

DOI: 10.11918/j.issn.0367-6234.201806014

# 基于 SIFT 字典学习的引导滤波多聚焦图像融合

陈蔓<sup>1,2</sup>, 钟勇<sup>1</sup>, 李振东<sup>1,2</sup>

(1.中国科学院 成都计算机应用研究所, 成都 610041; 2.中国科学院大学, 北京 100049)

**摘要:** 目前多数多聚焦图像融合算法仅仅针对解决某一类问题,如融合结果的局部细节保留能力差、空间连续性不足和对未配准的源图像鲁棒性差等. 为能够同时解决以上问题,提出了一种基于 SIFT 字典学习的引导滤波多聚焦图像融合算法. 该算法通过学习子字典克服了图像低秩表示具有全局性而局部细节描述不足的缺陷,同时子字典的分类利用图像 SIFT 特征的平移不变、尺度不变等特性,消除了未配准源图像融合结果出现伪影的现象. 此外,在源图像的低秩表示系数融合过程中引入引导滤波,增加了融合图像的空间连续性. 引导滤波的窗口大小是根据特征内容和非特征内容进行自适应选取,即属于特征内容的点选取较小的窗口,而属于非特征内容的点选取较大的窗口. 为验证算法的有效性,实验过程中选取 6 组数据,包括 3 组广泛应用于研究的多聚焦图像以及 3 组实际拍摄的多聚焦图像. 实验结果表明,该算法从主观视觉效果的定性分析和客观融合质量评价的定量分析都优于当前主流的多聚焦图像融合算法.

**关键词:** 图像融合; SIFT; 低秩表示; 字典学习; 引导滤波

**中图分类号:** TP399

**文献标志码:** A

**文章编号:** 0367-6234(2018)11-0059-08

## Multi-focus image fusion based on SIFT dictionary learning and guided filtering

CHEN Man<sup>1,2</sup>, ZHONG Yong<sup>1</sup>, LI Zhendong<sup>1,2</sup>

(1. Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, 610041;

2. University of Chinese Academy of Science, Beijing 100049)

**Abstract:** In order to solve the problem that the local detail retention ability, spatial continuity and non-registration problems of most multi-focus image fusion algorithms cannot be improved at the same time, this paper proposes a multi-focus image fusion algorithm based on SIFT dictionary learning and guided filtering. The algorithm overcomes the problem that the low rank representation of image can capture the global structure but could not preserve the local structure by learning sub-dictionaries. The classification of the sub-dictionaries utilizes the translation invariance and the scale invariance etc. of SIFT to eliminate the fusion artifacts of unregistered images. In addition, the adaptive-window guided filtering is performed during the low rank representation coefficients fusion progress, which increases the spatial continuity of fused image. Pixels with rich texture assign to small window, while weak texture pixels choose large window. We select 6 groups of data, including 3 groups of widely used images and 3 groups of real-world images for verifying the validity of the proposed algorithm. Experimental results show that this algorithm outperforms the current mainstream multi-focus image fusion algorithms from qualitative analysis and quantitative analysis.

**Keywords:** image fusion; SIFT; low rank representation; dictionary learning; guided filtering

图像融合是计算机视觉中的一个重要领域,其目标在于提取 2 幅及以上源图像的有效信息,合成为 1 幅包含所有有效信息的图像. 图像融合技术应用广泛,如遥感图像融合<sup>[1]</sup>,医学图像融合<sup>[2]</sup>,多聚焦图像融合<sup>[3-4]</sup>等. 本文主要针对的是金属表面纹理的多聚焦图像融合. 由于镜头景深受到限制,使 CCD 图像传感器获得的图像呈现部分位置聚焦、其他位置虚焦的状态. 因此需要使用图像融合算法提

取出所有聚焦的部分合成一幅完全聚焦的图像,使其能够更加方便人眼观察或者计算机处理.

常用的图像融合算法主要分为 2 大类:基于空间域和基于变换域. 基于空间域的算法主要通过提取图像的空间特征得到对应的权重,然后加权源图像得到最终的融合图像. 其代表算法有形态学滤波图像融合<sup>[3]</sup>、引导滤波图像融合<sup>[4]</sup>、脉冲耦合神经网络图像融合<sup>[5]</sup>等. 其优点在于能够保持较好的空间连续性,但是细节还原能力较差. 基于变换域的算法,其基本思想在于通过对源图像进行多尺度变换,然后对细节部分和近似部分采用不同的融合规则得到融合系数,再进行逆变换得到融合图像. 其代表算法有基于 Laplacian 金字塔分解<sup>[6]</sup>、基于离散

**收稿日期:** 2018-06-04

**基金项目:** 四川省科技厅科技成果转化项目(2014CC0043); 四川省科技创新苗子工程项目(SCMZ2006012)

**作者简介:** 陈蔓(1991—),女,博士研究生;  
钟勇(1966—),男,研究员,博士生导师

**通信作者:** 陈蔓, chenman13@mails.ucas.edu.cn

小波变换<sup>[7]</sup>、基于非下采样轮廓波变换<sup>[8]</sup>的图像融合等. 近年来, 基于稀疏表示<sup>[9-11]</sup>图像融合的出现, 使得图像融合效果有了进一步的提升, 但是稀疏表示不具有全局性, 容易产生块效应. 具有全局性的低秩表示算法<sup>[12]</sup>, 能够有效解决块效应的问题, 但是局部细节保持能力欠缺.

结合已有算法的优缺点, 提出一种基于 SIFT 字典学习的引导滤波图像融合算法. 图像的低秩表示具有全局表示能力, 但是局部细节描述不足, 本文通过学习不同类别的子字典, 增强了融合图像中的细节部分, 并且能够适用于未配准的源图像. 基于 SIFT 字典学习的引导滤波图像融合算法首先使用滑动窗口技术对图像进行分块, 再提取每个块的 SIFT 特征, 接着使用 k-means 对 SIFT 特征进行聚类, 然后对每个类别学习子字典, 最后合成一个整体字典. 得到字典之后, 首先将每幅源图像进行低秩表示, 按列向量取 1 范数最大者为融合图的低秩表示系数, 最后乘以字典得到融合图像. 为增加融合图的空间连续性, 在融合阶段引入引导滤波<sup>[13]</sup>, 并使用熵对内容(特征内容和非特征内容)进行区分, 根据不同内容类别选取不同大小的引导滤波窗口. 实验结果表明, 基于 SIFT 字典学习的引导滤波图像融合在局部细节保留能力、空间连续性和未配准问题方面都取得了非常好的结果.

## 1 相关技术

### 1.1 低秩表示

低秩表示<sup>[12]</sup>(Low-Rank Representation, LRR)是将数据集  $S$  表示为字典  $D$  和一个系数矩阵  $W$  的乘积, 即  $S = DW$ . 其中  $W$  是低秩的, 则称  $W$  是数据集  $S$  在字典  $D$  下的低秩表示, 可转换为求解如下的优化问题

$$\min_{W, E} (\|W\|_* + \lambda \|E\|_{2,1}), \text{ s.t. } S = DW + E. \quad (1)$$

式中:  $\lambda > 0$  为平衡因子,  $E$  为误差矩阵,  $\|X\|_*$  为  $X$  的核范数, 即求  $X$  奇异值之和,  $\|X\|_{2,1}$  为 21 范数, 其计算公式如下

$$\|X\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m [X]_{ij}^2}. \quad (2)$$

式中  $X \in \mathbf{R}^{m \times n}$ . 式(1)由拉格朗日乘子相关算法<sup>[14]</sup>计算即可.

### 1.2 引导滤波

引导滤波<sup>[13]</sup>(Guided Filtering, GF)是一种线性移可变的滤波操作, 并希望输出图像能与原始图像尽量接近. 假设引导图像  $G$  和输出图像  $q$  间是一种局部线性关系, 即

$$q_i = a_k G_i + b_k, \forall i \in \omega_k. \quad (3)$$

式中:  $a_k, b_k$  为线性系数,  $\omega_k$  是以像素  $k$  为中心的局部窗口, 可见在局部窗口内, 线性系数为常量. 为求解线性系数, 需要最小化如下的误差函数  $E_r$ .

$$E_r(a_k, b_k) = \sum_{i \in \omega_k} ((a_k G_i + b_k - p_i)^2 + \varepsilon a_k^2). \quad (4)$$

式中:  $p$  为输入图像,  $\varepsilon$  为衰减系数,  $\varepsilon a_k^2$  为衰减项, 防止参数过大. 由求解结果可知, 当引导图像  $G$  为输入图像本身时, 引导滤波则起到边缘保持滤波的作用, 即图像平滑的同时, 边缘能够保持不变.

## 2 本文算法

### 2.1 基于 SIFT 的字典学习

基于 SIFT 的字典学习方法主要包括以下几个部分: 基于滑动窗口技术分块、SIFT 特征提取、k-means 聚类、子字典学习和字典合成.

1) 分块. 将源图像  $I_i (i = 1, 2, \dots, N)$  使用滑动窗口技术进行分块,  $N$  为源图像总数. 设  $W, H$  分别为源图像宽高,  $s$  为滑动窗口大小, 步长为  $t$ , 则每幅源图像有  $Q = (\lfloor (W-s)/t \rfloor + 1) \times (\lfloor (H-s)/t \rfloor + 1)$  个块用于字典学习, 即学习总块数为  $NQ$ , 其中参数  $s$  和  $t$  的选取将在实验部分进行说明.

2) SIFT 特征提取. 计算第  $j$  块的 SIFT 特征, 记为  $B_j (j = 1, 2, \dots, NQ)$ . 为能够满足 k-means 聚类输入要求, 每个块的 SIFT 特征维度应该保持一致. 提取 SIFT 特征前需先提取每个块的关键点, 即为 DoG 尺度空间共 26 个邻域(见图 1)的极大、极小值点, 并去除部分不合格的点所剩下的特征点. DoG 有效层数  $S = 3$ , 初始尺度  $\delta_0 = 1.6$ , 参数和去除条件均参考文献[15]. 由于每个块的关键点个数不一致, SIFT 特征提取将分为以下 3 种情况:

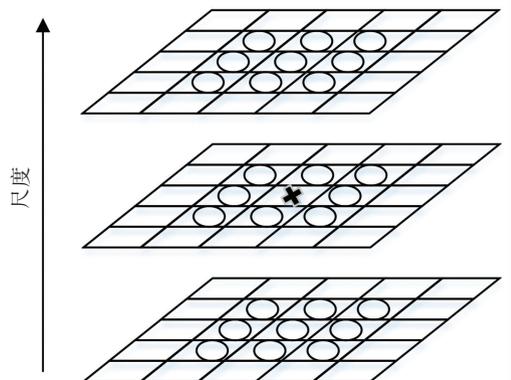


图 1 DoG 尺度空间及 26 个邻域

Fig.1 DoG scale space and 26 neighborhoods

情况 1: 无关键点.

若第  $j$  块无关键点, 则直接将其分为第 0 类, 记  $U_j = 0, U_j$  为第  $j$  块的类别.

情况 2: 关键点个数为 1.

若第  $j$  块只有 1 个关键点, 则直接使用该关键点的特征作为第  $j$  块的特征, 即

$$B_j = P_j. \quad (5)$$

式中:  $P_j$  为第  $j'$  个关键点的 SIFT 特征,  $P_j \in \mathbf{R}^T$ ,  $T$  为 SIFT 特征的维度, 本文采用 8 个方向和统计 16 个小块的梯度方向直方图, 因此  $T = 128$ .

情况 3: 关键点个数大于 1.

若第  $j$  块大于 1 个关键点, 则取特征最明显的关键点, 本文使用了 1 范数, SIFT 特征的 1 范数越大表示方向越多或越明显, 即

$$B_j = \max_{P_j} \| P_j \|_1. \quad (6)$$

情况 2、3 的块将用于第 3) 步的聚类算法.

3) k-means 聚类. 假设聚类总数为  $K$ , 若按照 SIFT 特征方向进行聚类, 则有  $K = 8$ , 具体流程如下:

S1: 首先初始化  $K$  个聚类中心, 分别为  $\{\mu_1^0, \mu_2^0, \dots, \mu_K^0\}$ , 并设置迭代次数  $it = 0$  和最大迭代次数  $T_{\max} = 100$ .

S2: 使用欧氏距离作为特征向量间距离的度量标准, 并对所有特征向量进行初始划分

$$U_j^{it} = \min_k \| B_j - \mu_k^{it} \|_2, \quad (7)$$

$$C_k^{it} = C_k^{it-1} \cup \{B_j\}. \quad (8)$$

式中:  $k = \{1, 2, \dots, K\}$  为类别标签,  $\mu_k^{it}$  为第  $k$  类的聚类中心在第  $it$  次迭代的结果,  $U_j^{it}$  为第  $j$  块在第  $it$  次迭代的类别,  $C_k^{it}$  为第  $k$  类的特征向量在第  $it$  次迭代的集合.

S3:  $it + 1 \rightarrow it$ , 更新每个类别样本的聚类中心为

$$\mu_k^{it} = \frac{1}{|C_k^{it-1}|} \sum_{B_j \in C_k^{it-1}} B_j. \quad (9)$$

如果  $*$  为标量, 则  $\| * \|$  为求绝对值; 若  $*$  为向量, 则  $\| * \|$  为求向量元素个数.

S4: 如果满足式 (10) 或者  $it \geq T_{\max}$ , 则聚类结束, 否则转到 S2 继续执行.

$$\| \mu_k^{it} - \mu_k^{it-1} \|_2 < \zeta, \quad (10)$$

式中  $\zeta$  为误差允许范围, 取  $\zeta = 10^{-6}$ .

4) 子字典学习和字典合成

采用一种标准无监督字典学习方法 K 奇异值分解算法<sup>[16]</sup> (K-Singular Value Decomposition, K-SVD) 对  $K + 1$  (包括无关键点的块) 个类别的块分别进行字典学习. 学习之前, 先将每个块按照行优先顺序展开成列向量的形式. 通过 K-SVD 算法学习之后分别得到子字典  $D_0, D_1, \dots, D_K$ , 最终合成一个完整的字典  $D = \{D_0, D_1, \dots, D_K\}$ , 字典大小设置为 256, 字典学习过程见图 2.

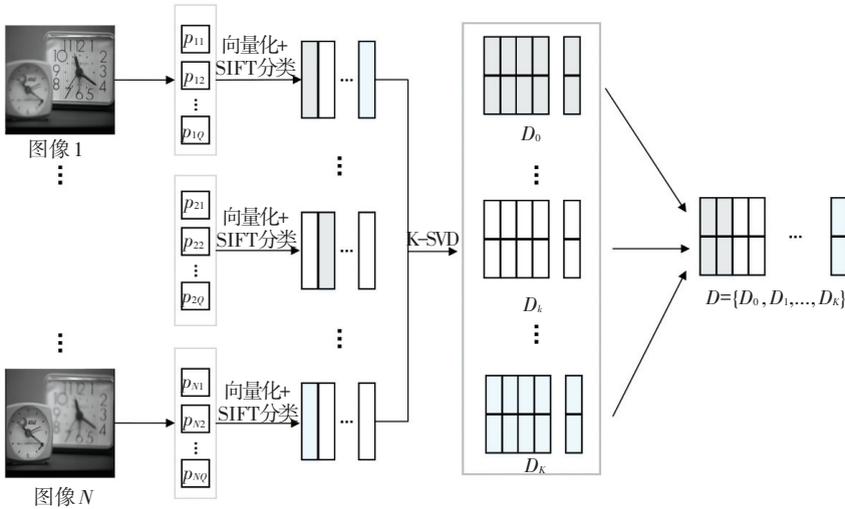


图 2 基于 SIFT 的字典学习过程

Fig.2 SIFT-based dictionary learning process

## 2.2 引导滤波低秩系数融合

字典学习完毕之后, 对单张源图像  $I_i$  进行低秩表示. 即将求取的  $I_i$  的分块向量化, 得到对应的  $V_i$ , 然后结合式 (1) 得到如下优化公式

$$\min(\| W_i \|_* + \lambda \| \tilde{E}_i \|_{2,1}), \text{ s.t. } V_i = D W_i + \tilde{E}_i. \quad (11)$$

式中:  $i = \{1, 2, \dots, N\}$ ,  $W_i$  为第  $i$  幅图像的低秩表示

系数,  $\tilde{E}_i$  为第  $i$  幅图像的误差矩阵, 若误差较小则  $\lambda$  取较大值, 否则取较小值<sup>[12]</sup>, 取  $\lambda = 4.5$ . 使用拉格朗日乘子法即可求解式 (11), 得到每幅源图像的低秩表示, 然后对其使用引导滤波相关的融合规则, 具体流程如下:

S1: 构建输入图像. 若  $W_i$  的第  $q$  列的 1 范数沿着  $z$  轴方向取得最大值, 则该列置 1, 否则置 0, 公式为

$$G_i(:,q) = \begin{cases} 1, & \text{s.t. } i = \max_{i'} \{ \| W_{i'}(:,q) \|_1 \}; \\ 0, & \text{其他.} \end{cases} \quad (12)$$

式中  $G_i$  为输入图像,  $G_i(:,q)$  为  $G_i$  的第  $q$  列.

S2: 进行引导滤波. 以  $G_i$  作为输入图像, 对应的  $W_i$  作为引导图像进行引导滤波, 得到引导滤波后的结果  $r_i$

$$r_i = f_{s_\omega, \varepsilon}(G_i, W_i). \quad (13)$$

式中:  $f$  为引导滤波操作,  $s_\omega, \varepsilon$  是  $f$  的输入参数,  $s_\omega$  是引导滤波的窗口大小, 即式(3)中的  $\omega_i$  的大小,  $\varepsilon = 0.01$  对应式(4)中的衰减系数.

S3: 计算窗口自适应条件.  $s_\omega$  会根据内容 (特征内容和非特征内容) 进行自适应变化, 规则如下: 沿着  $z$  轴方向求各列的熵 (本文使用列的 2 范数), 然后设置阈值, 大于阈值者为非特征内容, 引导滤波时使用较大窗口; 否则为特征内容, 引导滤波时使用较小窗口. 熵计算公式为

$$\varphi_i(:,q) = \| W_i(:,q) \|_2 / \sum_{j=1}^{j=N} \| W_j(:,q) \|_2, \quad (14)$$

$$E_n(p,q) = \sum_{i=1}^{i=N} -\varphi_i(p,q) \log \varphi_i(p,q). \quad (15)$$

式中:  $\varphi$  为概率, 是直接第  $q$  列的低秩表示系数的

2 范数沿  $z$  轴方向归一化得到,  $\| * \|_2$  为 2 范数,  $p, q$  为低秩系数的坐标,  $E_n$  为熵. 将  $E_n$  除以最大值进行归一化:

$$E_{\max} = \max(E_n), \quad (16)$$

$$E_n(p,q) = E_n(p,q) / E_{\max}, \quad (17)$$

式中  $E_{\max}$  为  $E_n$  的最大值.  $s_\omega$  将满足如下等式

$$s_\omega(p,q) = \begin{cases} s_b, & \text{s.t. } E_n(p,q) > \tau; \\ s_s, & \text{s.t. } E_n(p,q) \leq \tau. \end{cases} \quad (18)$$

式中:  $\tau$  为阈值, 取  $\tau = 0.95$ ,  $s_s, s_b$  为两个正整数, 分别为特征内容和非特征内容区域引导滤波的窗口大小, 并满足  $s_s < s_b$ , 取  $s_s = 3, s_b = 7$ .

S4: 求取融合系数. 使用  $r_i$  作为权重对  $W_i$  进行加权, 得到最终的融合低秩表示系数  $W_f$

$$W_f(p,q) = \sum_{i=1}^{i=N} r_i(p,q) W_i(p,q). \quad (19)$$

S5: 获取融合图. 结合式(1), 可求出融合图的块向量展开形式  $V_f$ :

$$V_f = D W_f. \quad (20)$$

将  $V_f$  还原成图像的形式, 重叠部分取平均值即可, 得到融合图  $F$ . 图像融合的整体流程图见图 3.

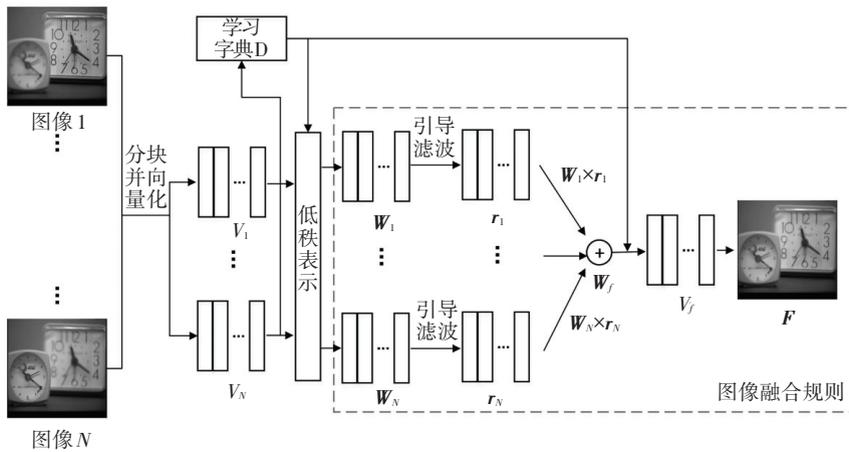


图 3 整体图像融合框架

Fig.3 The framework of proposed algorithm

### 3 实验及分析

实验使用 Genie Nano 系列的 200 万像素的黑白相机, 分辨率  $5.5 \mu\text{m}$ , 同轴光照明. Mitutoyo 的物镜和目镜, 物镜放大倍率为 20X, 景深  $1.8 \mu\text{m}$ . 日本 Oriental Motor 公司的 PMM33BH2 步进电机, 每次步进  $1 \mu\text{m}$ . 连续拍摄 100 帧图像, 图像大小为  $1472 \times 1472$  像素, 实验装置见图 4. 实验运行环境为 MATLAB R2016a, 2.4 GHz, 16 GB 内存.

#### 3.1 实验数据和相关设置

为验证提出算法的有效性, 使用 6 组实验数据,

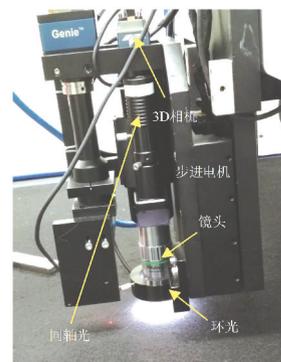


图 4 实验装置

Fig.4 Experimental device

包括 3 组广泛用于研究的多聚焦图像(2 帧/组, 见图 5)和 3 组使用图 4 装置实际拍摄的图像(100 帧/组, 见图 6). 为说明本文算法的有效性, 将其与以下算法进行对比分析: 基于引导滤波的图像融合<sup>[4]</sup>(Image Fusion based on Guided Filtering, IF-GF)、基于离散小波变换的图像融合<sup>[7]</sup>(Image Fusion based on Discrete Wavelet Transform, IF-DWT)、基于轮廓波变换的图像融合<sup>[8]</sup>(Image Fusion based on Non-Subsampled Contourlet Transform, IF-NSCT)、基于稀疏表示的图像融合<sup>[10]</sup>(Image Fusion based on Sparse Representation, IF-SR).

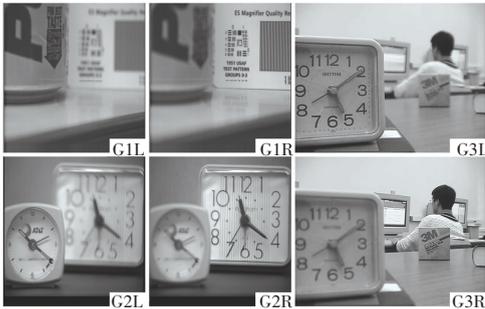


图 5 广泛应用于研究的多聚焦图像数据

Fig.5 Examples of widely used multi-focus images

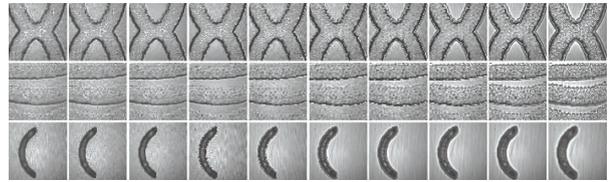


图 6 实际拍摄的多聚焦图像数据

Fig.6 Examples of real-world multi-focus images

### 3.2 参数设置

为给块大小  $s$  和步长  $t$  设置一个合理的值, 本节随机选取 1 组图像进行仿真实验分析, 为方便计算, 参数都取 2 的幂次方. 表 1 显示了子字典的块数分布, 0~8 分别代表 9 个子字典. 块越小越能学习更多的细节, 但是块太小不能有效提取 SIFT 特征, 从表 1 中可看出, 取  $s=4, 8$  基本不能提取块的 SIFT 特征, 当  $s=16, t=1, 2$  块数分布已经基本合理(如黑体字), 而  $s=16, t=4$  块数太少, 不符合学习条件. 若块数过于冗余, 会导致字典求解过程中冗余信息反复迭代, 使图像有用信息不能完全表达, 表 2 显示了  $s=16, t=1, 2$  重构子字典的误差分布,  $t=2$  误差更小(如黑体字). 综上,  $s=16, t=2$  为较为合理的选择.

表 1 分块大小对子字典学习的块数分布比较

Tab.1 The comparison of block size on the distribution of sub-dictionary learning blocks

分块大小/pixel	步长/pixel	子字典块数/块								
		0	1	2	3	4	5	6	7	8
4	1	518 162	0	0	0	0	0	0	0	0
	2	130 050	0	0	0	0	0	0	0	0
	4	32 768	0	0	0	0	0	0	0	0
8	1	506 526	189	214	1 447	44	129	238	745	518
	2	126 859	188	196	490	56	43	53	77	56
	4	31 957	19	69	24	14	9	142	17	7
16	1	<b>412 710</b>	<b>10 253</b>	<b>14 148</b>	<b>9 373</b>	<b>8 329</b>	<b>11 535</b>	<b>9 877</b>	<b>9 948</b>	<b>7 845</b>
	2	<b>104 917</b>	<b>3 032</b>	<b>2 335</b>	<b>3 354</b>	<b>1 946</b>	<b>1 796</b>	<b>2 297</b>	<b>2 311</b>	<b>2 014</b>
	4	26 424	699	409	440	515	690	896	796	381

表 2 步长对子字典学习的残差分布比较

Tab.2 The comparison of step size on the distribution of sub-dictionary learning residual

步长/pixel	子字典块数/块								
	0	1	2	3	4	5	6	7	8
1	1.200 9	1.322 0	2.215 2	2.085 0	1.518 6	1.626 2	2.288 8	1.132 6	1.635 0
2	<b>0.977 3</b>	<b>0.964 3</b>	<b>0.246 6</b>	<b>0.422 8</b>	<b>0.444 4</b>	<b>0.878 5</b>	<b>0.517 0</b>	<b>0.811 6</b>	<b>0.053 3</b>

### 3.3 融合图像质量评价

为更加直观的说明实验的效果, 使用 5 类质量评价函数<sup>[17]</sup>(以 2 幅源图像为例, 可推广到多幅), 每类评价函数都是值越大, 表示融合质量越好.

1) 归一化的互信息 (Normalized Mutual Information,  $Q_{NMI}$ ) 来源于信息论, 是测量源图像的信息在融合图像中的保留程度, 式(21)使用其改进形式

$$Q_{\text{NMI}} = 2 \left[ \frac{MI(S_1, F)}{H(S_1) + H(F)} + \frac{MI(S_2, F)}{H(S_2) + H(F)} \right]. \quad (21)$$

式中:  $H(S_1)$ 、 $H(S_2)$ 、 $H(F)$  分别为源图像  $S_1$ 、 $S_2$  和融合图像  $F$  的信息熵,  $MI(S_i, F)$  为  $S_i$  和  $F$  间的互信息, 其计算公式如下:

$$MI(S_i, F) = H(S_i) + H(F) - H(S_i, F), \quad (22)$$

式中  $i = \{1, 2\}$ ,  $H(S_i, F)$  为  $S_i$  和  $F$  的联合熵.

2) 梯度相关性 (Gradient-based Fusion Metric,  $Q_G$ ) 是度量图像间边缘信息的相关性为

$$Q_G = \frac{\sum_{x=1}^H \sum_{y=1}^W [Q_{xy}^{S_1 F} \omega_{xy}^{S_1} + Q_{xy}^{S_2 F} \omega_{xy}^{S_2}]}{\sum_{x=1}^H \sum_{y=1}^W [\omega_{xy}^{S_1} + \omega_{xy}^{S_2}]}, \quad (23)$$

式中:  $x, y$  为图像中的坐标,  $Q^{S_i F}$ ,  $i = \{1, 2\}$  为边缘信息保留值,  $\omega^{S_i}$  为对应的权重, 反映了  $Q^{S_i F}$  的重要程度.

3) 相位一致性 (Phase Congruency,  $Q_P$ ) 定义了和矩相关的度量, 因为矩包含角点和边缘的信息, 定义如下:

$$Q_P = (P_p)^\alpha (P_M)^\beta (P_m)^\gamma. \quad (24)$$

式中:  $P_*$ ,  $\{ * = p, M, m \}$  和相关系数有关, 下标  $p, M, m$  分别为相位一致、最大和最小矩,  $\alpha, \beta, \gamma$  分别为 3 个部分的重要程度.

4) 结构相似度 (Structural Similarity,  $Q_S$ ) 的测量能够近似的反映出图像的形变情况, 公式如下:

$$Q_S = \frac{1}{|\dot{W}|} \sum_{w \in \dot{W}} [\lambda_w Q(S_1, F | w) + (1 - \lambda_w) Q(S_2, F | w)]. \quad (25)$$

式中:  $\dot{W}$  为所有局部窗口的集合,  $\lambda_w$  为局部窗口显著度的比值,  $Q(S_i, F | w)$ ,  $i = \{1, 2\}$  为通用质量指数 (Universal Image Quality Index, UIQI).

5) Chen-Blum 度量法 (Chen-Blum Metrics,  $Q_{\text{CB}}$ ) 计算对比度特征在融合图像中的保留程度, 公式如下:

$$Q_{\text{CB}} = \frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W [\lambda_{xy}^{S_1} Q_{xy}^{S_1 F} + \lambda_{xy}^{S_2} Q_{xy}^{S_2 F}], \quad (26)$$

式中:  $x, y$  为图像中的下标,  $\lambda^{S_i}$ ,  $i = \{1, 2\}$  为源图像  $S_i$  的显著特征,  $Q^{S_i F}$  为信息保留值.

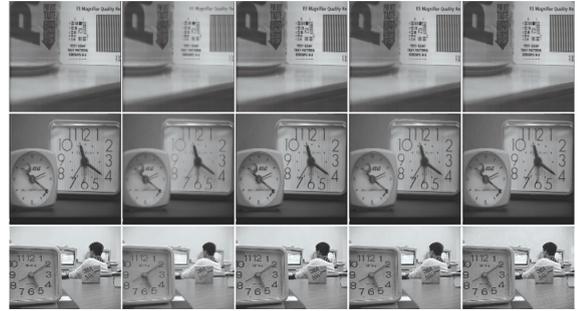
### 3.4 融合结果分析

将实验结果分成 2 部分进行分析: 广泛用于研究的多聚焦图像和实际拍摄的多聚焦图像. 前者验证提出算法的有效性, 后者说明提出算法的泛化能力及有效性.

#### 3.4.1 广泛用于研究的多聚焦图像分析

图 5 中每组图都是左幅图左部分清晰, 右幅图右部分清晰, 其中第 3 组为 2 幅未完全配准的图像.

图 7 分别显示了 IF-GF、IF-DWT、IF-NSCT、IF-SR 和本文算法对 3 组图像的融合结果. 为便于分析, 图 8 显示了对应的部分细节放大图 (图 8 的列对应图 7 的行). 从放大的细节图中可看出 IF-GF 的空间连续性较好, 没有明显的块效应, 但是细节部分较其他变化域算法更为模糊, 由 A1 的“倒影字母”可明显看出. IF-DWT 整体较差, 融合图基本不能反映出源图像的有效信息. IF-NSCT 和本文算法对未配准图像鲁棒性较高, 从 A13、A15 的头发边缘可明显看出, 而 IF-GF、IF-DWT、IF-SR 都有明显的伪影产生, 对应矩形框所示. IF-SR 因为不具备全局性, 所以边缘部分产生了块效应, A9 较明显. 综上, IF-NSCT 和本文算法在细节保留和克服未配准问题都取得了较好的结果, 但是 IF-NSCT 在空间连续性上比较欠缺, 如 A13 矩形框所示.



(a) IF-GF (b) IF-DWT (c) IF-NSCT (d) IF-SR (e) 本文算法

图 7 广泛应用于研究的多聚焦图像融合结果

Fig.7 Widely used multi-focus image fusion results

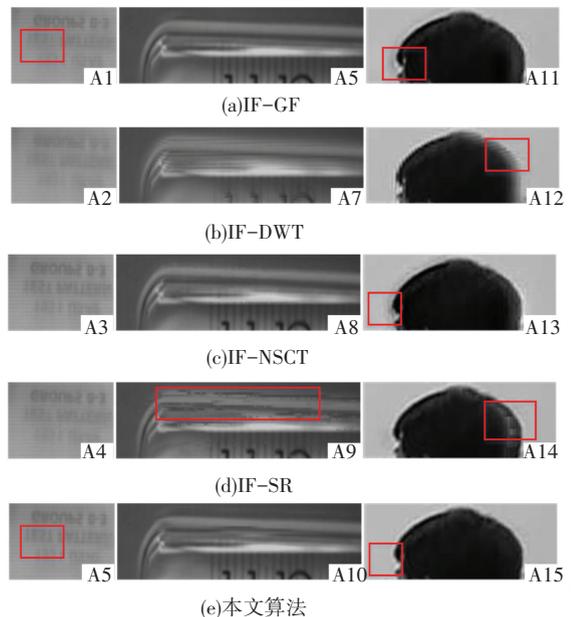


图 8 部分细节放大图 A

Fig.8 Magnified partial details A

表 3 用具体的数据进行客观质量评价, 利用 5 个融合质量评价函数, 其中加粗的数字为所有算法中的

最大值. 表中的数据可看出, 本文算法除了第 2 组的  $Q_S$  不是最大值, 其他都保持领先, 说明该算法在原始信息保留、细节保留、结构保持、对比度保持等方面都比其他主流融合算法更好. 此外, IF-GF 在  $Q_{NMI}$  和  $Q_S$  指标下取得较好结果, 说明基于空域的算法在空间连续性和结构保持方面更占优势. 而 IF-NSCT 在  $Q_G$ 、 $Q_{CB}$  指标下的值更高, 说明细节保持能力更强. 因此, 实验结果与理论分析基本达成一致.

表 3 不同融合算法的融合质量比较 A

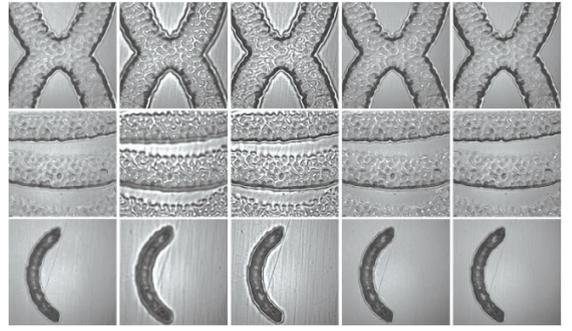
Tab.3 Comparison of fusion quality of 5 fusion algorithms A

指标	IF-GF	IF-DWT	IF-NSCT	IF-SR	本文算法
$Q_{NMI}$	0.99	0.90	0.95	0.96	<b>1.04</b>
$Q_G$	0.65	0.64	0.74	0.66	<b>0.76</b>
$Q_P$	0.74	0.41	0.83	0.63	<b>0.84</b>
$Q_S$	<b>0.97</b>	0.90	0.91	0.95	<b>0.97</b>
$Q_{CB}$	0.60	0.52	0.61	0.63	<b>0.70</b>
$Q_{NMI}$	1.13	0.91	0.96	1.05	<b>1.15</b>
$Q_G$	0.59	0.50	0.65	0.58	<b>0.66</b>
$Q_P$	0.68	0.37	0.73	0.52	<b>0.78</b>
$Q_S$	<b>0.96</b>	0.91	0.95	0.95	0.95
$Q_{CB}$	0.67	0.54	0.70	0.71	<b>0.73</b>
$Q_{NMI}$	1.05	0.93	1.01	1.01	<b>1.13</b>
$Q_G$	0.65	0.51	0.66	0.56	<b>0.71</b>
$Q_P$	0.70	0.42	0.72	0.59	<b>0.78</b>
$Q_S$	<b>0.96</b>	0.91	<b>0.96</b>	0.95	<b>0.96</b>
$Q_{CB}$	0.62	0.53	0.64	0.68	<b>0.69</b>

### 3.4.2 实际拍摄图像分析

实际拍摄图像的每 1 组都是由凹坑部分(前景)和平面部分(背景)组成, 由于拍摄是从上至下进行的, 因此都呈现平面部分先聚焦, 凹坑后聚焦的状态.

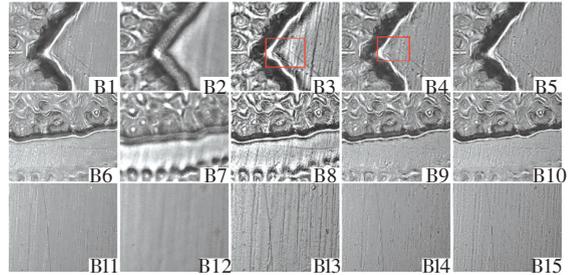
图 9 显示了 3 组图像使用不同融合算法的融合结果图, 图 10 则显示了对应的部分细节放大图(分别选择边缘、凹坑和平面). 从整体融合图中可发现 IF-DWT 算法存在明显模糊, 同广泛应用于研究的图像结果达成一致. 图 10 的 B1~B5 选择边缘进行分析, 其中 B1 整体比较平滑, 但是细节存在丢失的现象, 从边缘近平面的纹理较明显看出. B3 整体对比度非常强烈, 但是由边缘旁边的“光晕”可看出(矩形框所示) IF-NSCT 算法的结果并不符合人眼的主观视觉感受. B4~B5 则表现较好, 但是 B4 仍存在局部块效应(矩形框所示). B6~B10、B11~B15 分别选择凹坑细节部分和平面部分进行分析, 结果与 B1~B5 保持一致, 需要补充的是 B10、B15 较 B6、B11 分别在凹坑和平面上的细节更加丰富, 且都保持着较高的空间连续性.



(a) IF-GF (b) IF-DWT (c) IF-NSCT (d) IF-SR (e) 本文算法

图 9 实际拍摄多聚焦图像融合结果

Fig.9 Real-world multi-focus image fusion results



(a) IF-GF (b) IF-DWT (c) IF-NSCT (d) IF-SR (e) 本文算法

图 10 部分细节放大图 B

Fig.10 Magnified partial details B

表 4 对 5 种不同的融合算法在 5 个质量评价指标下的结果进行比较分析, 其中加粗的数字为所有算法中的最大值. 本文算法在大多数情况下都取得了最优的结果, 其他算法的趋势也与广泛应用于研究的图像结果高度一致, 验证了本文提出算法在细节保留能力和空间连续性方面的有效性以及对多幅实际拍摄图像具有较好的泛化能力.

表 4 不同融合算法的融合质量比较 B

Tab.4 Comparison of fusion quality of 5 fusion algorithms B

指标	IF-GF	IF-DWT	IF-NSCT	IF-SR	本文算法
$Q_{NMI}$	47.78	42.51	44.14	47.53	<b>52.98</b>
$Q_G$	0.25	0.20	0.30	0.25	<b>0.35</b>
$Q_P$	0.39	0.28	0.51	0.41	<b>0.60</b>
$Q_S$	0.50	0.34	0.43	0.50	<b>0.55</b>
$Q_{CB}$	0.34	0.30	0.36	0.35	<b>0.37</b>
$Q_{NMI}$	43.71	36.99	40.73	42.89	<b>46.75</b>
$Q_G$	0.23	0.18	0.28	0.24	<b>0.29</b>
$Q_P$	0.39	0.24	0.50	0.44	<b>0.59</b>
$Q_S$	<b>0.49</b>	0.26	0.42	0.47	<b>0.49</b>
$Q_{CB}$	0.34	0.33	0.36	0.36	<b>0.37</b>
$Q_{NMI}$	<b>57.31</b>	42.36	46.77	51.58	56.22
$Q_G$	0.22	0.15	0.28	0.24	<b>0.33</b>
$Q_P$	0.40	0.27	0.51	0.49	<b>0.58</b>
$Q_S$	0.46	0.35	0.45	0.45	<b>0.48</b>
$Q_{CB}$	0.39	0.19	0.44	0.40	<b>0.45</b>

## 4 结 论

本文提出了一种基于 SIFT 字典学习的引导滤波图像融合算法. 该算法使用低秩表示, 使融合图像具有全局性, 并且通过 SIFT 进行分类学习字典, 克服了低秩表示局部细节保持能力弱和对未配准源图像鲁棒性差的缺点. 另外, 通过加入自适应窗口的引导滤波, 增强了融合图像的空间连续性. 通过选取 3 组广泛应用于研究的多聚焦图像和 3 组实际拍摄的多聚焦图像, 使用 5 类质量评价指标验证了提出算法在细节保留、空间连续性和克服未配准问题方面的有效性.

## 参考文献

- [1] GHASSEMIAN H. A review of remote sensing image fusion methods [J]. *Information Fusion*, 2016, 32(PA):75. DOI: 10.1016/j.inffus.2016.03.003
- [2] EL-GAMAL E Z A, ELMOGY M, ATWAN A. Current trends in medical image registration and fusion[J]. *Egyptian Informatics Journal*, 2016, 17(1):99. DOI: 10.1016/j.eij. 2015.09.002
- [3] LI S, KANG X, HU J, et al. Image matting for fusion of multi-focus images in dynamic scenes[J]. *Information Fusion*, 2013, 14(2): 147. DOI: 10.1016/j.inffus.2011.07.001
- [4] LI S, KANG X, HU J. Image fusion with guided filtering[J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 2013, 22(7):2864. DOI: 10.1109/TIP.2013. 2244222
- [5] XU X, SHAN D, WANG G, et al. Multimodal medical image fusion using PCNN optimized by the QPSO algorithm [J]. *Applied Soft Computing*, 2016, 46:588. DOI: 10.1016/ j.asoc.2016.03.028
- [6] WANG W, CHANG F. A multi-focus image fusion method based on laplacian pyramid[J]. *Journal of Computers*, 2011, 6(12):2559. DOI: 10.4304/jcp.6.12.2559-2566
- [7] PAJARES G, CRUZ J M D L. A wavelet-based image fusion tutorial [J]. *Pattern Recognition*, 2004, 37(9):1855. DOI: 10.1016/ S0031-3203(04)00103-7
- [8] ZHANG Q, GUO B L. Multifocus image fusion using the nonsubsam-pled contourlet transform [J]. *Signal Processing*, 2009, 89(7): 1334. DOI: 10.1016/j.sigpro.2009.01.012
- [9] YIN H, LI S, FANG L. Simultaneous image fusion and super-reso- lution using sparse representation[J]. *Information Fusion*, 2013, 14 (3):229. DOI: 10.1016/j.inffus.2012.01. 008
- [10] NEJATI M, SAMAVI S, SHIRANI S. Multi-focus image fusion u- sing dictionary-based sparse representation[J]. *Information Fusion*, 2015, 25:72. DOI: 10.1016/j.inffus.2014. 10.004
- [11] LIU Y, CHEN X, WARD R K, et al. Image fusion with convolu- tional sparse representation [J]. *IEEE Signal Processing Letters*, 2016, 23(12):1882. DOI: 10.1109/ LSP.2016.2618776
- [12] LIU G, LIN Z, YAN S, et al. Robust recovery of subspace struc- tures by low-rank representation.[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2010, 35(1):171. DOI: 10. 1109/TPAMI.2012.88
- [13] HE K, SUN J, TANG X. Guided image filtering[C]// *European Conference on Computer Vision*. Springer-Verlag, 2010;1. DOI: 10. 1109/TPAMI.2012.213
- [14] LIN Z, CHEN M, MA Y. The augmented lagrange multiplier meth- od for exact recovery of corrupted low-rank matrices[J]. *Eprint Arxiv*, 2010, 9. DOI: 10.1016/j.jsb.2012.10. 010
- [15] LOWE D G. Distinctive image features from scale-invariant key- points[J]. *International Journal of Computer Vision*, 2004, 60(2): 91. DOI: 10.1023/B:VISI. 0000029664.99615.94
- [16] AHARON M, ELAD M, BRUCKSTEIN A. -SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311. DOI: 10.1109/TSP.2006.881199
- [17] LIU Z, BLASCH E, XUE Z, et al. Objective assessment of multi- resolution image fusion algorithms for context enhancement in night vision; a comparative study[J]. *IEEE Transactions on Pattern Anal- ysis & Machine Intelligence*, 2011, 34(1):94. DOI: 10.1109/ TPAMI.2011.109

(编辑 苗秀芝)