A NEW QUALITY METRIC FOR IMAGE FUSION

Gemma Piella and Henk Heijmans*

CWI, Kruislaan 413, 1098 SJ Amsterdam
The Netherlands

ABSTRACT

We present three variants of a new quality metric for image fusion. The interest of our metrics, which are based on an image quality index recently introduced by Wang and Bovik in [1], lies in the fact that they do not require a ground-truth or reference image. We perform several simulations which show that our metrics are compliant with subjective evaluations and can therefore be used to compare different image fusion methods or to find the best parameters for a given fusion algorithm.

1. INTRODUCTION

Image fusion is a methodology concerned with the integration of multiple images, e.g. derived from different sensors, into a composite image that is more suitable for the purposes of human visual perception or computer-processing tasks. The widespread use of image fusion methods, in military applications, in surveillance, in medical diagnostics, etc, has led to a rising demand of pertinent quality assessment tools in order to compare the results obtained with different algorithms or to obtain an optimal setting of parameters for a specific fusion algorithm.

Quality assessment of fused images is often carried out by human visual inspection [2]. Objective performance assessment is a difficult issue due to the variety of different application requirements and the lack of a clearly defined ground-truth. Indeed, various fusion algorithms presented in the literature (see Piella [3] for an overview) have been evaluated by constructing some kind of ideal fused image and using it as a reference for comparing with the experimental fused results [4, 5]. Mean squared error (MSE) based metrics are widely used for these comparisons.

A restricted number of objective fusion performance measures have been proposed where the knowledge of ground-truth is not assumed. Xydeas and Petrović [6] propose a metric that evaluates the relative amount of edge information that is transferred from the input images to the fused image. In [7], mutual information is employed for evaluating fusion performance.

This paper discusses a novel objective non-reference quality assessment algorithm for fused images that utilizes local measures to estimate how well the salient information from the inputs is present in the fused images. It is based on an image quality index recently introduced by Wang and Bovik in [1].

2. THE IMAGE QUALITY INDEX OF WANG AND BOVIK

We present a brief introduction to the image quality index that was recently introduced by Wang and Bovik in [1]. Given two real-valued sequences $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$, let $\bar{x}$ denote the mean of x, let σ_x^2 and σ_{xy} be the variance of x and covariance of x, y, respectively, i.e.,

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\,, \qquad \sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\,.$$

Now we compute

$$Q_0 = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\bar{x}^2+\bar{y}^2)(\sigma_x^2+\sigma_y^2)}\,, \tag{1}$$

which can be decomposed as

$$Q_0 = \frac{\sigma_{xy}}{\sigma_x\sigma_y}\cdot\frac{2\bar{x}\bar{y}}{\bar{x}^2+\bar{y}^2}\cdot\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\,. \tag{2}$$

Note that the first component in (2) is the correlation coefficient between x and y. The value $Q_0 = Q_0(x, y)$ is a measure for the similarity of the vectors x and y and takes values between -1 and 1.

Since image signals are generally non-stationary, it is more appropriate to measure the image quality index Q_0 over local regions and then combine the different results into a single measure. Note that in this case the values x_i, y_i are positive grey-scale values. Now the second component in (2) corresponds to the luminance distortion and it has a dynamic range of $[0, 1]$. The third factor in (2) measures the contrast distortion and its range is also $[0, 1]$. The maximum value $Q_0 = 1$ is achieved when x and y are identical.

In [1] the authors propose to use a sliding window approach: starting from the top-left corner of the two images a, b, a sliding window of fixed size (with n pixels) moves pixel by pixel over the entire image until the bottom-right corner is reached. For each window w, the local quality index $Q_0(a, b \mid w)$ is computed for the values $a(i, j)$ and $b(i, j)$ where pixels (i, j) lie in the sliding window w. Finally, the overall image quality index Q_0 is computed by averaging all local quality indices:

$$Q_0(a,b) = \frac{1}{|W|}\sum_{w\in W} Q_0(a, b \mid w)\,, \tag{3}$$

where W is the family of all windows and $|W|$ is the cardinality of W.

Wang and Bovik [1] have compared (e.g. through subjective testing using different types of distortions) their quality index with existing image measures such as the MSE. Their main conclusion was that their new index outperforms the MSE, and they believe this to be due to the index's ability of measuring structural distortions, in contrast to the MSE which is highly sensitive to the L^2 energy of errors.

*The work of Piella is supported by the Dutch Technology Foundation STW, project number CWI 4616.

3. A NEW FUSION QUALITY INDEX

We use the Wang-Bovik image quality index Q_0 defined in (3) to define a quality index $Q(a, b, f)$ for image fusion. Here a, b are two input images and f is the fused image. The index $Q(a, b, f)$ should express the 'quality' of the fused image given the inputs a, b.

We denote by $s(a|w)$ some saliency of image a in window w. It should reflect the local relevance of image a within the window w, and it may depend on e.g. contrast, sharpness, or entropy. Given the local saliencies $s(a|w)$ and $s(b|w)$ of the two input images a and b, we compute a local weight $\lambda(w)$ between 0 and 1 indicating the relative importance of image a compared to image b: the larger $\lambda(w)$, the more weight is given to image a. A typical choice for $\lambda(w)$ is

$$\lambda(w) = \frac{s(a|w)}{s(a|w) + s(b|w)} . \tag{4}$$

Now we define the fusion quality index $Q(a, b, f)$ as

$$Q(a,b,f) = \frac{1}{|W|} \sum_{w \in W} \Big(\lambda(w) Q_0(a, f|w) + \big(1 - \lambda(w)\big) Q_0(b, f|w) \Big) . \tag{5}$$

Thus, in regions where image a has a large saliency compared to b, the quality index $Q(a, b, f)$ is mainly determined by the input image a. On the other hand, in regions where the saliency of b is much larger than that of a, the index $Q(a, b, f)$ is determined mostly by input image b.

At this point, our model has produced a quality index which gives an indication of how much of the salient information contained in each of the input images has been transferred into the fused image without introducing distortions. However, the different quality measures obtained within each window have been treated equally. This is in contrast with the human visual system (HVS) which is known to give higher importance to visually salient regions in an image. We now define another variant of the fusion quality index by giving more weight to those windows where the saliency of the input images is higher. These correspond to areas which are likely to be perceptually important parts of the underlying scene. Therefore the quality of the fused image in those areas is of more importance when determining the overall quality index. The overall saliency of a window is defined as $C(w) = \max\big(s(a|w), s(b|w)\big)$. The *weighted fusion quality index* is then defined as

$$Q_W(a,b,f) = \sum_{w \in W} c(w) \Big(\lambda(w) Q_0(a, f|w) + \big(1 - \lambda(w)\big) Q_0(b, f|w) \Big) , \tag{6}$$

where $c(w) = C(w) / \big(\sum_{w' \in W} C(w') \big)$. There are various other ways to compute the weights $c(w)$ (for example, we could define $C(w) = s(a|w) + s(b|w)$), but we have found that the choice made here is a good indicator of important areas in the input images.

We define one final modification of the fusion quality index that takes into account some aspect of the HVS, namely the importance of edge information. Note that we can evaluate Q_W in (6) using 'edge images' (e.g., the Euclidean norm of the horizontal and vertical gradient images) instead of the original grey-scale images a, b and f. Let us denote the edge image corresponding with a by a'. Now we combine $Q_W(a, b, f)$ and $Q_W(a', b', f')$ into a so-called *edge-dependent fusion quality index* by

$$Q_E(a, b, f) = Q_W(a, b, f) \cdot Q_W(a', b', f')^{\alpha} , \tag{7}$$

where α is a parameter that expresses the contribution of the edge images compared to the original images.

4. EXPERIMENTS

In this section we use the proposed fusion quality indices defined in (5), (6) and (7) to evaluate different multiresolution (MR) image fusion schemes. The MR-based image fusion approach consists of performing a MR transform on each input image and, following some specific rules, combining them into a composite MR representation. The fused image is obtained by applying the inverse transform on this composite MR representation [3].

In this paper we only use the Laplacian pyramid, the ratio pyramid and the spatially-invariant discrete wavelet transform (SIDWT), and in all cases we perform a 3-level decomposition. We combine the coefficients of the MR decompositions of each input by selecting at each position the coefficient with a maximum absolute value, except for the approximation coefficients from the lowest resolution where we take the average. For comparison, we also use the simple fusion method of averaging the input images. In all displayed images, we perform a histogram stretching and scale the grey values of the pixels between 0 (black) and 255 (white).

In the computation of the quality measures defined in last section, we take $\lambda(w)$ as in (4), with $s(a|w)$, $s(b|w)$ being the variance of images a and b, respectively, within the window w of size 8×8.

First, we take as input images the complementary pair shown in the top row of Fig. 1. They have been created by blurring the original 'cameraman' image of size 256×256 with a disk of diameter of 11 pixels. The images are complementary in the sense that the blurring occurs at the left half and the right half, respectively. In the second row we display their total weights used to compute Q_W in (6). More specifically, each pixel (i, j) in the left image contains the value $c(w)\lambda(w)$ with w being the window whose top-left corner corresponds to (i, j). Similarly, the right image displays $c(w)(1 - \lambda(w))$ for every $w \in W$. The fused images obtained by the Laplacian pyramid, the ratio pyramid, the SIDWT and the average are depicted in the third and fourth row, from left to right. Table 1 compares the quality of these fused images using our proposed quality measures. The first row corresponds to the fusion index quality Q defined in (5), the second row to the weighted fusion quality index Q_W in (6) and the third row to the edge-dependent fusion quality index Q_E in (7) with $\alpha = 1$. For comparison, we also compute the root mean squared error between the original 'cameraman' image and each of the fused images. Note that in 'real' fusion scenarios we do not have access to the original image. The resulting errors are shown in the last row of Table 1.

Fig. 1 shows that the Laplacian and SIDWT methods are comparable and that they outperform the other two schemes. Note, for instance, the blurring (e.g., in the buildings) and the loss of texture (e.g., in the grass) of the fused images obtained by the ratio pyramid and averaging. Furthermore, in the ratio-pyramid fused image, the details of the man's face has been cleared out, and in the average fused image, the loss of contrast is evident. These subjective visual comparisons are corroborated by the results in Table 1. The fact that the Laplacian method has a higher Q_E than the SIDWT is most likely due to its known ability to preserve edges and reduce the ringing artifacts around them.

Consider now the input images in the top row of Fig. 2. They correspond to a computer tomography (CT) image and a magnetic resonance image (MRI). We repeat the same computations as described above. The results are shown in Fig. 2 and Table 2. Here, however, as we do not have a reference image to compare with, we cannot compute the MSE. Instead, we use a measure based on mutual information. More precisely, the results in the last row of Table 2 have been obtained by adding the mutual information between the fused image and each of the inputs, and dividing it by the sum of the entropies of the inputs [7].

In Fig. 2, once again we can see that the Laplacian and SIDWT methods clearly outperform the other two methods for both of which many details (specially the brain tissue in the magnetic resonance image) have been lost. Moreover, due to the low redundancy between the input images, the ratio pyramid (which is nonlinear) blows up the dynamic range for some pixels, which makes it necessary to clip them in order to be able to 'visualise' the image. Again, the subjective visual analysis is consistent with the new quality indices, as shown in Table 2. In both experiments, the edge-dependent fusion quality index gives a stronger separation between the good results (Laplacian and SIDWT) and the bad results (ratio and average). Note that the last row, where mutual information has been used, gives the best ranking to the average fusion method. However, mutual information has been shown to be a good indicator of the quality of MR fused images (as long as the average is not taken in all levels for the construction of the composite MR decomposition).

measures	Laplacian	Ratio	SIDWT	Average
Q	0.9034	0.7642	**0.9300**	0.8305
Q_W	0.9618	0.8270	**0.9649**	0.8741
Q_E	**0.9339**	0.6113	0.9276	0.6248
MSE	**8.4154**	164.3551	13.0387	30.6665

Table 1. *Comparison between different quality measures for the fused images in Fig. 1.*

measures	Laplacian	Ratio	SIDWT	Average
Q	0.6617	0.6018	**0.6991**	0.6361
Q_W	**0.7992**	0.6735	0.7705	0.6426
Q_E	**0.6959**	0.4178	0.6641	0.3699
MI	0.3372	0.2212	0.4092	**0.6913**

Table 2. *Comparison between different quality measures for the fused images in Fig. 2.*

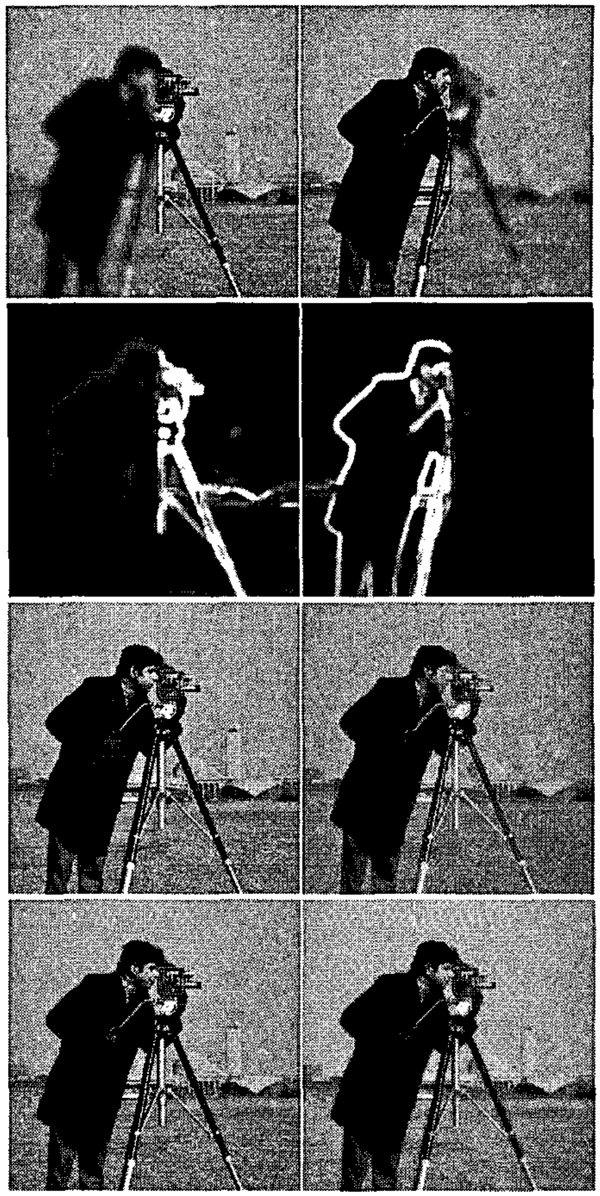

Fig. 1. *Experiment 1. Top: input images a (left) and b (right); second row: total weights $c \cdot \lambda$ (left) and $c \cdot (1 - \lambda)$ (right); third row: fused images with a Laplacian (left) and a ratio (right) pyramid decompositions; bottom: fused images with a SIDWT (left) decomposition and averaging (right).*

5. CONCLUSIONS

In this paper we have discussed some new objective quality measures for image fusion which do not require a reference image and correlate well with subjective criteria as well as with other existing performance measures. Our measures are easy to calculate and applicable to various input modalities (and hence to different fusion applications). In particular, our measures give good results on variable quality input images since it takes into account the locations as well as the magnitude of the distortions.

There are several areas in which our quality measures can be improved or extended. We currently consider grey-scale images, so inclusion of colour is an obvious extension. Other visual mechanisms of our HVS may also be included. In addition, we plan to include some information-theory metrics such as mutual information and entropy to better estimate the information content of the fused image. We also plan to study how our objective measures

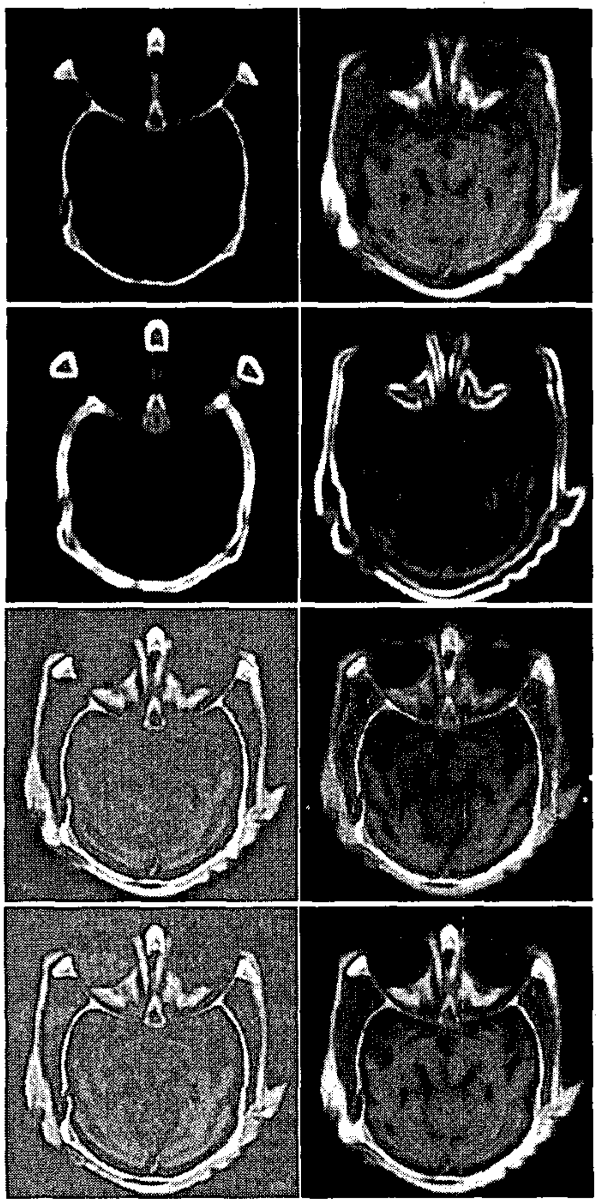

Fig. 2. *Experiment 2. Top: input images a (CT image, left) and b (MRI image, right); second row: total weights $c \cdot \lambda$ (left) and $c \cdot (1 - \lambda)$ (right); third row: fused images with a Laplacian (left) and a ratio (right) pyramid decompositions; bottom: fused images with a SIDWT (left) decomposition and averaging (right).*

can be used to guide a fusion algorithm and improve the fusion performance.

6. ACKNOWLEDGMENT

The authors would like to thank Alexander Toet for his valuable feedback and interesting discussions on the subject.

7. REFERENCES

[1] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.

[2] A. Toet and E. M. Franken, "Perceptual evaluation of different image fusion schemes," *Displays*, vol. 24, no. 1, pp. 25–37, February 2003.

[3] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," to appear in Information Fusion, 2003.

[4] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, May 1995.

[5] O. Rockinger, "Image sequence fusion using a shift invariant wavelet transform," in *Proceedings of IEEE International Conference on Image Processing*, 1997, vol. 13, pp. 288–291.

[6] C. Xydeas and V. Petrović, "Objective pixel-level image fusion performance measure," in *Proceedings of SPIE*, April 2000, vol. 4051, pp. 88–99.

[7] G. H. Qu, D. L. Zhang, and P. F. Yan, "Medical image fusion by wavelet transform modulus maxima," *Journal of the Optical Society of America*, vol. 9, no. 4, pp. 184–190, 2001.