

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2023)01-0144-12

论文引用格式: Sun B, Gao Y X, Zhuge W W and Wang Z X. 2023. Analysis of quality objective assessment metrics for visible and infrared image fusion. Journal of Image and Graphics, 28(01): 0144-0155 (孙彬, 高云翔, 诸葛吴为, 王梓萱. 2023. 可见光与红外图像融合质量评价指标分析. 中国图象图形学报, 28(01): 0144-0155) [DOI: 10.11834/jig.210719]

可见光与红外图像融合质量评价指标分析

孙彬*, 高云翔, 诸葛吴为, 王梓萱

1. 电子科技大学航空航天学院, 成都 611731; 2. 飞行器集群智能感知与协同控制四川省重点实验室, 成都 611731

摘要: 目的 客观评价作为图像融合的重要研究领域, 是评价融合算法性能的有力工具。目前, 已有几十种不同类型的评价指标, 但各应用领域包括可见光与红外图像融合, 仍缺少统一的选择依据。为了方便比较不同融合算法性能, 提出一种客观评价指标的通用分析方法并应用于可见光与红外图像融合。**方法** 将可见光与红外图像基准数据集中的客观评价指标分为两类, 分别是基于融合图像的评价指标与基于源图像和融合图像的评价指标。采用 Kendall 相关系数分析融合指标间的相关性, 聚类得到指标分组; 采用 Borda 计数排序法统计算法的综合排序, 分析单一指标排序和综合排序的相关性, 得到一致性较高的指标集合; 采用离散系数分析指标均值随不同算法的波动程度, 选择充分体现不同算法间差异的指标; 综合相关性分析、一致性分析及离散系数分析, 总结具有代表性的建议指标集合。**结果** 在 13 对彩色可见光与红外和 8 对灰度可见光与红外两组图像源中, 分别统计分析不同图像融合算法的客观评价数据, 得到可见光与红外图像融合的建议指标集(标准差、边缘保持度), 作为融合算法性能评估的重要参考。相较于现有方法, 实验覆盖 20 种融合算法和 13 种客观评价指标, 并且不依赖主观评价结果。**结论** 针对可见光与红外图像融合, 提出了一种基于统计分析的客观评价指标分析方法, 该方法可以推广至更多的图像融合应用, 指导选择具有代表性的客观评价指标。

关键词: 图像融合; 客观评价指标; 相关性分析; 一致性分析; 离散系数

Analysis of quality objective assessment metrics for visible and infrared image fusion

Sun Bin*, Gao Yunxiang, Zhuge Wuwei, Wang Zixuan

1. School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China;

2. Aircraft Swarm Intelligent Sensing and Cooperative Control Key Laboratory of Sichuan Province, Chengdu 611731, China

Abstract: Objective As a research branch in the field of image fusion, objective assessment metrics can overcome these shortcomings of subjective evaluation methods that are easily affected by human psychological interference, surrounding environment, and visual characteristics. It can be utilized to evaluate algorithms and design parameters. Our algorithms advantages proposed can be demonstrated via objective assessment metrics. However, there is still a lack of benchmarks and metrics in various application fields like visible and infrared image fusion. A couple of objective assessment metrics can

收稿日期: 2021-08-16; 修回日期: 2022-03-10; 预印本日期: 2022-03-17

* 通信作者: 孙彬 sunbinhust@uestc.edu.cn

基金项目: 国家自然科学基金项目(62020106011); 四川省科技计划(2020YFG0231); 四川省中央引导地方科技发展专项(2020ZYD016); 电子科技大学基于项目的研究生创新培养计划建设项目(XMZ20203-9)

Supported by: National Natural Science Foundation of China(62020106011); Sichuan Science and Technology Program(2020YFG0231); Science & Technology Department of Sichuan Province(2020ZYD016); Postgraduate Course-Teaching Research and Innovation Project of UESTC(XMZ20203-9)

be selected based on prior experience. To facilitate the comparative analysis for different fusion algorithms, our research is focused on a general option method for objective assessment metrics and a set of recommended metrics for the fusion of visible and infrared images. **Method** A new selecting method for objective assessment metrics is built. Our method consists of three parts: 1) correlation analysis, 2) consistency analysis and 3) discrete analysis. The Kendall correlation coefficient is utilized to perform correlation analysis for all objective assessment metrics. All the objective assessment metrics are clustered according to the value of the correlation coefficient: if the Kendall value of two metrics is higher than the threshold, the two metrics will be put into the same group. The Borda voting method is used in the consistency analysis. There is a ranking for all algorithms in terms of each metric value. An overall ranking is also generated by Borda voting method based on each single ranking of different metrics. The correlation coefficient is used to analyze the consistency between each single ranking and the overall ranking. The objective assessment metric has higher consistency if its correlation coefficient value is higher. Such experiments showed that the metric value will be fluctuated if the fusion quality is changed. A good metric should reflect the fusion quality of different algorithms clearly, so the metric value will cause a large fluctuation in terms of different fusion quality. The different fusion quality we illustrated is originated from multiple algorithms. The coefficient of variation is used to interpret the fluctuation because different objective assessment metrics match different measurement scales. The coefficient of variation reflects overall fluctuations under the influence of the measurement scale. Therefore, the final selected objective assessment metrics set has the following three characteristics: 1) high consistency, 2) high coefficient of variation and 3) non-same group. **Result** The experiments are conducted on the visible and infrared fusion benchmark (VIFB) dataset. The experiments are segmented into two groups in terms of the visible images in the dataset in related to grayscale images and RGB color images. The recommended objective assessment metric set is under the fusion of visible and infrared image, color visible and infrared image fusion: {standard deviation (SD), $Q^{AB/F}$ } or {SD, Q_{CB} }; gray visible and infrared image fusion: {SD, $Q^{AB/F}$ } or { Q_{CB} }. Under the fusion of color visible and infrared image, both of $Q^{AB/F}$ and Q_{CB} had good consistency and coefficient of variation within the same group. It did not make much difference to choose each of them. Combining the results of the two sets of experiments, {SD, $Q^{AB/F}$ } is applied to visible and infrared image fusion. SD is focused on evaluating the contrast information of the fused image, which can intuitively reflect the quality of the fusion of visible and infrared images. $Q^{AB/F}$ is focused on evaluating the edge details. The comparative analysis shows that the algorithms filtered by two objective assessment metrics are similar to the individual subjective evaluation results. The objective assessment metric set selected by our method can be used as a basis for evaluating the performances of visible and infrared fusion algorithms. Compared to the existing methods, this method covers more fusion algorithms and objective assessment metrics without subjective evaluation results. **Conclusion** A general selecting method for objective assessment metrics is proposed. The method is not only matched for the fusion of visible and infrared images, but also applies to image fusion in other scenes. Our quick-response method can screen out the most representative objective assessment metric in a scene. Based on the benchmark of visible and infrared image fusion, the recommended representative objective assessment metrics of visible and infrared image fusion are SD and $Q^{AB/F}$.

Key words: image fusion; objective assessment metrics; correlation analysis; consistency analysis; coefficient of variation

0 引言

图像融合是利用多幅图像在时空上的相关性和互补性形成更清晰全面的图像(Li等,2011)。其中,红外和可见光图像融合广泛应用于遥感探测、安全导航、医学图像和交通监测等领域(Li和Wu,2019)。目前,不同融合算法之间没有统一的评价标准,多数情况下研究人员是凭经验选择客观评价

指标(宫睿和王小春,2019;陈木生,2016)衡量红外光与可见光融合算法。如何遴选有效的评价指标是亟待解决的问题。

图像融合质量评价方法可以分为两类(Liu等,2018)。第1类为主观评价方法,主要以评价人的肉眼观察结果为准,但不同级别之间的界限较模糊且主观性太强,易使结果受干扰(Yang等,2007)。第2类为客观评价方法,即根据一个特定算法提供量化指标,从不同角度评估融合的质量和性能。

杨艳春等人(2018)根据不同参考对象将融合评价指标分为3类,分别是基于融合图像统计特征、基于融合图像与参考图像和基于源图像与融合图像的客观评价指标。Liu等人(2012)将指标细分为基于信息论、图像特征、结构相似性和人类视觉系统4类,并分析指标间的相关性、一致性以及失真情况下与图像质量评价(image quality measurement, IQM)的关系,提出了在夜视应用中不同可见光和红外融合模型的建议指标集合。Liu等人(2018)总结了基于源图像与融合图像的评价指标,度量从输入图像到融合结果中传递的图像特征、内容或信息数量,但不包含基于融合图像统计特征的指标。张小利等人(2014)结合相关性分析和主客观一致性排序遴选适合遥感图像融合评价的指标集合,但无法推广至缺少主观评价结果的应用。

可见光和红外图像融合缺少主观评价结果和统一的基准。Zhang等人(2020)提出一个可见光和红外融合基准(visible and infrared fusion benchmark, VIFB),包括21个图像对、20种融合算法以及13种指标。本文依据参与评估的对象不同,将13种指标分为基于融合图像质量的客观评价和基于融合图像与源图像的客观评价指标,利用相关性分析、一致性分析以及离散程度分析,提出一种通用的客观评价指标筛选方法,并给出适用于红外光与可见光融合的建议指标集合。

1 图像融合客观评价指标

目前,已经提出了几十种图像融合客观评价指标。设参与融合的源图像有 n 个,图像融合的客观评价指标定义为

$$S = m(F, I_1, I_2, \dots, I_n, R) \quad (1)$$

式中, m 为评价指标, F 表示融合图像, I_n 表示源图像, R 表示参考图像。为简化表达,以2个待融合图像为例,源图像分别记为 A 和 B ,图像尺寸均为 $M \times N$ 。

大多数图像融合应用,包括红外和可见光融合一般并不存在参考图像。Zhang等人(2020)总结了4类13种客观评价指标,包括基于信息论的熵(entropy, EN)、交叉熵(cross entropy, CE)、互信息(mutual information, MI)和峰值信噪比(peak signal-to-noise ratio, PSNR);基于图像特征的平均梯度(aver-

age gradient, AG)、边缘强度(edge intensity, EI)、空间频率(space frequency, SF)、标准差(standard deviation, SD)和边缘保持度($Q^{AB/F}$);基于结构相似性的均方根误差(root mean squared error, RMSE)和结构相似度(structural similarity index measure, SSIM);基于人类感知的 Q_{CV} (Chen-Varshney)和 Q_{CB} (Chen-Blum)。

本文依据参与评估的对象不同将指标分为基于融合图像质量客观评价指标以及基于融合图像和源图像的客观评价指标两类。基于融合图像质量的客观评价指标包括SD、EI、EN、AG和SF等5个指标,基于融合图像和源图像的客观评价指标包括CE、MI、RMSE、PSNR、 $Q^{AB/F}$ 、SSIM、 Q_{CV} 和 Q_{CB} 等8个指标。基于融合图像质量的评价指标主要是针对融合后图像本身的特性或统计特征,没有考虑融合图像与源图像之间的关系,而融合是从融合源中继承传递有价值信息的过程,基于源图像和融合图像的客观指标是更为普遍使用的评价指标。

1.1 基于融合图像质量的评价指标

基于融合图像质量的评价指标计算仅利用融合图像即可获得指标评分。

1) 标准差(SD)。标准差反映的是融合图像分布和对比度的统计概念(Ma等, 2020)。数学定义为

$$SD = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - \mu)^2} \quad (2)$$

式中, μ 为融合图像的灰度值均值,由于人类视觉会对对比度较高的区域更敏感,因此,融合图像的对比度越高,SD越大,意味着融合图像融合效果越好。

2) 边缘强度(EI)。边缘强度使用Sobel算子提取融合图像的边缘信息,其表达式为

$$EI = \frac{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (s_x(i,j)^2 + s_y(i,j)^2)}}{M \times N} \quad (3)$$

式中, $s_x = F \times h_x$, $s_y = F \times h_y$, h_x 和 h_y 为 x 和 y 方向上的Sobel算子。边缘强度大,说明图像细节清晰,图像质量好。

3) 熵(EN)。熵是度量图像信息量的客观评价指标(刘子闻等, 2020),融合图像 F 的熵定义为

$$EN_F = - \sum_{f=1}^n p_f \log p_f \quad (4)$$

式中, $p = \{p_1, p_2, \dots, p_n\}$ 表示图像的灰度分布, p_f

为图像中灰度值为 f 的像素所占的比例, n 为灰度等级。该指标关注融合图像的灰度值分布, 可以反映融合图像携带的信息量。

4) 平均梯度 (AG)。平均梯度主要反映图像中细节间的区别 (Ma 等, 2019) 和纹理变化, 其表达式为

$$AG = \frac{1}{(M-1)(N-1)} \times \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \times \sqrt{\frac{(F(i+1, j) - F(i, j))^2 + (F(i, j+1) - F(i, j))^2}{2}} \quad (5)$$

通常情况下, 平均梯度越大, 图像的清晰度越高, 融合质量越好。

5) 空间频率 (SF)。空间频率用于测量图像的总活动水平 (Zheng 等, 2007)。图像 F 的空间频率表达式为

$$SF = \sqrt{(F_R)^2 + (F_C)^2 + (F_{MD})^2 + (F_{SD})^2} \quad (6)$$

式中, F_R, F_C, F_{MD} 和 F_{SD} 分别对应 4 个方向 (左、上、左上、右上) 上的一阶梯度。一般情况下, 空间频率越大, 融合质量越好。

1.2 基于源图像和融合图像的评价指标

基于源图像和融合图像的评价指标计算需要人为选择要衡量的有效信息, 如信息量、边缘和结构相似度等, 先分别计算源图像 A 、 B 与融合图像 F 之间的指标值, 然后根据规律加权出指标最终评分。

1) 均方根误差 (RMSE)。均方根误差计算融合图像与源图像的误差 (Ma 等, 2019), 从而测量融合图像和源图像之间的差异, 表达式为

$$RMSE = \frac{RMSE_{AF} + RMSE_{BF}}{2} \quad (7)$$

式中, $RMSE_{AF} = \sqrt{MSE_{AF}}$ 表示源图像 A 与融合图像 F 的差异, $MSE_{AF} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (A(i, j) - F(i, j))^2$ 。较小的 $RMSE$ 表示融合图像具有较少量的误差和失真。

2) 峰值信噪比 (PSNR)。峰值信噪比是融合图像中峰值功率与噪声功率的比值 (Ma 等, 2019), 反映了融合过程中的失真情况。数学表达式为

$$PSNR = 10 \lg \frac{[\max(F(i, j)) - \min(F(i, j))]^2}{MSE} \quad (8)$$

$PSNR$ 越大, 说明融合图像与源图像越接近, 即融合产生的误差越小。

3) 交叉熵 (CE)。交叉熵反映的是融合图像与源图像灰度信息分布的差异性, 可作为确定各种融合算法优劣的依据。在熵的基础上加入源图像 A 和 B 的灰度分布, 融合图像为 F 。以源图像 A 与融合图像 F 为例, 交叉熵计算式为

$$CE_F^A = \sum_{i=1}^n q_i \log_2 \left(\frac{q_i}{p_i} \right) \quad (9)$$

式中, q 为源图像 A 的灰度分布。

总体交叉熵计算式为

$$CE = \sqrt{\frac{(CE_F^A)^2 + (CE_F^B)^2}{2}} \quad (10)$$

交叉熵值越小, 图像间差异性越小, 则融合方法从源图像中提取的信息量越多, 理论上融合效果会比较好。

4) 互信息 (MI)。互信息可定量估计源图像 A 与融合图像 F 之间的依赖程度 (Qu 等, 2002)。源图像 A 与融合图像 F 的互信息可以由边缘概率密度 p_F, p_A 以及联合概率密度 p_{FA} 定义为

$$I_{FA}(f; a) = \sum_{f, a} p_{FA}(f, a) \log_2 \frac{p_{FA}(f, a)}{p_F(f) p_A(a)} \quad (11)$$

则总的互信息为

$$MI_F^{AB} = I_{FA}(f; a) + I_{FB}(f; b) \quad (12)$$

其中, 3 幅图像的灰度值范围分别为 $I_A(i, j) \in [0, I]$ 、 $I_B(i, j) \in [0, J]$ 和 $I_F(i, j) \in [0, K]$, 则 $a = 0, 1, \dots, I$ 、 $b = 0, 1, \dots, J$ 和 $f = 0, 1, \dots, K$ 。

融合图像的 MI 值越大, 说明融合图像从源图像处继承的信息越多, 融合图像的质量较好。

5) 边缘保持度 ($Q^{AB/F}$)。边缘保持度 (Xydeas 和 Petrović, 2000) 反映的是融合图像中源图像边缘信息的融入度。利用 Sobel 边缘检测算子提取该点的边缘强度 $g(i, j)$ 和方向信息 $a(i, j)$, A 的边缘强度和方向信息分别为 $g_A(i, j) = \sqrt{s_A^x(i, j)^2 + s_A^y(i, j)^2}$ 和 $a_A(i, j) = \tan^{-1} \left(\frac{s_A^y(i, j)}{s_A^x(i, j)} \right)$, 其中, $s_A^x(i, j)$ 和 $s_A^y(i, j)$ 是 Sobel 模板以像素点 $p(i, j)$ 为中心, 对源图像 A 的对应像素沿水平和垂直方向做卷积的结果。输入图像 A 和融合图像 F 的 $G^{AF}(i, j)$ 和 $A^{AF}(i, j)$ 为

$$G^{AF}(i, j) =$$

$$\begin{cases} \frac{g_F(i,j)}{g_A(i,j)} & g_A(i,j) > g_F(i,j) \\ \frac{g_A(i,j)}{g_F(i,j)} & \text{其他} \end{cases} \quad (13)$$

$$A^{AF}(i,j) = 1 - \frac{|a_A(i,j) - a_F(i,j)|}{\pi/2} \quad (14)$$

边缘强度和方向的保留信息为 $Q_g^{AF}(i,j) = \frac{\Gamma_g}{1 + e^{k_g(G^{AF}(i,j) - \sigma_g)}}$ 和 $Q_a^{AF}(i,j) = \frac{\Gamma_a}{1 + e^{k_a(A^{AF}(i,j) - \sigma_a)}}$ 。系数 Γ_g, k_g, σ_g 和 Γ_a, k_a, σ_a 决定了边缘强度和方向保留信息函数的形状。所以,边缘保持度可以定义为 $Q^{AF}(i,j) = Q_g^{AF}(i,j)Q_a^{AF}(i,j)$, $0 \leq Q^{AF}(i,j) \leq 1$, 0 表示在该像素点处完全丢失了边缘信息,1 表示融合过程中没有边缘信息丢失。

对于两幅源图像 A 和 B 融合出 F 的情况,其边缘保持度定义为

$$Q^{AB/F} = \frac{\sum_{i=1}^N \sum_{j=1}^M Q^{AF}(i,j)\omega^A(i,j) + Q^{BF}(i,j)\omega^B(i,j)}{\sum_{i=1}^N \sum_{j=1}^M (\omega^A(i,j) + \omega^B(i,j))} \quad (15)$$

通常, $\omega^A(i,j) = [g_A(i,j)]^L$, $\omega^B(i,j) = [g_B(i,j)]^L$, L 是一个常数。 $0 \leq Q^{AB/F}(i,j) \leq 1$, 指标值接近 1 时,说明边缘保持特性越好,融合的效果越好。实验中常用参数取值为 $L = 1$, $\Gamma_g = 0.9994$, $k_g = -15$, $\sigma_g = 0.5$, $\Gamma_a = 0.9879$, $k_a = -22$, $\sigma_a = 0.8$ 。

6) 结构相似度。结构相似度 (SSIM) 的测量可以对图像的失真做出近似 (Martinez 等, 2019), SSIM 的取值范围为 $[-1, 1]$, 值越大说明源图像和融合图像的相似程度越高。图像 A 和 F 的结构相似度为

$$SSIM_{AF} = l(A, F) \times c(A, F) \times s(A, F) \quad (16)$$

式中, $l(A, F) = \frac{2u_A u_F}{u_A^2 + u_F^2}$ 表示两幅图像的亮度相似性, $c(A, F) = \frac{2\sigma_A \sigma_F}{\sigma_A^2 + \sigma_F^2}$ 表示两幅图像的对比度相似性, $s(A, F) = \frac{\sigma_{AF}}{\sigma_A \sigma_F}$ 表示两幅图像的灰度值离散程度的结构相似性。 u_A, u_F 和 σ_A, σ_F 分别为 A 和 F 的灰度均值和标准差, σ_{AF} 表示 A 和 F 的协方差。结构相似性越高,理论上融合图像的质量越好。

7) Chen-Varshney 指标 (Q_{CV})。Chen-Varshney 方法是在人类视觉系统模型的基础上,使用 Sobel 算子提取源图像以及融合图像的边缘信息 (Chen 和 Varshney, 2007), 获得边缘强度图 G 。首先,采用局部区域的边缘强度的平方和对显著性进行度量, $\lambda(A^W) = \sum_{w \in W} G_A(w)^\alpha$, α 为常数。接着,用 CSF (contrast sensitivity function) 滤波器 $\hat{f}^W$ 得到区域相似性 $D(A^W, F^W) = \frac{1}{|W|} \sum_{w \in W} \hat{f}_r^W(i,j)^2$, r 为局部区域中像素的数量。全局质量评价为

$$Q_{CV} = \frac{\sum_{l=1}^L (\lambda(A^{W_l})D(A^{W_l}, F^{W_l}) + \lambda(B^{W_l})D(B^{W_l}, F^{W_l}))}{\sum_{l=1}^L (\lambda(A^{W_l}) + \lambda(B^{W_l}))} \quad (17)$$

式中, L 表示局部窗口的数量, W^l 为第 l 个局部窗口对应的区域。此方法得到的值越小,说明融合效果越好。

8) Chen-Blum 指标 (Q_{CB})。Chen 和 Blum (2009) 指出 Chen-Varshney 指标没有归一化处理以及没有充分利用人类视觉系统的感知特性,因此在 Chen-Varshney 指标的基础上,提出一种 Chen-Blum 模仿人类视觉系统。首先,使用 Peli's 对比灵敏度滤波器 C 对图像进行滤波。然后,对所有图像中每个像素计算对比度保留度 $C'_A(i,j) = \frac{t(C_A(i,j))^p}{h(C_A(i,j))^q + Z}$, t, h, p, q 和 Z 是确定掩蔽函数非线性形状的实数标量参数。接着,生成像素的显著图 $\lambda_A(i,j) = \frac{C'^2_A(i,j)}{C'^2_A(i,j) + C'^2_B(i,j)}$ 以及源图像到融合图像转化的保留信息量。具体为

$$Q_{AF}(i,j) = \begin{cases} \frac{C'_A(i,j)}{C'_F(i,j)} & C'_A(i,j) < C'_F(i,j) \\ \frac{C'_F(i,j)}{C'_A(i,j)} & \text{其他} \end{cases} \quad (18)$$

最后,得到全局质量图,即

$$Q_{GQM}(i,j) = \lambda_A(i,j)Q_{AF}(i,j) + \lambda_B(i,j)Q_{BF}(i,j) \quad (19)$$

Chen-Blum 指标结果为全部像素全局质量图的均值。实验结果表明,该方法与主观评价结果具有高度一致性,值越大说明融合效果越好。

2 客观评价指标分析

Zhang 等人(2020)提出的 VIFB 包含 13 对彩色可见光(RGB 多通道)和红外图像对,以及 8 对灰度可见光(单通道)和红外图像对,涵盖广泛的场景和工作条件,例如室内、室外、低照度和过度曝光等。Zhang 等人(2020)总结了 20 种开源图像融合方法在 VIFB 数据集的融合实验,包括基于深度学习的卷积神经网络(convolutional neural network, CNN)、深度学习框架(deep learning framework, DLF)和残差网络(residual network, ResNet)融合算法;基于多尺度的多分辨率奇异值分解(multi-resolution singular value decomposition, MSVD)、各向异性扩散变换(anisotropic diffusion fusion, ADF)和交叉双边滤波器(cross bilateral filter, CBF);基于引导滤波器的内容增强(guided filter context enhancement, GFCE)和引导滤波融合(guided filtering fusion, GFF);基于引导滤波器的混合多尺度分解(hybrid multi-scale decomposition with guided filter, HMSD_GF)、混合多尺度分解(hybrid multi-scale decomposition, Hybrid_MSD)和多尺度引导滤波(multi-scale guided image

and video fusion, MGFF)融合算法;基于显著性的双尺度图像融合(two-scale image fusion, TIF)和潜在低秩表示(latent low-rank representation, LatLRR)融合算法;基于子空间的四阶偏微分方程(fourth order partial differential equations, FPDE)融合算法;基于混合方法的多尺度稀疏(multi-scale transform and sparse representation, MST_SR)、非下采样轮廓波稀疏(nonsubsampled contourlet transform and sparse representation, NSCT_SR)、比率金字塔稀疏(ratio pyramid and sparse representation, RP_SR)、视觉显著图以及加权最小二乘(visual saliency map and weighted least square, VSMWLS),以及梯度转移融合(gradient transfer fusion, GTF)、红外特征提取与视觉信息保存(infrared feature extraction and visual information preservation, IFEVIP)等方法。

21 组图像分为 13 组灰度可见光与红外图像对(Gray 组)和 8 组彩色可见光与红外图像对(RGB 组),分别统计上述 20 种融合算法在分组数据实验中 13 种指标的平均值,结果如表 1 和表 2 所示。其中,彩色可见光与红外图像融合指标结果为 RGB 3 个通道分别与红外图像融合计算指标值后的平均结果。

表 1 在 VIFB 数据集中不同算法融合 13 组彩色(RGB)可见光图像与红外图像的 13 种指标均值

Table 1 Average of 13 metrics about 13 sets of color image and infrared image fused by different algorithms in VIFB dataset

算法	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
ADF	5.073	1.534	51.884	6.904	2.072	57.143	0.474	0.469	901.377	0.132	15.382	1.325	38.459
CBF	8.138	1.002	85.379	7.385	1.989	56.374	0.529	0.493	2 123.500	0.158	22.129	1.069	50.694
CNN	6.526	0.986	68.178	7.338	2.474	56.593	0.612	0.563	711.692	0.150	20.311	1.324	63.689
DLF	4.320	1.413	43.975	6.836	2.194	57.191	0.419	0.457	858.831	0.131	13.502	1.396	37.937
FPDE	5.114	1.368	52.254	6.884	2.065	57.130	0.439	0.463	896.877	0.132	14.811	1.303	38.258
GFCE	7.721	1.589	80.613	7.219	1.887	55.482	0.467	0.504	1 083.892	0.191	22.602	1.165	53.852
GFF	5.784	1.189	60.427	7.177	2.445	56.861	0.554	0.555	1 230.438	0.140	17.938	1.334	52.278
GTF	4.807	1.086	49.242	6.864	2.347	56.640	0.420	0.445	2 245.385	0.148	15.706	1.301	45.574
HMSD_GF	7.113	1.255	74.599	7.282	2.219	56.668	0.573	0.554	740.585	0.147	21.687	1.332	59.448
Hybrid_MSD	7.005	1.324	73.092	7.308	2.313	56.889	0.585	0.567	699.992	0.140	21.590	1.344	56.155
IFEVIP	5.929	1.125	62.002	7.098	2.524	55.943	0.517	0.473	672.331	0.174	17.753	1.326	52.105
LatLRR	10.005	1.560	104.586	6.880	1.826	55.134	0.413	0.485	858.377	0.207	31.580	1.138	63.297
MGFF	6.673	1.351	69.808	7.279	1.948	56.927	0.534	0.526	884.423	0.138	19.759	1.338	49.893
MST_SR	6.593	0.869	68.974	7.355	2.501	56.625	0.616	0.589	691.208	0.148	20.308	1.324	60.068
MSVD	4.054	1.434	41.667	6.819	2.124	57.163	0.339	0.447	889.554	0.131	13.716	1.349	37.640
NSCT_SR	7.401	0.832	78.088	7.457	2.573	56.154	0.590	0.556	2 021.423	0.166	21.009	1.179	55.095
ResNet	4.130	1.345	42.302	6.836	2.121	57.189	0.387	0.456	852.746	0.131	12.627	1.394	38.100
RP_SR	7.344	0.944	75.508	7.370	2.044	56.389	0.492	0.561	1 234.500	0.156	23.772	1.238	57.648
TIF	6.210	1.460	65.216	7.223	1.935	56.963	0.537	0.526	857.208	0.137	18.916	1.329	47.239
VSMWLS	6.334	1.550	65.073	7.127	2.167	56.967	0.511	0.500	1 040.638	0.137	19.104	1.355	48.190

表 2 在 VIFB 数据集中不同算法融合 8 组灰度可见光图像与红外图像的 13 种指标均值

Table 2 Average of 13 metrics about 8 sets of gray image and infrared image fused by different algorithms in VIFB

算法	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
ADF	3.784	1.349	37.826	6.601	1.675	60.456	0.594	0.481	577.038	0.060	12.102	1.523	29.865
CBF	5.556	0.982	57.058	7.225	2.440	59.579	0.658	0.579	684.063	0.074	17.538	1.337	45.050
CNN	4.640	1.101	47.343	7.291	2.944	60.108	0.732	0.716	188.988	0.065	16.379	1.500	54.204
DLF	3.020	1.413	29.785	6.542	1.763	60.480	0.459	0.426	598.900	0.059	10.848	1.568	29.486
FPDE	3.602	1.363	35.894	6.576	1.693	60.468	0.556	0.455	590.363	0.060	11.286	1.524	29.524
GFCE	7.135	2.486	72.351	7.342	1.772	56.682	0.476	0.584	598.425	0.143	22.236	1.084	47.842
GFF	4.580	1.190	46.702	7.263	2.952	60.114	0.738	0.723	314.800	0.065	16.188	1.502	46.455
GTF	3.483	1.609	34.599	5.929	1.412	59.844	0.469	0.364	1964.463	0.069	13.177	1.484	18.160
HMSD_GF	4.838	1.018	49.491	7.261	2.882	60.008	0.706	0.684	195.563	0.066	17.006	1.496	54.642
Hybrid_MSD	4.697	1.148	47.891	7.299	3.117	60.261	0.719	0.714	203.550	0.062	16.522	1.505	52.918
IFEVIP	3.450	1.687	35.175	6.673	1.799	59.173	0.436	0.445	413.600	0.080	12.746	1.497	42.619
LatLRR	7.267	1.887	73.682	6.955	1.372	57.879	0.478	0.516	435.513	0.107	26.216	1.260	47.118
MGFF	4.484	1.204	45.656	6.847	1.476	60.300	0.636	0.569	339.663	0.062	14.922	1.516	35.185
MST_SR	4.647	1.100	47.466	7.314	3.310	60.105	0.734	0.735	248.838	0.065	16.368	1.497	52.837
MSVD	2.717	1.508	27.323	6.519	1.679	60.450	0.320	0.393	678.113	0.060	10.590	1.549	29.061
NSCT_SR	5.016	1.012	51.491	7.297	3.663	59.516	0.736	0.715	514.450	0.075	16.756	1.435	48.216
ResNet	2.933	1.396	29.055	6.570	1.773	60.475	0.440	0.426	516.963	0.059	10.289	1.568	29.805
RP_SR	4.771	1.074	48.501	7.325	2.810	60.033	0.686	0.679	327.150	0.066	16.944	1.486	52.817
TIF	4.498	1.227	45.850	6.835	1.494	60.276	0.660	0.575	216.175	0.062	15.826	1.511	35.175
VSMWLS	4.439	1.179	44.543	6.866	1.820	60.187	0.623	0.490	290.063	0.064	15.319	1.518	43.104

2.1 相关性分析

相关性分析是通过计算指标变量间的相关系数分析不同指标间的潜在联系,融合客观评价指标基于相关程度而不是指标类别进行聚类分析。常用的相关性分析方法包括 Pearson、Spearman 和 Kendall 等。其中, Pearson 法用于衡量两变量间的线性相关性,仅适用于服从或近似服从正态分布的情况(张小利等,2014)。本文考察不同算法的多个评价指标间的相关性,选择 Kendall 方法进行分析。

令 X_i 和 Y_i 表示两个随机变量 X 和 Y 的第 i 个值 ($0 \leq i \leq N$), 这里 N 为变量长度。Kendall 计算式为

$$\tau = 2 \times \frac{C - D}{N(N-1)} \quad (20)$$

式中, C 为集合中一致的元素个数即 $(X_i > X_j, Y_i > Y_j)$ 或 $(X_i < X_j, Y_i < Y_j)$, D 为集合中非一致的元素个数即 $(X_i > X_j, Y_i < Y_j)$ 或 $(X_i < X_j, Y_i > Y_j)$ 。

Kendall 相关系数的取值范围为 $[-1, 1]$ 。当 $\tau = 1$ 时, 即 X 与 Y 具有一致的等级相关性; 当 $\tau = -1$ 时, 即 X 与 Y 具有完全相反的等级相关性; 当 $\tau = 0$ 时, 表示 X 与 Y 相互独立。

对表 1 和表 2 中 13 项指标之间分别进行 Kendall 相关系数分析, 结果如表 3 和表 4 所示。当指标 Kendall 相关性 $|\tau| > 0.7$, 则将指标视为同一组, 在表中以粗体表示。

彩色 (RGB 组) 可见光与红外图像融合指标聚类分组结果如图 1 所示, 其中 AG 与 EI、SF 的 Kendall 系数分别为 0.99 和 0.84, 指标侧重于反映邻域的灰度变化; PSNR 与 RMSE 高度负相关, 其中 RMSE 是负向指标, PSNR 是正向指标; $Q^{AB/F}$ 与 EN、 Q_{CB} 的系数分别为 0.70 和 0.74, 该组指标的定义有较大区别, EN 反映的是信息熵, 而 $Q^{AB/F}$ 和 Q_{CB} 主要反映边缘和对比度的保留程度。

表3 彩色(RGB)可见光图像与红外图像融合评价指标之间的 Kendall 相关系数

Table 3 Kendall correlation coefficient between color (RGB) visible image and infrared image fusion assessment metrics

	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
AG	1.00	-0.15	0.99	0.63	-0.16	-0.63	0.34	0.41	0.09	0.63	0.84	-0.45	0.56
CE	-0.15	1.00	-0.16	-0.44	-0.53	0.33	-0.39	-0.29	0.06	-0.33	-0.12	0.23	-0.27
EI	0.99	-0.16	1.00	0.64	-0.17	-0.64	0.35	0.42	0.08	0.64	0.83	-0.46	0.55
EN	0.63	-0.44	0.64	1.00	0.16	-0.46	0.70	0.69	0.04	0.46	0.58	-0.28	0.55
MI	-0.16	-0.53	-0.17	0.16	1.00	-0.06	0.32	0.16	-0.20	0.06	-0.15	0.14	0.09
PSNR/dB	-0.63	0.33	-0.64	-0.46	-0.06	1.00	-0.24	-0.27	-0.06	-1.00	-0.64	0.65	-0.59
$Q^{AB/F}$	0.34	-0.39	0.35	0.70	0.32	-0.24	1.00	0.74	-0.15	0.24	0.28	-0.04	0.53
Q_{CB}	0.41	-0.29	0.42	0.69	0.16	-0.27	0.74	1.00	-0.16	0.27	0.44	-0.07	0.62
Q_{CV}	0.09	0.06	0.08	0.04	-0.20	-0.06	-0.15	-0.16	1.00	0.06	0.11	-0.26	-0.12
RMSE	0.63	-0.33	0.64	0.46	0.06	-1.00	0.24	0.27	0.06	1.00	0.64	-0.65	0.59
SF	0.84	-0.12	0.83	0.58	-0.15	-0.64	0.28	0.44	0.11	0.64	1.00	-0.44	0.67
SSIM	-0.45	0.23	-0.46	-0.28	0.14	0.65	-0.04	-0.07	-0.26	-0.65	-0.44	1.00	-0.33
SD	0.56	-0.27	0.55	0.55	0.09	-0.59	0.53	0.62	-0.12	0.59	0.67	-0.33	1.00

注:加粗字体表示 Kendall 值大于阈值的指标。

表4 灰度可见光图像与红外图像融合评价指标之间的 Kendall 相关系数

Table 4 Kendall correlation coefficient between gray visible image and infrared image fusion assessment metrics

	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
AG	1.00	-0.48	0.99	0.63	0.22	-0.60	0.47	0.54	-0.24	0.57	0.91	-0.72	0.55
CE	-0.48	1.00	-0.47	-0.45	-0.46	0.19	-0.59	-0.48	0.38	-0.16	-0.41	0.24	-0.45
EI	0.99	-0.47	1.00	0.64	0.23	-0.61	0.46	0.55	-0.25	0.58	0.89	-0.71	0.56
EN	0.63	-0.45	0.64	1.00	0.48	-0.42	0.53	0.72	-0.40	0.39	0.62	-0.45	0.75
MI	0.22	-0.46	0.23	0.48	1.00	-0.07	0.52	0.58	-0.35	0.06	0.19	-0.06	0.51
PSNR/dB	-0.60	0.19	-0.61	-0.42	-0.07	1.00	-0.16	-0.26	0.03	-0.97	-0.63	0.86	-0.38
$Q^{AB/F}$	0.47	-0.59	0.46	0.53	0.52	-0.16	1.00	0.79	-0.47	0.17	0.38	-0.25	0.51
Q_{CB}	0.54	-0.48	0.55	0.72	0.58	-0.26	0.79	1.00	-0.52	0.25	0.44	-0.34	0.72
Q_{CV}	-0.24	0.38	-0.25	-0.40	-0.35	0.03	-0.47	-0.52	1.00	-0.02	-0.27	0.02	-0.57
RMSE	0.57	-0.16	0.58	0.39	0.06	-0.97	0.17	0.25	-0.02	1.00	0.62	-0.83	0.37
SF	0.91	-0.41	0.89	0.62	0.19	-0.63	0.38	0.44	-0.27	0.62	1.00	-0.73	0.58
SSIM	-0.72	0.24	-0.71	-0.45	-0.06	0.86	-0.25	-0.34	0.02	-0.83	-0.73	1.00	-0.39
SD	0.55	-0.45	0.56	0.75	0.51	-0.38	0.51	0.72	-0.57	0.37	0.58	-0.39	1.00

注:加粗字体表示 Kendall 值大于阈值的指标。

灰度可见光(Gray 组)与红外图像融合指标聚类分组结果如图2所示。单通道情况下,不同指标之间表现出较多通道更强的相关性,其中SD与EN、 Q_{CB} 表现出一定的相关性;SSIM利用均值和协方差从亮度、对比度、结构等方面度量与源图像间的

相关性,一定程度上与反映亮度空间变化的AG、SF和EI正相关,与PSNR、RSME负相关。

2.2 一致性分析

由于红外和可见光融合缺少公开的主观评价结果,无法进行主客观一致性分析。本文将客观评价

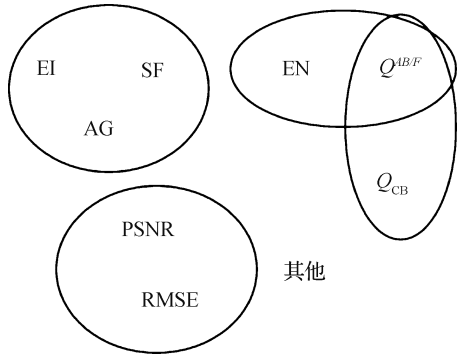


图1 指标聚类分组图(RGB组)

Fig. 1 Schematic diagram of metrics group(RGB)

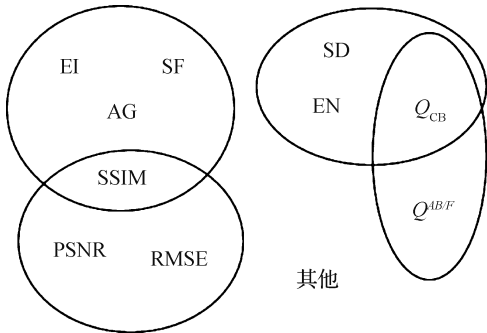


图2 指标聚类分组图(灰度组)

Fig. 2 Schematic diagram of metrics group(gray)

指标视为分类器(Liu等,2012),平等对待所有指标,并利用排序式的Borda投票法(Emerson,2013),综合大多数分类器(客观评价指标)的选择得到算法的排名。表5是利用Borda计数排序法分别统计RGB组和Gray组的总体评价结果,记为RGB排序/Gray排序。表中各列为表1和表2中特定指标在不同算法间的排序,1代表该指标下相对最优的算法,排名最高可获得20积分,其余依次递减,综合所有指标排序得到的积分可得最后一列的算法Borda rank排序结果。若存在积分/排序相同情况,以所排序列号的均值统一排序,如表中CNN和HMSD_GF的RGB综合排序相同,因此以均值3.5作为最终的综合排序。

所有评价指标会对融合方法产生一个总体评价。某指标与总体评价相关性越高,说明该指标与总体评价的一致性越高。为了体现单一指标与Borda排序总体评价的一致性,计算算法各指标排序与Borda排序的Kendall相关系数,结果如表6所示。RGB组和Gray组一致性较高的4个指标均为Q_{CB}、Q^{AB/F}、EN和SD。

表5 Borda计数排序总体评价结果(RGB排序/Gray排序)

Table 5 The overall evaluation result of Borda count sorting(RGB rank/Gray rank)

算法	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD	Borda rank
ADF	16/14	17/12	16/14	14/15	13/16	4/4	13/12	15/14	13/14	4/4	16/16	11/5	16/15	17/14
CBF	2/3	5/1	2/3	2/9	16/8	16/16	9/9	12/9	19/19	16/16	4/3	20/18	11/10	9/8
CNN	10/9	4/6	10/9	5/6	4/5	14/11	2/4	3/3	4/1	14/11	8/8	13/11	1/2	3.5/3
DLF	18/18	14/15	18/18	18/18	9/13	1/1	17/17	17/17	9/17	1/1	19/18	1/2	19/18	14.5/17
FPDE	15/15	13/13	15/15	15/16	14/14	5/3	15/13	16/15	12/15	5/3	17/17	14/4	17/17	18/15
GFCE	3/2	20/20	3/2	10/1	19/12	19/20	14/15	10/8	15/16	19/20	3/2	18/20	8/7	14.5/12
GFF	14/10	8/9	14/10	11/7	5/4	10/10	6/1	6/2	16/7	10/12	13/10	7/10	9/9	8/7
GTF	17/16	6/17	17/17	17/20	6/19	12/15	16/16	20/20	20/20	12/15	15/14	15/16	15/20	20/20
HMSD_GF	6/5	9/3	6/5	7/8	8/6	11/14	5/6	7/6	5/2	11/14	5/4	8/14	4/1	3.5/4
Hybrid_MSD	7/7	10/7	7/7	6/4	7/3	9/8	4/5	2/5	3/3	9/8	6/7	5/9	6/3	1.5/1
IFEVIP	13/17	7/18	13/16	13/14	2/10	18/18	10/19	14/16	1/10	18/18	14/15	10/13	10/12	12/18
LatLRR	1/1	19/19	1/1	16/10	20/20	20/19	18/14	13/12	8/11	20/19	1/1	19/19	2/8	13/13
MGFF	8/12	12/10	8/12	8/12	17/18	8/6	8/10	9/11	10/9	8/6	10/13	6/7	12/13	7/11
MST_SR	9/8	2/5	9/8	4/3	3/2	13/12	1/3	1/1	2/5	13/10	9/9	12/12	3/4	1.5/2
MSVD	20/20	15/16	20/20	20/19	11/15	3/5	20/20	19/19	11/18	3/5	18/19	4/3	20/19	19/19
NSCT_SR	4/4	1/2	4/4	1/5	1/1	17/17	3/2	5/4	18/12	17/17	7/6	17/17	7/6	5/5
ResNet	19/19	11/14	19/19	18/17	12/11	2/2	19/18	18/18	6/13	2/2	20/20	2/1	18/16	16/16
RP_SR	5/6	3/4	5/6	3/2	15/7	15/13	12/7	4/7	17/8	15/13	2/5	16/15	5/5	6/6
TIF	12/11	16/11	11/11	9/13	18/17	7/7	7/8	8/10	7/4	7/7	12/11	9/8	14/14	10/10
VSMWLS	11/13	18/8	12/13	12/11	10/9	6/9	11/11	11/13	14/6	6/9	11/12	3/6	13/11	11/9

表 6 单一指标排序与 Borda 计数排序的相关系数

Table 6 Correlation coefficient between single metric sorting and Borda count sorting

组别	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
RGB 组	0.41	0.37	0.42	0.67	0.23	-0.21	0.72	0.81	0.23	-0.21	0.42	-0.01	0.58
Gray 组	0.48	0.60	0.49	0.71	0.57	-0.21	0.74	0.80	0.57	-0.18	0.43	-0.22	0.75

注:加粗字体为一致性排名前 4 的指标。

2.3 离散程度分析

除了分析指标的相关性和一致性外,还需要考察指标受融合质量影响的波动情况。图 3 为 21 组图像对数据中 10 种不同算法的 PSNR 和 $Q^{AB/F}$ 指标的量化折线图(Zhang 等,2020)。可以看出,不同指标的取值范围差异较大,表现为表 1 和表 2 中的指标取值范围,即图中折线的纵坐标(指标值)范围有较大差异,折线图上方数值为对应算法的指标均值;

其次,随着图像内容(横坐标为不同图像对)的变化,指标波动较大,表现为不同指标的折线图均有明显的峰值和谷值;再次,图像内容变化时,不同算法的指标数值具有大致相同的变化规律,表现为不同折线有近似的峰谷位置以及变化趋势,因此可以用多个图像对的指标均值近似表征指标的变化。而当图像内容相同时,不同算法产生的指标数值差异可以反映不同算法融合质量的相对差异。

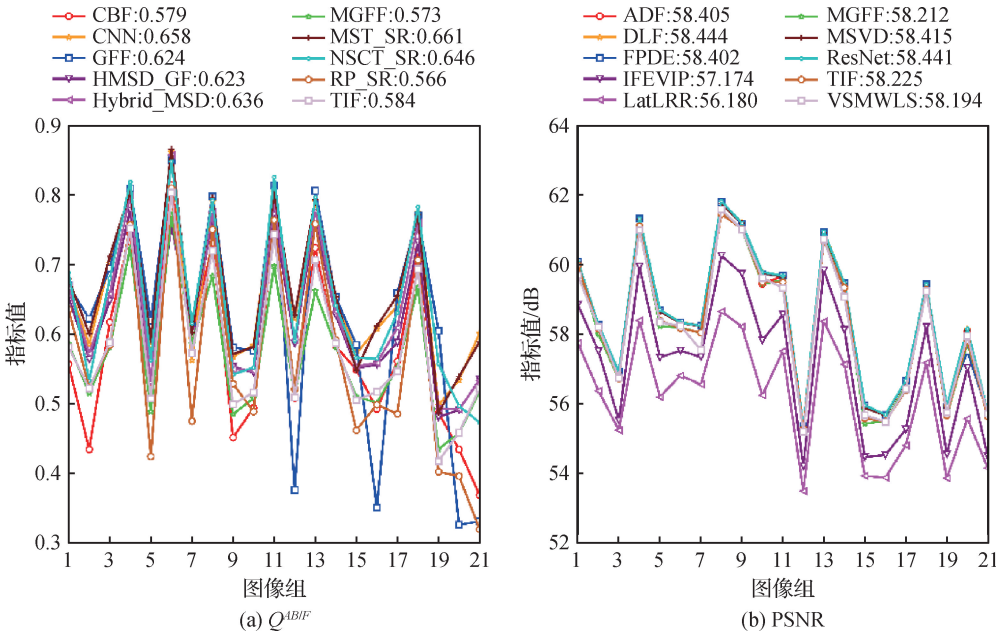


图 3 10 种算法在 21 对图像上指标的折线图

Fig. 3 Line chart of metric of 10 algorithms on 21 pairs of images((a) $Q^{AB/F}$;(b) PSNR)

不同于 Liu 等人(2012)通过计算不同输入图像对的指标方差来度量指标随图像内容变化的规律,本文度量均值随不同算法的波动情况。由于不同客观指标的定义不同,值域范围有较大差异,不适合用方差进行分析,所以利用离散系数分析指标随算法的波动情况。其中,离散系数定义为标准差与均值之比。表 7 为表 1 和表 2 中不同指标的均值、标准差以及离散系数。通过对比可以看出,离散系数值越大,说明指标在不同算法间分布差异越大,越能体现不同算法间的性能差异。

相关性分析的目的是得到聚类分组,一致性分析的目的是得到指标评估的有效性,离散程度分析则可以通过评估指标的分布情况,选择能够充分体现不同算法间差异的指标。在实验设定下,一致性分析得到的候选指标集为 $\{Q^{AB/F}, EN, SD, Q_{CB}\}$,对应的离散系数在表 7 中以粗体显示。RGB 分组实验中 EN 和 $Q^{AB/F}$, $Q^{AB/F}$ 和 Q_{CB} 属于同组指标,其中 EN 的离散系数明显小于其他 3 项指标, $Q^{AB/F}$ 的离散系数高于 Q_{CB} 。Gray 分组实验中, $Q^{AB/F}$ 和 Q_{CB} , Q_{CB} 、SD 和 EN 为同组指标,且 EN 的离散系

表 7 VIFB 数据集下评价指标的均值、标准差以及离散系数

Table 7 The mean, standard deviation and coefficient of variation of the objective metrics on the VIFB dataset

组别	对比内容	AG	CE	EI	EN	MI	PSNR/dB	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
彩色	均值	6.314	1.261	65.643	7.132	2.189	56.621	0.500	0.510	1 074.749	0.150	19.210	1.293	50.281
	标准差	1.451	0.236	15.525	0.216	0.223	0.558	0.077	0.046	470.781	0.020	4.289	0.086	8.491
	离散系数	0.230	0.187	0.237	0.030	0.102	0.010	0.154	0.090	0.438	0.136	0.223	0.066	0.169
灰度	均值	4.478	1.347	45.384	6.927	2.192	59.820	0.593	0.563	494.834	0.071	15.463	1.468	41.204
	标准差	1.173	0.352	12.143	0.384	0.715	0.932	0.125	0.122	374.853	0.020	3.801	0.113	10.527
	离散系数	0.262	0.262	0.268	0.055	0.326	0.016	0.211	0.217	0.758	0.275	0.246	0.077	0.255

注:加粗字体为一致性排名前 4 的指标对应的离散系数。

数远小于其他 3 项指标, Q_{CB} 的离散系数略高于 $Q^{AB/F}$ 。综合上述实验分析得到, RGB 组的客观评价建议指标集为 $\{SD, Q^{AB/F}\}$ 或 $\{SD, Q_{CB}\}$; Gray 组的客观评价建议指标集为 $\{SD, Q^{AB/F}\}$ 或 $\{Q_{CB}\}$ 。分组建议指标集合均包含 SD 、 Q_{CB} 和 $Q^{AB/F}$ 3 项指标, 其中不同分组结果中 SD 指标与其他两项指标的组合方式不同。原因是 Gray 组和 RGB 组分别基于相关性得到的指标聚类分组结果有差异, 可参见图 1 和图 2。依据 Zhang 等人(2020)中的计算方式, RGB 组评价指标是 RGB 三个通道分别与红外图像融合客观评价结果的平均值, 相较于单一的灰度通道, SD 指标与其他指标的相关性更低。

综合两组的結果, 取交集为 $\{SD, Q^{AB/F}\}$, 作为可见光与红外图像融合评价的建议指标集。其中, 标准差 SD 属于基于融合图像质量的评价指标, 通过与均值的差分反映融合结果中的对比度信息; 边缘保持度 $Q^{AB/F}$ 属于基于源图像和融合图像的评价指标, 通过度量可见光和红外源图像传递到融合图像的边缘保留程度, 反映融合结果图像对可见光和红外源图像的融入度。

红外和可见光融合应用中, 红外图像中显著性通常表现为对比度较大的局部区域, 而可见光中通常表现为边缘、纹理等细节信息。建议指标集中的两项指标从属于不同的指标类型, 分别从对比度和边缘细节保留度对融合结果进行评价, 具有一定的互补性。

综合 SD 和 $Q^{AB/F}$ 指标, VIFB 数据集下排名靠前的算法包括 CNN、HMSD_GF、Hybrid_MSD 以及 MST_SR, 在定性分析结果中也表现出较好的融合效果。

需要强调的是, 1) 目前统计的融合评估指标中

没有专门针对色彩失真方面的评价, 与主观感受并不绝对一致, 例如 MST_SR 融合彩色图像出现的伪影现象。研究更多主客观一致的评价指标仍是图像融合领域需要持续关注的重点。2) 针对有限样本在实验设定下得到的建议指标集合是非排他性的指标建议, 即选择多个指标从不同角度综合评价融合结果时, 建议选择而非只选择的指标集合。基于统计的指标分析方法可以推广至更多的融合数据源、融合算法以及评估指标, 融合建议指标集合也会有新的发展。

3 结 论

本文在讨论图像融合客观评价指标分类基础上, 结合 VIFB 数据集和代表性图像融合算法, 提出一种通用的客观评价指标分析方法, 将 VIFB 数据分为灰度/彩色可见光与红外图像融合两组, 分别进行相关性、一致性以及离散系数统计分析, 基于客观统计结果, 得到适用于可见光与红外图像融合质量评价的建议指标集合 $\{SD, Q^{AB/F}\}$ 。其中, SD 属于基于融合图像质量的评价指标, $Q^{AB/F}$ 属于基于源图像和融合图像的评价指标, 分别侧重从对比度和细节信息传递融入度评价融合结果, 具有一定的互补性。相较于依赖主观评价或聚焦特定类型融合算法的分析方法, 本文方法基于客观统计数据且涵盖不同类型的融合算法, 具有更好的适用性和可推广性。但是, 目前的客观评价仅限于 VIFB 中涉及的 13 种指标, 且仅讨论针对红外和可见光图像融合下的应用。下一步工作将研究覆盖更多融合源、融合算法以及客观评价指标, 并将方法推广到多聚焦图像、医

学图像以及遥感图像融合,得到适用于不同应用场景的图像融合客观评价指标建议。

参考文献 (References)

- Chen H and Varshney P K. 2007. A human perception inspired quality metric for image fusion based on regional information. *Information Fusion*, 8(2): 193-207 [DOI: 10.1016/j.inffus.2005.10.001]
- Chen M S. 2016. Image fusion of visual and infrared image based on NSCT and compressed sensing. *Journal of Image and Graphics*, 21(1): 39-44 (陈木生. 2016. 结合 NSCT 和压缩感知的红外与可见光图像融合. *中国图象图形学报*, 21(1): 39-44) [DOI: 10.11834/jig.20160105]
- Chen Y and Blum R S. 2009. A new automated quality assessment algorithm for image fusion. *Image and Vision Computing*, 27(10): 1421-1432 [DOI: 10.1016/j.imavis.2007.12.002]
- Emerson P. 2013. The original Borda count and partial voting. *Social Choice and Welfare*, 40(2): 353-358 [DOI: 10.1007/s00355-011-0603-9]
- Gong R and Wang X C. 2019. Infrared and visible image fusion based on BEMD and W-transform. *Journal of Image and Graphics*, 24(6): 987-999 (宫睿, 王小春. 2019. BEMD 分解和 W 变换相结合的红外与可见光图像融合. *中国图象图形学报*, 24(6): 987-999) [DOI: 10.11834/jig.180530]
- Li H and Wu X J. 2019. DenseFuse: a fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5): 2614-2623 [DOI: 10.1109/TIP.2018.2887342]
- Li S T, Yang B and Hu J W. 2011. Performance comparison of different multi-resolution transforms for image fusion. *Information Fusion*, 12(2): 74-84 [DOI: 10.1016/j.inffus.2010.03.002]
- Liu Z, Blasch E, Bhatnagar G, John V, Wu W and Blum R S. 2018. Fusing synergistic information from multi-sensor images: an overview from implementation to performance assessment. *Information Fusion*, 42: 127-145 [DOI: 10.1016/j.inffus.2017.10.010]
- Liu Z, Blasch E, Xue Z Y, Zhao J Y, Laganieri R and Wu W. 2012. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1): 94-109 [DOI: 10.1109/TPAMI.2011.109]
- Liu Z W, Luo X Q and Zhang Z C. 2020. Multi-focus image fusion with a self-learning fusion rule. *Journal of Image and Graphics*, 25(8): 1637-1648 (刘子闻, 罗晓清, 张战成. 2020. 自学习规则下的多聚焦图像融合. *中国图象图形学报*, 25(8): 1637-1648) [DOI: 10.11834/jig.190614]
- Ma J Y, Liang P W, Yu W, Chen C, Guo X J, Wu J and Jiang J J. 2020. Infrared and visible image fusion via detail preserving adversarial learning. *Information Fusion*, 54: 85-98 [DOI: 10.1016/j.inffus.2019.07.005]
- Ma J Y, Ma Y and Li C. 2019. Infrared and visible image fusion methods and applications: a survey. *Information Fusion*, 45: 153-178 [DOI: 10.1016/j.inffus.2018.02.004]
- Martinez J, Pistonesi S, Maciel M C and Flesia A G. 2019. Multi-scale fidelity measure for image fusion quality assessment. *Information Fusion*, 50: 197-211 [DOI: 10.1016/j.inffus.2019.01.003]
- Qu G H, Zhang D L and Yan P F. 2002. Information measure for performance of image fusion. *Electronics Letters*, 38(7): 313-315 [DOI: 10.1049/el:20020212]
- Xydeas C S and Petrović V. 2000. Objective image fusion performance measure. *Electronics Letters*, 36(4): 308-309 [DOI: 10.1049/el:20000267]
- Yang M H, Cao Y D, Tan L, Zhang C Y and Yu J. 2007. A new multi-quality image fusion method in visual sensor network//*Proceedings of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. Kaohsiung, China: IEEE: 667-670 [DOI: 10.1109/IIH-MSP.2007.42]
- Yang Y C, Li J and Wang Y P. 2018. Review of image fusion quality evaluation methods. *Journal of Frontiers of Computer Science and Technology*, 12(7): 1021-1035 (杨艳春, 李娇, 王阳萍. 2018. 图像融合质量评价方法研究综述. *计算机科学与探索*, 12(7): 1021-1035) [DOI: 10.3778/j.issn.1673-9418.1710001]
- Zhang X C, Ye P and Xiao G. 2020. VIFB: a visible and infrared image fusion benchmark//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle, USA: IEEE: 468-478 [DOI: 10.1109/CVPRW50498.2020.00060]
- Zhang X L, Li X F and Li J. 2014. Validation and correlation analysis of metrics for evaluating performance of image fusion. *Acta Automatica Sinica*, 40(2): 306-315 (张小利, 李雄飞, 李军. 2014. 融合图像质量评价指标的相关性分析及性能评估. *自动化学报*, 40(2): 306-315) [DOI: 10.3724/SP.J.1004.2014.00306]
- Zheng Y F, Essock E A, Hansen B C and Haun A M. 2007. A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. *Information Fusion*, 8(2): 177-192 [DOI: 10.1016/j.inffus.2005.04.003]

作者简介

孙彬,通信作者,女,副教授,主要研究方向为信息融合、图像处理 and 导航定位。E-mail: sunbinhust@uestc.edu.cn

高云翔,男,硕士研究生,主要研究方向为图像处理。

E-mail: 202022100530@std.uestc.edu.cn

诸葛吴为,男,硕士研究生,主要研究方向为图像处理。

E-mail: 202022100531@std.uestc.edu.cn

王梓萱,女,硕士研究生,主要研究方向为图像处理。

E-mail: zixuan_wang0716@163.com