.9607	4965.74

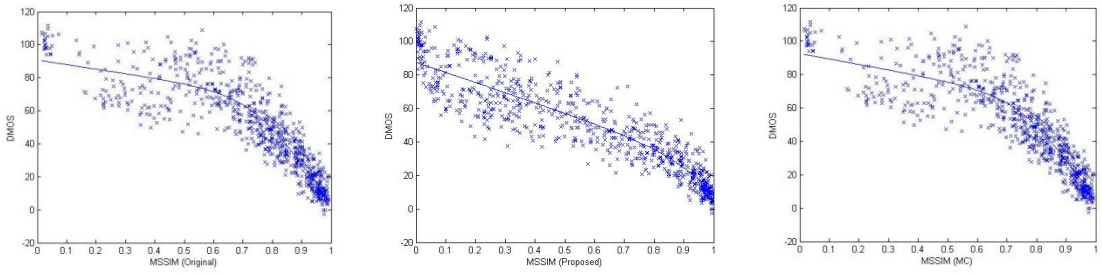

Fig. 1. The scatter plots of DMOS versus the original MSSIM (left), the proposed method (middle), and the MSSIM with global motion compensation (MC) (right) for all distorted images in the LIVE database.

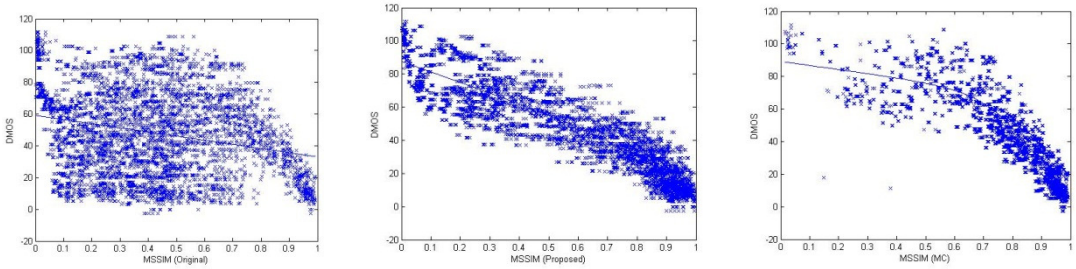

Fig. 2. The scatter plots of DMOS versus the original MSSIM (left), the proposed method (middle), and the MSSIM with global motion compensation (MC) (right) for the augmented LIVE database obtained by shifting every image in the LIVE database by 0, 2, 4, 6, 8, 10 pixels in both the horizontal and vertical directions.