

一种基于 EASSF 的红外与可见光图像视觉保真度融合

李 威, 李忠民

(南昌航空大学 信息工程学院, 江西 南昌 330063)

摘要: 最近, 多尺度特征提取被广泛应用于红外与可见光图像融合领域, 但是大多数提取过程过于复杂, 并且视觉效果不佳。为了提高融合结果的视觉保真度, 本文提出一种基于边缘感知平滑锐化滤波器 (Edge-Aware Smoothing-Sharpener Filter, EASSF) 的多尺度图像融合模型。首先, 提出一种基于 EASSF 的多尺度水平图像分解方法对源图像进行分解, 得到水平方向上的多尺度纹理成分和基础成分; 其次, 采用最大值融合规则 (Max-Fusion, MF) 融合纹理成分, 避免图像细节信息的丢失; 然后, 通过一种感知融合规则 (Perceptual-Fusion, PF) 融合基础成分, 捕获显著性目标信息; 最后, 通过整合融合后的多尺度纹理成分和基础成分得到融合图像。实验通过分析感知融合系数, 对比融合结果的客观数据得出红外与可见光图像融合在多尺度 EASSF 下较为合适的取值范围; 在该取值范围内, 本文提出的融合模型同一些较为经典和流行的融合方法相比, 不仅解决了特征信息提取的复杂性, 而且通过整合基础成分的显著性光谱信息, 有效地保证了融合结果的视觉保真度。

关键词: 红外与可见光图像融合; 边缘感知平滑滤波器; 视觉保真度; 感知融合

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-8891(2022)07-0686-07

Visual Fidelity Fusion of Infrared and Visible Image Using Edge-Aware Smoothing-Sharpener Filter

LI Wei, LI Zhongmin

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: Recently, multi-scale feature extraction has been widely used in the field of infrared and visible image fusion; however, most extraction processes are too complex, and the visual effect is not good. To improve the visual fidelity of the fusion result, a multi-scale horizontal image fusion model based on an edge-aware smoothing-sharpening filter (EASSF) is proposed for infrared and visible images. First, to obtain multi-scale texture components and basic components in the horizontal direction, a multi-scale horizontal image decomposition method based on the EASSF is proposed to decompose the source image. Second, the maximum fusion rule is used to merge texture components, which can avoid loss of information detail. Then, to capture salient target information, the basic components are fused via the perceptual-fusion rule. Finally, the fused image is obtained by integrating the fused multi-scale texture components and basic components. By analyzing the perceptual fusion coefficient of PF, the appropriate range of infrared and visible image fusion in the multi-scale EASSF is obtained through the objective data of the fusion results. In this range, compared with several classical and popular fusion methods, the proposed fusion model not only avoids the complexity of feature information extraction, but also effectively ensures the visual fidelity of fusion results by integrating the significant spectral information of basic components.

Key words: infrared and visible image fusion, edge-aware smoothing-sharpening filter, visual fidelity, perceptual fusion

收稿日期: 2021-10-24; 修订日期: 2021-11-09.

作者简介: 李威 (1997-), 男, 硕士, 主要从事图像融合的研究, E-mail: lv2881314@163.com.

通信作者: 李忠民 (1975-), 男, 博士, 副教授, 主要从事图像融合和人工智能方向的研究, E-mail: zhongmli@nchu.edu.cn.

基金项目: 国家自然科学基金 (61263040, 61861033); 江西省自然科学基金 (20202BABL202005)。

0 引言

可见光图像在弱光条件下不能很好地显示场景信息,但可以通过反射光清晰地呈现细节和背景;相比之下,红外图像中的物体可以通过热辐射来显示,但红外光图像无法显示细节和清晰的背景。整合两者的优势信息,能够充分地描述特定场景的环境和背景,这对于目标识别、目标跟踪和遥感遥测方面有着极为重要的作用^[1-6]。

近年来,基于图像特征信息提取技术在图像融合领域取得了较为明显的进步。2018年, Du 等人提出了一种用于融合不同分辨率的红外与可见光图像的融合算法(Different Resolutions Total Variation, DRTV^[7]),分别利用红外图像的像素灰度与融合图像在数据保真度上的相似性和可见光图像和融合图像在梯度上的相似性,将融合问题表述为全变差(Total Variation, TV)最小化问题;为了提高收敛速度, Zhao 等人在 2020 年提出了一种分层 Bayesian 模型^[8],该模型引入 TV,利用期望最大化算法对模型进行了高效的推断,从而将图像融合任务转化为回归问题。虽然基于 TV 的融合方法对图像特征提取有很大的效果,能够保留并且整合较多的结构特征信息,但是在视觉感知效果方面稍显不足。

随着神经网络的不断发展,深度学习慢慢地成为了图像处理领域较为活跃的研究工具。Li 等人在 2018 年利用深度学习框架视觉几何组(Visual Geometry Group, VGG^[9])整合红外与可见光图像的特征信息;随后, Wu 等人在 2019 年提出了一种基于深度学习残差网络(Residual-network, Resnet^[10])和零相位分量分析的新颖融合框架。这些深度学习网络能够很好地提高图像视觉效果,但是由于算法过于复杂,在细节特

征提取方面存在一定的人为影响。为了实现深度次的特征提取, Tan 等人在 2021 年基于多尺度边缘滤波和脉冲耦合神经网络(pulse coupled neural network, PCNN)提出了一种多模态融合方法(multi-level edge-preserving filtering, MLEPF^[11]),该方法虽然达到了多层次上的特征信息提取,但是受限于源图像曝光度,容易使得融合结果形成局部斑块。

通过上述的分析可知,红外与可见光图像融合目前的研究目标主要有两个:①提高源图像特征信息的提取能力,同时避免算法的复杂性带来的人为影响;②在保证融合过程的可行性的同时,进一步优化和巩固融合结果的保真度,从而契合人体的视觉效果。为了实现上述两个目标,本文提出一种基于 EASSF 的多尺度红外与可见光图像融合模型,该模型通过 EASSF 实现多尺度分解,提取源图像水平方向上的信息。

1 边缘感知平滑锐化滤波器

平滑可以去除图像中的噪声干扰,而锐化可以突出图像的边缘。2021 年 3 月, Deng 等人提出了一种边缘感知平滑锐化滤波器(Edge-Aware Smoothing-Sharpening Filter, EASSF^[12])用于图像处理, EASSF 通过调节平滑(锐化)参数 L 就可以实现图像特征信息的提取。图 1 为可见光图像“Road”在不同参数下的平滑锐化结果,从图 1 中容易看出,当 $L \leq 1$ 时, EASSF 进行图像平滑;而 $L > 1$ 时, EASSF 进行图像锐化。由于 EASSF 的简便性和通用性,一种适用于多尺度水平方向的图像分解方法被提出并且应用于图像的特征提取。图 2 为红外图像“Road”的分解示例图,其中 V 为输入可见光图像, EF 为 EASSF 图像平滑锐化过程, V_1, V_2 和 V_n 分别是 EF 平滑锐化后的

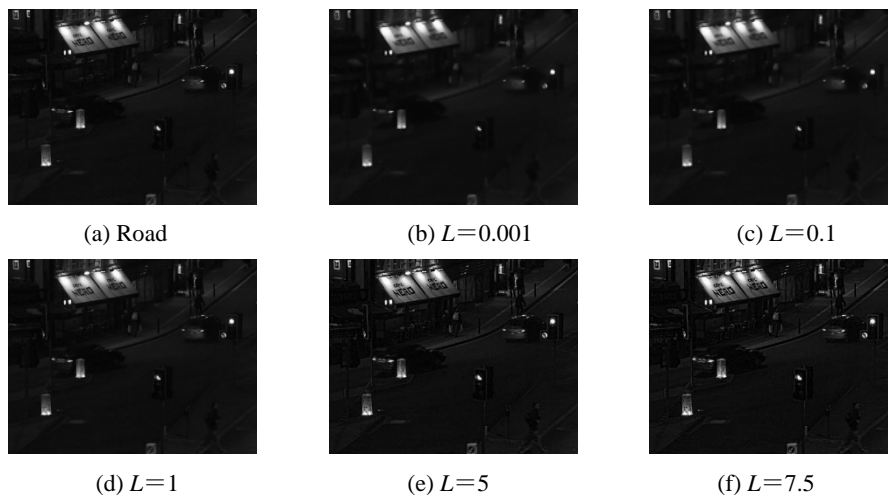


图 1 不同参数 L 的 EASSF 成像结果

Fig.1 The results of images by EASSF with different L

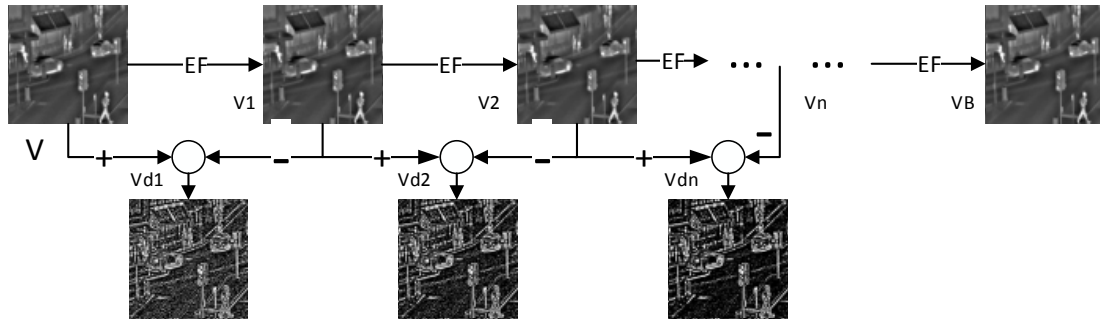


图2 EASSF 图像分解

Fig.2 The image decomposition by EASSF

图像; V_{d1} , V_{d2} 和 V_{dn} 分别是分解后结果对数放大的纹理成分; V_B 为输入图像的基础成分。具体的多尺度水平分解过程可以用如下公式表示:

$$V_{d1} = V - V_1 \quad (1)$$

$$V_{dn} = (V_n - 1) - V_n \quad (2)$$

$$V_B = EF(V_n - 1) \quad (3)$$

2 基于 EASSF 的红外与可见光图像融合

图3为本文提出的基于 EASSF 的融合算法框架,其中 V 和 R 分别代表可见光图像和红外图像;可见光图像 V 和红外图像 R 分解过程及得到的各个成分图像参考图2; F_{d1} , F_{d2} 和 F_{dn} 分别为各尺度上融合后的纹理成分; F_B 和 F 分别为融合后的基础成分和融合图像; Max-fusion 和 Perceptual-fusion 分别代表最大值融合规则和感知融合规则。具体的融合模型可以细分为4部分:

1) 基于 EASSF 的红外与可见光图像分解

利用 EASSF 分别对可见光图像 V 和红外光图像 R 进行分解得到分解后的纹理成分和基础成分。

2) 纹理成分融合

利用 Max-fusion 融合规则对红外图像的纹理成分和可见光图像的纹理成分进行融合,得到融合后的纹理成分 F_{d1} , F_{d2} 和 F_{dn} 。

3) 基础成分融合

利用 Perceptual-fusion 感知融合规则对红外图像的基础成分 R_B 和可见光图像的基础成分 V_B 进行感知融合,捕获光谱信息,得到融合的基础成分 F_B 。

4) 重构源图像

通过叠加融合后的纹理成分和基础成分实现源图像的重构,得到融合图像 F 。

2.1 基于 EASSF 的红外与可见光图像分解

基于 EASSF 的图像分解过程可以参照第一章 EASSF, 其中数学公式如(1)~(3)所示。通过 EASSF

的多尺度水平分解分别得到对应源图像的纹理成分和基础成分。

2.2 纹理成分融合

通过 EASSF 的多尺度水平分解,我们可以得到较为细腻的纹理细节信息,从图3可以明显地看出纹理细节信息。众所周知,如果图像当中某点的像素较大,反映在图像上是人体视觉感知最为明显的一个区域或者一个位置,这也是图像当中较为重要的一个信息点。为了尽量保留较多的细节信息,Max-fusion 最大值融合规则被应用于整合纹理成分,融合后的纹理成分 F_{dn} 可以通过公式(4)得到:

$$F_{dn} = \begin{cases} V_{dn}, & |V_{dn}| \geq |R_{dn}| \\ R_{dn}, & \text{otherwise} \end{cases} \quad (4)$$

2.3 基础成分融合

基础成分是源图像的一个粗略缩写,它大致上包含源图像当中的基础信息,其中以显著性光谱信息尤为突出。为了使融合图像达到突出显著性目标效果,同时使得融合结果较为契合人体的视觉感知,一种称为 Perceptual-fusion 的感知融合规则被应用于对红外图像的基础成分 R_B 和可见光图像的基础成分 V_B 进行感知融合,从而得到融合的基础成分 F_B 。

Perceptual-fusion 的融合规则具体如下:

1) 通过像素值得到初始权重值 IW ;

$$IW = \begin{cases} V_B, & |V_B| \geq |R_B| \\ R_B, & \text{otherwise} \end{cases} \quad (5)$$

2) 一种非线性函数 $N_\lambda(x)$ 被引入用于调节红外与可见光图像融合的光谱信息:

$$N_\lambda(x) = \arctan(\lambda x) / \arctan(\lambda) \quad (6)$$

3) 利用高斯滤波器对 $N_\lambda(IW)$ 进行平滑,排除融合图像数据因噪声干扰而形成的不必要影响,得到最终的融合权重图:

$$F_W = \text{Gaussian}(N_\lambda(IW)) \quad (7)$$

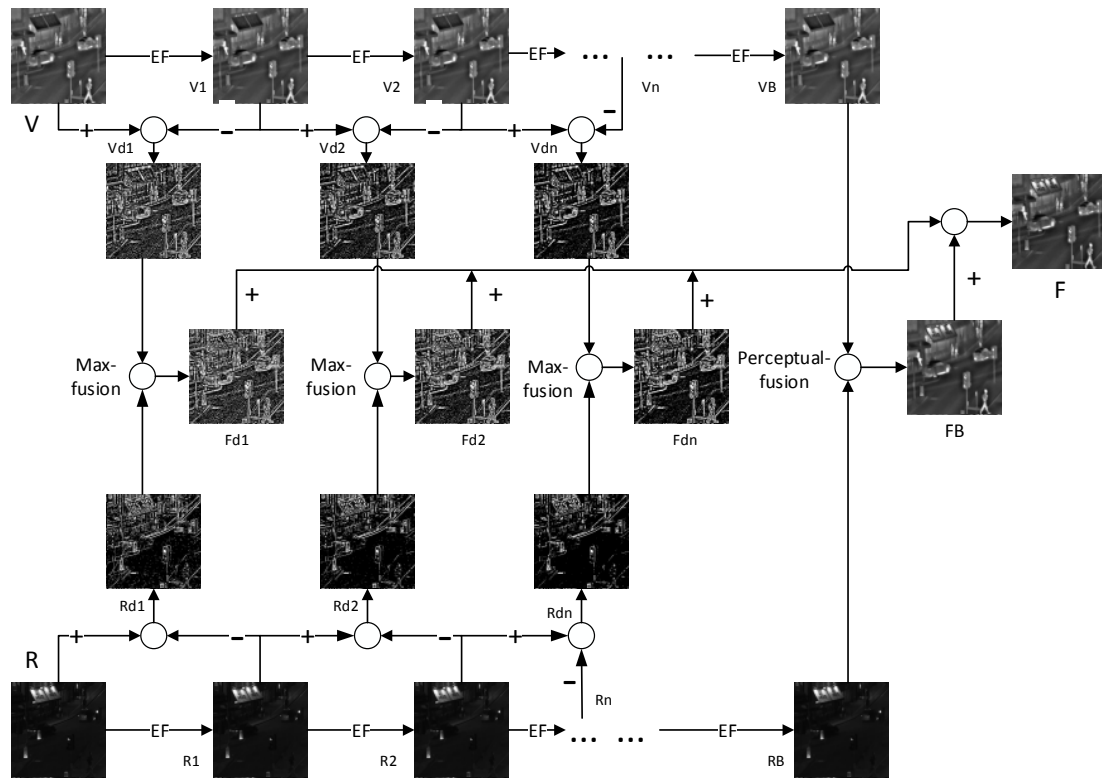


图3 基于 EASSF 的图像融合框架

Fig.3 The EASSF-based image fusion framework

4) 重构融合后的基础成分 F_B 。

$$F_B = F_W * V_B + (1 - F_W) * R_B \quad (8)$$

2.4 重构源图像

融合后的图像 F 通过公式(9)得到:

$$F = F_B + \sum_{i=1}^n F_{di} \quad (9)$$

3 仿真结果分析

仿真实验选取的数据集都来自文献[13], 我们选取了3对图像数据进行实验。所有的实验平台都是基于MATLAB R2018b, 电脑配置为3.6GHz AMD Ryzen 5 2600X Six-Core Processor。5种经典并且流行的融合算法被用来验证本文提出的融合算法, 他们分别为DRTV、Bayesian、VGG、Resnet和MLEPF, 其中VGG, Resnet和MLEPF是当前较为流行的融合算法, 并且MLEPF是将多尺度与深度学习相结合的融合算法, 而Bayesian和MLEPF分别在2020年和2021年被提出, 所有的融合算法都是近3年内的。为了验证本文基于EASSF融合算法的优势, 首先, 对EASSF下的感知融合系数 λ 进行一定的实验, 获取在EASSF融合模型下的最优融合结果; 其次, 基于该感知系数下, 从主客观层面对融合结果进行有效性分析, 其中在客观定量分析方面, 7种融合图像的指标被用于定量分析实验融合效果, 它们分别为信息熵EN、互信息量

MI、基于人为影响的度量 N_{abr} ^[14]、基于噪音干扰的度量 Q_{abr} ^[15]、差异相关性的总和SCD^[16]、基于结构相似性的度量MS-SIM^[17]和视觉保真度VIFF^[1], 上述定量指标除 Q_{abr} 之外, 数值越大表明融合结果表现越好^[1,9]。

3.1 感知系数分析

感知融合系数 λ 用来捕获源图像之间的光谱信息, 实现人体视觉效果。实验通过调节 λ 系数, 得到的融合结果如图4所示。其中实线方框标记的是行人, 虚线方框标记的是指示灯。通过图4可以很明显地观察到随着 λ 值的变大, 融合图像能够显著性地突出行人; 但是, 被蓝色标记的指示灯越来越不模糊, 丢失可见光图像的纹理信息, 使得融合图像视觉效果不佳。为了实现最佳的视觉感知融合效果, 我们分别对图4当中的图像进行了客观定量分析, 得到的定量结果如表1所示。从表1中定量数据可以清楚地看出图4(d)和图4(e)两幅图的客观数据最佳, 而随着 λ 值的增大, 除了 N_{abr} 指标变差外, 其他的客观数据都在 $\lambda = [10, 20]$ 能够有较好的表现。为了使得融合效果最佳, 从全面且综合的角度考虑, 令 $\lambda = 10$, 进行后续分析。

3.2 主观分析

实验选取的图像是来自经过配准好的TNO图像数据库, 在本节选择3对图像组进行主客观分析, 分别是“Road”、“Kaptein_1654”和“Kaptein_1123”, 它们的图像融合结果如图5~7所示。

图5反映的是一个路口交通情况，选取被黑色实线方标记的行人进行分析，可以很明显的看出，与 Bayesian、VGG、Resnet 和 MLEPF 得到的融合结果相比，我们的方法 EASSF 融合可以很明显的突出图像目标人物；但是对于 DRTV，我们的优势在于融合图像较为细腻，图像数据保真度较高，感知视觉效果

好。图6呈现的是一个工作人员进门时的场景，从标记的目标与融合图像的背景信息看出，我们提出的方法 EASSF 很好地定位了工作人员，融合效果较好。图7展示的是一个举伞路人路过凉亭的画面，从图7中我们能够较为直观地看出我们的融合结果最佳，能够较为形象地描述出该特定场景。

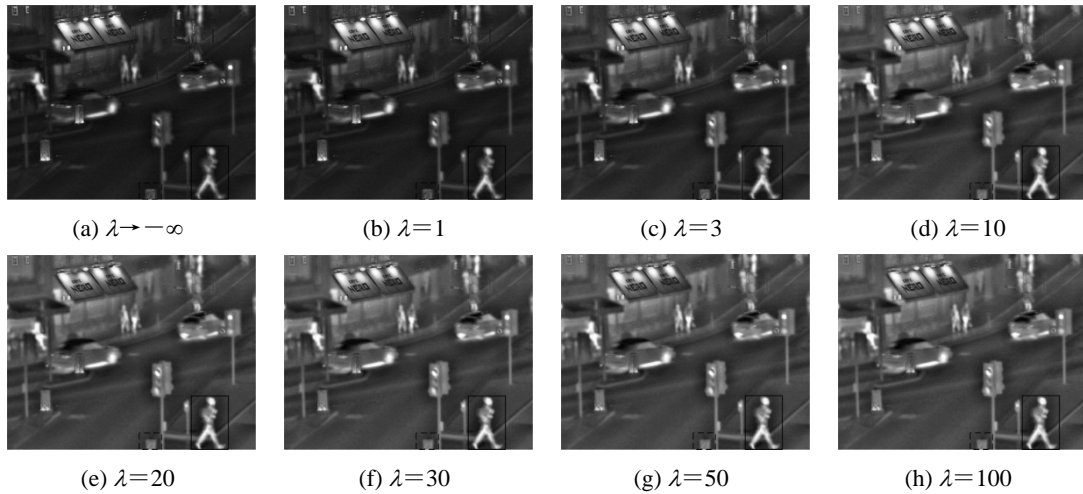


图4 不同参数 λ 的EASSF成像结果

Fig.4 The result of images by EASSF with different λ

表1 不同感知系数下的融合定量数据

Table 1 Fusion of quantitative data under different perception coefficients

	EN	MI	Q_{abf}	N_{abf}	SCD	MS-SIM	VIFF
(a)	6.0190	12.0380	0.5339	0.0458	1.7619	0.9088	0.4558
(b)	6.1825	12.3651	0.5638	0.0505	1.7885	0.9301	0.4970
(c)	6.5425	13.0851	0.6187	0.0684	1.8166	0.9613	0.5727
(d)	6.7653	13.5307	0.6243	0.1090	1.7887	0.9667	0.5887
(e)	6.7762	13.5524	0.6175	0.1218	1.7700	0.9649	0.5801
(f)	6.7717	13.5434	0.6099	0.1091	1.7624	0.9643	0.5764
(g)	6.7753	13.5506	0.6039	0.1229	1.7555	0.9633	0.5713
(h)	6.7622	13.5244	0.5691	0.1358	1.7495	0.9628	0.5680

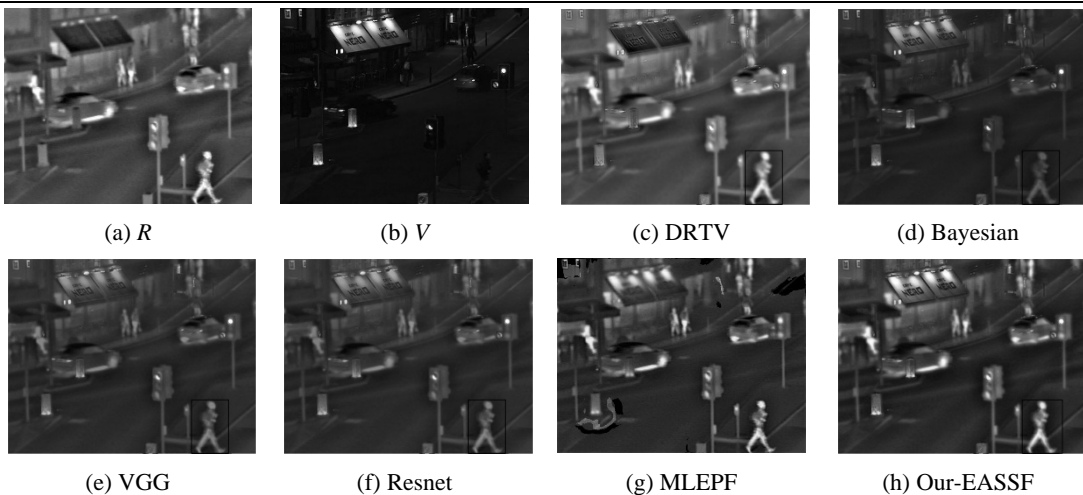


图5 不同方法在“Road”图像下的融合结果

Fig.5 The fusion results by different methods under “Road”

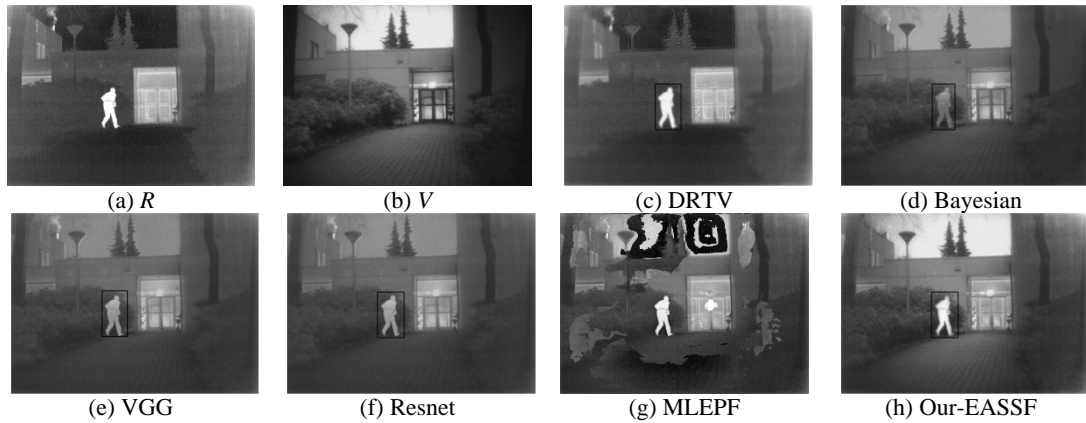


图6 不同方法在“Kaptein_1654”图像下的融合结果

Fig.6 The fusion results by different methods under “Kaptein_1654”

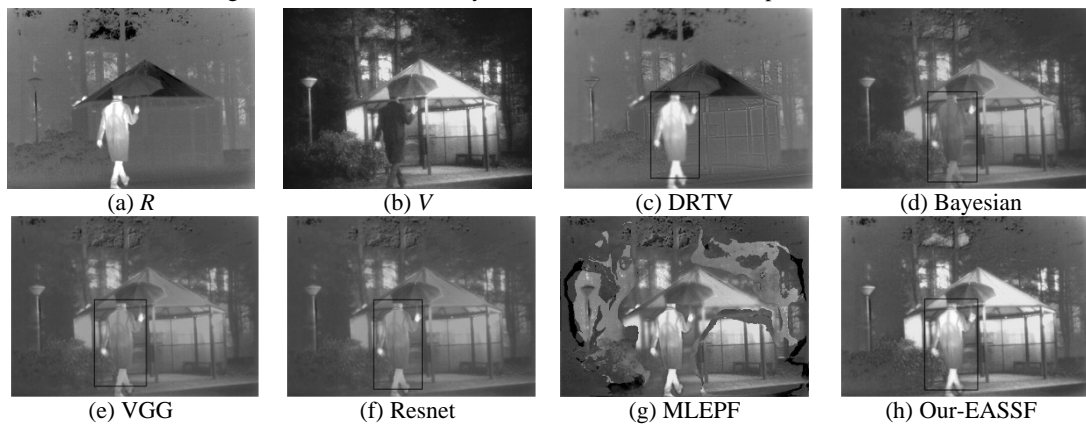


图7 不同方法在“Kaptein_1123”图像下的融合结果

Fig.7 The fusion results by different methods under “Kaptein_1123”

3.3 客观分析

客观分析是从数据的角度定量衡量融合图像的质量。选取的17组实验图像数据组进行分析,我们得到的结果见表2~表4。表2~表4分别是“Road”、“Kaptein_1654”和“Kaptein_1123”的客观评价。在表2中我们的融合方法在MI、 Q_{abf} 、SCD、MS-SIM和VIFF数值较高,而在EN和 N_{abf} 方面略低于DRTV和

Resnet;而在表3当中,我们的方法EASSF在EA、MI、SCD和VIFF占据较好的优势,其他指标略低于算法Bayesian和Resnet;另外在表4中,我们在MI和 N_{abf} 指标上略低于MLEPF和VGG,其他的数值指标都占优势。而在SCD和VIFF指标上面,我们的融合方法都有良好的表现。而针对于个别如 Q_{abf} 和 N_{abf} 指标的劣势,这主要原因在于感知系数 λ 的变化。

表2 “Road”的客观评价数据

Table 2 Objective evaluation data of "Road"

	EN	MI	Q_{abf}	N_{abf}	SCD	MS-SIM	VIFF
DRTV	6.7132	13.0263	0.3982	0.1437	1.1212	0.9177	0.4249
Bayesian	5.5489	11.0977	0.3295	0.0034	1.4586	0.81	0.2566
VGG	5.9387	11.8774	0.4144	0.0038	1.4889	0.8933	0.3772
Resnet	5.9322	11.8644	0.3823	0.0022	1.4762	0.8889	0.3634
MLEPF	6.2354	12.4707	0.4911	0.1379	1.565	0.8808	0.4712
Ours	6.5425	13.0851	0.6187	0.0684	1.8166	0.9613	0.5727

表3 “Kaptein_1654”的客观评价数据

Table 3 Objective evaluation data of "Kaptein_1654"

	EN	MI	Q_{abf}	N_{abf}	SCD	MS-SIM	VIFF
DRTV	7.2171	14.4343	0.3368	0.1296	1.0802	0.6963	0.1312
Bayesian	6.9898	13.9795	0.5255	0.0016	1.6979	0.8648	0.2578
VGG	6.7667	13.5333	0.3614	0.0006	1.6992	0.8516	0.2842
Resnet	6.7802	13.5604	0.3493	0.0003	1.6941	0.8510	0.2819
MLEPF	7.2391	14.4781	0.4921	0.1874	1.204	0.6755	0.222
Ours	7.2296	14.4592	0.4448	0.0438	1.7434	0.8442	0.2974

表4 “Kaptein_1123”的客观评价数据
Table 4 Objective evaluation data of "Kaptein_1123"

	EN	MI	Q_{abf}	N_{abf}	SCD	MS-SIM	VIFF
DRTV	6.6741	13.3482	0.2392	0.11	1.2149	0.8055	0.296
Bayesian	6.4215	12.843	0.2711	0.0035	1.8185	0.8048	0.1505
VGG	6.4876	12.9753	0.3248	0.000896	1.82	0.8782	0.272
Resnet	6.5069	13.0139	0.3171	0.0007	1.8192	0.8781	0.2679
MLEPF	7.1401	14.2803	0.411	0.1744	1.3662	0.5958	0.1951
Ours	7.1576	14.1151	0.4484	0.0728	1.8204	0.8945	0.3414

从表1得出, λ 越大, N_{abf} 值越大, 带入的噪音越大, 使得融合效果略显不佳; 从客观分析总体上来说, 我们的方法 EASSF 具有一定的优势, 能够实现红外与可见光图像在视觉保真度融合。

4 结论

针对图像特征提取复杂度问题, 为了实现图像视觉保真度融合, 本文基于 EASSF 提出了一种较为简易且能够满足视觉感知融合的红外与可见光图像融合模型。利用 EASSF 平滑锐化的简便性, 实现了源图像的多尺度水平特征提取, 得到纹理成分和基础成分; 采用 Max-fusion 最大值融合规则实现了纹理成分的融合, 而基础成分利用 Perceptual-fusion 感知融合规则实现图像的视觉保真度融合; 最后叠加融合后的各个成分重构源图像。实验通过对感知系数 λ 的分析, 得出较为合适的感知融合系数区间 λ 取值在[10,20]; 通过主客观的分析表明, 本文的融合方法 EASSF 与现在较为经典和流行的算法相比具有一定的优势, 能够保留源图像数据的真实性, 整合较多的源图像信息, 并且契合人体的视觉感官。

参考文献:

[1] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: a survey [J]. *Information Fusion*, 2019, **45**: 153-178.
 [2] Singh R, Vatsa M, Noore A. Integrated multilevel image fusion and match score fusion of visible and infrared face image for robust recognition[J]. *Pattern Recogn.* 2008, **41**(3): 880-893.
 [3] Reinhard E, Ashikhmin M, Gooch B, et al. Color transfer between images[J]. *IEEE Computer. Graph. Appl.*, 2002, **21**(5): 34-41.
 [4] Somone G, Farina A, Morabito F C, et al. Image fusion techniques for remote sensing applications [J]. *Inform. Fusion*, 2002, **3**(1): 3-15.
 [5] 陈峰, 李敏, 马乐, 等. 基于滚动引导滤波的红外与可见光图像融合算法[J]. *红外技术*, 2020, **42**(1): 54-61.
 CHEN Feng, LI Min, MA Le, et al. Infrared and visible image fusion

algorithm based on the rolling guidance filter[J]. *Infrared Technology*, 2020, **42**(1): 54-61.
 [6] 杨九章, 刘炜剑, 程阳. 基于对比度金字塔与双边滤波的非对称红外与可见光图像融合[J]. *红外技术*, 2021, **43**(9): 840-844.
 YANG Jiuzhang, LIU Weijian, CHENG Yang. A symmetric infrared and visible image fusion based on contrast pyramid and bilateral filtering [J]. *Infrared Technology*, 2021, **43**(9): 840-844.
 [7] DU Qinglei, XU Han, MA Yong, et al. Fusing infrared and visible images of different resolutions via total variation model[J]. *Sensors*, 2018, **18**(11): 3827.
 [8] ZHAO Z, XU S, ZHANG C, et al. Bayesian fusion for infrared and visible images[J]. *Signal Processing*, 2020, **177**: 165-168
 [9] LI H, WU X, Kittler J. Infrared and visible image fusion using a deep learning framework[C]//*The 24th International Conference on Pattern Recognition (ICPR)*, 2018: 2705-2710.
 [10] LI Hui, WU Xiaojun, Tariq S. Durrani. Infrared and visible image fusion with ResNet and zero-phase component analysis[J]. *Infrared Physics & Technology*, 2019, **102**: 1030390.
 [11] TAN Wei, Thitn William, XIANG Pei, Zhou, Huixin. Multi-modal brain image fusion based on multi-level edge-preserving filtering[J]. *Biomedical Signal Processing and Control*, 2021, **64**(11): 102280.
 [12] DENG Guan, Galetto Fernando J, Al-Nasrawi M, et al. A guided edge-aware smoothing-sharpening filter based on patch interpolation model and generalized Gamma distribution[J]. *IEEE Open Journal of Signal Processing*, 2021, **2**: 119-135.
 [13] Toet A. TNO Image Fusion Dataset[EB/OL]. [2021-10-01]. http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.
 [14] Xydeas C S, Petrovic V. Objective image fusion performance measure[J]. *Electron. Lett.*, 2000, **36**(4): 308-309.
 [15] Aslantas V, Bendes E. A new image quality metric for image fusion: the sum of the correlations of differences[J]. *International Journal of Electronics and Communications*, 2015, **69**(12): 1890-1896.
 [16] LI H, WU X J, Kittler J. Infrared and visible image fusion using a deep learning framework[C]//*The 24th International Conference on Pattern Recognition (ICPR)*, 2018: 2705-2710.
 [17] YAN Huibin, LI Zhongmin. A general perceptual infrared and visible image fusion framework based on linear filter and side window filtering technology[J]. *IEEE Access*, 2020, **8**: 3029-3041.