

Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study

Zheng Liu, *Senior Member, IEEE*, Erik Blasch, *Senior Member, IEEE*, Zhiyun Xue,
Jiying Zhao, *Member, IEEE*, Robert Laganière, *Member, IEEE*, and Wei Wu

Abstract—Comparison of image processing techniques is critically important in deciding which algorithm, method, or metric to use for enhanced image assessment. Image fusion is a popular choice for various image enhancement applications such as overlay of two image products, refinement of image resolutions for alignment, and image combination for feature extraction and target recognition. Since image fusion is used in many geospatial and night vision applications, it is important to understand these techniques and provide a comparative study of the methods. In this paper, we conduct a comparative study on 12 selected image fusion metrics over six multiresolution image fusion algorithms for two different fusion schemes and input images with distortion. The analysis can be applied to different image combination algorithms, image processing methods, and over a different choice of metrics that are of use to an image processing expert. The paper relates the results to an image quality measurement based on power spectrum and correlation analysis and serves as a summary of many contemporary techniques for objective assessment of image fusion algorithms.

Index Terms—Night vision, context enhancement, pixel-level image fusion, multiresolution analysis, objective fusion assessment, performance metric, image quality.

1 INTRODUCTION

ADVANCED surveillance applications employ multimodal imaging sensors to enhance the exploitation performance and expand the capability of vision systems under varied environmental conditions, target variations, and viewpoint obscurations. An important technique for surveillance applications is the “multisensor image fusion,” from which a composite (fused) image from multiple sensor inputs can be generated. The fused image provides comprehensive information about the scene such that the operator

does not have to check each image separately. One example is a night vision application of context enhancement where an infrared (IR) surveillance camera enhances objects for detection in a specific environment from a visual spectrum background. According to Kirchhoff’s law, the emissivity quantifies the energy emitting characteristics of different materials and surfaces [1]. The emitted energy of an object reaches the infrared sensor and is converted into an electrical signal. The electrical signal can be further converted into a temperature value based on the sensor’s calibration equation and object’s emissivity. Thus, the thermography can “see in the night” without infrared illumination. However, the infrared sensor is not sensitive to the “cool” background, while the visible image (VI) can provide a relatively clear perspective of the environment. Visible images are subject to illumination levels and are not able to present contrasts between hot and cool surfaces. The purpose of image fusion is to fuse the information across the electromagnetic spectrum, (e.g., the visible and infrared band) for applications such as context enhancement in night vision.

The framework for implementing image fusion may vary from application to application. One of the most important techniques for night vision is *multiresolution image fusion* (MIF) [2] (presented in the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>). The idea of MIF is to retain the most significant features from the input images using the multiresolution transform domain coefficients. MIF research focuses on accessible multiresolution feature representations and an image fusion rule to guide the combination of coefficients in the transform domain. Numerous MIF solutions have been proposed; however, the effectiveness and efficiency of the proposed algorithms need

- Z. Liu is with the School of Information Technology and Engineering, University of Ottawa, 800 King Edward Ave., Ottawa, ON K1N 6N5, Canada and also with the National Research Council Canada, 1200 Montreal Road, Bld M-20, Ottawa, ON K1A 0R6, Canada.
E-mail: zheng.liu@ieee.org, zheng.liu@nrc-cnrc.gc.ca.
- E. Blasch is with the US Air Force Research Laboratory (AFRL), AFRL/RYAA, 2241 Avionics Cir, WPAFB, OH 45433.
E-mail: erik.blasch@wpafb.af.mil.
- Z. Xue is with the National Library of Medicine, NIH, Communications Engineering Branch/MSC 3824, Bldg. 38A, Room B1N30A16, 8600 Rockville Pike, Bethesda, MD 20894. *E-mail:* xuez@mail.nih.gov.
- J. Zhao is with the School of Information Technology and Engineering (SITE), University of Ottawa, 800 King Edward Ave., Ottawa, ON K1N 6N5. *E-mail:* jyzhao@site.uottawa.ca.
- R. Laganière is with the School of Information Technology and Engineering (SITE), University of Ottawa, 800 King Edward Ave., PO Box 450, Stn A, Ottawa, ON K1N 6N5. *E-mail:* laganier@site.uottawa.ca.
- W. Wu is with the College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, P.R. China.
E-mail: wuwei@scu.edu.cn.

Manuscript received 21 June 2010; revised 3 Mar. 2011; accepted 15 Apr. 2011; published online 13 May 2011.

Recommended for acceptance by R. Ramamoorthi.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2010-06-0466.

Digital Object Identifier no. 10.1109/TPAMI.2011.109.

to be verified and collectively evaluated [2], [3]. Although an adaptive fusion strategy is preferred, how the fusion algorithm adapts to different object-to-background situations is still not well understood.

In order to objectively assess the performance of an MIF algorithm, a number of evaluation metrics, either objective or subjective, have been proposed [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. The problem is that certain fusion algorithms may work for one application but do not have the same performance for another application. Each application varies based on the sensors used, the targets of interest, and the environmental conditions. Studies on image fusion lack information that explicitly defines the applicability and feasibility of a specific fusion algorithm for a given application. The same problem also exists in the research on information fusion performance evaluation, with the difficulty being how to define and validate objective evaluation metrics. Usually a subjective evaluation is carried out to validate an objective assessment [18]. However, identifying a reliable subjective score needs extensive experiments, which is expensive and cannot cover all possible conditions of interest. Typically, a robust performance model is required to account for the critical image fusion parameters and better assess the trend of image fusion performance quality.

The objective of this Multi-Image Fusion for Context Enhancement (MIF-CE) work is to carry out a comparative study of the objective image fusion assessment metrics and investigate their effectiveness for context enhancement in a night vision application. The MIF-CE contributes to:

- understanding the relationship between the image fusion metrics,
- demonstrating the effectiveness of these metrics by referencing the image quality measurement, and
- learning the difference between the fusion of heterogeneous and homogeneous images from fusion metrics.

In this study, 12 fusion metrics have been implemented with Matlab¹ and applied to six multiresolution fusion algorithms. The 12 metrics are categorized into four groups:

1. information theory based metrics,
2. image feature based metrics,
3. image structural similarity based metrics, and
4. human perception inspired fusion metrics.

A direct image fusion and a modified image fusion method are considered in the MIF-CE study. Detailed information on the image fusion methods are available in the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>. The selected image fusion metrics are presented in Section 2 and Section 3 explains the methodology for the MIF-CE comparative study. Experimental results are presented in Section 4. Discussions and conclusions can be found in Sections 5 and 6, respectively.

1. The Matlab implementation is available upon request.

2 ALGORITHMS FOR OBJECTIVE IMAGE FUSION PERFORMANCE ASSESSMENT

Two types of fusion schemes were considered in the MIF-CE study. The first one is *direct (heterogeneous) image fusion* of IR and visible images with a multiresolution approach at the pixel level. The other method is a *modified (homogeneous) image fusion* as described in [19], where the visible image enhanced from the IR image is fused with the original visible image. (Please refer to the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>.) Utilizing the fusion of heterogeneous and homogeneous images will help to understand how the fusion metrics perform over various applications. For the rest of the paper, we will use “VI-IR direct fusion” and “VI-EVI modified fusion” to refer to the two image fusion schemes.

The assessment of a fused image can be carried out in two different ways. The first method is to compare the fusion result with a known reference image (or ground truth). However, a reference image is not always available in a practical application. The second implementation, a blind or nonreferenced assessment, is generally preferred. In this paper, we will focus on the blind assessment. Different approaches have been proposed for blind assessment so far [4], [6], [7], [10], [14], [15], [20], [21], [22], [23], [24]. In this study, 12 most representative metrics are used for the comparative study and each metric is briefly described below.

2.1 Information Theory-Based Metrics

2.1.1 Normalized Mutual Information (Q_{MI})

Mutual information (MI) is a quantitative measure of the mutual dependence of two variables. The definition of mutual information for two discrete random variables U and V is

$$MI(U; V) = \sum_{v \in V} \sum_{u \in U} p(u, v) \log_2 \frac{p(u, v)}{p(u)p(v)}, \quad (1)$$

where $p(u, v)$ is the joint probability distribution function of U and V , and $p(u)$ and $p(v)$ are the marginal probability distribution functions of U and V , respectively. Actually, MI quantifies the distance between the joint distribution of U and V , i.e., $p(u, v)$, and the joint distribution when U and V are independent, i.e., $p(u)p(v)$. Mutual information can be equivalently expressed with joint entropy $\{H(U, V)\}$ and marginal entropy $\{H(U), H(V)\}$ of the two variable U and V as

$$MI(U, V) = H(U) + H(V) - H(U, V), \quad (2)$$

where

$$H(U) = - \sum_u p(u) \log_2 p(u),$$

$$H(V) = - \sum_v p(v) \log_2 p(v),$$

$$H(U, V) = - \sum_{u,v} p(u, v) \log_2 p(u, v).$$

Qu et al. used the summation of the MI between the fused image $F(i, j)$ and two input images, $A(i, j)$ and $B(i, j)$, to

represent the difference in quality [4]. The expression of the MI-based fusion performance measure M_F^{AB} is

$$\begin{aligned} M_F^{AB} &= MI(A, F) + MI(B, F) \\ &= \sum_{i,j} \left(h_{AF}(i, j) \log_2 \frac{h_{AF}(i, j)}{h_A(i)h_F(j)} \right. \\ &\quad \left. + h_{BF}(i, j) \log_2 \frac{h_{BF}(i, j)}{h_B(i)h_F(j)} \right), \end{aligned} \quad (3)$$

where $h_{AF}(i, j)$ indicates the normalized joint gray level histogram of images $A(i, j)$ and $F(i, j)$; $h_K(i, j)$ ($K = A, B$, and F) is the normalized marginal histogram of images A , B , or F , respectively.

One problem with (3) is that it mixes two joint entropies measured at different scales. This may cause instability of the measure and bias the measure toward the source image with the highest entropy [14]. Thus, Hossny et al. modified (3) as

$$Q_{MI} = 2 \left[\frac{MI(A, F)}{H(A) + H(F)} + \frac{MI(B, F)}{H(B) + H(F)} \right]. \quad (4)$$

We will use Hossny's definition in our experiments.

2.1.2 Fusion Metric-Based on Tsallis Entropy (Q_{TE})

Cvejic et al. and Nava et al. suggested using Tsallis entropy to define the fusion metric [10], [20]. Tsallis entropy is a divergence measure of the degree of dependence between two discrete random variables. For the input image $A(i, j)$ and fused image $F(i, j)$, the Tsallis entropy is ([10] and [20])

$$I^q(A, F) = \frac{1}{1-q} \left(1 - \sum_{i,j} \frac{h_{AF}(i, j)^q}{h_F(j)h_A(i)^{q-1}} \right), \quad (5)$$

where q is real value and $q \neq 1$. A quality metric of order q can be defined as [10]

$$Q_{TE}^q = I^q(A, F) + I^q(B, F), \quad (6)$$

or as a normalized value [20]

$$Q_{TE} = \frac{I^q(A, F) + I^q(B, F)}{H^q(A) + H^q(B) - I^q(A, B)}. \quad (7)$$

The use of Renyi entropy was also suggested by Zheng et al. [9]. However, the MI-based metric still needs a reference value to compare with. We cannot tell in advance if a fused image with a given MI value is good or not, so a reference point is a must. Moreover, the MI-based approach is sensitive to impulsive noise and is subject to significant change in the presence of additive Gaussian noise.

2.1.3 Nonlinear Correlation Information Entropy (Q_{NCIE})

For two discrete variables $U = \{u_i\}_{1 \leq i \leq N}$ and $V = \{v_i\}_{1 \leq i \leq N}$, the nonlinear correlation coefficient (NCC) is defined as [12]

$$NCC(U, V) = H'(U) + H'(V) - H'(U, V), \quad (8)$$

which is similar to the definition of the mutual information between U and V in (2). Considering $NCC(A, B)$ for images A and B , the entropies are defined as

$$H'(A, B) = - \sum_{i=1}^b \sum_{j=1}^b h_{AB}(i, j) \log_b h_{AB}(i, j), \quad (9)$$

$$H'(A) = - \sum_{i=1}^b h_A(i) \log_b h_A(i), \quad (10)$$

$$H'(B) = - \sum_{i=1}^b h_B(i) \log_b h_B(i), \quad (11)$$

where b is determined by the intensity level, i.e., $b = 256$. A nonlinear correlation matrix of the input image $A(i, j)$, $B(i, j)$, and fused image $F(i, j)$ is defined as

$$\begin{aligned} R &= \begin{pmatrix} NCC_{AA} & NCC_{AB} & NCC_{AF} \\ NCC_{BA} & NCC_{BB} & NCC_{BF} \\ NCC_{FA} & NCC_{FB} & NCC_{FF} \end{pmatrix} \\ &= \begin{pmatrix} 1 & NCC_{AB} & NCC_{AF} \\ NCC_{BA} & 1 & NCC_{BF} \\ NCC_{FA} & NCC_{FB} & 1 \end{pmatrix}. \end{aligned} \quad (12)$$

The eigenvalue of the nonlinear correlation matrix R is λ_i ($i = 1, 2, 3$). Therefore, the nonlinear correlation information entropy Q_{NCIE} can be obtained:

$$Q_{NCIE} = 1 + \sum_{i=1}^3 \frac{\lambda_i}{3} \log_b \frac{\lambda_i}{3}. \quad (13)$$

2.2 Image Feature-Based Metrics

Another type of assessment is implemented by measuring how the features are transferred from the input images to the fused one.

2.2.1 Gradient-Based Fusion Performance (Q_G)

Xydeas and Petrovic proposed a metric to evaluate the amount of edge information, which is transferred from input images to the fused image [6]. A Sobel edge operator is applied to get the edge strength of input image $A(i, j)$, $g_A(i, j)$, and orientation $\alpha_A(i, j)$:

$$g_A(i, j) = \sqrt{s_A^x(i, j)^2 + s_A^y(i, j)^2}, \quad (14)$$

$$\alpha_A(i, j) = \tan^{-1} \left(\frac{s_A^x(i, j)}{s_A^y(i, j)} \right), \quad (15)$$

where $s_A^x(i, j)$ and $s_A^y(i, j)$ are the convolved results with the horizontal and vertical Sobel templates [6]. The relative strength (G^{AF}) and orientation values (Δ^{AF}) between input image A and fused image F are

$$G^{AF}(i, j) = \begin{cases} \frac{g_F(i, j)}{g_A(i, j)}, & g_A(i, j) > g_F(i, j), \\ \frac{g_A(i, j)}{g_F(i, j)}, & \text{Otherwise,} \end{cases} \quad (16)$$

$$\Delta^{AF}(i, j) = 1 - \frac{|\alpha_A(i, j) - \alpha_F(i, j)|}{\pi/2}. \quad (17)$$

The edge strength and orientation preservation values can be derived:

$$Q_g^{AF}(i, j) = \frac{\Gamma_g}{1 + e^{\kappa_g(G^{AF}(i, j) - \sigma_g)}}, \quad (18)$$

$$Q_\alpha^{AF}(i, j) = \frac{\Gamma_\alpha}{1 + e^{\kappa_\alpha(\Delta^{AF}(i, j) - \sigma_\alpha)}}. \quad (19)$$

The constants $\Gamma_g, \kappa_g, \sigma_g$ and $\Gamma_\alpha, \kappa_\alpha, \sigma_\alpha$ determine the shape of the sigmoid functions used to form the edge strength and orientation preservation value [6]. See Table 2 in the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>, for the values used in our implementation. Edge information preservation value is then defined as

$$Q^{AF}(i, j) = Q_g^{AF}(i, j)Q_\alpha^{AF}(i, j). \quad (20)$$

The final assessment is obtained from the weighted average of the edge information preservation values.

$$Q_G = \frac{\sum_{n=1}^N \sum_{m=1}^M [Q^{AF}(i, j)w^A(i, j) + Q^{BF}(i, j)w^B(i, j)]}{\sum_{n=1}^N \sum_{m=1}^M (w^A(i, j) + w^B(i, j))}, \quad (21)$$

where the weighting coefficients are defined as: $w^A(i, j) = [g_A(i, j)]^L$ and $w^B(i, j) = [g_B(i, j)]^L$, respectively. Here, L is a constant value.

2.2.2 Image Fusion Metric-Based on a Multiscale Scheme (Q_M)

Wang and Liu proposed a metric which is implemented with a two-level Haar wavelet [25]. The edge information is retrieved from the high and band-pass components of the decomposition. At each level s , for input image $A(i, j)$ and fused image $F(i, j)$, there is

$$\begin{aligned} \psi_{H_s}^{AF}(m, n) &= \exp(-|LH_s^A(m, n) - LH_s^F(m, n)|), \\ \psi_{V_s}^{AF}(m, n) &= \exp(-|HL_s^A(m, n) - HL_s^F(m, n)|), \\ \psi_{D_s}^{AF}(m, n) &= \exp(-|HH_s^A(m, n) - HH_s^F(m, n)|). \end{aligned}$$

Then, the global edge preservation value at scale s can be derived as

$$EP_s^{AF} = \frac{\psi_{H_s}^{AF}(m, n) + \psi_{V_s}^{AF}(m, n) + \psi_{D_s}^{AF}(m, n)}{3}. \quad (22)$$

A normalized performance metric weighted by $w_s^A(m, n)$ and $w_s^B(m, n)$ at scale s is defined as

$$\begin{aligned} Q_s^{AB/F} &= \frac{\sum_m \sum_n (EP_s^{AF}(m, n)w_s^A(m, n) + EP_s^{BF}(m, n)w_s^B(m, n))}{\sum_m \sum_n (w_s^A(m, n) + w_s^B(m, n))}. \end{aligned} \quad (23)$$

The high-frequency energy of the input images is used as a weight coefficient:

$$w_s^A(m, n) = LH_{A_s}^2(m, n) + HL_{A_s}^2(m, n) + HH_{A_s}^2(m, n). \quad (24)$$

Similarly, $w_s^B(m, n)$ can be derived for image B . The overall metric is obtained by combining the measurement at different scales using [25]:

$$Q_M = \prod_{s=1}^N (Q_s^{AB/F})^{\alpha_s}, \quad (25)$$

where α_s is a constant to adjust the relative importance of different scales.

2.2.3 Image Fusion Metric-Based on Spatial Frequency (Q_{SF})

Zheng et al. used “spatial frequency” to measure the activity level of an image $I(i, j)$ as defined below [8]:

$$SF = \sqrt{(RF)^2 + (CF)^2 + (MDF)^2 + (SDF)^2}. \quad (26)$$

For an image $I(i, j)$, there are

$$RF = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=2}^N [I(i, j) - I(i, j-1)]^2}, \quad (27)$$

$$CF = \sqrt{\frac{1}{MN} \sum_{j=1}^N \sum_{i=2}^M [I(i, j) - I(i-1, j)]^2}, \quad (28)$$

$$MDF = \sqrt{w_d \cdot \frac{1}{MN} \sum_{i=2}^M \sum_{j=2}^N [I(i, j) - I(i-1, j-1)]^2}, \quad (29)$$

$$SDF = \sqrt{w_d \cdot \frac{1}{MN} \sum_{j=1}^{N-1} \sum_{i=2}^M [I(i, j) - I(i-1, j+1)]^2}. \quad (30)$$

Here, RF, CF, MDF, and SDF are the four first-order gradients along four directions. Distance weight w_d is $1/\sqrt{2}$. The four reference gradients are obtained by taking the maximum of absolute gradient values between input images A and B along four directions [8]:

$$\begin{aligned} \text{Grad}^D(I_R(i, j)) \\ = \max\{\text{abs}[\text{Grad}^D(I_A(i, j))], \text{abs}[\text{Grad}^D(I_B(i, j))]\}, \end{aligned} \quad (31)$$

where there is $D = \{H, V, MD, SD\}$, which represents horizontal, vertical, main diagonal, and secondary diagonal, respectively. With the reference gradients substituting the differences in (27)-(30), the four directional references, RF_R , CF_R , MDF_R , and SDF_R , can be calculated. Thus, SF_R can be derived from (26). Finally, the ratio of SF error (metric Q_{SF}) is defined as

$$Q_{SF} = (SF_F - SF_R)/SF_R. \quad (32)$$

2.2.4 Image Fusion Metric-Based on Phase Congruency (Q_P)

Zhao et al. and Liu et al. used the phase congruency, which provides an absolute measure of image feature, to define an evaluation metric [7], [15]. In [7], the principal (maximum and minimum) moments of the image phase congruency were employed to define the metric because the moments contain the information for corners and edges. The metric is defined as a product of three correlation coefficients,

$$Q_P = (P_p)^\alpha (P_M)^\beta (P_m)^\gamma, \quad (33)$$

where p, M, m refers to phase congruency (p), maximum, and minimum moments, respectively, and there are

$$\begin{aligned} P_p &= \max(C_{AF}^p, C_{BF}^p, C_{SF}^p), \\ P_M &= \max(C_{AF}^M, C_{BF}^M, C_{SF}^M), \\ P_m &= \max(C_{AF}^m, C_{BF}^m, C_{SF}^m). \end{aligned}$$

Herein, C_{xy}^k , $\{k|p, M, m\}$ stands for the correlation coefficients between two sets x and y :

$$C_{xy}^k = \frac{\sigma_{xy}^k + C}{\sigma_x^k \sigma_y^k + C}, \quad (34)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (35)$$

The suffixes A , B , F , and S correspond to the two inputs, fused image, and maximum-select map. The exponential parameters α , β , and γ can be adjusted based on the importance of the three components [7].

2.3 Image Structural Similarity-Based Metrics

The image similarity measurement is based on the evidence that the human visual system is highly adapted to structural information and a measurement of the loss of structural information can provide a good approximation of the perceived image distortion. Wang proposed a structural similarity index measure (SSIM) for images A and B defined as [26]

$$\begin{aligned} SSIM(A, B) &= [l(A, B)]^\alpha [c(A, B)]^\beta [s(A, B)]^\gamma \\ &= \left(\frac{2\mu_A\mu_B + C_1}{\mu_A^2 + \mu_B^2 + C_1} \right)^\alpha \left(\frac{2\sigma_A\sigma_B + C_2}{\sigma_A^2 + \sigma_B^2 + C_2} \right)^\beta \left(\frac{\sigma_{AB} + C_3}{\sigma_A\sigma_B + C_3} \right)^\gamma, \end{aligned} \quad (36)$$

where μ_A and μ_B are the average values of images $A(i, j)$ and $B(i, j)$, σ_A , σ_B , and σ_{AB} are the variance and covariance, respectively [26]. $l(A, B)$, $c(A, B)$, and $s(A, B)$ in (36) are the luminance, contrast, and correlation components, respectively. The parameters α , β , and γ are used to adjust the relative importance of the three components. The constant values C_1 , C_2 , and C_3 are defined to avoid the instability when the denominator are very close to zero. By setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, (36) becomes

$$SSIM(A, B) = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)}. \quad (37)$$

A previous version of this index is known as the universal image quality index (UIQI) and is written as [27]

$$\begin{aligned} Q(A, B) &= \frac{\sigma_{AB}}{\sigma_A \sigma_B} \cdot \frac{2\mu_A\mu_B}{\mu_A^2 + \mu_B^2} \cdot \frac{2\sigma_A\sigma_B}{\sigma_A^2 + \sigma_B^2} \\ &= \frac{4\sigma_{AB}\mu_A\mu_B}{(\sigma_A^2 + \sigma_B^2)(\mu_A^2 + \mu_B^2)}. \end{aligned} \quad (38)$$

The following image structural similarity fusion metrics are based on these two definitions. The calculation of the

quality indices is based on a sliding window approach, which moves from top-left to bottom-right. The SSIM and Q value can be calculated locally, summed, and averaged to get the overall index. See [28] for the detailed implementation of the SSIM algorithm.

2.3.1 Piella's Metric (Q_S)

Piella and Heijmans defined three fusion quality index based on Wang's UIQI method [5]. Assume the local $Q(A, B|w)$ value is calculated in a sliding window w . There are

$$Q_S = \frac{1}{|W|} \sum_{w \in W} [\lambda(w)Q_0(A, F|w) + (1 - \lambda(w))Q_0(B, F|w)], \quad (39)$$

$$Q_W = \sum_{w \in W} c(w)[\lambda(w)Q_0(A, F|w) + (1 - \lambda(w))Q_0(B, F|w)], \quad (40)$$

$$Q_E = Q_W(A, B, F) \cdot Q_W(A', B', F')^\alpha, \quad (41)$$

where the weight $\lambda(w)$ is defined as

$$\lambda(w) = \frac{s(A|w)}{s(A|w) + s(B|w)}. \quad (42)$$

Herein, $s(A|w)$ is a local measure of image salience. In Piella's implementation, $s(A|w)$ and $s(B|w)$ are the variance of images A and B within the window w , respectively. The coefficient $c(w)$ in (40) is [5]

$$c(w) = \frac{\max[s(A|w), s(B|w)]}{\sum_{w' \in W}[s(A|w'), s(B|w')]} \cdot \quad (43)$$

In (41), $Q_W(A', B', F')$ is the Q_w calculated with the edge images, i.e., A' , B' , and F' , and α is a manually adjustable parameter to weight the edge-dependent information.

2.3.2 Cvejic's Metric Q_C

Cvejic et al. defined a performance measure as [21]

$$\begin{aligned} Q_C &= \sum_{w \in W} \text{sim}(A, B, F|w)Q(A, F|w) \\ &\quad + (1 - \text{sim}(A, B, F|w))Q(B, F|w), \end{aligned} \quad (44)$$

where the function $\text{sim}(A, B, F|w)$ is [21]

$$\text{sim}(A, B, F|w) = \begin{cases} 0, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF} + \sigma_{BF}} < 0, \\ \frac{\sigma_{AF}}{\sigma_{AF} + \sigma_{BF}}, & \text{if } 0 \leq \frac{\sigma_{AF}}{\sigma_{AF} + \sigma_{BF}} \leq 1, \\ 1, & \text{if } \frac{\sigma_{AF}}{\sigma_{AF} + \sigma_{BF}} > 1. \end{cases} \quad (45)$$

The weighting factor depends on the similarity in spatial domain between the input images and the fused image. The higher the similarity between the input and fused image, the larger the corresponding weighting factor.

2.3.3 Yang's Metric Q_Y

Yang et al. proposed another way to use SSIM for fusion assessment [11]:

$$Q_Y = \begin{cases} \lambda(w)\text{SSIM}(A, F|w) + (1-\lambda(w))\text{SSIM}(B, F|w), & \text{SSIM}(A, B|w) \geq 0.75, \\ \max\{\text{SSIM}(A, F|w), \text{SSIM}(B, F|w)\}, & \text{SSIM}(A, B|w) < 0.75. \end{cases} \quad (46)$$

The local weight $\lambda(w)$ is as the definition in (42).

2.4 Human Perception Inspired Fusion Metrics

2.4.1 Chen-Varshney Metric (Q_{CV})

The Chen-Varshney metric consists of five steps [24]:

- *Extract edge information:* The extraction of edge information is implemented by applying Sobel edge detector to get the edge strength map $G_K(i, j)$.
- *Partition images into local regions:* The images are divided into nonoverlapped local regions (windows).
- *Calculate local region saliency:* The local region saliency is calculated as the summation of squares of edge intensities in the local region: $\lambda(A^W) = \sum_{w \in W} G_A(w)^\alpha$. Here, α is a constant.
- *Similarity measure in the local region:* The measure is the mean squared value of the contrast sensitive function (CSF) filtered image \hat{f}_A^W :

$$D(I_A^W, I_F^W) = \frac{1}{|W|} \sum_{w \in W} \hat{f}_r^W(i, j)^2, \quad (47)$$

where r refers to the input images A and B and $|W|$ is the number of pixels in local region W .

- *Global quality measure:* The global quality measure is the weighted summation over all of the nonoverlapping regions (windows):

$$Q_{CV} = \frac{\sum_{l=1}^L (\lambda(I_A^{W_l}) D(I_A^{W_l}, I_F^{W_l}) + \lambda(I_B^{W_l}) D(I_B^{W_l}, I_F^{W_l}))}{\sum_{l=1}^L (\lambda(I_A^{W_l}) + \lambda(I_B^{W_l}))}. \quad (48)$$

2.4.2 Chen-Blum Metric Q_{CB}

There are five steps involved in Chen-Blum metric [23]:

- *Contrast sensitivity filtering:* Filtering is implemented in the frequency domain. Image $I_A(i, j)$ is transformed into the frequency domain and get $I_A(m, n)$. The filtered image is obtained: $\tilde{I}_A(m, n) = I_A(m, n)S(r)$, where $S(r)$ is the CSF filter in polar form with $r = \sqrt{m^2 + n^2}$. In [23], there are three choices suggested for CSF, which include Mannos-Sakrison, Barton, and DoG filter.
- *Local contrast computation:* Peli's contrast is defined as

$$C(i, j) = \frac{\phi_k(i, j) * I(i, j)}{\phi_{k+1}(i, j) * I(i, j)} - 1. \quad (49)$$

A common choice for ϕ_k would be

$$G_k(x, y) = \frac{1}{(\sqrt{2\pi}\sigma_k)} e^{-\frac{x^2+y^2}{2\sigma_k^2}}, \quad (50)$$

with a standard deviation $\sigma_k = 2$.

- *Contrast preservation calculation:* The masked contrast map for input image $I_A(i, j)$ is calculated as

$$C'_A = \frac{t(C_A)^p}{h(C_A)^q + Z}. \quad (51)$$

Here, t , h , p , q , and Z are real scalar parameters that determine the shape of the nonlinearity of the masking function [23].

- *Saliency map generation:* The saliency map for $I_A(i, j)$ is defined as

$$\lambda_A(i, j) = \frac{C'_A^2(i, j)}{C'_A^2(i, j) + C'_B^2(i, j)}. \quad (52)$$

The information preservation value is computed as

$$Q_{AF}(i, j) = \begin{cases} \frac{C'_A(i, j)}{C'_F(i, j)}, & \text{if } C'_A(i, j) < C'_F(i, j), \\ \frac{C'_B(i, j)}{C'_F(i, j)}, & \text{otherwise.} \end{cases} \quad (53)$$

- *Global quality map:*

$$Q_{GQM}(i, j) = \lambda_A(i, j)Q_{AF}(i, j) + \lambda_B(i, j)Q_{BF}(i, j). \quad (54)$$

The metric value is obtained by average the global quality map, i.e., $Q_{CB} = \overline{Q_{GQM}(i, j)}$.

3 THE COMPARATIVE STUDY

The assessment of fusion metrics remains a challenge in a comparative analysis of image fusion algorithms. A well-adopted method is to validate the fusion metric with subjective results, which can be costly and time-intensive to obtain [18]. The duality between image fusion algorithms and metrics was studied in [29]. However, humans are limited by their capacity for processing information [30]. Some methods to compare the subjective and objective assessment were presented in previous studies [23], [31], [32]. Correlation and root mean square error (RMSE) between the human evaluation scores and fusion quality measurements were calculated to depict the relation between the subjective and objective assessment [23]. Chen and Blum also investigated the performance of the fusion metrics under additive Gaussian noise. Kaplan et al. [31], [32] fitted an arbitrary monotonic curve to the subjective and objective results and defined a monotonic correlation (MC), i.e., ρ_{mono} , as the coefficient of determination (R^2), which indicates the goodness of fit with this curve. A larger ρ_{mono} value means a better fit and thus a better fusion metric. Kaplan et al. found that Q_{CB} was a leading candidate metric.

In Section 2, the 12 metrics considered in this study were described. The major purpose of this work is not to rate the fusion metrics, but to compare them in a single experiment as well as present them in a unified fashion. Different metrics would work best in varied conditions and the data available in this study are not extensive. All metrics should be considered together until further evaluations and

documentation are considered for various applications. For the MIF-CE study, we are only concerned with the night vision application and tested six algorithms for the heterogeneous and homogeneous image fusion methods. Using the methods as described in Section 2, we computed 12 generalized MIF quality metrics.

The experiments presented in the next section investigate the following issues:

- the correlation between different MIF metrics,
- the consistency of the MIF metrics in rating fusion algorithms,
- the impact of image distortion to MIF metrics,
- the relation of MIF metrics with an image quality measurement,
- the MIF metrics' performance on assessing heterogeneous and homogeneous image fusion.

A correlation method, Borda count majority ranking, and an objective image quality measurement are used to evaluate the MIF metrics.

Kendall τ rank correlation coefficient is adopted to measure the similarity of fusion metrics (e.g., variables X and Y for two metrics) [33]. We rearrange variables X and Y in a set of elements $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Any pair (x_i, y_i) and (x_j, y_j) are concordant if both $x_i > x_j$ and $y_i > y_j$ or both $x_i < x_j$ and $y_i < y_j$. This pair is said to be concordant; otherwise, it is said to be discordant. If $x_i = x_j$ or $y_i = y_j$, this pair is neither concordant nor discordant. Thus, the Kendall τ correlation is defined as

$$\tau = \frac{2[(\text{number of concordant pairs}) - (\text{number of discordant pairs})]}{\{n(n - 1)\}}. \quad (55)$$

Correlation coefficients can range from +1.00 to -1.00, which represent a perfect agreement or disagreement between two fusion metrics.

Each fusion metric gives a quantified value to each fused result. A rank can be derived from a comparison of the metric value for each fusion algorithm, which is similar to a pattern classification process. The fusion metrics can be treated as a classifier and the ranking corresponds to the classes. To determine the consistency of these "classifiers," a reference is needed. The Borda count method was introduced as a reference to map a set of individual rankings to a combined ranking and is a generalization of the majority vote [34]. The Borda count for a class is the sum of the number of classes ranked below it by each classifier [34]. In our study, each fusion metric is treated equally, even though they may have a varied performance.

A close topic to fusion metric is the image quality assessment. Nill and Bouzas proposed an objective image quality measurement (IQM) based on the digital image power spectrum of normally acquired arbitrary scenes [35]. The IQM measure incorporates a modulation transfer function representing human visual system. The IQM does not need any reference image for comparison and has demonstrated good correlation with objective (visual) assessments [35]. The implementation by MITRE is used in this study [36].

The input visible and IR images are subject to certain distortion. To mimic such situations, the input images are

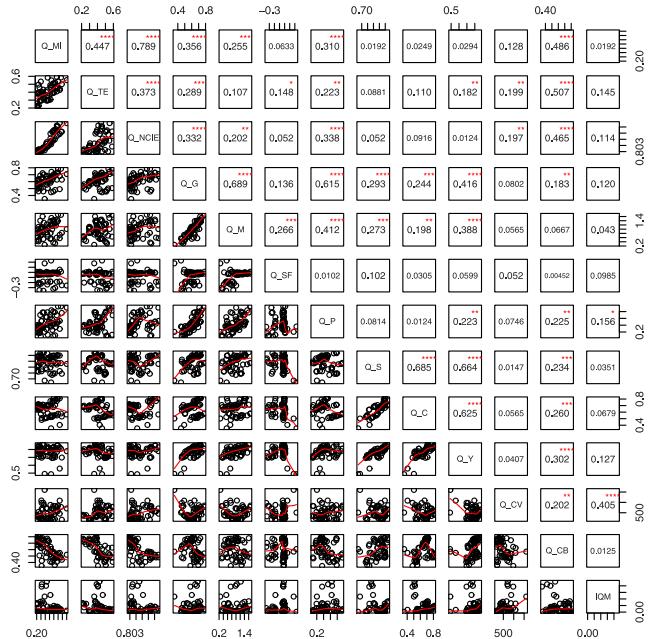


Fig. 1. The correlation matrix of fusion metrics for the VI-IR direct fusion scheme.

added with Gaussian white noise and blurred with a Gaussian low-pass filter, respectively. We will investigate how the distortion affects the fusion metrics in the experiments.

4 EXPERIMENTS

4.1 Experimental Setup

The two fusion schemes adopted in the experiments are identified as *VI-IR direct (heterogeneous) fusion* and *VI-EVI (homogeneous) modified fusion*. Six multiresolution-based fusion algorithms were considered and image pairs in the supplemental material of this paper, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>, were used in the experiments. The selected 12 fusion metrics gave a metric value for each fusion implementation. The parameters for each fusion metric used in the experiment are provided in a table in the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>.

4.2 Experimental Results

4.2.1 The Correlation between Assessment Metrics

The correlation matrices for different fusion metrics are plotted in Figs. 1 and 2, respectively. The star on the top indicates the significance of the P-value for the correlation test. For the VI-IR direct fusion, the three largest values 0.789, 0.689, and 0.685 correspond to the correlation of Q_{MI} and Q_{NCIE} , Q_G and Q_M , Q_S and Q_C , respectively. For the VI-EVI modified fusion, the three largest values 0.854, 0.835, and 0.765 represent the correlation of Q_{MI} and Q_{NCIE} , Q_C and Q_Y , Q_G and Q_C . Obviously, Q_{MI} and Q_{NCIE} show a higher correlation in both cases as the two approaches are based on mutual information and are quite similar. The

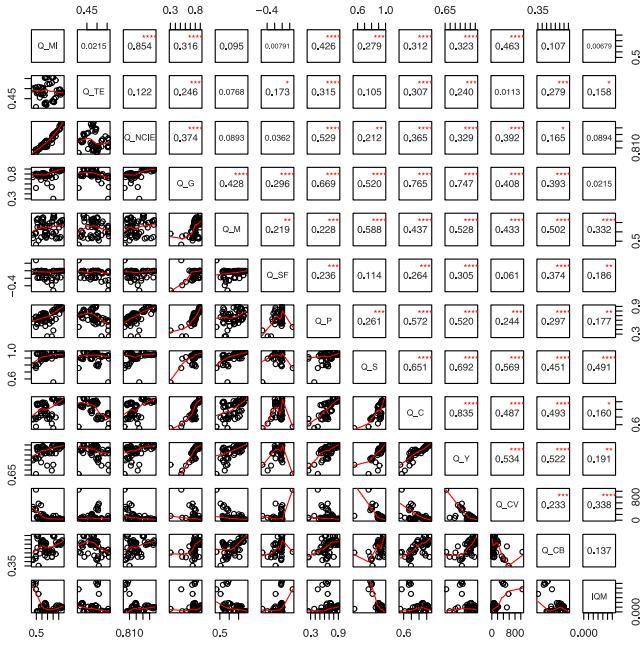


Fig. 2. The correlation matrix of fusion metrics for the VI-EVI modified fusion scheme.

other metrics do not have the same correlation when applied to the results obtained by the two different fusion schemes. However, the fusion metrics with a higher correlation generally come from a same category. The correlation analysis reveals the similarities between the same types of fusion metrics.

A dendrogram plot can be created from the similarity matrix with a tool called “DendroUPGMA” [37], [38]. The dendrogram tool transforms similarity coefficients into distances and clusters the coefficients using the unweighted pair group method with arithmetic mean (UPGMA) algorithm. The dendrogram plots are given in Fig. 3. The local topological relationships are identified in order of similarity (Kendall correlation), and the phylogenetic tree is built in a stepwise manner. The length represents the correlation between these fusion metrics and the difference between two fusion schemes can be observed. The visible and infrared images employ different intensity tables and this makes the joint gray-level histogram in VI-IR fusion quite different from that of the VI-EVI fusion. And, the information theory-based metrics use such information (e.g., joint histogram) to calculate the metric values. This partially explains the different tree structures. However, this may not

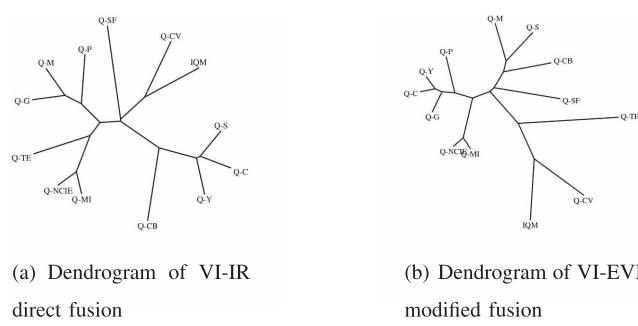


Fig. 3. The dendrogram of image fusion metrics.

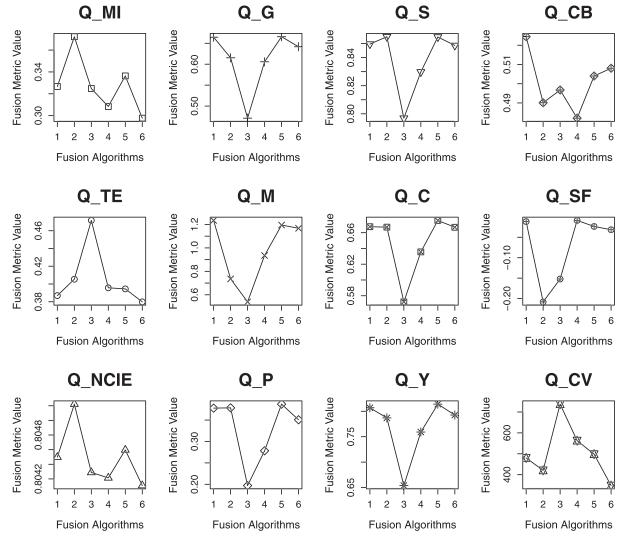


Fig. 4. Fusion metric values for VI-IR direct fusion. Numbers 1 to 6 refer to the fusion algorithm: LAP, GRAD, RoLP, DB4, SIDW, and STEER, respectively.

be enough to justify the shifting of Q_{TE} and should be further investigated. Besides, image contents may also have an impact on the metric value as well. The dendrogram is meaningful as the assessment rates the fusion algorithms based on a relative value. The fusion metrics are clustered based on similarity rather than their categories since the image structural similarity-based metrics may also depend on image features.

4.2.2 The Consistency of Assessment Metrics

An intuitive illustration of the change in the fusion metric value against fusion algorithms can be found in Figs. 4 and 5. All the metrics will assign a larger value to the better fusion result. These figures illustrate how the metric values change with the algorithms. To better understand the consistency of one metric with the others, the ranks of the fusion results/algorithms with an integer number from 1 to

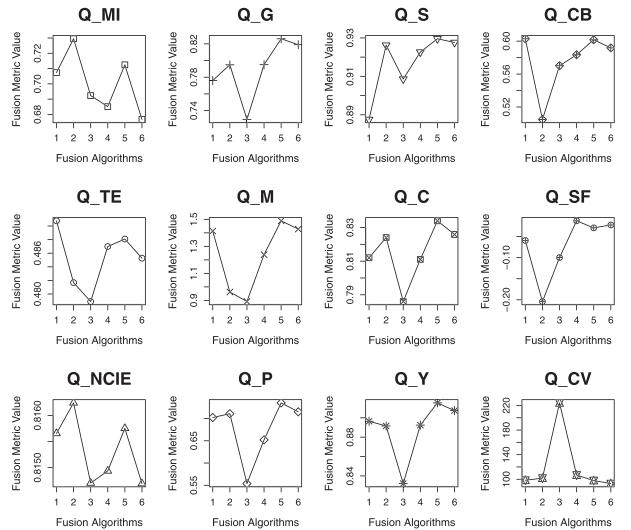


Fig. 5. Fusion metric values for VI-EVI modified fusion. Numbers 1 to 6 refer to the fusion algorithm: LAP, GRAD, RoLP, DB4, SIDW, and STEER, respectively.

TABLE 1
Algorithm Rank for Fusion Schemes

	VI-IR direct fusion						VI-EVI Modified fusion					
	LAP	GRAD	RoLP	DB4	SIDW	STEER	LAP	GRAD	RoLP	DB4	SIDW	STEER
Q_{MI}	4	6	3	2	5	1	4	6	3	2	5	1
Q_{TE}	2	2	6	4	3	1	6	2	1	4	5	3
Q_{NCIE}	4	6	3	2	5	1	4	6	2	3	5	1
Q_G	5	3	1	2	6	4	2	3	1	4	6	5
Q_M	6	2	1	3	5	4	4	2	1	3	6	5
Q_{SF}	5	1	2	6	4	3	3	1	2	6	4	5
Q_P	4	5	1	2	6	3	3	4	1	2	6	5
Q_S	4	6	1	2	5	3	1	4	2	3	6	5
Q_C	5	4	1	2	6	3	3	4	1	2	6	5
Q_Y	5	3	1	2	6	4	4	2	1	3	6	5
Q_{CV}	3	2	6	5	4	1	3	4	6	5	2	1
Q_{CB}	6	2	3	1	4	5	6	1	2	3	5	4
Borda count result	5	4	1	2.5	6	2.5	4	2	1	3	6	5

TABLE 2
The Variance of Fusion Metrics across Images

	VI-IR direct fusion						VI-EVI Modified fusion					
	LAP	GRAD	RoLP	DB4	SIDW	STEER	LAP	GRAD	RoLP	DB4	SIDW	STEER
Q_{MI}	7.69E-03	12.6E-03	7.09E-03	8.23E-03	9.20E-03	<u>5.39E-03</u>	2.49E-02	<u>1.78E-02</u>	2.06E-02	2.70E-02	2.57E-02	2.48E-02
Q_{TE}	9.66E-03	9.72E-03	<u>4.75E-03</u>	10.2E-03	10.8E-03	9.57E-03	1.87E-03	1.88E-03	1.99E-03	1.49E-03	1.55E-03	<u>1.38E-03</u>
Q_{NCIE}	2.07E-06	3.37E-06	1.90E-06	2.07E-06	2.36E-06	<u>1.56E-06</u>	4.46E-05	<u>3.42E-05</u>	4.04E-05	4.73E-05	4.73E-05	4.23E-05
Q_G	5.39E-03	3.29E-03	<u>2.93E-03</u>	5.43E-03	4.51E-03	4.14E-03	30.4E-03	3.35E-03	18.6E-03	4.34E-03	<u>2.73E-03</u>	3.16E-03
Q_M	9.88E-02	<u>3.05E-02</u>	3.91E-02	6.65E-02	7.55E-02	7.71E-02	1.34E-01	1.05E-01	1.79E-01	1.11E-01	<u>1.00E-01</u>	1.08E-01
Q_{SF}	73.5E-04	10.1E-04	497E-04	2.91E-04	3.68E-04	<u>1.88E-04</u>	269E-04	10.5E-04	337E-04	<u>1.71E-04</u>	4.31E-04	3.93E-04
Q_P	12.4E-03	10.6E-03	2.92E-03	8.06E-03	<u>1.03E-02</u>	9.02E-03	2.33E-02	<u>1.02E-02</u>	3.81E-02	1.87E-02	1.11E-02	1.32E-02
Q_S	3.38E-03	3.23E-03	5.32E-03	3.29E-03	<u>3.20E-03</u>	3.22E-03	17.2E-03	<u>2.44E-03</u>	4.42E-03	2.87E-03	2.70E-03	2.92E-03
Q_C	9.27E-03	9.34E-03	14.0E-03	9.10E-03	<u>8.97E-03</u>	10.2E-03	15.2E-03	<u>8.12E-03</u>	16.8E-03	10.0E-03	8.23E-03	9.40E-03
Q_Y	1.68E-03	<u>1.58E-03</u>	10.3E-03	1.66E-03	1.59E-03	1.78E-03	4.28E-03	<u>1.41E-03</u>	10.8E-03	2.30E-03	1.60E-03	1.89E-03
Q_{CV}	6.10E+04	5.84E+04	16.8E+04	10.0E+04	6.59E+04	<u>2.87E+04</u>	<u>1.01E+04</u>	2.86E+04	10.6E+04	1.89E+04	1.35E+04	1.28E+04
Q_{CB}	4.81E-03	3.61E-03	<u>1.52E-03</u>	5.19E-03	5.68E-03	5.12E-03	6.23E-03	<u>5.77E-03</u>	5.99E-03	6.89E-03	6.40E-03	6.59E-03

6 are given in Table 1. The results from Borda count method are listed on the last column in each table. A larger Borda count number indicates a better result.

For VI-IR direct fusion, the SIDW is ranked the best algorithm and next are the LAP and GRAD algorithms. The DB4 and STEER are given a number 2.5, which means they are equal and between the ranks 2 and 3. The last algorithm is the RoLP. Compared with the Borda count result, metrics Q_C , Q_P , and Q_Y show a reasonable consistency. In the results of VI-EVI modified fusion, Borda count ranks STEER the second and LAP the third. The rank for DB4 and GRAD are three and two, respectively. RoLP is again ranked last. Among all the metrics, Q_Y and Q_M show a perfect consistency with the Borda count result while Q_G and Q_P give a reasonably consistent result.

To understand the performance of a fusion metric across different inputs, the variance is calculated and listed in Table 2. As far as the fusion algorithm is concerned, an ideal fusion metric should not change with the contents of input images because the fusion metric evaluates the fusion algorithms rather than the image contents. A lower variance indicates a good stability of a fusion metric for a specific fusion algorithm. For example, metric Q_{NCIE} is most stable for fusion algorithm STEER in the VI-IR direct fusion, while, in the VI-EVI modified fusion, algorithm GRAD still prefers Q_{NCIE} for the stability.

4.2.3 The Impact of Image Distortion

The input UN camp images are distorted by additive white noise and blurring operation, respectively, in the experiment. The visible and IR images are evaluated with the IQM [35] and plotted in Fig. 6. The image quality degrades with the variance of Gaussian white noise and the standard deviation of the Gaussian filter. However, the IQM does not discriminate the severe degradation between images, which does not necessarily mean those images are of the same quality. Another observation is that the IR image is of lower quality in comparison with the corresponding visible image

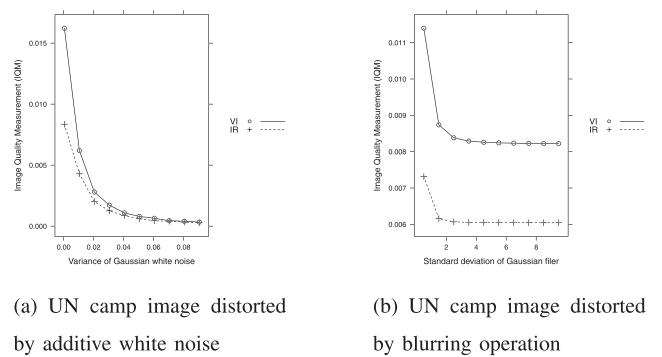


Fig. 6. The distorted UN camp images assessed by IQM.

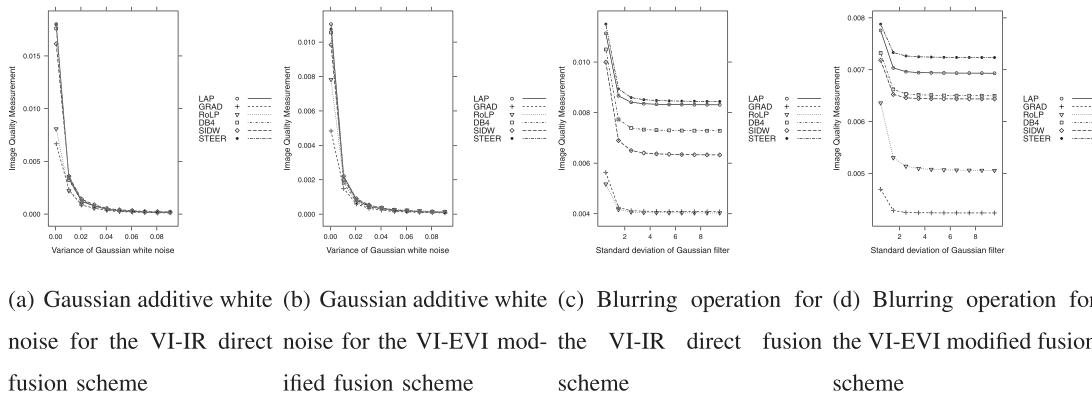


Fig. 7. The assessment of fusion algorithms under distortions with IQM.

in terms of IQM. And the IR image is more sensitive to the blurring operation.

We first look at how the IQM changes with the image quality. Fig. 7 indicates that the quality of fused images degrade with the decrease in image quality. When the Gaussian additive noise is severe, all the fused images are almost of the same quality, regardless of the fusion scheme. For blurred input images, varied fusion algorithms generate the fused images with different qualities in terms of IQM. The VI-EVI modified fusion has a relatively higher quality,

which means the degradation of the infrared image has a larger impact on the fusion result.

The impact of Gaussian additive white noise to fusion metrics is illustrated in Figs. 8 and 9, respectively. An example of theoretical analysis of correlation-based quality measures for weighted averaging image fusion was reported in [39]. The computation of a new diffuse prior monotonic likelihood ratio was further proposed in [40]. For MIF, a complicated theoretical analysis has not been reported. For the VI-IR direct fusion scheme, metrics Q_G, Q_M, Q_S, Q_C, Q_Y ,

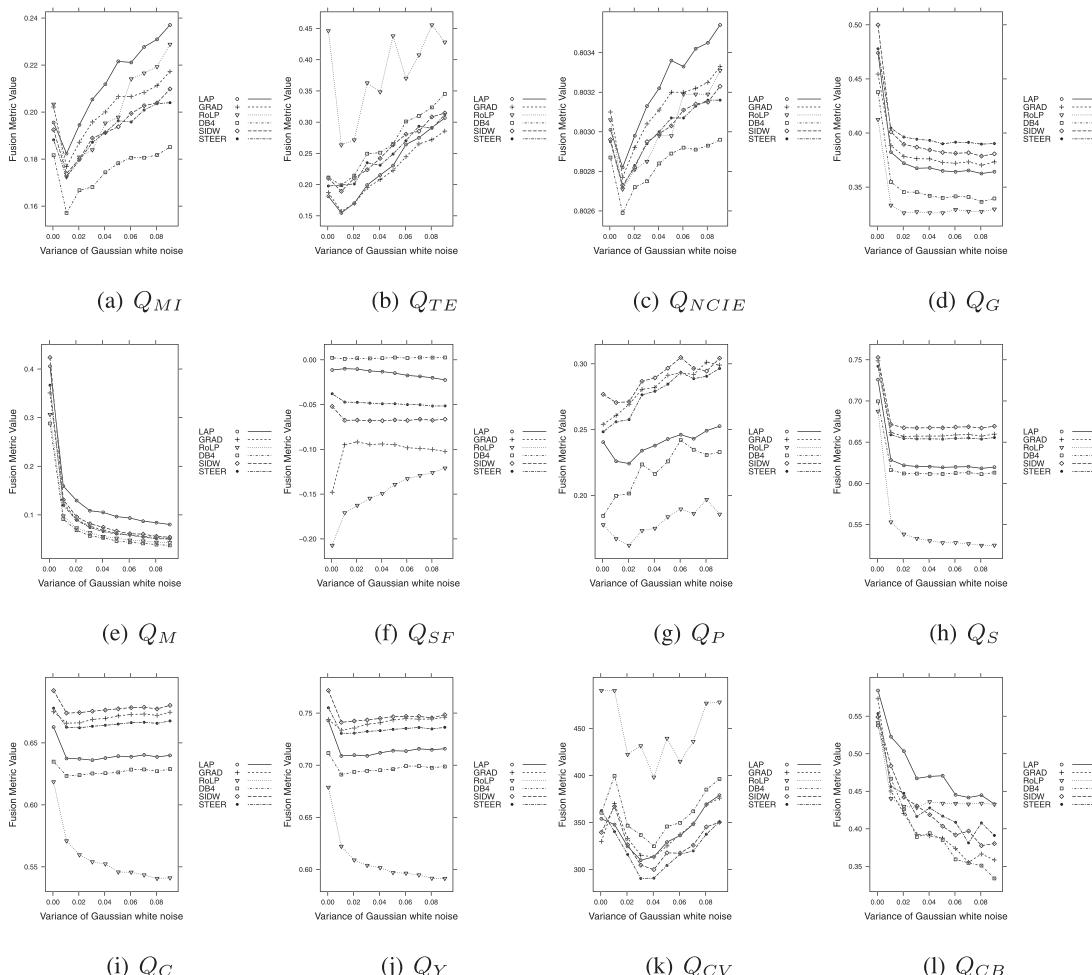


Fig. 8. The impact of Gaussian additive noise to the fusion metrics for the VI-IR direct fusion scheme.

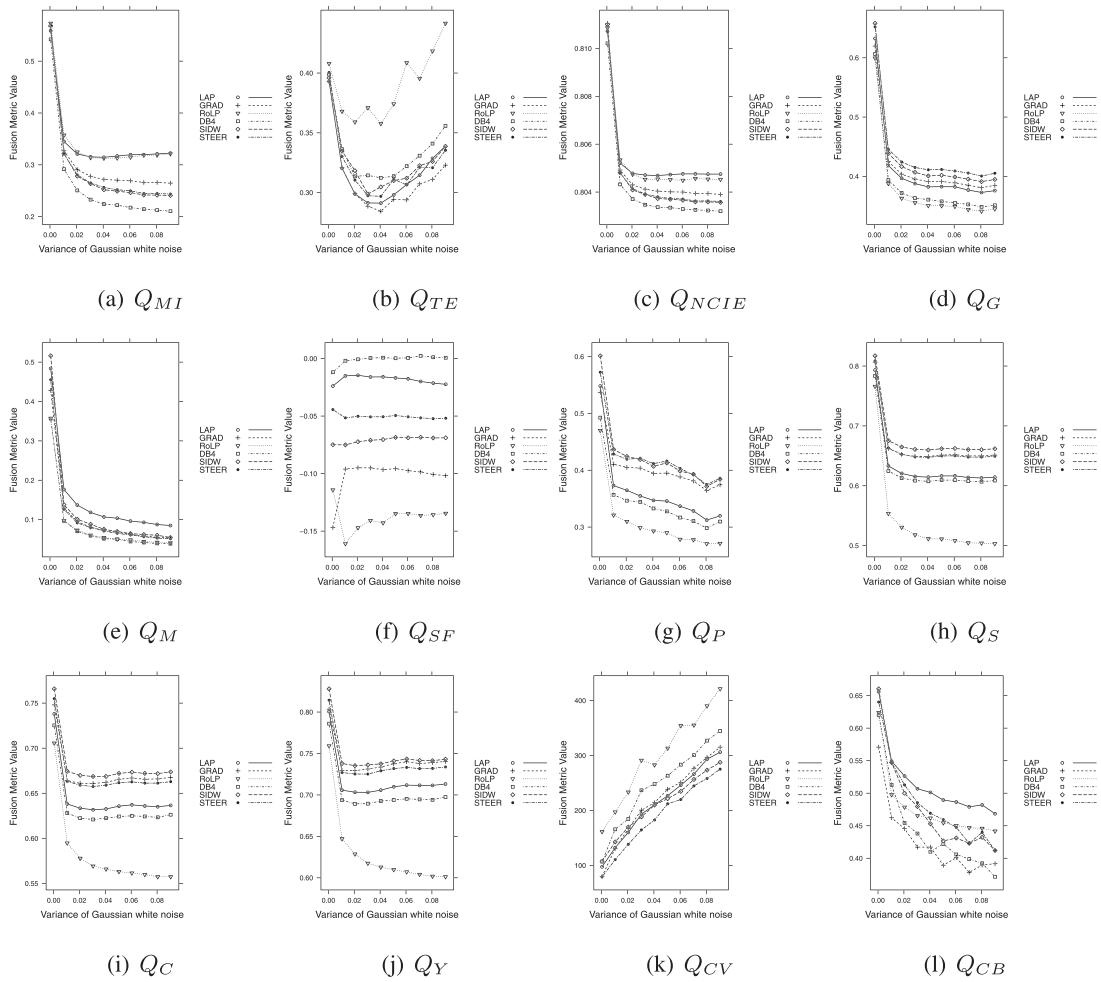


Fig. 9. The impact of Gaussian additive noise to the fusion metrics for the VI-EVI modified fusion scheme.

and Q_{CB} show a general decreasing trend with the degradation of image quality. An inflection appears around 0.01 variance value. In contrast, the value of metrics Q_{MI} , Q_{TE} , Q_{NCIE} , and Q_P increase because these metrics count the additive noises as part of the input “features” or “information.” When the variance value goes beyond 0.01 or the noise becomes significant for the Q_{MI} , Q_{TE} , Q_{NCIE} , and Q_P metrics. Q_{SF} demonstrates a relative stability since the Q_{SF} metric considers four directional gradients, which are not greatly affected by additive noises. The last metric Q_{CV} decreases at the beginning and increases around 0.04. In the Q_{CV} metric, a contrast sensitive filtering is applied to input images. The CSF carries out a band-pass filtering operation, which may suppress the noises to some extent (which depends on the specific CSF operation).

One difference between the direct and the modified image fusion schemes is the input image. In VI-IR direct fusion, an infrared image is input with a visible image while an enhanced visible image is used instead in the VI-EVI modified fusion. Thus, the changes of the results in Fig. 9 are subject to one of the input images. Most of the metric values decrease except Q_{TE} and Q_{CV} , while Q_{SF} does not show any significant change. Among all the metrics, the Q_{MI} , Q_{NCIE} , Q_{TE} , Q_P , and Q_{CV} are subject to such changes. As the IR and visible images have a different intensity definition, the three information theory-based approaches,

which calculate the fusion metric with pixel values, reflect the difference. The phase congruency-based metric Q_P incorporates a feature extraction function, which is sensitive to noise. While noise may have an impact on the fusion metric trend, it is not clear how phase congruency changes with noise.

The impact of blurring operation is illustrated in Figs. 10 and 11, respectively. As the infrared imaging measures the emitted energy of an object, the different regions in an IR image indicate the variance in temperature. IR images do not show a sharp edge or boundary as a visible image. The multiresolution analysis represents image features, like edges and boundaries, with larger coefficients. The blurring operation does not greatly change the temperature regions. Thus, in the VI-IR direct fusion, the fusion metrics give a relatively stable value for each fusion algorithm. The initial blurring may also serve as a low-pass filtering operation to the IR image and this may lead to a “better” result in terms of some metrics. For the VI-EVI modified fusion, all the metrics except Q_{SF} and Q_{CV} decrease when the standard deviation of the Gaussian filter increases.

A fusion metric can be affected by the quality change (i.e., pixel value, numerical rounding, etc.) at a certain step of its calculation. Thus, a metric may exhibit a different performance to such changes. It should also be noted that the fusion algorithms are subject to the quality of input

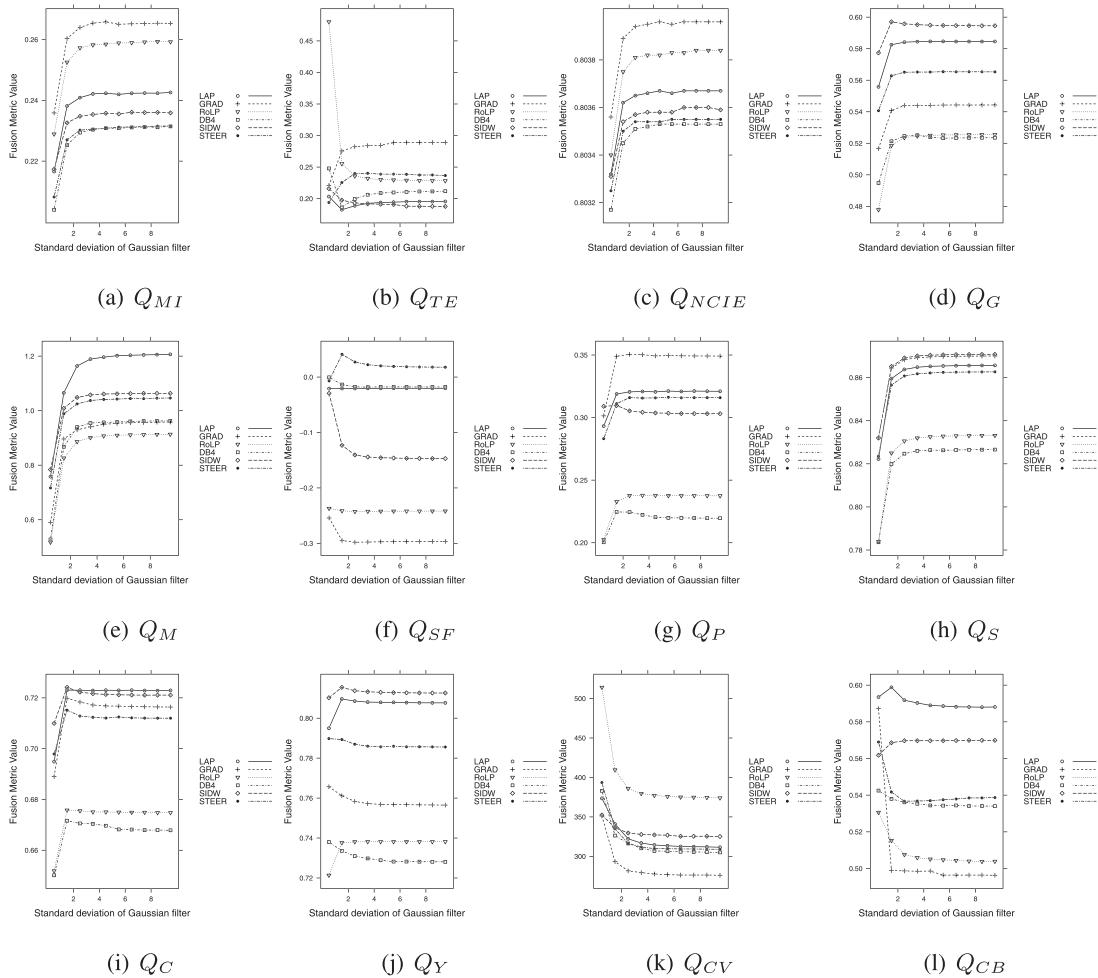


Fig. 10. The impact of blurring operation to the fusion metrics for the VI-IR direct fusion scheme.

images as well. Some are sensitive to the quality change and some are not, which is beyond the discussion of this paper. Readers can refer to the experimental results for further particulars.

To figure out how the fusion metrics are related to the IQM, the correlation is calculated and given in Fig. 12. No obvious correlation is observed for most metrics, although Q_{CV} obtained a larger correlation value. The lack of IQM-to-MIF metric correlation is because most fusion metrics count on how the input images are fused together rather than the quality of the fused image. Note: When the input images are of significantly different quality, we found that a fusion metric may lead to a confused judgment.

5 DISCUSSION

What does a fusion metric tell? If there are two pairs of multisensor images fused with the same algorithm, the performance of the algorithm should be the same, but the fusion metric values are not equal in most cases. Fusion metrics are calculated differently, measure various contextual details, and provide a relative value for comparison. From the experiments, we find that all fusion metrics considered in this study vary with the image contents. An MIF metric value is only meaningful in reference to the MIF goals when evaluating a specific image pair. For

example, mutual information gives a coarse estimate of the similarity between images and has been used for both image registration and image fusion. An MIF metric value is only a relative ranking of various fusion algorithms for a specific application. In general, we desire that the image fusion metric remain descriptive and consistent for algorithm selection over various sensors and environmental conditions.

The role of image fusion metrics is important for applications, user acceptance, and image fusion algorithm improvements. For a specific application and a specific fusion algorithm, should the fusion metric keep a constant value or need to vary with the quality of the input images? If the fusion metric only considers how much information is transferred from inputs to the fused result, the quality of the input images will not have an impact on the measure, as the metric should only reflect the capability of the fusion algorithms. However, the change of image contents (e.g., degradation) may change the amount of information transfer and thus change the fusion metric value. For example, if the two input images are blank, it does not mean a failure of the fusion algorithm. However, the night vision application requires a fused image be of a “good” quality as related to some standards. According to the “National Imagery Interpretability Rating Scale” [41], the different rating scales define the capability to

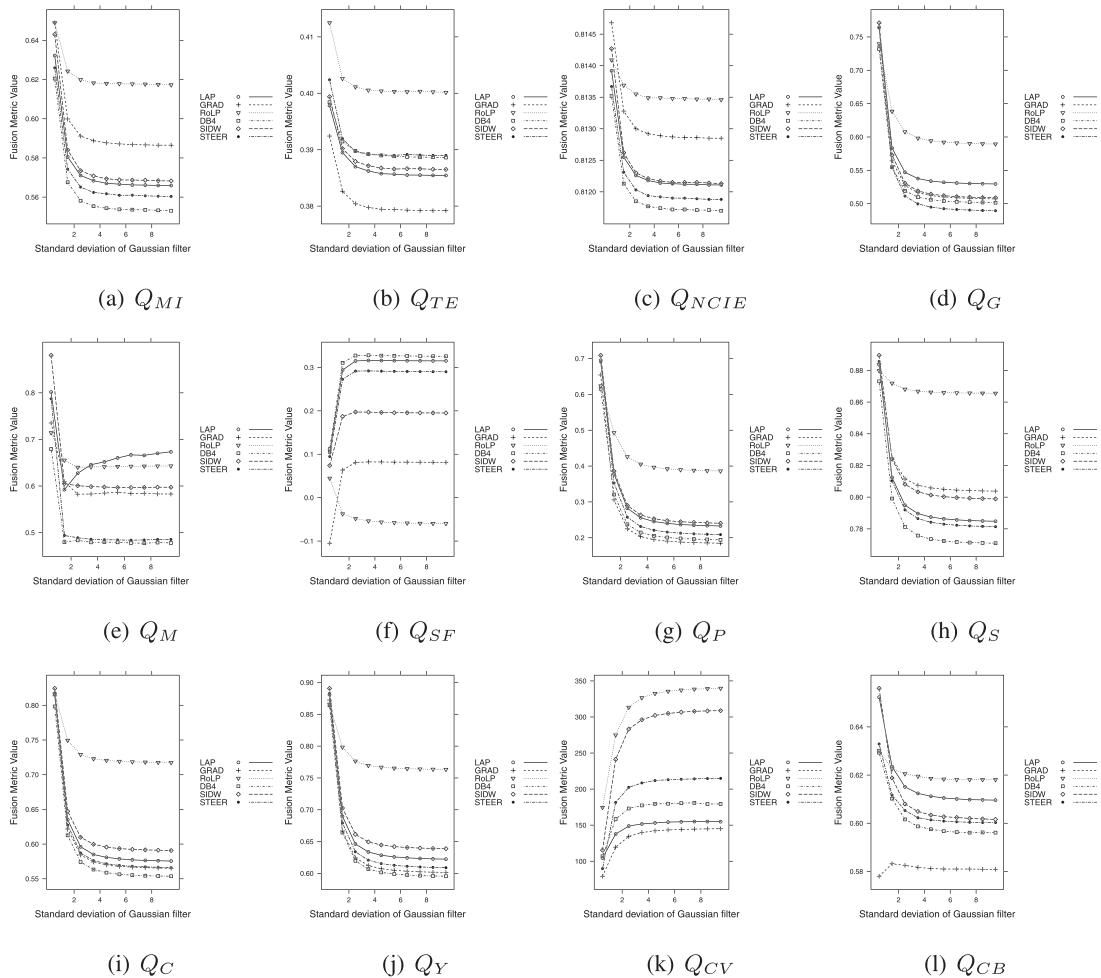


Fig. 11. The impact of blurring operation to the fusion metrics for the VI-EVI modified fusion scheme.

identify varied objects from the image. Therefore, the quality of fused image needs to be considered.

There are a host of metrics and an equivalent variation in how to measure the parameters that compose the metric. An obvious relation between the image quality measurement and image fusion metrics considered in this study was not observed. However, each image would have a quality rating and fusing the image qualities does not compute an

image fusion quality metric. In this study, we assume that the image fusion results in a composite image from which the image fusion quality metric is determined. As stated, the “enhancement” comes from the fused images (mostly of different modalities) and the motivation is to improve night vision imaging for improved image analysis.

The value of a metric varies from its definition and implementation. It could be in the range of $(-\infty, \infty)$, $[-1, 1]$,

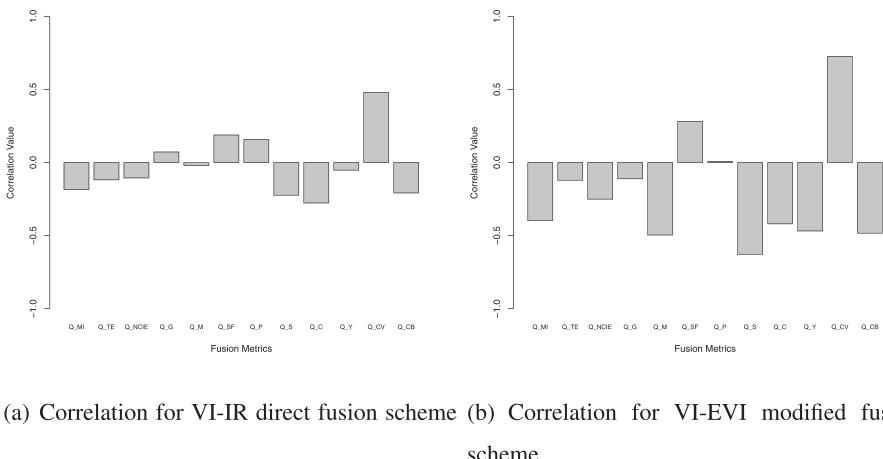


Fig. 12. The correlation between fusion metrics and image quality measurement.

[0, 1], or [0, +∞), but none of them gives an absolute measurement. In other words, how the fusion performance is distributed in the given range is not clear. Given two metric values 0.99 and 0.96, for instance, we do not know how significant the difference (0.03) is with these two results for a specific application. Thus, the normalization of the metric value to the range [0, 1] does not make any sense because the metric value is a relative result.

The fusion metrics can be applied to different image modalities. However, it also depends on what is expected from the fused image. A single fusion metric is not sufficient to justify all the requirements for the fusion applications like multifocus imaging, surveillance, and medical imaging, etc. In multifocus imaging, the input images are of the same modality, but for the night vision application, heterogeneous images are fused to highlight the “hot” human beings and the “cool” background. This MIF-CE study investigates the fusion of heterogeneous images (VI-IR direct fusion) and homogeneous images (VI-EVI modified fusion). According to Figs. 1 and 2 and results in Table 1, most fusion metrics rate the fusion algorithms differently in the two fusion schemes.

The choice of a fusion algorithm and fusion metric are application dependent. The application drives the requirements from which metric selection follows. As far as the night vision context enhancement application is concerned, the VI-EVI modified fusion scheme creates a fused image more suitable for human perception. In addition to the fusion metrics presented in this paper, a successful fusion of the IR and visible images can also be learned from the segmentation of the input and fused images [42]. To find a representative measure from the multiple fusion metrics, a hierarchical cluster analysis can be applied [43].

6 CONCLUSION

In this paper, we described 12 metrics used to assess the fusion performance of multimodal images. These fusion metrics are categorized into four groups:

1. information theory based metrics,
2. image feature based metrics,
3. image structural similarity based metrics, and
4. human perception inspired fusion metrics.

With the infrared and visible image pairs from night vision application of context enhancement, we investigated these fusion metrics over six multiresolution image fusion algorithms for two fusion schemes. One is the VI-IR direct image fusion (heterogeneous sensor images) and the other is the VI-EVI modified image fusion (homogeneous sensor images). The major difference between these two schemes is that the VI-IR direct fusion fuses two heterogeneous images while the VI-EVI modified fusion fuses two homogeneous images. Six multiresolution pixel-level fusion algorithms were considered in this study. The impact of image quality to the fusion metric was studied by applying Gaussian additive white noise and blurring operation to the input images. Various comparative approaches were conducted such as correlation, Borda count, and IQM metric-to-image quality relations. In addition, an image quality measurement based on image power spectrum was computed as a reference and compared with the fusion metrics.

The fusion metrics demonstrate its diversities due to the different mechanisms in its implementation.

From the experiments, we understand that the fusion metrics considered in this study only provide a relative assessment (value) on how the input images are fused together rather than the quality of the fused image. If a fusion metric only counts how the information is transferred to a fused image from inputs, the image quality should not affect metric value. Meanwhile, the metric value varies with image contents and is subject to distortion like additive noise and blurring. With this knowledge, a fusion metric can be selected based on application requirements, which is paramount for multisensor image fusion, or a representative measurement can be derived from multiple metrics with a hierarchical cluster analysis. We demonstrate a possible method to select image fusion metrics derived from two phenograms in the supplement, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.109>.

This study considers a night vision application. Metrics Q_G (gradient-based fusion metric), Q_C (Cvejic's metric), and Q_Y (Yang's metric)² are suggested for the VI-IR direction image fusion, while metrics Q_M (multiscale metric) and Q_Y are recommended to the VI-EVI modified image fusion. When other applications are considered, the use of a certain fusion metric will depend on operational requirements. A more reliable and universal fusion metric is expected from future research.

ACKNOWLEDGMENTS

The images used in the experiments are obtained from <http://www.imagefusion.org>. The authors wish to pay tribute to the contributors of these images for their valuable support to image fusion research.

REFERENCES

- [1] X.P.V. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*, K. Chang, ed. John Wiley and Sons, Inc., 2001.
- [2] R.S. Blum and Z. Liu, Eds., *Multi-Sensor Image Fusion and Its Applications*. Taylor and Francis, 2005.
- [3] G. Piella, “A General Framework for Multiresolution Image Fusion: From Pixels to Regions,” *Information Fusion*, vol. 4, no. 4, pp. 259-280, Dec. 2003.
- [4] G. Qu, D. Zhang, and P. Yan, “Information Measure for Performance of Image Fusion,” *Electronics Letters*, vol. 38, no. 7, pp. 313-315, 2002.
- [5] G. Piella and H. Heijmans, “A New Quality Metric for Image Fusion,” *Proc. Int'l Conf. Image Processing*, 2003.
- [6] C.S. Xydeas and V. Petrovic, “Objective Image Fusion Performance Measure,” *Electronics Letters*, vol. 36, no. 4, pp. 308-309, 2000.
- [7] J. Zhao, R. Laganiere, and Z. Liu, “Performance Assessment of Combinative Pixel-Level Image Fusion Based on an Absolute Feature Measurement,” *Int'l J. Innovative Computing, Information and Control*, vol. 3, no. 6(A), pp. 1433-1447, Dec. 2007.
- [8] Y. Zheng, E.A. Essock, B.C. Hansen, and A.M. Haun, “A New Metric Based on Extended Spatial Frequency and Its Application to DWT Based Fusion Algorithms,” *Information Fusion*, vol. 8, no. 2, pp. 177-192, Apr. 2007.
- [9] Y. Zheng, Z. Qin, L. Shao, and X. Hou, “A Novel Objective Image Quality Metric for Image Fusion Based on Renyi Entropy,” *Information Technology J.*, vol. 7, no. 6, pp. 930-935, 2008.
2. Q_C and Q_Y are similar.

- [10] N. Cvejic, C.N. Canagarajah, and D.R. Bull, "Image Fusion Metric Based on Mutual Information and Tsallis Entropy," *Electronics Letters*, vol. 42, no. 11, pp. 626-627, May 2006.
- [11] C. Yang, J. Zhang, X. Wang, and X. Liu, "A Novel Similarity Based Quality Metric for Image Fusion," *Information Fusion*, vol. 9, pp. 156-160, 2008.
- [12] Q. Wang, Y. Shen, and J. Jin, "Performance Evaluation of Image Fusion Techniques," *Image Fusion: Algorithms and Applications*, ch. 19, T. Stathaki, ed., pp. 469-492. Elsevier, 2008.
- [13] M. Hossny, S. Nahavandi, and D. Creighton, "A Quadtree Driven Image Fusion Quality Assessment," *Proc. Fifth IEEE Int'l Conf. Industrial Informatics*, vol. 1, pp. 419-424, July 2007.
- [14] M. Hossny, S. Nahavandi, and D. Vreighton, "Comments on 'Information Measure for Performance of Image Fusion,'" *Electronics Letters*, vol. 44, no. 18, pp. 1066-1067, Aug. 2008.
- [15] Z. Liu, D.S. Forsyth, and R. Laganiere, "A Feature-Based Metric for the Quantitative Evaluation of Pixel-Level Image Fusion," *Computer Vision and Image Understanding*, vol. 109, no. 1, pp. 56-68, Jan. 2008.
- [16] E. Blasch, X. Li, G. Chen, and W. Li, "Image Quality Assessment for Performance Evaluation of Image Fusion," *Proc. 11th Int'l Conf. Information Fusion*, June/July 2008.
- [17] M. Hossny, S. Nahavandi, D. Creighton, and A. Bhatti, "Image Fusion Performance Metric Based on Mutual Information and Entropy Driven Quadtree Decomposition," *Electronics Letters*, vol. 46, no. 18, pp. 1266-1268, Sept. 2010.
- [18] V. Petrovic, "Subjective Tests for Image Fusion Evaluation and Objective Metric Validation," *Information Fusion*, vol. 8, no. 2, pp. 208-216, 2007.
- [19] Z. Liu and R. Laganiere, "Context Enhancement through Infrared Vision: A Modified Fusion Scheme," *Signal, Image and Video Processing*, vol. 1, no. 4, pp. 293-301, Oct. 2007.
- [20] R. Nava, G. Cristóbal, and B. Escalante-Ramírez, "Mutual Information Improves Image Fusion Quality Assessments," SPIE News Room, <http://spie.org/documents/Newsroom/Imported/0824/0824-2007-08-30.pdf>, Sept. 2007.
- [21] N. Cvejic, A. Loza, D. Bul, and N. Canagarajah, "A Similarity Metric for Assessment of Image Fusion Algorithms," *Int'l J. Signal Processing*, vol. 2, no. 3, pp. 178-182, 2005.
- [22] G. Piella, "New Quality Measures for Image Fusion," *Proc. Int'l Conf. Information Fusion*, 2004.
- [23] Y. Chen and R.S. Blum, "A New Automated Quality Assessment Algorithm for Image Fusion," *Image and Vision Computing*, vol. 27, pp. 1421-1432, 2009.
- [24] H. Chen and P.K. Varshney, "A Human Perception Inspired Quality Metric for Image Fusion Based on Regional Information," *Information Fusion*, vol. 8, pp. 193-207, 2007.
- [25] P. Wang and B. Liu, "A Novel Image Fusion Metric Based on Multi-Scale Analysis," *Proc. IEEE Int'l Conf. Signal Processing*, pp. 965-968, 2008.
- [26] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 1, pp. 1-14, 2004.
- [27] Z. Wang and A.C. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, Mar. 2002.
- [28] "Dr. Zhou Wang's Website," <http://www.ece.uwaterloo.ca/~z70wang/research/ssim/>, Aug. 2009.
- [29] M. Hossny and S. Nahavandi, "Image Fusion Algorithms and Metrics Duality Index," *Proc. 16th IEEE Int'l Conf. Image Processing*, pp. 2169-2172, 2009.
- [30] G.A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *The Psychological Rev.*, vol. 63, pp. 81-97, 1956.
- [31] L.M. Kaplan and R.S. Blum, "Evaluation of Image Quality Features via Monotonic Analysis," Technical Report XAARLA-DELPHI, US Army Research Laboratory Adelphi, MD 20783, 2008.
- [32] L.M. Kaplan, S.D. Burks, R.S. Blum, R.K. Moore, and Q. Nguyen, "Analysis of Image Quality for Image Fusion via Monotonic Correlation," *IEEE J. Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 222-235, Apr. 2009.
- [33] M.G. Kendall, "A New Measure of Rank Correlation," *Biometrika*, vol. 30, nos. 1/2, pp. 81-93, <http://biomet.oxfordjournals.org/content/30/1-2/81.short>, 1938.
- [34] T.K. Ho, J.J. Hull, and S.N. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66-75, Jan. 1994.
- [35] N.B. Nill and B. Bouzas, "Objective Image Quality Measure Derived from Digital Image Power Spectra," *Optical Eng.*, vol. 31, no. 4, pp. 813-825, 1992.
- [36] MITRE, "Image Quality Evaluation," <http://www.mitre.org/tech/mtf/>, 2011.
- [37] S. Garcia-Vallve, J. Palau, and A. Romeu, "Horizontal Gene Transfer in Glycosyl Hydrolases Inferred from Codon Usage in Escherichia Coli and Bacillus Subtilis," *Molecular Biology and Evolution*, vol. 16, no. 9, pp. 1125-1134, Sept. 1999.
- [38] S. Garcia-Vallve and P. Puigbo, "Dendrogramma: A Dendrogram Construction Utility," <http://genomes.urv.cat/UPGMA/>, June 2010.
- [39] C. Wei and R.S. Blum, "Theoretical Analysis of Correlation-Based Quality Measures for Weighted Averaging Image Fusion," *Information Fusion*, vol. 11, pp. 301-310, June 2009.
- [40] C. Wei, L. Kaplan, S. Burks, and R. Blum, "Diffuse Prior Monotonic Likelihood Ratio Test for Evaluation of Fused Image Quality Measures," *IEEE Trans. Image Processing*, vol. 20, no. 2, pp. 327-344, Feb. 2011.
- [41] J.M. Irvine, "National Imagery Interpretability Rating Scale (NIIRS)," *Encyclopedia of Optical Eng.*, pp. 1442-1456, 2003.
- [42] A. Toet, M.A. Hogervorst, S.G. Nikolov, J.J. Lewis, T.D. Dixon, D.R. Bull, and C.N. Canagarajah, "Towards Cognitive Image Fusion," *Information Fusion*, vol. 11, no. 2, pp. 95-113, June 2009.
- [43] S. Li, Z. Li, and J. Gong, "Multivariate Statistical Analysis of Measures for Assessing the Quality of Image Fusion," *Int'l J. Image and Data Fusion*, vol. 1, no. 1, pp. 47-66, Mar. 2010.



Zheng Liu received the doctorate in engineering from Kyoto University, Japan, in 2000. From 2000 to 2001, he was a research fellow with the control and instrumentation division of Nanyang Technological University, Singapore. He then joined the Institute for Aerospace Research (IAR), National Research Council Canada, Ottawa, as a governmental laboratory visiting fellow in 2001. After being with IAR for five years, he transferred to the NRC Institute for Research in Construction, where he currently holds a research officer position. He holds an adjunct professorship at the University of Ottawa. His research interests include image/data fusion, computer vision, pattern recognition, sensor/sensor network, structural health monitoring, and nondestructive inspection and evaluation. He cochairs the IEEE IMS TC-36. He is a senior member of the IEEE and a member of SPIE.

Research in Construction, where he currently holds a research officer position. He holds an adjunct professorship at the University of Ottawa. His research interests include image/data fusion, computer vision, pattern recognition, sensor/sensor network, structural health monitoring, and nondestructive inspection and evaluation. He cochairs the IEEE IMS TC-36. He is a senior member of the IEEE and a member of SPIE.



Erik Blasch received the BS degree in mechanical engineering from the Massachusetts Institute of Technology in 1992 and master's degrees in mechanical engineering (1994), health science (1995), and industrial engineering (human factors) (1995) from the Georgia Institute of Technology and attended the University of Wisconsin for the MD/PhD degree in mechanical engineering/neurosciences until being called to active duty in 1996 for the United States Air Force.

He also received the MBA (1998), MSEE (1998), MS Econ (1999), MS/PhD psychology (ABD), and the PhD degrees in electrical engineering from Wright State University and is a graduate of the Air War College. He is currently a United States Air Force Research Laboratory (AFRL) exchange scientist to Defence R&D Canada (DRDC) at Valcartier, Quebec, in the Future Command and Control (C2) Concepts and Structures Group of the C2 Decision Support Systems Section. Prior to this sabbatical, he was the Information Fusion Evaluation Tech lead for the AFRL Sensors Directorate—COMprehensive Performance Assessment of Sensor Exploitation (COMPASE) Center and an adjunct electrical engineering and biomedical engineering professor at Wright State University (WSU) and the Air Force Institute of Technology (AFT) in Dayton, Ohio. He is also a reserve major with the US Air Force Office of Scientific Research (AFRL/AFOSR) in Washington, DC. He is currently a member of the IEEE AEES BoG, associate editor for the *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, and a member of the IEEE AEES Track Standards committee. He received the 2009 IEEE Russ Bioengineering award and supported the IEEE 2005 and 2008 Sections Congress meetings. He was a founding member of the International Society of Information Fusion (ISIF) in 1998 and the 2007 ISIF president. He began his career in the IEEE Robotics and Automation Society, compiling more than 30 top 10 finishes as part of robotic teams in international competitions, including winning the 1991 American Tour del Sol solar car competition, the 1994 AIAA mobile robotics contest, and the 1993 Aerial Unmanned Vehicle competition, where his team was first in the world to automatically control a helicopter. He has focused on automatic target recognition, targeting tracking, and information fusion research, compiling 300+ scientific papers and book chapters. He is a fellow of SPIE and a senior member of the IEEE.



Zhiyun Xue received the bachelor's and master's degrees in electrical engineering from Tsinghua University, China, in 1996 and 1998, respectively, and the PhD degree in electrical engineering from Lehigh University in 2006. She joined the Lister Hill National Center for Biomedical Communications at the National Library of Medicine (NLM) in 2006. Her research interests are in the areas of medical image analysis, computer vision, and pattern recognition.

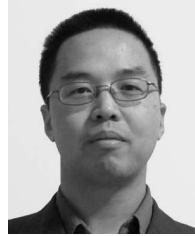


Jiyong Zhao received the PhD degree in electrical engineering from North China Electric Power University, and the PhD degree in computer engineering from Keio University. He is a professor with the School of Information Technology and Engineering, University of Ottawa, Canada. His research interests include image and video processing and multimedia communications. He is a member of the IEEE, the Institute of Electronics, Information and Communication Engineers (IEICE), and is also a member of the Professional Engineers Ontario (PEO).



Robert Laganière received the PhD degree from INRS-Telecommunications in Montreal in 1996. He is a full professor and a faculty member of the VIVA research lab at the School of Information Technology and Engineering at the University of Ottawa, Canada. His research interests are in computer vision and image processing with applications to visual surveillance, driver assistance, image-based modeling, and content-based video interpretation. He is the

author of *OpenCV Computer Vision Application Programming* (Packt Publishing) and coauthor of *Object-Oriented Software Development* (McGraw Hill). He is a member of the IEEE.



Wei Wu received the BS degree from Tianjin University, China, in 1998 and received the MS and PhD degrees in communication and information system from Sichuan University, China, in 2003 and 2008, respectively. He now is a faculty member at Sichuan University, China. His current research interests are image process and video communication, wireless communication, and super-resolution.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.