

引用格式: Zhou Yi, Ma Jiayi, Huang Jun. Infrared and Visible Image Fusion based on Mutually Guided Filtering and Saliency Map[J]. Remote Sensing Technology and Application, 2020, 35(6): 1404-1413. [周怡, 马佳义, 黄珺. 基于互导滤波和显著性映射的红外可见光图像融合[J]. 遥感技术与应用, 2020, 35(6): 1404-1413.]
doi: 10.11873/j.issn.1004-0323.2020.6.1404

基于互导滤波和显著性映射的红外可见光图像融合

周 怡, 马佳义, 黄 珺

(武汉大学 电子信息学院, 湖北 武汉 430072)

摘要:为了解决传统多尺度红外可见光融合图像边缘模糊、对比度低和目标不显著的问题,提出一种基于互导滤波和显著性映射的红外可见光图像融合算法。由于互导滤波器能将图像一致结构和不一致结构分离并且具有尺度和保边意识,因此首先利用互导滤波器将原图像分解为具有冗余信息的结构层和不同尺度上具有互补信息的纹理层;其次根据过明或过暗区域更容易引起注意的视觉特点构造图像显著性映射函数对结构层和不同尺度的纹理层进行显著性映射;最后根据不同尺度的结构和纹理特性对图像进行融合重构。在两个数据集上的实验结果表明与传统的多尺度融合方法相比提出的方法在保持图像边缘、增强图像对比度、突出目标方面具有较好的主客观评价效果。

关键词:互导滤波器;显著性映射;多尺度融合;红外

中图分类号: TP75 **文献标志码:** A **文章编号:** 1004-0323(2020)06-1404-10

1 引 言

图像融合是将同一场景下不同传感器或同一传感器在不同条件下获得的多幅图像合成一幅复合图像的技术。其意义在于将输入图像的冗余信息和互补信息整合到最终的融合图像中去,使得融合图像比单张输入图像具有更丰富的信息便于后续图像处理任务,例如数字摄影、遥感、农业、医学和生物识别等领域^[1-2]。其中红外可见光融合是图像融合中一个重要的话题。红外图像捕获物体的热辐射特性,受环境影响(恶劣天气、低照度环境)较小且图像目标显著,但是与视觉特性不一致。而可见光图像恰恰与之相反,形成了互补特性。红外可见光融合由于其天然的互补特性可以得到信息丰富、鲁棒性强的融合图像,因此被广泛应用于隐蔽武器探测、目标检测和跟踪、视频监控等领域^[3-4]。

目前已经提出了许多有效的融合算法,根据融合过程中抽象层次的不同,图像融合算法可以分为

3类:像素级、特征级和决策级^[5]。像素级融合利用单个像素形成融合策略从而对图像进行组合;特征级融合根据特定需求和图像特征将图像划分为不同区域,并对感兴趣的区域进行特征分析从而形成融合策略实现特征级融合;决策级融合输入的不是原图像而是经过分析、识别后提取的决策结果,并对其进行统计推理和逻辑分析从而实现图像融合任务。在这3类融合方法中,像素级融合的抽象层次最低,得到的融合图像最丰富,因此被众多学者广泛研究,而包含在像素级融合中的多尺度融合方法由于算法设计灵活,融合效果与视觉特性一致等优点受到了重点关注。多尺度方法一般包括两个关键步骤:子图像分解和融合重构。子图像分解可以将原图像信息从不同角度出发进行表征,从而为融合原图像不同信息提供了可能性。融合重构是根据分解子图像表征的信息特点进行重新整合以实现最终的图像融合。因此除了原图像质量外这

收稿日期:2019-11-28;修订日期:2020-04-20

基金项目:国家自然科学基金项目(61605146、61773295),湖北省自然科学基金项目(2019CFA037)。

作者简介:周 怡(1994—),女,福建南平人,硕士研究生,主要从事红外可见光图像融合技术研究。E-mail: 2017202120122@whu.edu.cn

通讯作者:黄 珺(1985—),男,湖南祁阳人,副教授,主要从事红外图像处理方面的研究。E-mail: junhwong@whu.edu.cn

两个步骤是决定融合图像质量的关键所在。

Sharma等^[6]对各种变换分解方法进行了探讨和比较。根据分解方法的不同多尺度方法可以分为基于金字塔的方法、基于小波的方法、基于轮廓波的方法、基于保边滤波器的方法和一些其他方法。基于金字塔的方法通过滤波和下采样的不断迭代获得具有金字塔形分辨率的子图像,并对这些子图像进行融合和重构。如Toet^[7]提出了一种基于多分辨率对比度分解的分层图像融合方案RP(Ratio of low-pass Pyramid),有效保留了与视觉感知最相关的图像细节。Bulanon等^[8]提出了拉普拉斯金字塔变换LP(Laplacian Pyramid)和模糊逻辑相结合的图像融合方法,提高了水果检测效果。Chen等^[9]基于LP的分解特点提出了一种用于红外和可见光图像融合的目标增强多尺度变换分解模型,以同时增强红外图像中的目标并保留可见光图像中的纹理细节。基于小波分解的方法从频率角度出发将原图像分解为高低频子图像再进行后续的融合处理。如John等^[10]用双树复小波变换DTCWT(Dual-Tree Complex Wavelet Transform)分割输入图像特征,以此产生区域映射用于区域级融合,提高了每个区域的特征。基于轮廓波的方法结合拉普拉斯金字塔和方向滤波器组,能够很好地捕捉图像边缘的几何形状,如Zhang等^[11]提出了一种基于非下采样轮廓波变换NSCT(Nonsubsampled Contourlet Transform)的图像融合算法从原图像中提取更重要的视觉信息,该方法在未精确配准情况下的融合效果要高于离散小波变换图像融合方法。由于保边滤波器的兴起,近年来还提出了一些基于保边滤波器的方法,如Bavirisett等^[12]提出了一种基于四阶偏微分方程的融合算法FPDES(Fourth order Partial Differential Equations),有效地将有用的信息融合到图像中,具有较小的融合伪影,同时该方法计算时间较少适合于实时处理。

常用的融合规则有最大值法,直接平均法,加权平均法和基于视觉显著性的方法等^[1]。其中视觉显著性机制在人类视觉系统中起着至关重要的作用,对目标检测和识别具有重要意义,被广泛研究。如Yan等^[13]基于低秩稀疏矩阵分解算法提出了目标检测的显著性映射估计模型,该模型在视觉显著性检测任务上具有竞争力。Bavirisetti等^[14]基于中值滤波器和均值滤波器提出了一种显著性映射方法并将其用于两尺度图像分解的图像融合方法TS

(Two-scale image fusion)使得融合图像具有良好的视觉显著性效果。此外还存在很多其他类型的多尺度图像融合方法,如Naidu^[15]提出了一种基于多分辨率奇异值分解MSVD(Multi-resolution Singular Value Decomposition)的图像融合算法,该算法的融合效果和小波相当。Nencini等^[16]将曲波变换CVT(Curvelet transform)用于遥感图像融合。

除了多尺度的方法之外还有许多其他类型的方法,例如Liu等^[17]提出了一种自适应稀疏表示ASR(Adaptive Sparse Representation)模型用于图像融合和去噪。Ma等^[18]考虑了可见光和红外的不同特性提出了梯度转移与总梯度最小化的融合方法GTF(Gradient Transfer Fusion)用于红外可见光融合,有效提高了融合图像目标显著性。Liu等^[19]提出了一种基于卷积神经网络的红外与可见光图像融合方法CNN(Convolutional Neural Networks)能够对场景进行更加全面的描述。Ma等^[20]提出了一种生成对抗网络FusionGAN用于红外可见光融合,该网络是端到端的模型,避免了传统融合策略中人工设计复杂的有效性测量和融合规则,同时将其推广到不同分辨率的红外可见光融合,拓宽了算法是适用范围。在FusionGAN的基础上Ma等^[21]为了进一步增强融合图像细节,提出了基于细节保持对抗学习的红外与可见光图像融合方法DPALF(Detail Preserving Adversarial Learning Fusion),该方法设计了两个损失函数,即细节损失和目标边缘增强损失,有效提高了融合图像细节信息。Xu等^[22]提出LBP-BEGAN,通过局部二值模式定义损失函数使得融合图像具有丰富的边界信息,同时由于采用分布的鉴别器对融合后的图像与原始的红外和可见光图像进行鉴别保证了融合结果的质量。这些方法都为红外可见光融合开拓了新的思路。

以上算法虽然都完成了基本的图像融合任务但是还存在一些问题,如传统的多尺度融合方法都存在边缘失真问题,这会影响图像的边缘检测和图像分割。大多数方法没有考虑目标的显著性,从而导致图像对比度下降,目标不够突出,影响整体的视觉效果和目标检测。此外几乎所有基于多尺度的方法在图像分解时都没有充分考虑两幅图像结构之间的冗余互补问题,因此可能会导致在融合时一些互补细节信息不能在融合图像中得到较好的表现。而基于学习的方法虽然在大多数情况下有较好的表现,但是也存在一定的缺陷。如CNN的

融合结果在某些情况下会出现类似过曝光的情况,导致高亮处的细节损失。而 FusionGAN 和 DPALF 以及 LBP-BEGAN 倾向于保留红外图像亮度和可见光图像梯度,并且后二者还特意突出了边界信息使得融合结果看起来像是锐化的红外图像这和视觉特性不一致。

为了解决以上问题,本文基于互导滤波器和图像显著性提出了一种简单有效的融合算法。实验结果表明,该方法是可行的并且与传统的多尺度融合方法相比,该图像融合算法有以下几个突出的优点:

(1) 从图像之间信息冗余互补的角度出发利用互导滤波器进行图像分解,将原图像分解为具有冗余信息的结构层和具有互补信息的多尺度纹理层,以便重构时能够充分融合图像互补信息。

(2) 根据视觉特性构建显著性映射,并结合结构纹理特性对不同尺度图像进行融合和重构,有效提升融合图像对比度,突出目标。

(3) 与传统的多尺度方法相比本文提出的融合算法从主客观上具有较好的融合效果。

2 互导滤波器

由于传统高斯滤波器用于多尺度融合存在边缘模糊问题,因此一类基于保边滤波器的多尺度方法被提出,较好地改善了边缘模糊问题。然而大多数的保边滤波分解没有充分考虑两幅图像之间结构的差异,只根据单幅图像进行分解。因此,这样的滤波器不能很好地将两幅图像之间的冗余和互补信息进行分离。考虑到同一场景下不同模式的图像存在相同和不同的信息 Guo 等^[23]提出了一种互导滤波器,该滤波器不仅有保边意识、尺度意识、还能将两幅图像之间不一致的纹理进行移除,较好地保留图像的冗余结构。

互导滤波器是通过设计相关结构来实现的。首先定义了 3 种结构:互结构、不一致结构和平坦区域。互结构:两幅图像在同一坐标位置的像素梯度都足够大;不一致结构:两幅图像在同一坐标位置的像素梯度一大一小;平坦区域:两幅图像在同一坐标位置的像素梯度都很小。互导滤波器的目的在于保持图像互结构,对不一致结构和平坦区域进行平滑。若将互结构和平坦区域统称为一致结构,也可以将互导滤波器的滤波目的理解为:保持图像的一致结构,对不一致结构进行移除。为此提出了

相关结构的概念用于制定滤波规则,相关结构的定义如下:

$$S(\mathbf{O}, \mathbf{R}, \varepsilon_r) = \sum_i \sum_{d \in \{h, v\}} \frac{|\nabla_d \mathbf{O}_i|}{\max(|\nabla_d \mathbf{R}_i|, \varepsilon_r)} \quad (1)$$

其中: \mathbf{O} 为目标图像, \mathbf{R} 为同分辨率的参考图像, i 表示像素坐标, h 表示行, v 表示列, ∇_d 表示一阶行梯度或列梯度, ε_r 是为了避免分母为 0 而引入一个小的正常数,绝对值 $|\cdot|$ 符是为了消除方向的影响。相关结构 $S(\mathbf{O}, \mathbf{R}, \varepsilon_r)$ 从梯度角度测量了目标图像相对于参考图像的不一致性。若目标图像相对于参考图像在 i 处属于不一致结构,那么对相关结构 $S(\mathbf{O}, \mathbf{R}, \varepsilon_r)$ 求局部像素平均将远离 1,反之若参考图像在 i 处属于一致结构,那么相关结构 $S(\mathbf{O}, \mathbf{R}, \varepsilon_r)$ 求局部像素平均接近 1。基于相关结构的定义,建立互导滤波器的规则如下:

$$\arg \min_{\mathbf{O}, \mathbf{R}} \alpha_o S(\mathbf{O}, \mathbf{R}, \varepsilon_r) + \beta_o \|\mathbf{O} - \mathbf{O}_0\|_2^2 + \alpha_r S(\mathbf{R}, \mathbf{O}, \varepsilon_o) + \beta_r \|\mathbf{R} - \mathbf{R}_0\|_2^2 \quad (2)$$

其中: $S(\mathbf{O}, \mathbf{R}, \varepsilon_r)$ 和 $S(\mathbf{R}, \mathbf{O}, \varepsilon_o)$ 作为平滑项,是保留一致结构同时移除不一致结构的关键。 $\|\mathbf{O} - \mathbf{O}_0\|_2^2$ 和 $\|\mathbf{R} - \mathbf{R}_0\|_2^2$ 作为数据保真项,通过约束 \mathbf{O} 和 \mathbf{R} ,使其不与输入产生较大偏差,从而避免了平凡解。 $\alpha_o, \alpha_r, \beta_o, \beta_r$ 是非负常数,可以对数据保真项和平滑项进行调整。 $\|\cdot\|_2$ 表示 l_2 范数。以 $S(\mathbf{O}, \mathbf{R}, \varepsilon_r)$ 为例,从最优化的角度来看,当目标图像 \mathbf{O} 像素坐标对应参考图像 \mathbf{R} 边缘时由于 $\nabla_d \mathbf{R}_i$ 很大,惩罚系数 $1/\max(|\nabla_d \mathbf{R}_i|, \varepsilon_r)$ 很小,因此目标图像 \mathbf{O} 梯度 $\nabla_d \mathbf{O}_i$ 变化很小,梯度得以保留,反之亦然。在 $S(\mathbf{R}, \mathbf{O}, \varepsilon_o)$ 中也是如此。因此只有当 $\nabla_d \mathbf{R}_i$ 和 $\nabla_d \mathbf{O}_i$ 都很大时该像素坐标的梯度才得以保留,而在双方梯度大小不一致的情况下会对梯度进行移除,从而实现保留图像参考图像和目标图像之间一致结构,移除不一致结构的目的。

互导滤波模型(2)中有 6 个自由参数: $\varepsilon_r, \varepsilon_o, \alpha_o, \alpha_r, \beta_o, \beta_r$ 。从模型中可以看出目标图像和参考图像具有相同的地位,因此通常情况下设置 $\varepsilon_r = \varepsilon_o = \varepsilon$ 、 $\alpha_r = \alpha_o = \alpha$ 、 $\beta_r = \beta_o = \beta$,并且依照经验通常 ε 设置为 0.01, β 设置为 1,而 α 是具有尺度意识的尺度因子,可根据需要进行调节。本文中考虑到参考图像和目标图像地位相当,因此采用 \mathbf{S}_1 和 \mathbf{S}_2 表示输入的两幅图像,同时将互导滤波器表示为: $MuGIF(\mathbf{S}_1, \mathbf{S}_2, \alpha)$ 。

3 图像融合方法

提出的红外与可见光图像融合框架流程如图1所示。首先利用互导滤波器将原图像分解为冗余

结构层和多尺度的互补细节层,然后构建显著映射函数对冗余结构层和互补细节层进行显著性映射,最后,基于得到的分解子图像和显著性映射对图像进行融合和重构得到最终的融合图像。

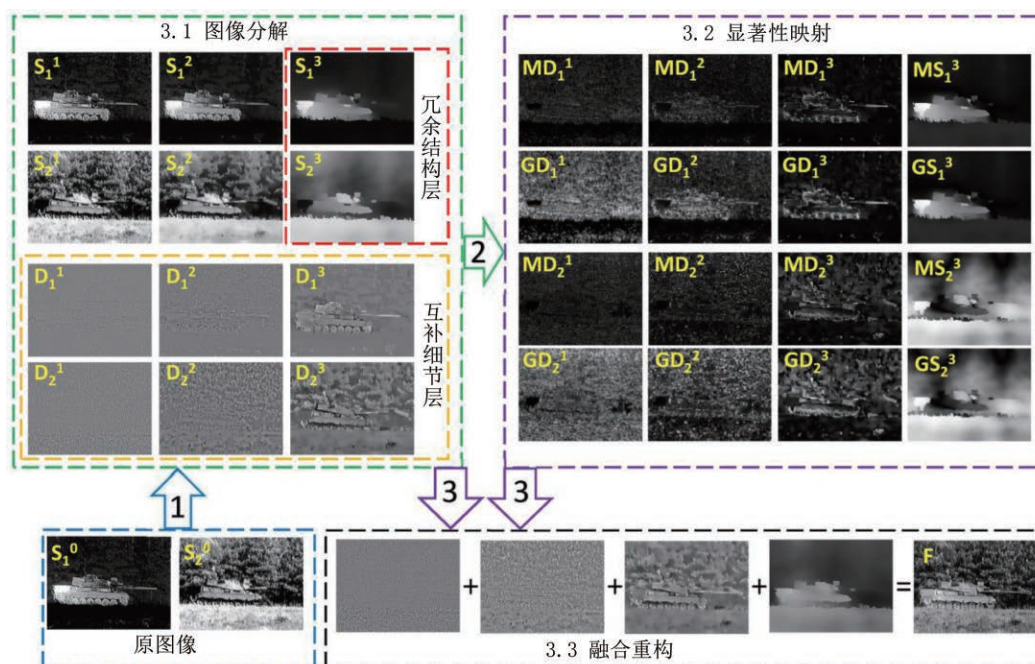


图1 融合算法流程框架图

Fig.1 The frame diagram of the fusion algorithm flow

3.1 图像分解

利用互导滤波器对原图像 S_j^0 进行滤波:

$$[S_1^t, S_2^t] = \text{MuGIF}(S_1^{t-1}, S_2^{t-1}, \alpha^t) \quad t \in \{1, 2, 3\} \quad (3)$$

其中: S_j^t 表示第 j 幅图像第 t 次滤波的结果, 当 $t=0$ 时表示原图像。同时可以通过相减的方法获得不同尺度的残差层:

$$D_j^t = S_j^t - S_j^{t-1} \quad t \in \{1, 2, 3\} \quad (4)$$

其中: D_j^t 表示第 j 幅图像第 t 个尺度的残差层。图1中图像分解模块显示了通过该方法获得的不同尺度子图像。其中前两行是原图滤波的结果, 从左到右滤波尺度因子不断增大, 本文设置尺度参数因子 $\alpha = [0.0005, 0.005, 0.05]$ 。由于最后一次滤波得到的两幅图像从表达的信息量来看是冗余的并且主要反映图像的结构信息, 因此称之为冗余结构层。后两行是原图和滤波结果或滤波结果之间做差得到的残差层, 主要包含不同尺度的细节信息。由于这些细节信息是互补的因此称之为互补细节层。由互导滤波器和做差法分解得到的冗余结构层和互补细节层不仅不存在边缘模糊效应, 还有效将图像的冗余和互补信息在不同的子图像中进行了表

征, 为后续更好进行子图像互补信息融合提供了可能。

3.2 显著性映射

视觉显著性是指图像中某些区域比其他区域更突出从而表现出容易引起关注的图像视觉特性。在现实中发现相较于图像中等灰度区域, 过明或过暗的区域更能引起注意, 基于这一事实本文提出了一种像素级的显著性映射。用像素值偏离均值的距离来度量该像素点的显著性, 像素值偏离均值越远该点的图像显著性越强。基于以上定义对于输入图像 X 可以得到相应的显著性映射 M 如下:

$$M = \text{normal}(|X - \bar{X}|) \quad (5)$$

其中: \bar{X} 表示图像的均值矩阵, 采用绝对值符号 $|\cdot|$ 消除梯度方向的影响, 最后用 $\text{normal}(\cdot)$ 函数拉开各个显著值的距离, 使得显著性映射更加有效。

$$\text{normal}(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

通过对冗余结构层和各个尺度互补细节层进行显著性映射可以得到对应的显著图: $\{MS_1^3, MD_1^3, MD_1^2, MD_1^1; MS_2^3, MD_2^3, MD_2^2, MD_2^1\}$ 。为

了使得融合图像边缘过渡自然,但是不至于产生晕影,本文用一个窗口较小的高斯滤波对显著图进行了处理即:

$$\mathbf{G} = \text{Gaussian}(\mathbf{Y}, \delta, r) \quad (7)$$

其中: $\text{Gaussian}(\cdot)$ 表示高斯滤波, \mathbf{G} 表示滤波之后显著图, \mathbf{Y} 表示滤波之前各尺度显著图, δ 为高斯滤波器的标准差, γ 为高斯滤波器的直径, δ 和 γ 都设置为一个较小值: 5。经过高斯滤波可以得到: $\{\mathbf{GS}_1^3, \mathbf{GD}_1^3, \mathbf{GD}_1^1, \mathbf{GD}_1^2; \mathbf{GS}_2^3, \mathbf{GD}_2^3, \mathbf{GD}_2^2, \mathbf{GD}_2^1\}$ 。图 1 中显著性映射模块显示了高斯滤波前后的显著性映射图,从中可以看出它们较好地突出了各尺度子图像中容易引起视觉关注的图像信息。

3.3 融合重构

由于冗余结构层所包含的结构信息是冗余的,因此我们可以选择冗余层中的任一子图像作为最终的结构层,而最常见的结构层融合方法是直接平均法。本文考虑到冗余结构层较为平滑且子图像之间存在亮度差异,采用显著性映射对结构层进行融合可以提升整体的视觉效果,因此在直接平均法的基础上加上了显著性映射的影响。考虑到冗余结构层较为平滑而互补的细节层较为尖锐。因此对二者采用不同的权重矩阵构造方式,冗余结构层权重矩阵 \mathbf{WS}_j^3 :

$$\mathbf{WS}_1^3 = 0.5 + 0.5(\mathbf{GS}_1^3 - \mathbf{GS}_2^3) \quad (8)$$

$$\mathbf{WS}_2^3 = 1 - \mathbf{WS}_1^3 \quad (9)$$

互补细节层权重矩阵 \mathbf{WD}_j^t :

$$\mathbf{WD}_1^t = \frac{\mathbf{GD}_1^t}{\mathbf{GD}_1^t + \mathbf{GD}_2^t} \quad t \in \{1, 2, 3\} \quad (10)$$

$$\mathbf{WD}_2^t = 1 - \mathbf{WD}_1^t \quad t \in \{1, 2, 3\} \quad (11)$$

通过各尺度的互补细节和冗余结构融合重构得到最终融合图像 \mathbf{F} :

$$\mathbf{F} = \sum_{j \in \{1, 2\}} \mathbf{WS}_j^3 \cdot \mathbf{S}_j + \sum_{t \in \{1, 2, 3\}} \sum_{j \in \{1, 2\}} \mathbf{WD}_j^t \cdot \mathbf{D}_j^t \quad (12)$$

至此融合任务结束。

4 实验分析与讨论

4.1 实验设置

为了验证融合方法的有效性,选取了 8 种算法进行了定性分析和定量比较。对比方法包括 7 种多尺度的方法和一种其他类型的方法。其中包括两种基于金字塔的方法: $\text{RP}^{[7]}$ 、 $\text{LP}^{[8]}$; 一种基于小波的方法: $\text{DTCWT}^{[10]}$; 一种基于轮廓波的方法: $\text{NSCT}^{[11]}$; 一种基于保边滤波器的方法: $\text{FPDES}^{[12]}$; 一种基于显著性的方法 $\text{TS}^{[14]}$; 一种其他类型的多

尺度方法: $\text{MSVD}^{[15]}$ 。而非多尺度的融合算法为: $\text{ASR}^{[17]}$ 。这些算法都是基于公开代码实现的,其中前 4 种方法采用的是 *Liu* 等的工具包^[24],分解层数设置为 4 层,其余采用的是原文作者的代码,并将相应的参数设置为默认值。而本文方法采用上一节提到的固定参数。实验均在 2.5 GHz Intel Core CPU and 8 GB memory 的笔记本电脑上进行,所有算法都在 Matlab2016a 上运行。

实验在两个数据集上进行,即 TNO^[25]数据集和 RoadScene^[26]数据集。其中 TNO 数据集是红外可见光融合常用的数据集,该数据集主要包含了军事等各种场景下已配准的红外可见光图像对,被广泛应用于红外可见光融合算法的对比实验中。RoadScene 数据集是新公布的数据集包含了不同光照条件下配准的各种街道场景的红外可见光图像对。为了验证算法对不同场景的适用性,本文在两个数据集上分别挑选了 3 幅典型场景用于视觉上的定性分析,同时还分别在每个数据集上随机挑选 20 个场景用于定量比较。

4.2 评价方法

图像的融合结果可以通过主观的视觉判断(定性分析)或客观指标(定量比较)进行评价。其中主观方法基于人类视觉和心理进行判断,当图像差异较明显时可以较好地进行主观评价。但是当融合图像差异较小时,很难通过主观判断进行评价,为此根据不同的融合目的提出了各种客观评价指标。考虑到一种指标不能对图像进行全面的评价,因此为了全面度量图片的质量本文采用信息熵 $\text{EN}(\text{Entropy})^{[27]}$ 、归一化互相信息 $\text{NMI}(\text{Normalized Mutual Information})^{[28]}$ 、图像标准差 $\text{STD}(\text{Standard Deviation})^{[29]}$ 和平均结构相似度 $\text{MSSIM}(\text{Mean Structural Similarity})^{[30]}$ 来度量图像的质量。其中 EN 反映了图像的信息量,值越大表示包含的信息越丰富。 NMI 反映了原图像到目标图像的信息转移量,值越大表明越多的信息转移到了融合图像中去。 STD 反映了图像的对比度信息, STD 越大表明图像的明暗差距越大,目标越突出。 MSSIM 将图像畸变建模为相关性损失、亮度畸变和对比度畸变 3 个因素的组合,较为全面地度量了结构信息的转移, MSSIM 越大,表明融合图像与源图像的结构越相似融合效果越好。

4.3 实验结果和讨论

4.3.1 定性分析

图 2 显示了 TNO 和 RoadScene 中 6 对场景图及



(a) TNO和RoadScene数据集的6对原图像



(b) TNO场景各算法融合结果

(c) RoadScene场景各算法融合结果

图2 不同场景下各算法的融合结果图

Fig.2 The fusion images of different algorithms in different scenarios

各算法的融合结果,左上角数值为对应的场景号。其中图 2(a)为场景原图,第一行为红外图像用 I 表示,第二行为可见光图像用 V 表示。前 3 个场景来自 TNO,后 3 个场景来自 RoadScene。图 2(b)和 2(c)分别显示了各算法在 TNO 和 RoadScene 场景下的融合结果。左上角的数值和图 2(a)中的场景号对应,字母和算法对应,本文算法用 OURS 表示。

在 TNO 的 3 个场景中红外图像目标突出但背景细节信息几乎不存在,而可见光是在不同光照条件下获取的,图像目标不显著但背景细节丰富。图 3 中各种融合算法融合了红外可见光图像信息,使得融合图像信息更加丰富,一定程度上提升了目标的可见性,然而传统多尺度算法和 ASR 还存在边缘模糊或过渡不自然的问题。如在图 2(b)的场景 1 和场景 3 中:RP 融合图像存在白色斑点,在屋顶和树边缘存在模糊边缘甚至产生光晕,大大降低了图像质量;LP、DTCWT、NSCT、TS、ASR 在屋顶边缘存在黑色晕影,在树边缘存在模糊白边;MSVD 则在两个场景中都表现出锯齿边缘;这些算法存在边缘问题多是因为算法在分解过程中没有考虑保边问题。本文算法在这两个场景中都不存边缘问题,原因在于本文使用保边滤波器进行图像分解,较好地保留了图像边缘特性,同时为了使得边缘不太突兀,融合时使用了小窗口高斯滤波器对显著性映射图进行滤波,使得边缘不存在晕影或锯齿的同时过渡自然。同样考虑到分解时保边的 FPDEF 算法在场景 1 中不存在边缘问题而在场景 3 中却由于噪声的影响而丢失了树的轮廓信息,相比而言本文方法却能较好地保留该信息。这是由于本文采用的分解方法在保边的同时有效将各尺度的互补细节分离,同时根据各尺度显著性映射进行融合,使得互补细节在各尺度上得以保留,而噪声在最小的尺度中,只在最小尺度中发挥作用,减小了噪声的影响。在 TNO 的 3 个场景中还可以发现从整体来看本文方法与其他方法相比对比度更高,目标更加突出,例如场景 1 中的房屋和云彩,场景 2 和场景 3 中的人物目标。这是由于本文对冗余结构层和互补细节

层采用了显著性映射的融合方法,使得融合图像较好地保留了视觉显著特性,从而提升了图像的对比度和目标的显著性。

RoadScene 数据集与 TNO 数据集具有不同的特点,在 RoadScene 场景中红外图像除了目标显著之外还存在各种突出的光源点信息,并且包含了一定的场景细节信息。而不同光照条件下的可见光图像虽然包含了一些场景细节,但是由于点光源的存在会产生过曝光现象,从而导致周围细节丢失,同时目标也不够突出。在图 2(c)中各种算法都较好地完成了融合任务,但是从局部放大图中还是可以发现某些算法存在的缺陷。如在场景 4 中本文算法得到的车牌轮廓比 RP、DTCWT、FPDEF、TS、ASR 的更清晰,并且除了 RP 和 LP 之外本文算法得到的融合图像人物目标更突出。在场景 5 中,RP、FPDEF 存在一些噪声干扰,MSVD 存在锯齿边缘。而本文算法如之前所述有效克服了这些问题。在场景 6 中 RP 和 DTCWT 的白色竖条纹变浅,TS 和 FPDEF 甚至丢失了竖条纹,而 MSVD 仍然存在锯齿边缘。而本文算法较好地保留了这一信息。但是从整体上来看在 RoadScene 的这 3 个场景中本文算法的融合结果从视觉上和其他算法差异较小,仅在某些局部可以看出一些微小的差异。因此为了更加客观全面地对算法进行评价还需要进行定量比较。

4.3.2 定量比较

为了更加客观全面地对算法进行评价,本文在 TNO 和 RoadScene 上分别随机抽取了 20 幅图像用于算法的定量比较。结果如图 3 所示,图中从上到下分别为 EN、NMI、STD、MSSIM 的指标图,其中横坐标表示不同场景,纵坐标表示指标值大小,每个指标图上方的题注标明了指标均值。图 3(a)为 TNO 数据集上的结果,图 3(b)为 RoadScene 数据集的结果。从图中可以看出本文方法在随机挑选的 20 个场景中基本都可以达到最高的指标,相较于其他算法有较强的优越性。为了更加直观地感受到本文方法相较于其他方法的提升长度在表 1 中

表 1 算法的时间对比

Table 1 Time comparison results of different algorithms

| 分辨率 | RP | LP | DTCWT | NSCT | FPDEF | MSVD | TS | ASR | OURS |
|---------|-------|-------|-------|--------|-------|-------|-------|---------|--------|
| 280×280 | 0.015 | 0.007 | 0.181 | 1.784 | 0.182 | 0.202 | 0.038 | 58.160 | 5.947 |
| 620×450 | 0.080 | 0.016 | 0.522 | 6.532 | 0.962 | 0.629 | 0.106 | 149.614 | 25.958 |
| 768×576 | 0.111 | 0.018 | 0.623 | 8.001 | 1.294 | 1.050 | 0.128 | 201.592 | 39.188 |
| 917×678 | 0.207 | 0.029 | 1.286 | 11.392 | 1.832 | 1.574 | 0.173 | 232.084 | 58.254 |

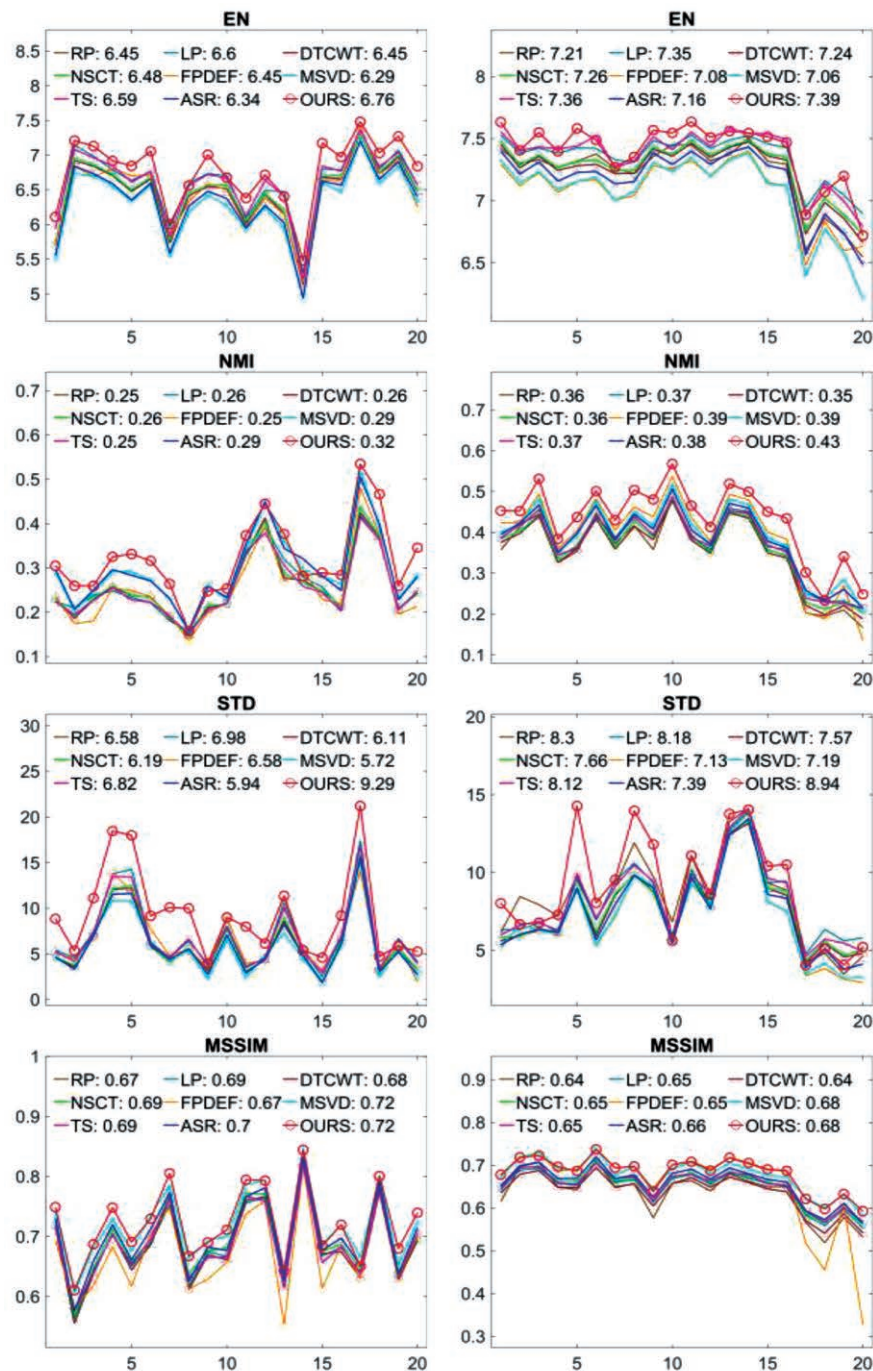


图 3 TNO 和 RoadScene 数据集下 40 幅场景的指标图

Fig.3 Indicator graphs of 40 scenarios under the TNO and RoadScene dataset

列出了本文算法相对于其他算法指标提升的百分比,其中加粗表示指标提升的最高百分比,下划线表示指标提升的最低百分比。T 代表 TNO 场景,R 代表 RoadScene 场景。图表的数据都保留小数点后两位。从图中可以看出:无论是 TNO 还是 RoadScene 本文算法的 4 个指标都以较为稳定的状态优于其他算法。如表 1 中所示在 TNO 数据集中本文算法的 EN 指标相对于第二名算法提升 2.42%,相

对于最差算法提升 7.47%;NMI 相对第二名算法提升 10.34%,相对最差算法提升 28%;STD 相对于第二名算法提升 33.09%,相对于最差算法提升了 62.41%;MSSIM 和 MSVD 持平,而相对于最差算法提升了 7.46%。而在 RoadScene 数据集集中的每个指标也都有所提升,但是提升的程度都小于 TNO 数据集,和定性分析的结果一致。

本文还对算法效率进行了比较,表 2 显示了每

表 2 指标提升百分比

Table 2 Percentage improvement of indicators

| 数据集 | 算法 | EN | NMI | STD | MSSIM |
|-----|-------|------|-------|-------|-------|
| T | RP | 4.81 | 28.00 | 41.19 | 7.46 |
| | LP | 2.42 | 23.08 | 33.09 | 4.35 |
| | DTCWT | 4.81 | 23.08 | 52.05 | 5.88 |
| | NSCT | 4.32 | 23.08 | 50.08 | 4.35 |
| | FPDEF | 4.81 | 28.00 | 41.19 | 7.46 |
| | MSVD | 7.47 | 10.34 | 62.41 | 0.00 |
| | TS | 2.58 | 28.00 | 36.22 | 4.35 |
| | ASR | 6.62 | 10.34 | 56.40 | 2.86 |
| R | RP | 2.50 | 19.44 | 7.71 | 6.25 |
| | LP | 0.54 | 16.22 | 9.29 | 4.62 |
| | DTCWT | 2.07 | 22.86 | 18.10 | 6.25 |
| | NSCT | 1.79 | 19.44 | 16.71 | 4.62 |
| | FPDEF | 4.38 | 10.26 | 25.39 | 4.62 |
| | MSVD | 4.67 | 10.26 | 24.34 | 0.00 |
| | TS | 0.41 | 16.22 | 10.10 | 4.62 |
| | ASR | 3.21 | 13.16 | 20.97 | 3.03 |

个算法在 4 种分辨率下运行 10 次的平均时间,单位为秒。随着原图像尺寸的增大,融合过程所花费的时间也增大。其中 LP 的算法效率最高,紧随其后的有 RP、TS、DTCWT、MSVD、FPDEF。这些算法可以在 1~2 s 之内完成融合任务,适合实时融合处理。而 NSCT 的方法由于采用了滤波器组进行分解相对较为复杂因此算法效率有所下降。而本文方法由于采用的互导滤波器是用迭代方式求解,因此在图像分解的部分效率较低,仅优于 ASR 方法,不具有实时性。综上所述本文算法在保留图像细节信息,防止边缘模糊,提升图像对比度和突出目标上有较好的综合性能,但是效率较低不适合实时处理。

5 结 语

本文提出了一种基于互导滤波器和图像显著性的融合方法,通过互导滤波将原图像进行冗余结构和互补细节的分解,较好地将冗余和互补信息分离,有效提高了互补信息转移到融合图像中的可能。同时提出了一种显著性映射方法用于各尺度融合从而提高了图像对比度。此外由于互导滤波器是保边滤波器因此也能避免边缘模糊问题。实验证明无论是从直观的视觉评价还是客观的指标比较,本文所提出的融合算法相较于传统的多尺度融合方法在保持图像信息,提升对比度,突出目标和防止边缘模糊方面具有一定的优越性。

参考文献 (References):

[1] Ma J, Ma Y, Li C. Infrared and Visible Image Fusion Methods and Applications: A Survey [J]. Information Fusion, 2019, 45: 153-178.doi:10.1016/j.inffus.2018.02.004.

[2] Zhang Xiaofeng, Lu Xiaoqi, Zhang Xinxue, *et al.* Data Fusion and Visualization of Multi-time Ocean Color Remote Sensing Data[J]. Remote Sensing Technology and Application, 2018, 33(5): 873-880.[[张晓峰, 吕晓琪, 张信雪, 等. 多时刻海色遥感数据融合及其可视化[J]. 遥感技术与应用, 2018, 33(5): 873-880.]. doi: 10.11873/j. issn. 1004-0323.2018.5. 0873.]

[3] Li Y, Jha D K, Ray A, *et al.* Information Fusion of Passive Sensors for Detection of Moving Targets in Dynamic Environments[J]. IEEE Transactions on Cybernetics, 2016, 47(1): 93-104.doi:10.1109/TCYB.2015.2508024.

[4] Muller A C, Narayanan S. Cognitively-engineered Multisensor Image Fusion for Military Applications [J]. Information Fusion, 2009, 10(2): 137-149.doi: 10.1016/j.inffus. 2008. 08.008.

[5] Meher B, Agrawal S, Panda R, *et al.* A Survey on Region based Image Fusion Methods[J]. Information Fusion, 2019, 48: 119-132. doi:10.1016/j.inffus.2018.07.010.

[6] Sharma A M, Vig R, Dogra A, *et al.* A Comparative Analysis of Transforms for Infrared and Visible Image Fusion[M]. Intelligent Communication, Control and Devices. Springer, Singapore, 2020: 85-93. doi: 10.1007/978-981-13-8618-3_10.

[7] Toet A. Image Fusion by a Ratio of Low-pass Pyramid [J]. Pattern Recognition Letters, 1989, 9(4): 245-253.

[8] Bulanon D M, Burks T F, Alchanatis V. Image Fusion of Visible and Thermal Images for Fruit Detection[J]. Biosystems Engineering, 2009, 103(1): 12-22.doi: 10.1016/j.biosystemseng.2009.02.009.

[9] Chen J, Li X, Luo L, *et al.* Infrared and Visible Image Fusion based on Target-enhanced Multiscale Transform Decomposition[J]. Information Sciences, 2020, 508:doi: 10.1016/j.ins.2019.08.066.

[10] Lewis J J, O'Callaghan R J, Nikolov S G, *et al.* Pixel-and Region-based Image Fusion with Complex Wavelets [J]. Information fusion, 2007, 8(2): 119-130. doi: 10.1016/j.inffus.2005.09.006.

[11] Zhang Q, Guo B. Multifocus Image Fusion Using the Nonsub-sampled Contourlet Transform [J]. Signal processing, 2009, 89(7): 1334-1346.doi:10.1016/j.sigpro.2009.01.012.

[12] Bavirisetti D P, Xiao G, Liu G. Multi-sensor Image Fusion based on Fourth Order Partial Differential Equations [C] // 2017 20th International Conference on Information Fusion (Fusion). IEEE, 2017: 1-9. doi: 10.23919/ICIF. 2017. 8009719.

[13] Yan J, Zhu M, Liu H, *et al.* Visual Saliency Detection Via Sparsity Pursuit [J]. IEEE Signal Processing Letters, 2010, 17(8): 739-742.doi:10.1109/LSP.2010.2053200.

[14] Bavirisetti D P, Dhuli R. Two-scale Image Fusion of Visible and Infrared Images Using Saliency Detection [J]. Infrared Physics & Technology, 2016, 76: 52-64.doi: 10.1016/j.infra-

- red.2016.01.009.
- [15] Naidu V P S. Image Fusion Technique Using Multi-resolution Singular Value Decomposition[J]. Defence Science Journal, 2011, 61(5): 479-484.
- [16] Nencini F, Garzelli A, Baronti S, *et al.* Remote Sensing Image Fusion Using the Curvelet Transform[J]. Information fusion, 2007, 8 (2) : 143-156. doi: 10.1016/j.inffus. 2006. 02.001.
- [17] Liu Y, Wang Z. Simultaneous Image Fusion and Denoising with Adaptive Sparse Representation[J]. IET Image Processing, 2014, 9(5): 347-357. doi:10.1049/iet-ipr.2014.0311.
- [18] Ma J, Chen C, Li C, *et al.* Infrared and Visible Image Fusion Via Gradient Transfer and Total Variation Minimization[J]. Information Fusion, 2016, 31. doi: 10.1016/j.inffus. 2016. 02.001.
- [19] Liu Y, Chen X, Cheng J, *et al.* Infrared and visible image fusion with convolutional neural networks[J]. International Journal of Wavelets, Multiresolution and Information Processing, 2018, 16(3): 1-20. doi: 10.1142/S0219691318500182.
- [20] Ma J, Yu W, Liang P, *et al.* FusionGAN: A Generative Adversarial Network for Infrared and Visible Image Fusion[J]. Information Fusion, 2019, 48: 11-26. doi: 10.1016/j.inffus. 2018.09.004.
- [21] Ma J, Liang P, Yu W, *et al.* Infrared and Visible Image Fusion Via Detail Preserving Adversarial Learning[J]. Information Fusion, 2020, 54: 85-98. doi: 10.1016/j.inffus. 2019. 07.005.
- [22] Xu J, Shi X, Qin S, *et al.* LBP-BEGAN: A Generative Adversarial Network Architecture for Infrared and Visible Image Fusion[J]. Infrared Physics & Technology, 2020, 104: 103144. doi: 10.1016/j.infrared.2019.103144.
- [23] Guo X, Li Y, Ma J, *et al.* Mutually Guided Image Filtering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42: 694-707. doi: 10.1109/TPAMI. 2018. 2883553.
- [24] Liu Y, Liu S, Wang Z. A General Framework for Image Fusion based on Multi-scale Transform and Sparse Representation [J]. Information Fusion, 2015, 24: 147-164. doi: 10.1016/j.inffus.2014.09.004
- [25] Toet A, The TNO Image Fusion Dataset[DB/OL], https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029, 26.04.2014, 06.10.2018
- [26] Xu H, Jiang X Y. RoadScene [DB/OL]. <https://github.com/hanna-xu/RoadScene/blob/master/README.md>, 08.2019, 06.09.2019.
- [27] Roberts J W, Van Aardt J A, Ahmed F B. Assessment of Image Fusion Procedures Using Entropy, Image Quality, and Multispectral Classification [J]. Journal of Applied Remote Sensing, 2008, 2(1): 023522. doi:10.1117/1.2945910.
- [28] Hossny M, Nahavandi S, Creighton D. Comments on 'Information Measure for Performance of Image Fusion' [J]. Electronics letters, 2008, 44 (18) : 1066-1067. doi: 10.1049/el: 20081754.
- [29] Jin X, Jiang Q, Yao S, *et al.* A Survey of Infrared and Visual Image Fusion Methods [J]. Infrared Physics & Technology, 2017, 85: 478-501. doi:10.1016/j.infrared.2017.07.010.
- [30] Wang Z, Bovik A C, Sheikh H R, *et al.* Image Quality Assessment: from Error Visibility to Structural Similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612. doi: 10.1109/TIP.2003.819861.

Infrared and Visible Image Fusion based on Mutually Guided Filtering and Saliency Map

Zhou Yi, Ma Jiayi, Huang Jun

(School of Electronic Information, Wuhan University, Wuhan 430072, China)

Abstract: In order to solve the problems of traditional multi-scale infrared visible fusion image with blurred edges, low contrast and inconspicuous targets, an infrared visible image fusion algorithm based on mutually guided filtering and saliency map is proposed. Firstly, the original images are decomposed into structure layers with redundant information and texture layers with complementary information at different scales by means of mutually guided filter, because the filter can separate the consistent structure from the inconsistent structure, and has the awareness of scale and edge preservation. Secondly, the visual saliency map function is constructed to map saliency of structure layers and texture layers of different scales according to the visual characteristics of the over-light or over-dark regions that are more likely to attract attention. Finally, according to the structure and texture characteristics of different scales, the final fusion image is reconstructed. The experimental results on two data sets show that compared with the traditional multi-scale fusion methods, the proposed method has a better subjective and objective evaluation effect in maintaining the image edge, enhancing the image contrast and highlighting the target.

Key words: Mutually guided filter; Saliency map; Multi-scale fusion; Infrared