Blind quality assessment of multi-focus image fusion algorithms

Rodrigo Nava[a], Boris Escalante–Ramírez[b] and Gabriel Cristóbal[c]

[a]Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, Mexico City, Mexico.
[b]Departamento de Procesamiento de Señales, Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico.
[c]Instituto de Óptica, Spanish National Research Council (CSIC), Serrano 121, Madrid 28006, Spain.

ABSTRACT

At present time, image fusion is widely recognized as an important aspect of information processing. It consists of combining information originated from several sources in order to improve the decision making process. In particular, multi-focus image fusion combines images that depict the same scene but they are not in-focus everywhere. The task seeks to reconstruct an image as sharp as possible by preserving in-focus areas while discarding blurred areas. The quality of fused images is of fundamental importance. Many objective quality metrics for image fusion have been proposed. However, the evaluation of fused images is still a difficult task, especially because there is no reference image to compare with. Blind image quality assessment refers to the problem of evaluating the visual quality of an image without any reference. In this paper, we describe a blind image fusion quality assessment procedure based on the use of mutual information (MI). This procedure is concise and explicit and will be useful in scenarios where the absence of a reference image can hamper the assessment of the results. Furthermore, several image fusion algorithms have been rated and they have shown that our metric is compliant with subjective evaluations. Consequently, it can be used to compare different image fusion methods or to optimize the parameter settings for a given fusion algorithm.

Keywords: Blind image quality assessment, mutual information, multi-focus image fusion

1. INTRODUCTION

Image quality is of fundamental importance in a large number of applications and it has been described in terms of the presence of visible distortions such as blur, block effects and noise. Although image quality assessment is a hardly new topic and plays an important role in various image–processing areas, only limited success has been achieved. A common way to measure image quality is based on early visual models but since human beings are the ultimate receivers in most applications, the most reliable way of assessing the quality of an image is by subjective evaluation.

In order to define what image quality is, we need to answer what images are used for. Some authors refer to image quality as fidelity and it is defined as the capacity of a process to render an image accurately. Both terms, image quality and image fidelity, are often used interchangeably. But this is a wrong idea. Image fidelity is the ability to detect differences between images. In contrast, image quality deals with subjective impressions and it is much more difficult to characterize. However, the most accepted idea about image quality is "the degree to which an image satisfies the requirements imposed on it"[1]

The subjective quality measure, mean opinion score (MOS), provides a numerical indication of the perceived quality. It has been used for many years and considered the best method for image quality. Nevertheless, MOS

Further author information: (Send correspondence to R.N.)
R.N.: E-mail: urielrnv@uxmcc2.iimas.unam.mx
B.E.: E-mail: boris@servidor.unam.mx, Telephone: (52)55-56161719
G.C.: E-mail: gabriel@optica.csic.es, Telephone: (34)91-5616800(x2313), Fax: (34)91-5645557

metric is inconvenient because it demands human observers, is expensive, and is usually too slow to apply in real–time scenarios. Moreover, quality perception is strongly influenced by a variety of factors that depend on the observer. For this reason, it is desirable to have an objective metric able to predict image quality automatically.

Many objective image quality metrics have been proposed. They are based on measuring physical characteristics and they intend to predict perceived image quality accurately and automatically. It means, that they should predict image's quality that an average human observer will report. Objective quality metrics attempt to quantify the visibility error between a distorted image and a reference image and they have several advantages. Firstly, they can be used to monitor quality control systems. Secondly, they can be employed to benchmark image processing algorithms and they are useful to optimize parameter settings.

Among the available ways to measure objective image quality, the mean square error (MSE) and peak signal–to–noise ratio (PSNR) are widely employed because they are easy to calculate and usually they have low computational cost, but such measures are not necessarily consistent with human observer evaluation.[2] Both MSE and PSNR reflect the global properties of the image quality but they are inefficient in predicting structural degradations. An illustrative example is shown in Fig. 1, where an original image was deformed by different types of distortion: Gaussian noise, impulse noise and Gaussian blur. PSNR and MSE values are given. Note that PSNR values regarding to the original image, Fig. 1(a), are nearly identical, even though the images present different visual quality.
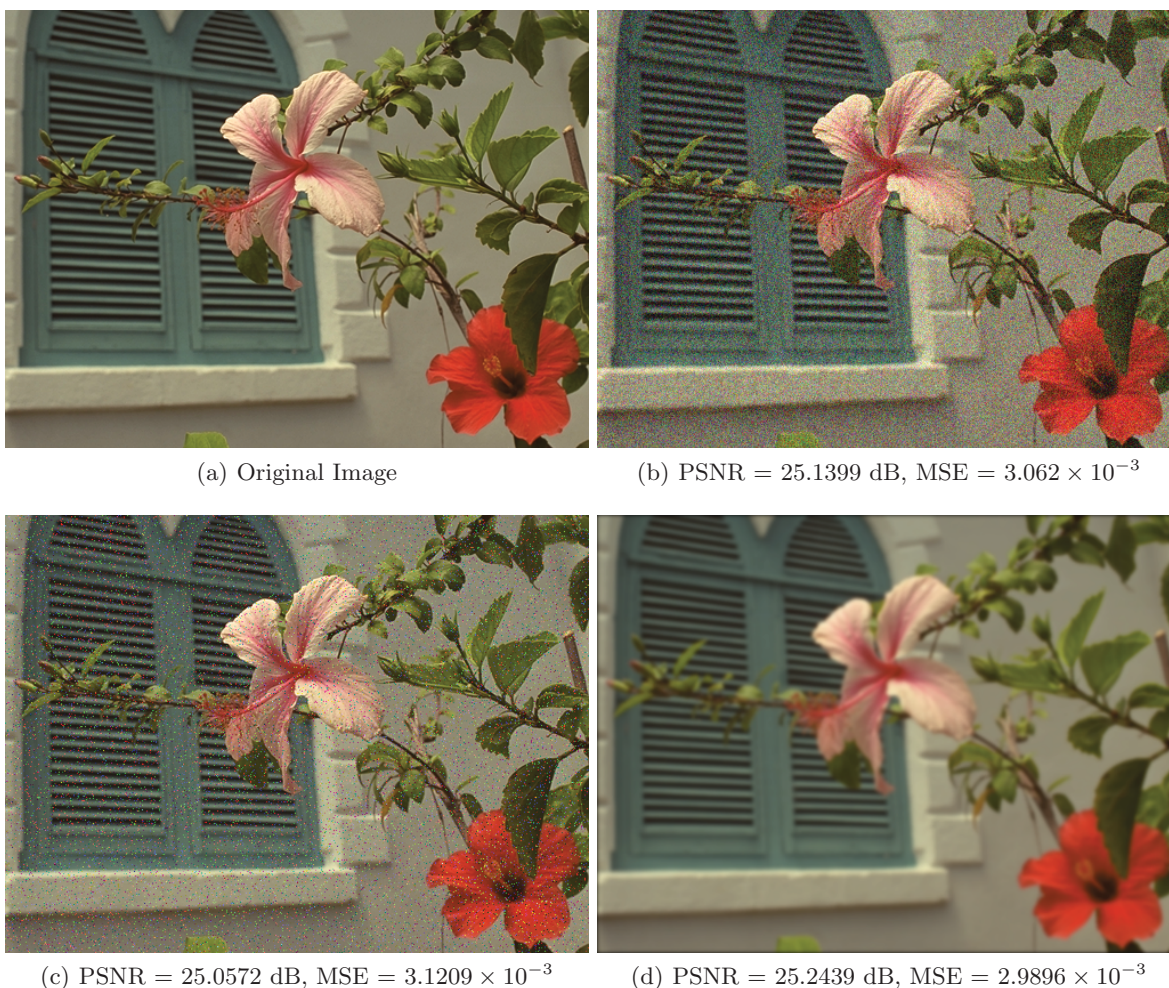

(a) Original Image

(b) PSNR = 25.1399 dB, MSE = 3.062×10^{-3}

(c) PSNR = 25.0572 dB, MSE = 3.1209×10^{-3}

(d) PSNR = 25.2439 dB, MSE = 2.9896×10^{-3}

Figure 1. Image quality comparison using PSNR for "flower" image deformed with different types of distortions. Reference image was taken from Ponomarenko's database.[3] Fig. 1(a) Reference image. Fig. 1(b) Gaussian noise. Fig. 1(c) Impulse noise. Fig. 1(d) Gaussian blur.

Other metrics have been proposed such as Wigner signal–to–noise ratio (SNR_W)[4] and structural similarity quality index (SSIM).[5] These approaches are based on some models derived from human visual system (HVS). However, all of these objective metrics require a reference image, (distortion free), together with the processed image in order to evaluate the visibility of artifacts. This imposes obvious limitations on the applications where such metrics can be used.

In many practical applications, image quality metrics do not have access to reference images. However, it is desirable to develop measurement approaches that can evaluate image quality blindly. Blind or non–reference image quality assessment turns out to be a very difficult task, because metrics are not related to the original image.[6]

In this paper, we considered multi-focus images, i.e, situations where two or more images that depict the same scene will not be in-focus everywhere (if one object is in–focus, another one will be out–of–focus). This occurs because there are sensors which cannot generate images of all objects at various distances with equal sharpness. The advantages of multi–focus images can be fully exploited by merging the sharply focused regions into one image that will be in-focus everywhere.[7]

The existing evaluation metrics of image fusion algorithms are broadly based on a measure of the fidelity from the input images to the fused output. Nevertheless, most methods are just informal tests over a few scenes and they did not consider HVS.[8–10]

In the following sections, we present an evaluation procedure for image fusion algorithms based on MI. Firstly, we give a brief definition of MI. Later, we present the proposed algorithm, and finally, we show some experimental results using both standard and blind image quality metrics.

2. MATHEMATICAL BACKGROUND

The introduction of MI in the field of images dates back to early 1997 when Viola *et al* introduced a registration measure for multi-modality images.[11] MI measures the degree of dependence between two random variables. The definition is related to the relative entropy or Kullback–Leibler Distance (KLD) as follows:

$$D_{KL}(A||B) = \sum_i A(i) \log \frac{A(i)}{B(i)} \tag{1}$$

for two random variables with distribution functions $A(i)$ and $B(i)$.

MI is a special case of the measured distance between the joint probability density function of two random variables and the product of their marginal probabilities. From (1) we have:

$$I(A, B) = \sum_{a,b} p(a, b) \log \frac{p(a, b)}{p(a)p(b)} \tag{2}$$

where $p(a, b)$ is the joint distribution function and $p(a)$ and $p(b)$ are the marginal probability functions.

The Viola's theory is that there is a maximal dependence between pixel gray values of two images when they are correctly aligned. The fewer registration errors, the greater the MI value.

2.1 Mutual Information Properties

MI has the following properties:

1. $I(A, B) \geq 0$ MI is non negative.

2. $I(A, B) = 0$ if and only if both random variables A and B are independent. When A and B are not in any way related, no knowledge is gained about one random variable when the other is given.

3. $I(A, B) = I(B, A)$ Symmetry property. However, MI is not symmetric in practice. Some techniques such as interpolation for registering different number of samples, can result in differences in outcome when registering A to B or B to A

4. $I(A, A) = H(A)$ The information content of a random variable A about itself is equal to the entropy of random variable A.

5. $I(A, B) \leq H(A)$ and $I(A, B) \leq H(B)$ The information the random variables contain about each other cannot be greater than the information from the random variables themselves.

2.2 Joint and Marginal Distributions

Let us consider the image intensity values a and b of a pair of corresponding pixels in two images as a random variables A and B. Marginal and joint distribution estimations $p(a)$, $p(b)$ and $p(a, b)$ respectively, can be obtained by normalization of the joint and marginal histograms of both images as follows:

$$p(a) = \sum_b p(a, b) \tag{3}$$

$$p(b) = \sum_a p(a, b) \tag{4}$$

$$p(a, b) = \frac{h(a, b)}{\sum_{a,b} h(a, b)} \tag{5}$$

where h es the joint histogram of the image pair given by a 2D matrix of the following form:

$$h(a, b) = \begin{pmatrix} h(0, 0) & h(0, 1) & \ldots & h(0, N-1) \\ h(1, 0) & h(1, 1) & \ldots & h(1, N-1) \\ \vdots & \vdots & & \vdots \\ h(M-1, 0) & h(M-1, 1) & \ldots & h(M-1, N-1) \end{pmatrix} \tag{6}$$

under the assumption that the intensity values in the first image may vary from 0 to $M-1$ and in the second image from 0 to $N-1$.

The value $h(a, b)$ is the number of corresponding pairs having intensity value a in the first image and intensity value b in the second image. From Eq. (3) to Eq. (5) we can be concluded that the joint histogram is the only requirement to determine the MI between two images.

3. DESCRIPTION OF THE METHOD

Let us consider the case of two input images, A and B. Fused image F should contain important information from the original image set. Since MI is the information amount that one image contains over the other, we can calculate the amount of information that image F contains from both input images A and B using Eq. (2).

$$I(F, A) = \sum_{f,a} p(f, a) \log \frac{p(f, a)}{p(f)p(a)} \tag{7}$$

$I(F, A)$ is the amount of information that fused image F contains from input image A. In a similar way, we can calculate $I(F, B)$.

$$I(F, B) = \sum_{f,b} p(f, b) \log \frac{p(f, b)}{p(f)p(b)} \tag{8}$$

$I(F, B)$ is the amount of information that fused image F contains from input image B. Therefore, the performance of image fusion algorithms can be defined as:

$$M(F, A, B) = I(F, A) + I(F, B) \tag{9}$$

Eq. (9) reflects the total amount of information that fused image F contains from both input images A and B. Nevertheless, the metric defined above is not enough, because it has significant disadvantages, mainly due to the fact that it assigns the highest value to the average method (AM). AM fuses images by averaging them, but is well known that AM is not correlated with subjective procedures, because introduces contrast loss.[12]

Cvejic et *et al* proposed a modified metric using Tsallis entropy (TE) in order to get a more accurate model that transfers information from input images into the fused output.[13] TE is a generalization of the standard Boltzmann–Gibbs entropy. It is defined as:

$$H^q = \frac{1}{1-q} \left(\sum_x p^q(x) - 1 \right) \tag{10}$$

in this case, p denotes the probability distribution function of interest and the parameter q is a measure of the non–extensitivity property of the system of interest.[*]

Besides that, Tsallis proposed a divergence measure that represents the degree of dependence between two discrete random variables defined by:

$$D\left(P||R\right) = \frac{1}{1-q} \left(1 - \sum_x \frac{p_x^q}{r_x^{q-1}} \right) \tag{11}$$

with $q \in \mathbb{R} - \{1\}$.

Replacing p_x and r_x in Eq. (11) by the joint probability function and the product of the marginal densities between fused image F and input images A and B respectively, we have:

$$I^q(F, A) = \frac{1}{1-q} \left(1 - \sum_{f,a} \frac{p(f,a)^q}{(p(f)p(a))^{q-1}} \right) \tag{12}$$

and in the same way

$$I^q(F, B) = \frac{1}{1-q} \left(1 - \sum_{f,b} \frac{p(f,b)^q}{(p(f)p(b))^{q-1}} \right) \tag{13}$$

Replacing Eqs. (12) and (13) into Eq. (9), the image fusion performance measure can be defined as:

$$M^q\left(F, A, B\right) = I^q(F, A) + I^q(F, B) \tag{14}$$

3.1 Normalized Version of the Proposed Method

Now, we turn into a normalized version definition of this metric. Image entropy measures how well we are able to predict the intensity at an arbitrary point within the image. If there is no uncertainty about pixel value then the entropy is zero. This means that the image is completely homogeneous (flat image). On the other hand, if the image has higher entropy value then the image consists of a large number of intensities, which all have the same probability. The MI between two images expresses how much the uncertainty on one of the images decreases when the other one is known.

We assumed that a good image fusion algorithm must preserve the information between source images. This metric represents how much information was obtained through the fusion process. Input images can be used to assess the performance of different image fusion rules. Since no assumptions is made, the measure is very general in the sense that it can be applied in many fusion scenarios. Besides that, the size of the overlapping part among

---

[*]An intensive property is a physical property of the system that does not depend on the system size or the amount of material in the system. By contrast, an extensive property of a system does depend on the system size or the amount of material in the system. Non-extensive systems are those for which the thermodynamic properties do not scale linearly with the system size.

the images influences the MI measure. A decrease in such overlapping part diminishes the number of samples, which reduces the statistical strength of the probability distribution estimation.

The image fusion process can be seen as an information transfer problem in which two images are combined into a new one that should contain all the information from the original inputs.[14]

However, not all of the available input information is transferred into a fused image and some loss of information from the input images take place. At the same time, image fusion process creates additional information. In order to get the maximal amount of information it is imperative to quantify the exclusive information gained by the fused image F.

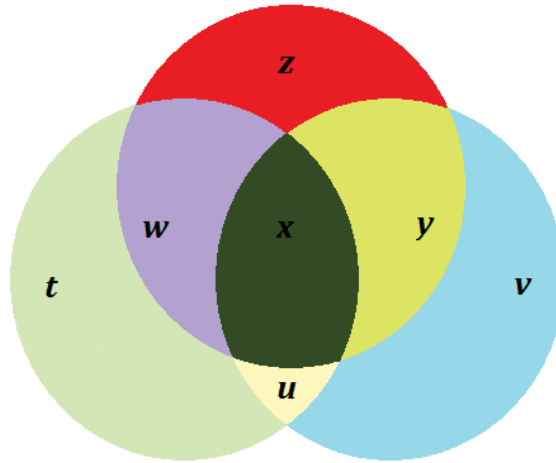For a better understanding of the method, we used Venn diagram notation, see Fig. 2.

Figure 2. z area represents gained information (GI) or artifacts introduced into the fused image by the process. This exclusive information gained by the fused image is given by the union of both input images A and B subtracted from fused image F. t, u and v areas represent loss information (LI) during the fusion process. LI is available information in both input images A and B but not in the fused image F. w area represents exclusive information from image A available in the fused image F. y area represents exclusive information from image B available in the fused image F. x area represents redundant information from both input images A and B.

From Fig. 2, we can get the total amount of information as follows:

$$MAX^q\,(F,A,B) = H^q\,(A) + H^q\,(B) + H^q\,(F) - I^q\,(A,B) - I^q\,(F,A) - I^q\,(F,B) + 2I^q\,(F,A,B) \qquad (15)$$

where $I^q\,(F,A,B)$ is the joint entropy between the input images and the fused image giving by:

$$I^q\,(F,A,B) = \frac{1}{1-q}\left(1 - \sum_{f,a,b}\frac{p(f,a,b)^q}{(p(f)p(a)p(b))^{q-1}}\right) \qquad (16)$$

Using the quotient between Eq. (14) and Eq. (15) we can define the normalized metric of order q.

$$NM^q\,(F,A,B) = \frac{M^q\,(F,A,B)}{MAX^q\,(F,A,B)} \qquad (17)$$

rewriting Eq. (17)

$$NM^q\,(F,A,B) = \frac{I^q(F,A) + I^q(F,B)}{H^q\,(A) + H^q\,(B) + H^q\,(F) - I^q\,(A,B) - I^q\,(F,A) - I^q\,(F,B) + 2I^q\,(F,A,B)} \qquad (18)$$

4. EXPERIMENTAL RESULTS

We present two experiments. In the first one, *experiment #1*, we used an artificial ground truth with the purpose of compare our method with other objective quality metrics. We considered grayscale images A and B of size 256×256 with 8 bits per pixel. In the second test, *experiment #2*, we do not have any reference images. We evaluated two different fusion schemes only with the proposed algorithm. This is a real scenario where our metric will be very useful due to the fact we do not have a reference image to compare with. Therefore we cannot apply standard objective metrics.

4.1 Experiment with an Artificial Reference Image

In *experiment #1*, we introduced defocus blur in two complementary regions of the original image Fig. 3(a). Fig. 3(b) and Fig. 3(c) represent both input images A and B respectively.

(a) Ground truth

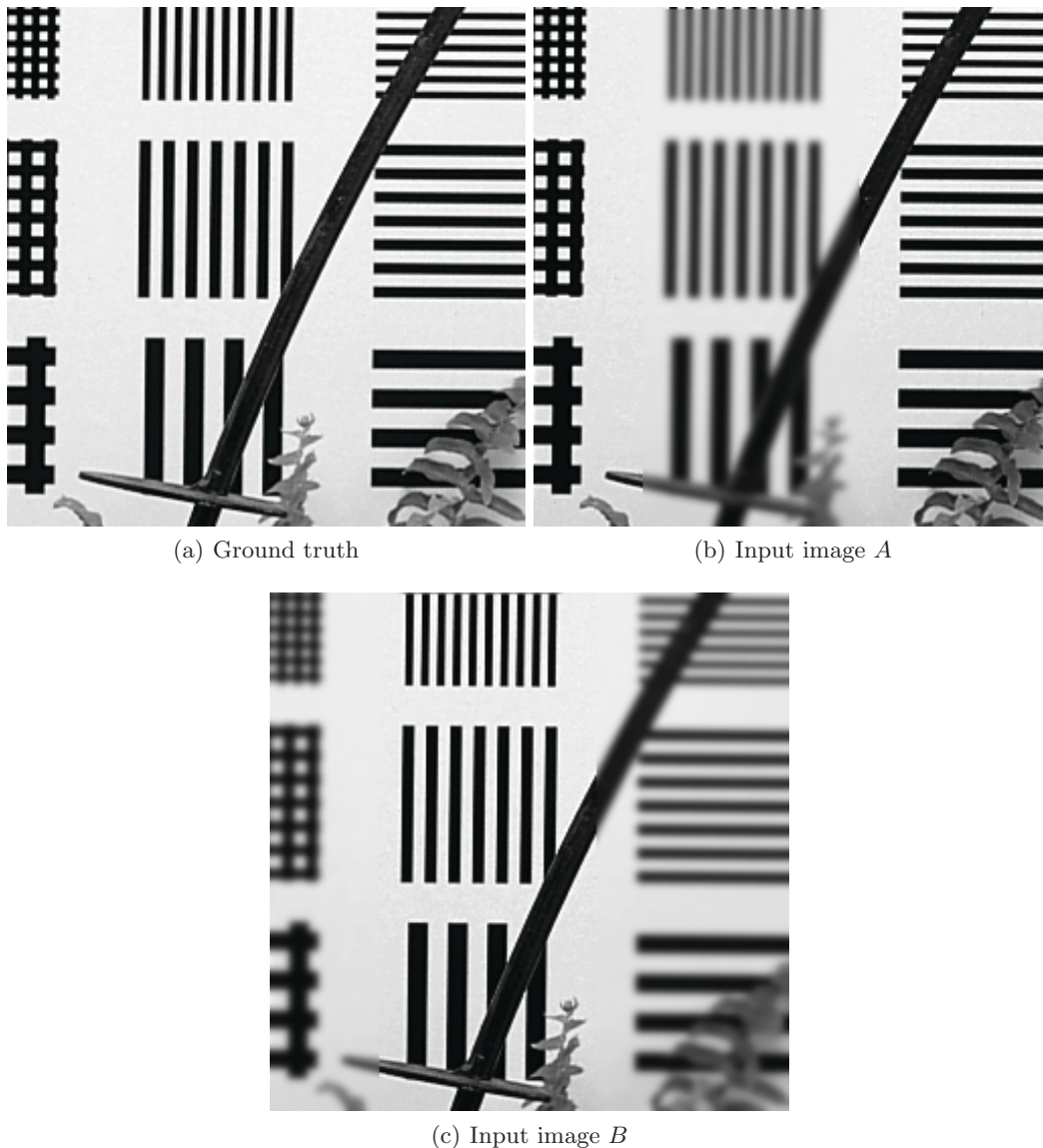(b) Input image A

(c) Input image B

Figure 3. Images for the experiment #1. Fig. 3(a), shows the artificial ground truth (distortion free), Fig. 3(b) and Fig. 3(c) are blurred versions of it.

After registration process, input images were combined using the discrete wavelet transform (DWT). We generated three fused images using three different wavelet functions. In *scheme #1.1*, we used Daubechies wavelet $(Db_2)^\dagger$, the result appears in Fig. 4(a). In *scheme #1.2*, we used Symlet wavelet $(Sym_4)^\ddagger$, the result appears in Fig. 4(b). In *scheme #1.3*, we used Coiflet wavelet $(Cf_6)^\S$, the fused image appears in Fig. 4(c). Details of the fusion algorithm can be consulted in Nava *et al.*[12]

(a) Result for *scheme #1.1*

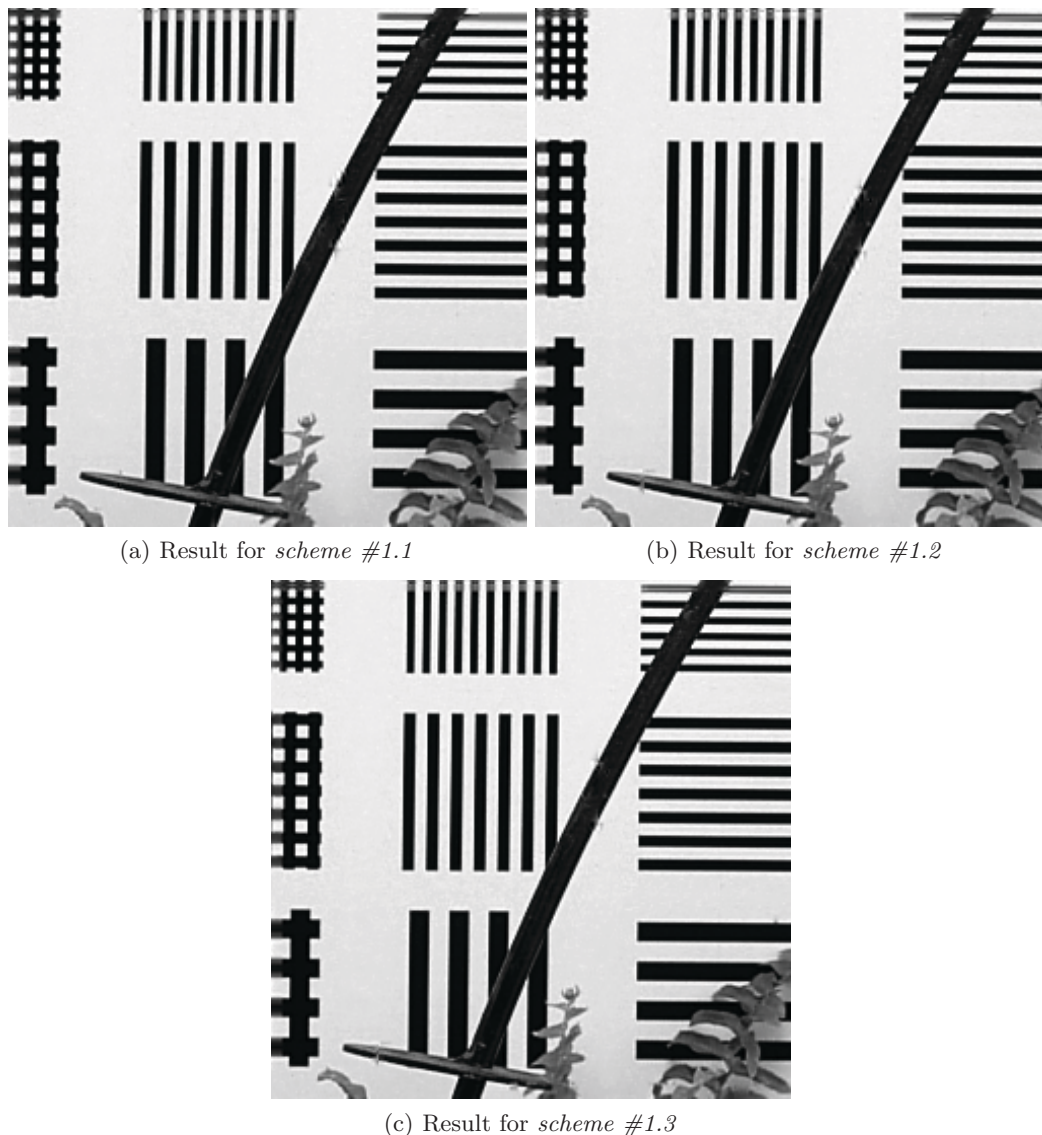(b) Result for *scheme #1.2*

(c) Result for *scheme #1.3*

Figure 4. Fused images for the *experiment #1*. Fig. 4(a) Fused image using Daubechies wavelet (DB_2). Fig. 4(b) Fused image using Symlet wavelet (Sym_4). Fig. 4(c) Fused image using Coiflet wavelet (Cf_6).

Table 1 presents fusion performance results of the three schemes in the *experiment #1* using both standard objective metrics such as PSNR and MSE and perceptual quality metrics like SNR_W,[4] SSIM[5] and Rényi Metric (RM).[15]

---

†Coefficients for Db$_2$: $a_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}$, $a_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}$, $a_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}$ and $a_4 = \frac{1-\sqrt{3}}{4\sqrt{2}}$.

‡Coefficients for Sym$_4$: $a_0 = -0.0758$, $a_1 = -0.0296$, $a_2 = 0.4976$, $a_3 = 0.8037$, $a_4 = 0.2979$, $a_5 = -0.0992$ and $a_6 = -0.0126$ and $a_7 = 0.0322$.

§Coefficients for Cf$_6$: $a_0 = -0.0157$, $a_1 = -0.0727$, $a_2 = 0.3849$, $a_3 = 0.8526$, $a_4 = 0.3379$ and $a_5 = -0.0727$.

Table 1. Results of the fusion performance for all schemes of the *experiment #1*.

Fusion	PSNR	MSE	SNR_W	SSIM	RM
#1.1	26.2837	0.0023	11.0328	0.9784	0.9950
#1.2	26.2222	0.0023	11.0803	0.9807	0.9974
#1.3	26.3830	0.0022	11.1387	0.9808	1

Table 2 presents results of the proposed metric for all fusion schemes. We have set the value of $q = 0.1$ to adjust the objective metric NM_{FAB}^q with subjective tests.

Table 2. Results of the fusion performance for the three fusion schemes in *experiment #1* using the proposed metric based on MI.

Fusion	$I^q(F,A)$	$I^q(F,B)$	$M^q(F,A,B)$	$NM^q(F,A,B)$
#1.1	0.8027	0.8184	1.6211	0.8781
#1.2	0.8033	0.8225	1.6258	0.8806
#1.3	0.8038	0.8223	1.6261	0.8808

From Table 2 we can see that scheme #1.3 got the highest value. This rate agrees with values in Table 1, where all metrics rated scheme #1.3 as the better fusion method.

4.2 Experiment with no Artificial Reference Image

In the next *experiment #2*, we used two grayscale images. Image A, Fig. 5(a), and image B, Fig. 5(b), that were taken in our lab, with different focal planes.

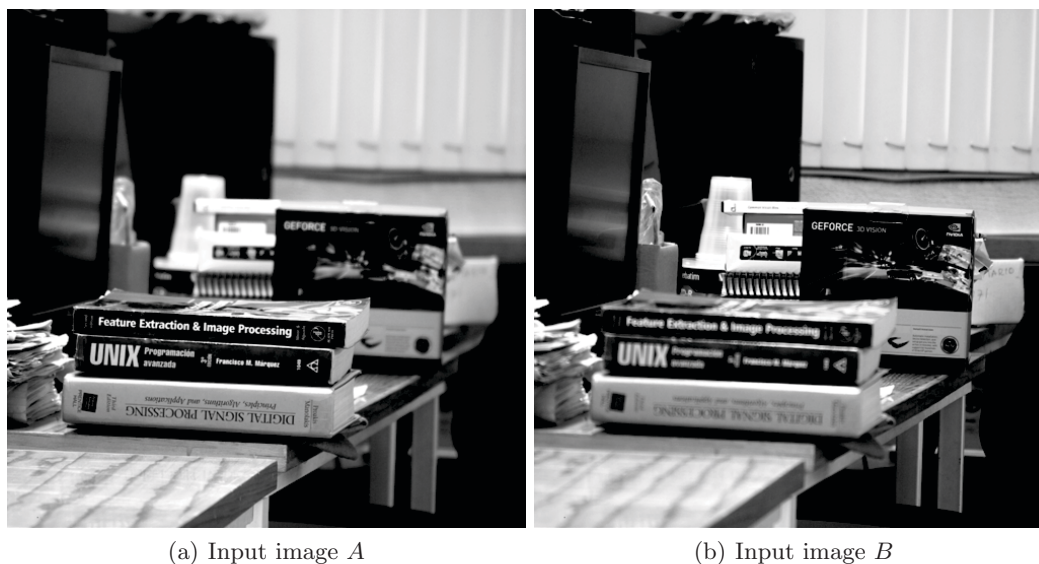(a) Input image A (b) Input image B

Figure 5. Multi-focal input images for the *experiment #2*. Fig. 5(a) has the three books in-focus. Fig. 5(b) shows three blurred books.

We merged both images using two different fusion algorithms. In *scheme #2.1* we used the algorithm which appears in Nava *et al*, Fig. 6(a).[12] This algorithm has shown better results in terms of visual quality than traditional fusion algorithms based on pixels. In *scheme #2.2* we fused images using traditional average

method, which introduces contrast loss, Fig. 6(b). Consequently, we expect that the proposed metric assigns a higher rate to *scheme #2.1*.

<table>
<tr><td>(a) Result for scheme#2.1</td><td>(b) Image result for scheme#2.2</td></tr>
</table>

Figure 6. Fused images for *experiment #2*. Fig. 6(a), shows result of the algorithm proposed by Nava *et al.* Fig. 6(b) was obtained by averaging pixels. Clearly, we can see contrast loss because of the pixel average.

Table 3 shows the evaluation of the *experiment #2*. We can see that edges in *scheme #2.1* are sharper than in *scheme #2.2*. Besides, contrast in *scheme #2.1* is better than in *scheme #2.2*. In fact, in *scheme #2.2* the whole image appears defocused because of the pixel average.

Table 3. Results of the fusion performance for *experiment #2*

Fusion	$I^q(F, A)$	$I^q(F, B)$	$M^q(F, A, B)$	$NM^q(F, A, B)$
#2.1	0.8927	0.8584	1.7511	0.9562
#2.2	0.8233	0.7825	1.6058	0.8768

As a result, the proposed method assigns a higher value to *scheme #2.1*. This rated is correlated with visual aspects. Fig. 6(a) is sharper than Fig. 6(b), then it has large number of intensities. On the contrary, Fig. 6(b) is a more homogeneous version, which produce less MI value.

5. CONCLUSIONS

In this paper we proposed a normalized metric for image fusion based on the mutual information and Tsallis entropy. The metric calculation is concise and explicit and it will be useful for real-time applications and industries. It corresponds to other perceptual quality metrics such as SNR_W, PSNR, SSIM and RM. Two experiments has been studied using the proposed metric and the results have shown that they agree with perceptual assessments. Lack of reference image or ground truth necessity is main advantage of this method. Another advantage of the proposed method is that calculation of MI is easy because only depends on the image histograms. We demonstrated by examples that MI increases with contrasts and resolution. One drawback of this method is the fact that works in a pairwise manner (two images). MI of multiple random variables requires further investigations.

ACKNOWLEDGMENTS

REFERENCES

[1] Silverstein, D. A. and Farrell, J. E., "The relationship between image fidelity and image quality," *International Conference on Image Processing* **1**, 881–884 (1996).

[2] Wang, Z. and Bovik, A. C., "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine* **26**(1), 98–117 (2009).

[3] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., and Battisti, F., "Tid2008 - a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics* **10**(4), 30–45 (2009).

[4] Beghdadi, A. and Iordache, R., "Image quality assessment using the joint spatial/spatial-frequency representation," *EURASIP Journal on Applied Signal Processing.* **2006**, 40–40 (2006).

[5] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[6] Nava, R., Cristóbal, G., and Escalante-Ramírez, B., "Nonreference image fusion evaluation procedure based on mutual information and a generalized entropy measure," *Bioengineered and Bioinspired Systems III* **6592**(1), 659215, SPIE (2007).

[7] Gabarda, S. and Cristóbal, G., "Multifocus image fusion through pseudo-wigner distribution," *Optical Engineering* **44**(4), 047001 (2005).

[8] Cvejic, N., Seppanen, T., and Godsill, S. J., "A nonreference image fusion metric based on the regional importance measure," *IEEE Journal of Selected Topics in Signal Processing* **3**(2), 212–221 (2009).

[9] Dumic, E., Grgic, S., and Grgic, M., "New image quality measure based on wavelets," *Wavelet Applications in Industrial Processing VI* **7248**(1), 72480G, SPIE (2009).

[10] Kaplan, L. M., Burks, S. D., Blum, R. S., Moore, R. K., and N., Q., "Analysis of image quality for image fusion via monotonic correlation," *IEEE Journal of Selected Topics in Signal Processing,* **3**(2), 222–235 (2009).

[11] Viola, P. and III, W. M. W., "Alignment by maximization of mutual information," *International Journal of Computer Vision* **24**(2), 137–154 (1997).

[12] Nava, R., Escalante-Ramírez, B., and Cristóbal, G., "A novel multi-focus image fusion algorithm based on feature extraction and wavelets," *Optical and Digital Image Processing* **7000**(1), 700028, SPIE (2008).

[13] Cvejic, N., Canagarajah, C. N., and Bull, D. R., "Image fusion metric based on mutual information and tsallis entropy," *Electronics Letters* **42**(11), 626–627 (2006).

[14] Petrovic, V. and Xydeas, C., "Objective image fusion performance characterisation," *IEEE International Conference on Computer Vision* **2**, 1866–1871 (2005).

[15] Gabarda, S. and Cristóbal, G., "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A* **24**(12), B42–B51 (2007).