

基于联合图像块聚类自适应字典学习的多模态医学图像融合

王丽芳 史超宇 蔺素珍 秦品乐 高 媛

(中北大学山西省生物医学成像与影像大数据重点实验室 太原 030051)

摘 要 针对多模态医学图像融合中过完备自适应字典存在的大量冗余信息会导致图像重建质量不佳的问题,文中提出了基于联合图像块聚类自适应字典学习的多模态医学图像融合方法(JCPD)。该方法首先计算图像块的欧氏距离,通过比较设定的阈值和图像块的最小距离来剔除冗余图像块,减少冗余图像块的数量。然后,使用局部调制核回归(SKR)提取图像块的局部梯度信息作为聚类中心,将具有相同局部梯度信息的两种模态的图像块进行联合图像块聚类。在联合图像块聚类的基础上使用改进的 K-SVD 算法对图像块聚类形成的类簇进行训练得到子字典,并将子字典合并成自适应字典。最后,在自适应字典的作用下用正交匹配追踪算法(OMP)计算得到稀疏表示系数,再使用“2 范数最大”的规则融合稀疏系数,之后通过重建得到融合图像。实验表明,与 2 种基于多尺度变换的方法和 6 种基于稀疏表示的方法相比,所提方法在保证字典信息的完整性和字典的紧凑性基础上使得融合的图像清晰度更高、对比度更强,便于临床诊断和辅助治疗。

关键词 多模态,医学图像融合,稀疏表示,图像块聚类,自适应字典学习

中图法分类号 TP391.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.07.036

Multi-modal Medical Image Fusion Based on Joint Patch Clustering of Adaptive Dictionary Learning

WANG Li-fang SHI Chao-yu LIN Su-zhen QIN Pin-le GAO Yuan

(The Key Laboratory of Biomedical Imaging and Imaging on Big Data, North University of China, Taiyuan 030051, China)

Abstract In view of the poor image reconstruction quality problem caused by a large amount of redundant information existing in over complete adaptive dictionaries in medical image fusion, this paper proposed a multi-modal medical image fusion method based on joint image patch clustering and adaptive dictionary learning. First, this method calculates the Euclidean distance of image patches and reduces redundant image patches by comparing the cut-off threshold and the minimum distance of image patches. Then, it extracts the local gradient information of image patches as the clustering center by local regression weight of steering kernel (SKR), and combines the two different modal image patches with the same local gradient information for image patch clustering. On the basis of joint image patch clustering, it uses the improved K-SVD algorithm to train the clusters formed by image patch clustering to get sub-dictionaries, and merges the sub-dictionaries into an adaptive dictionary. Finally, the sparse representation coefficients can be obtained by the orthogonal matching tracking algorithm (OMP) and the adaptive dictionary, and they are fused with the rule of “2-norm max”. Through the reconstruction, this paper obtained the fused image. Compared with two methods based on multi-scale transform and six methods based on sparse representation, experimental results show that the proposed method can construct a compact and informative dictionary, and endow the fused image with higher clarity and stronger contrast to facilitate clinical diagnosis and adjuvant treatment.

Keywords Multi-model, Medical image fusion, Spares representation, Image patch clustering, Adaptive dictionary learning

1 引言

在临床医学上,单模态医学图像所提供的信息有限,不能

反映相关组织的所有细节信息,可能会造成临床误诊,将不同模态的医学图像进行适当融合,可以使源图像的信息互补,从而获得信息更为丰富的图像^[1]。医学图像融合是图像融合研

收到日期:2018-06-04 返修日期:2018-09-30 本文受山西省青年基金项目(201601D021080),中北大学研究生科技立项自然科学基金项目(20171441)资助。

王丽芳(1977—),女,博士,副教授,CCF 会员,主要研究方向为机器视觉、大数据处理、医学图像处理,E-mail:wsm2004@nuc.edu.cn(通信作者);史超宇(1993—),男,硕士生,主要研究方向为医学图像融合、机器学习;蔺素珍(1966—),女,博士,教授,主要研究方向为图像处理、文物虚拟修复;秦品乐(1978—),男,博士,副教授,主要研究方向为机器视觉、大数据处理、三维重建;高媛(1972—),女,硕士,副教授,主要研究方向为机器视觉、大数据处理、三维重建。

究的重要方向之一,能为临床诊断提供更为可靠的依据。

目前,多模态图像融合方法分为两大类:基于变换域的融合方法和基于空间域的融合方法^[2]。基于空间域的多模态图像融合方法往往会导致融合图像的细节与对比度等重要信息丢失,无法得到理想结果^[2]。基于变换域的多模态图像融合方法在捕捉图像的边缘、轮廓等重要方向和结构信息方面具有优势,典型的变换域方法有基于多尺度变换的方法、基于稀疏表示的方法、基于金字塔变换的方法等^[3]。基于多尺度变换的方法有离散小波变换(Discrete Wavelet Transform, DWT)^[4]、平稳小波变换(Stationary Wavelet Transform, SWT)^[5]、双树复小波变换(Dual Tree Complex Wavelet Transform, DTCWT)^[6]、拉普拉斯金字塔(Laplacian Pyramid, LP)^[7]、非下采样轮廓波变换(Non-Subsampled Contourlet Transform, NSCT)^[8]。虽然,基于多尺度变换的分解方法可以很好地提取图像的显著特征,且计算效率高,但是多尺度分解水平不易确定,传统的融合规则也会导致融合结果不理想。在基于稀疏表示方法的过完备字典中,解析字典具有快速实现的优点,但这种字典仅限于特定类型的信号表示,并且在选择解析基时需要先验知识。相比解析字典,学习字典包含更丰富的图像特征信息,表示能力较强但时间效率较低。学习字典按照所使用的训练图像的差异可以进一步分为全局训练字典和自适应字典。全局训练字典使用的训练数据来自一些公开的图像数据库,是由外部预先收集的图像组成,在实际应用中,收集合适的图像数据集并不总是可行的。此外,不同类型图像的图像内容千差万别,全局训练字典对不同类型的图像表现出来的性能差异较大,对医学图像的自适应性不强。Yin 等^[9]提出基于稀疏表示的多聚焦图像融合方法,构建过完备自适应字典,此方法不需要先验知识也不需要外部预先收集的训练图像数据,但冗余的过完备自适应字典会影响图像的融合结果,导致重构结果不稳定,对医学图像会产生伪影。Kim 等^[10]提出基于联合图像块聚类字典学习的多模态图像融合方法,通过对不同模态的图像进行聚类训练从而到字典,提高字典学习的效率,但该方法将整个图像进行字典学习,得到的冗余字典在图像重建时会导致块效应。Zhu 等^[11]提出基于密度峰值聚类的字典学习方法用于图像块融合,通过预处理保留细节信息丰富的图像块用于字典学习,减少字典的冗余信息,但使用密度峰值聚类对图像块训练占用大量时间,导致字典的学习效率低。

针对上述问题,本文提出了基于联合图像块聚类自适应字典学习的多模态医学图像融合方法(Joint patch clustering based adaptive dictionary learning for multimodality medical image fusion, JCPD)。首先,通过预处理(即计算图像块的欧氏距离)来选择细节信息丰富的图像块,并比较设定的阈值和图像块的最小距离来剔除冗余图像块,减少冗余图像块的数量。其次,为了保证训练得到的字典信息的完整性和字典的紧凑性,本文直接从待融合的源图像块中使用局部调制核回归(Steering Kernel Regression, SKR)^[12]提取图像块的局部梯度信息作为聚类中心,将具有相同局部梯度信息的两种模态的图像块进行联合图像块聚类,并使用改进的 K-SVD^[13]算法对每个图像块类簇进行训练得到子字典,将子字典合并

成自适应字典。最后根据训练的自适应字典将稀疏系数融合,通过重构得到融合图像。

2 相关理论

稀疏表示理论的主要思想是图像信号 $y \in R^n$ 可以近似表示为一个过完备字典 $\Phi \in R^{n \times m}$ ($n < m$) 中“少量”原子的线性组合, n 和 m 分别表示信号的维数和字典的大小^[14]。也就是说,信号 y 可以表示为 $y \approx \Phi \alpha$, $\alpha \in R^m$ 是未知的稀疏表示系数。稀疏系数 α 可以由式(1)计算:

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s. t.} \quad \|y - \Phi \alpha\|_2 \leq \epsilon \quad (1)$$

其中, $\|\alpha\|_0$ 表示向量 α 中的非零个数, $\epsilon \geq 0$ 表示允许偏差的精度。式(1)的求解过程称为“稀疏编码”。

联合稀疏模型(Joint Sparsity Model, JSM)^[15]通过对多模态医学图像信息进行联合分解,可以充分保持多模态医学图像信息之间的相关性。在联合稀疏模型中,对同一病灶部位不同模态的信号可以组成一个信号整体。通过对这个信号整体进行稀疏表示,可以得到一个所有信号具有相同的共同稀疏部分和各信号独有的稀疏部分。假设信号整体 $\Gamma = \{x_1, x_2, \dots, x_k\}$, Γ 中的每个信号可以表示为:

$$x_i = D\alpha_c + D\alpha_i, i = 1, 2, \dots, k \quad \text{s. t.} \quad x_i \in R^n \quad (2)$$

其中, $D \in R^{n \times m}$ 表示字典, $\alpha_c \in R^m$ 为信号公共部分, $\alpha_i \in R^m$ 表示第 i 个信号的独特部分。信号整体 Γ 中的信号可以表示为:

$$X = (x_1^T, x_2^T, \dots, x_k^T)^T \in R^{Kn} \quad (3)$$

联合稀疏分解表达式(2)可以表示为:

$$X = \tilde{D}\Theta \quad (4)$$

其中:

$$\tilde{D} = \begin{pmatrix} D & D & 0 & \dots & 0 \\ D & 0 & D & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D & 0 & 0 & \dots & D \end{pmatrix} \in R^{Kn \times (K+1)m} \quad (5)$$

$$\Theta = (\alpha_c^T, \alpha_1^T, \alpha_2^T, \dots, \alpha_k^T)^T \in R^{(K+1)m} \quad (6)$$

$0 \in R^{n \times m}$ 表示零矩阵。结合稀疏表示理论,联合稀疏表示模型(4)中的稀疏系数的优化问题表示为:

$$\min_{\Theta} \|\Theta\|_0 \quad \text{s. t.} \quad \|X - \tilde{D}\Theta\|_2 \leq \epsilon \quad (7)$$

其中, $\epsilon \geq 0$ 表示容许误差。

在字典学习中^[16],假设有 J 个大小为 $\sqrt{n} \times \sqrt{n}$ 的训练图像块,将这些图像块拉成列向量并组成训练数据集 $\{x_i\}_{i=1}^J$, $x_i \in R^n$,则字典学习的模型可以表示为:

$$\min_{D, \{\alpha_i\}_{i=1}^J} \|x_i - D\alpha_i\|_F^2 \quad \text{s. t.} \quad \forall i \|\alpha_i\|_0 \leq \tau \quad (8)$$

其中, $D \in R^{n \times M}$, 符号 $\|\cdot\|_F$ 表示 F 范数,定义为: $\|X\|_F = \sqrt{\sum_{ij} X_{ij}^2}$, τ 是稀疏向量 $\{\alpha_i\}_{i=1}^J$ 的稀疏度。

3 基于联合图像块聚类自适应字典学习的多模态医学图像融合

3.1 预处理

将全部图像块进行聚类会得到冗余自适应字典,使用冗余的自适应字典进行稀疏表示和图像重建,会降低图像最后

的融合质量,而将细节丰富的图像块进行聚类 and 字典学习构造的低冗余自适应字典,不仅可以满足图像重建的精度要求还能保证字典的紧凑性。为了选取细节信息丰富的图像块,通过预处理将冗余图像块剔除,减少冗余图像块数量。如图1所示,在医学图像中,将医学图像分块后,图(c)和图(d)表示冗余信息,图(e)和图(f)则表示细节信息丰富的图像块。

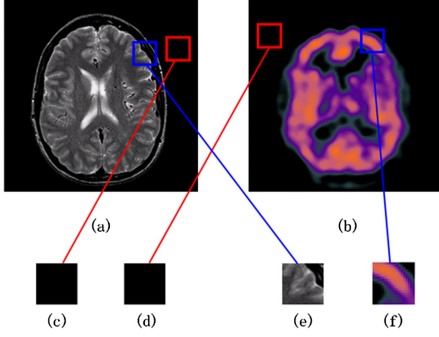


图1 医学图像中的信息图像块和冗余图像块

Fig.1 Informational image patches and redundant image patches in medical images

预处理的具体操作步骤如下:

(1)将源图像划分为若干图像块并计算图像块间的欧氏距离组成矩阵 D ,如式(9)所示:

$$D^{n \times n} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \quad (9)$$

其中, n 表示所有图像块的数目, $D^{n \times n}$ 表示所有图像块之间的相互距离。任意图像块 i 和图像块 j 之间的欧氏距离 $d_{ij} = \|p_i - p_j\|$, 其中 p_i, p_j 分别代表图像块 i, j , 且 $i, j \in (1, 2, \dots, n)$ 。

(2)计算图像块之间的最小距离,如式(10)所示:

$$s_i = \min_{i \neq k} (d_{ik}) \quad (10)$$

其中, s_i 表示目标图像块 p_i 与其他图像块 p_k 的最小距离。

(3)设置截断阈值 ϕ , 将 s_i 与阈值 ϕ 进行比较,如式(11)所示:

$$\begin{cases} s_i > \phi, & p_i = p_i \\ s_i \leq \phi, & p_i = 0 \end{cases} \quad (11)$$

对于每一个目标图像块 p_i , 如果 $s_i > \phi$, 则 $p_i = p_i$, 即目标图像块被保留, 否则 $p_i = 0$, 然后把所有为 0 的图像块去掉, 通过这种方法可以在图像块聚类之前, 剔除冗余的图像块, 保留信息丰富的图像块, 减少冗余图像块。根据文献[11]以及本文第4节的实验将截断阈值 ϕ 设置为 0.1, ϕ 的数值太大可能会剔除过多的图像块, 丢失重要细节, ϕ 的数值太小则会导致剔除的冗余图像块的数量不足。

3.2 联合图像块聚类

为了减少字典的数目, 保证训练得到的字典信息的完整性和紧凑性, 本文采用局部的调制核回归(Steering Kernel Regression, SKR)进行图像块特征提取, 并将提取的图像结构特征信息作为聚类的中心, 将具有相同局部梯度信息的两种模态的图像块进行联合图像块聚类。聚类过程中提取图像块

像素的局部梯度作为图像结构特征信息, 并将其表示为特征向量。对源图像预处理得到的第 k 个图像块的第 i 个像素, 通过调制内核计算得到其特征向量为:

$$\omega_k^i = [\cdots \omega_k^{ij} \cdots]^T \quad (12)$$

$$\omega_k^{ij} = \frac{\sqrt{\det(C_j)}}{2\pi h_{steer}^2} \exp\left[-\frac{(x_k^i - x_j^i)^T C_j (x_k^i - x_j^i)}{2h_{steer}^2}\right],$$

for $j=1, 2, \dots, M$ (13)

其中, ω_k^i 表示源图像预处理后的第 k 个图像块中第 i 个像素的特征向量, x_k^i 表示第 i 个像素 I_k 的空间位置, h_{steer} 是调整支持内核的全局平滑参数, C_j 是以第 j 个像素为中心的相邻图像块的局部梯度估计的转置协方差矩阵, M 是内核中的像素数。 C_j 的估计为:

$$\hat{C}_j \approx \begin{bmatrix} \sum_{x_j \in \omega_i} Z_{x_1}(x_j) Z_{x_1}(x_j) & \sum_{x_j \in \omega_i} Z_{x_1}(x_j) Z_{x_2}(x_j) \\ \sum_{x_j \in \omega_i} Z_{x_1}(x_j) Z_{x_2}(x_j) & \sum_{x_j \in \omega_i} Z_{x_2}(x_j) Z_{x_2}(x_j) \end{bmatrix} \quad (14)$$

其中, $Z_{x_1}(\cdot)$ 和 $Z_{x_2}(\cdot)$ 分别表示目标位置的局部分析窗口 ω_i 处在 x_1 方向的一阶导数和 x_2 方向的一阶导数。将提取出来的第 k 个图像块中的所有像素的特征向量合并为一个特征向量 w_k , 即 $w_k = [\cdots \omega_k^i \cdots]$, 再把所有图像块的特征向量合并存储为一个 k 维的特征矩阵, 即 $w = [w_1, w_2, \dots, w_k]$ 。

使用 K 均值算法对图像块中提取出的特征向量进行初步聚类。随机选取 N 个点作为初始聚类中心, 聚类中心的坐标值为 $\mu_1, \mu_2, \dots, \mu_N \in \mathbb{R}^n$ 。以聚类中心坐标 μ_c 为例, 通过将任意特征向量到初始聚类中心 μ_c 的 L_2 距离最小把每个图像块分成以下几个不相交的聚类:

$$I_k = \bigcup_{n=1}^N \{i | i \in \Omega_c\} \quad (15)$$

其中, I_k 表示第 k 个图像块, Ω_c 表示第 c 个聚类, N 表示上述 k 均值算法得到聚类的总数。将 k 维的特征矩阵中的特征向量 w_k 按照最小距离原则分配到最邻近聚类。

$$c^{(i)} = \arg \min_j \|w_k^{(i)} - \mu_j\|^2 \quad (16)$$

其中, $c^{(i)}$ 表示第 i 个聚类。计算每个聚类的平均值, 将计算得到的均值作为新的聚类中心并赋值给 μ_j 。

$$\mu_j = \frac{\sum_{i=1}^m \mathbf{1}\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m \mathbf{1}\{c^{(i)} = j\}} \quad (17)$$

对于每个聚类, 计算每个特征向量到所属类簇聚类中心的距离平方和 $J(c, \mu)$ 。

$$J(c, \mu) = \sum_{i=1}^m \|w_k^{(i)} - \mu_{c^{(i)}}\|^2 \quad (18)$$

其中, $\mu_{c^{(i)}}$ 表示第 i 个聚类 $c^{(i)}$ 的聚类中心, 将设定阈值 ϕ 与 $J(c, \mu)$ 做比较, 当 $J(c, \mu)$ 小于阈值 ϕ 时聚类过程不再迭代计算聚类中心, 得到 k 个图像块 $[\Omega_1, \Omega_2, \dots, \Omega_N]$ 组成的 N 个聚类。不同模态的图像块中, 有些图像块具有相同的局部梯度, 即具有相同的特征向量, 如图2所示, 不同模态的图像 I_1 和 I_2 中红色、蓝色和绿色框的图像块分别代表3种具有相同特征向量的图像块。将具有相同特征向量的图像块聚类合并为联合图像块集群, 表示生成联合图像块集群的过程为:

$$\Omega_j^{patch} = \bigcup_{k=1}^Q \{y_k | y_k \in \Omega_j\} \quad (19)$$

其中, Ω_j^{patch} 是第 j 个联合图像块簇, y_k 表示第 k 个图像块, 式(19)将属于聚类 Ω_j s. t. $j \in (1, 2, \dots, M)$ 的类簇合并, 即把属

于 Ω_j 的所有图像块聚类合并成为联合图像块聚类 Ω_j^{patch} , 同样的方法可以得到 $\Omega_1^{patch}, \Omega_2^{patch}, \dots, \Omega_M^{patch}$ 的联合图像块聚类。

图 2 表示脑部 PET/MR 生成联合图像块聚类, 用红、绿、蓝 3 种颜色分别表示 3 种图像特征的图像块集群, 不同模态的图像块中的同一种颜色表示具有相同特征向量的图像块, 将具有相同特征向量的图像块的类簇合并。图 2 为由 PET/MR 图像中根据 3 种特征向量而形成的 3 个图像块聚类的示例。

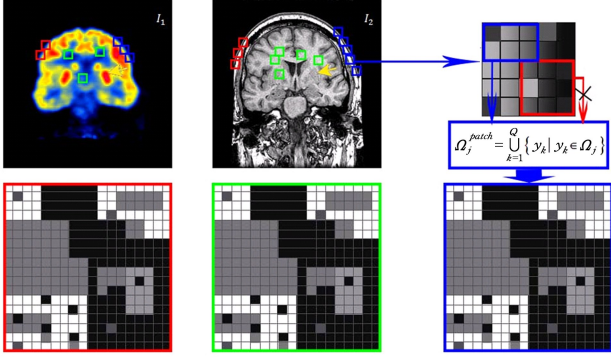


图 2 生成联合图像块聚类

Fig. 2 Generating joint image patches clustering

3.3 字典构造

将聚类得到的类簇进行训练得到子字典, 然后将子字典融合得到自适应字典, 如图 3 所示。

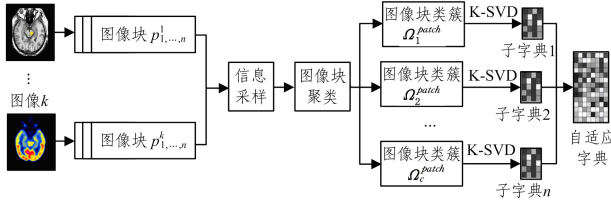


图 3 字典构建过程

Fig. 3 Procedure of dictionary construction

以类簇 Ω_1^{patch} ($n \times N$) 为例, 使用改进的 K-SVD 算法对联合图像块聚类进行训练得到子字典:

$$\begin{aligned} \min_{D_1, X} & \| \Omega_1^{patch} - D_1 X \|_F^2 \\ \text{s. t.} & \| X \|_0 \leq \tau, X \odot M = 0 \end{aligned} \quad (20)$$

其中, 过完备字典 $D_1 \in R^{n \times K}$, 稀疏表示系数 $X \in R^{K \times N}$, \odot 表示点乘运算, 掩膜矩阵 M 由 0 和 1 组成, 定义为 $M = \{|X| = 0\}$, 等价于若 $X(i, j) = 0$, 则 $M(i, j) = 1$, 否则为 0, 因此 $X \odot M = 0$ 能使 X 中的所有 0 项保持完备。式(20)的求解过程分为稀疏编码和字典更新两个阶段。

首先, 随机矩阵初始字典 $D_1^{(0)}$ 通过求解式(21)来实现对系数矩阵 X 的更新。

$$\begin{aligned} X = \arg \min_X & \| \Omega_1^{patch} - D_1 X \|_F^2 \\ \text{s. t.} & X \odot M = 0 \end{aligned} \quad (21)$$

分别对系数矩阵 X 中每一列的非零元素进行处理, 并保持零元素完备, 则式(21)可以转化为式(22)。

$$x_i = \arg \min_{x_i} \| \Omega_{1i}^{patch} - D_{1i} x_i \|_2^2 \quad (22)$$

其中, D_{1i} 是 D_1 对应 X 的非零子集的子矩阵, x_i 是 X 第 i 列的非零部分。使用正交匹配跟踪算法 (Orthogonal Matching Pursuit)^[16] 求解式(22), 得到更新的稀疏系数矩阵 X 。

其次, 在字典更新阶段, 式(21)的优化问题可以转化为:

$$\{D_1, X\} = \arg \min_{D_1, X} \| \Omega_1^{patch} - D_1 X \|_F^2 \quad (23)$$

$$\text{s. t. } X \odot M = 0$$

则式(23)的补偿项可以写为:

$$\begin{aligned} \| \Omega_1^{patch} - D_1 X \|_F^2 &= \| \Omega_1^{patch} - \sum_{j=1}^K d_{1j} x_j^T \|_F^2 \\ &= \| (\Omega_1^{patch} - \sum_{j \neq q} d_{1j} x_j^T) \odot (1_n \cdot m_j^T) - d_{1q} x_q^T \|_F^2 \\ &= \| E_q - d_{1q} x_q^T \|_F^2 \end{aligned} \quad (24)$$

其中, d_{1q} 表示字典 D_1 中待更新的第 q 列; x_q^T 表示稀疏系数矩阵 X 的第 q 列; m_j^T 表示掩膜矩阵 M 的第 j 行, 用来保证 x_q^T 中的零元素在正确位置。 $1_n \cdot m_j^T$ 是将行向量 m_j^T 复制 n 次得到尺寸为 $n \times N$ 的秩为 1 的矩阵, 利用掩膜矩阵 $1_n \cdot m_j^T$ 可以有效去除 $\Omega_1^{patch} - \sum_{j \neq q} d_{1j} x_j^T$ 中那些未用到第 q 个原子所对应样本的列。对误差矩阵 E_q 进行奇异值分解 (Singular Value Decomposition, SVD) 得到 $E_q = U \Delta V^T$, 使用矩阵 U 的第一列更新字典 D_1 中的原子 d_{1q} , 同时将稀疏系数矩阵 X 中的 x_q^T 更新为矩阵 V 的第一列与 $\Delta(1, 1)$ 的乘积。

最后, 循环执行上述两个步骤, 直至达到预设的迭代次数为止, 输出字典 D_1 , 字典 D_2, D_3, \dots, D_N 的训练过程同字典 D_1 , 将训练得到的 c 个子字典 $D_1, D_2, D_3, \dots, D_N$ 组合成自适应字典 $D = [D_1, D_2, D_3, \dots, D_N]$ 。由于在字典更新阶段同时更新字典和稀疏表示系数的非零元素, 使得字典的表示误差更小且字典收敛速度更快。

3.4 多模态图像融合过程

已经配准的脑部 MR/PET 源图像的融合过程如下 (见图 4):

(1) 使用步长为 1 的滑动窗口将源图像 I_c, I_p 分为 $\sqrt{m} \times \sqrt{m}$ 大小的图像块, 对于图像 I_c 和 I_p , 均有 $(M - \sqrt{m} + 1)(N - \sqrt{m} + 1)$ 个图像块。根据 3.1 节的预处理步骤剔除冗余的图像块, 将剩余的细节信息丰富的图像块编纂成 m 维列向量, 将这些图像块中的第 j 个图像块拉成的列向量记为 x_c^j, x_p^j 。

(2) 融合阶段。使用正交匹配跟踪算法 (OMP) 求解 $\{x_c^j, x_p^j\}$ 的稀疏表示系数, 如式(25)和式(26)所示:

$$\alpha_c^j = \arg \min_{\alpha} \| \alpha \|_0 \quad \text{s. t.} \quad \| x_c^j - D \alpha \|_2 \leq \epsilon \quad (25)$$

$$\alpha_p^j = \arg \min_{\alpha} \| \alpha \|_0 \quad \text{s. t.} \quad \| x_p^j - D \alpha \|_2 \leq \epsilon \quad (26)$$

其中, D 为 3.3 节中训练得到的联合图像块自适应字典。将稀疏系数的 l_2 范数作为源图像的活跃度测量, 则稀疏系数 α_c^j 和 α_p^j 通过式(27)的融合规则融合。

$$\alpha_f^j = \begin{cases} \alpha_c^j, & \| \alpha_c^j \|_2 \geq \| \alpha_p^j \|_2 \\ \alpha_p^j, & \text{其他} \end{cases} \quad (27)$$

则 x_c^j 和 x_p^j 的融合结果为:

$$x_f^j = D \alpha_f^j \quad (28)$$

(3) 重建阶段。对所有的图像块执行上述两个步骤以得到所有图像块的融合结果。对于每个块向量 x_f^j , 通过反滑动窗口的过程重塑成 $\sqrt{m} \times \sqrt{m}$ 的图像块并放回到对应的像素位置, 再对重复像素取平均得到最终的融合图像 I_f 。

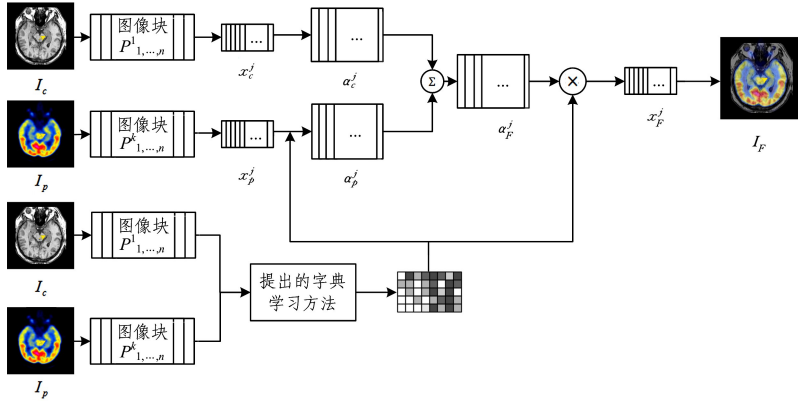


图4 多模态脑部图像融合过程

Fig. 4 Fusion procedure of multi-modal brain image

4 实验

本文的实验环境为 64 位 Windows10 操作系统, MatlabR2015b, Intel i5-6300HQ 2.30 GHz 处理器, 8.0 GB 运行内存。为了验证本文方法的有效性, 选取 3 组已经配准的脑部多模态医学图像进行融合, 这 3 组图像分别为正常脑部 MR/PET、脑部肿瘤 MR/SPECT、脑部肿瘤 CT/SPECT, 图片尺寸均为 256×256 。选取的对比算法有: 基于多尺度变换和稀疏表示的方法 (Multi-Scale Transform and Sparse Representation Based, MST-SR)^[17]、基于压缩感知的离散余弦变换 (Discrete Cosine Transform and Compressed Sensing Based, CS-DCT)^[18]、基于 K-SVD 和同步正交匹配追踪的稀疏表示方法 (Sparse Representation and K-SVD using Synchronization Orthogonal Matching Pursuit, SOMP K-SVD)^[11]、基于 K-SVD 和批量正交匹配追踪的自适应字典学习方法 (The Adaptive trained dictionary using K-SVD and batch orthogonal matching pursuit based, Batch-OMP K-SVD)^[8]、基于密度峰值聚类的字典学习方法 (Local density peaks clustering dictionary learning approach, LDPC)^[11]、基于主成分分析的联合聚类字典学习方法 (Principal Component Analysis using Joint patch clustering, JPC-PCA)^[10]、基于形态几何学的字典学习方法 (The Geometric Dictionary Learning, GDL)^[7]、基于主成分分析的稀疏表示方法 (Sparse Representation and Principal Component Analysis using Synchronous Orthogonal Matching Pursuit Based, SOMP-PCA)^[8]。

基于多尺度变换和稀疏表示的方法 MST-SR 中, 将其分解水平设为 3。基于稀疏表示的方法中其滑动步长设为 1, 图像块尺寸均为 8×8 , 字典大小均为 64×256 , 误差 $\epsilon = 0.01$, 稀疏度 $\tau = 6$ 。本文方法使用的字典构建算法执行 6 个字典更新周期 (Dictionary Update Cycle, DUC) 和 30 次迭代。

为进一步验证 JCPD 的有效性, 使用互信息 (Mutual Information, MI)、基于相位一致 Q_p 、基于梯度 $Q^{AB/F}$ 、基于通用图像质量指标 Q_w 以及视觉信息保真度 VIF (Visual Information Fidelity) 对融合图像进行客观评价^[19-20]。互信息用于衡量两个域变量间的依赖性, 可以度量融合图像与输入图像在灰度分布上的相似程度, 即融合图像保留源图像信息的多少。

分别计算融合图像与源图像之间的互信息。

$$I_{XY}(x, y) = \sum_{x, y} p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \quad (29)$$

之后, 将两者相加, 即可得到融合图像 F 的质量评价结果。

$$M_{AB}^F = I_{FA}(f, a) + I_{FB}(f, b) \quad (30)$$

MI 反映融合图像从源图像中提取信息量的多少, 提取的信息量越多, 融合图像的质量就越好; Q_p 反映融合图像多种类型的显著特征, 例如边缘和角点等, 该指标被定义为 3 个相关系数的乘积。

$$Q_p = (P_p)^a (P_M)^\beta (P_m)^\gamma \quad (31)$$

其中, p 表示相位一致性, M 表示最大值, m 表示最小值。

$$P_p = \max(C_{AF}^p, C_{BF}^p, C_{SF}^p)$$

$$P_M = \max(C_{AF}^M, C_{BF}^M, C_{SF}^M)$$

$$P_m = \max(C_{AF}^m, C_{BF}^m, C_{SF}^m)$$

其中, $C_{xy}^k, \{k | p, M, m\}$ 表示两组 x 和 y 之间的相关系数。

$$C_{xy}^k = \frac{\delta_{xy}^k + C}{\delta_x^k \delta_y^k + C}$$

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

角标 A 和 B 表示输入图像, F 表示融合图像, S 表示最大选择映射, 指数参数 α, β 和 λ 可以根据相位一致、最大值和最小值的重要程度进行调整; $Q^{AB/F}$ 反映融合图像对源图像边缘特性的保留情况, 融合图像 F 的质量评价结果的计算方式为:

$$Q_{AB}^F = \frac{\sum_{i=1}^M \sum_{j=1}^N (Q^{AF}(i, j) \omega^A(i, j) + Q^{BF}(i, j) \omega^B(i, j))}{\sum_{i=1}^M \sum_{j=1}^N (\omega^A(i, j) + \omega^B(i, j))} \quad (32)$$

其中, $\omega^A(i, j)$ 和 $\omega^B(i, j)$ 为 $Q^{AF}(i, j)$ 和 $Q^{BF}(i, j)$ 的权重; $Q^{AF}(i, j)$ 和 $Q^{BF}(i, j)$ 分别为图像 A 与 B 边缘宽度与方向的相似度; $Q^{AB/F}$ 取值范围为 $[0, 1]$, 越接近 1 表示融合图像的边缘越清晰; Q_w 反映融合图像在系数相关性、光照和对比度方面与源图像的关联, 符合人类视觉系统的特点, 考虑到每个图像块的重要程度存在差异, 按照重要程度设置图像块质量在质量评价中的比重, 计算方式为:

$$Q_w(A, B, F) = \sum_{\tau \in W} c(\tau) (\lambda_A(\tau) SSIM(A, F | \tau) + \lambda_B(\tau))$$

$$SSIM(B, F|w) \quad (33)$$

其中, $SSIM(A, F|w)$ 是融合图像 F 与源图像 A 在 w 处的子图像块(大小为 $block_size \times block_size$) 结构相似度。 $\lambda_A(w) = \frac{s(A|w)}{s(A|w) + s(B|w)}$, $s(A|w)$ 和 $s(B|w)$ 分别为图像 A 和图像 B 在窗口 w 处的显著性, $c(w)$ 是 w 处的图像块在整个图像中的重要程度: $c(w) = C(w) / \sum_{w' \in W} C(w')$, 这里, $C(w) = \max(s(A|w), s(B|w))$ 。

VIF 反映视觉信息保真度, 能较好地模拟人类视觉系统, 该指标先将源图像和融合图像进行多分辨率分解, 并对每一个子带进行区域分割; 之后分别模拟人类视觉系统(Human Visual System, HVS) 计算源图像与融合图像在某一分辨率下, 某一区域内提取的信息, 分别表示为扭曲的视觉信息(Visual Information with Distortion, VIND) 和非扭曲的视觉信息(Visual Information without Distortion, VID), 每一个子带的视觉保真度为:

$$VIF_k(A, B, F) = \frac{\sum_b FVID_{k,b}(A, B, F)}{\sum_b FVIND_{k,b}(A, B, F)} \quad (34)$$

则融合图像 F 的整体质量表示为:

$$VIF(A, B, F) = \sum_{k=1}^N p_k \cdot VIF_k(A, B, F) \quad (35)$$

实验所用图像均来自哈佛大学医学院^[21], 图 5 为脑肿瘤的 MR/SPECT 融合结果。图 5(a) 和图 5(b) 分别表示 MR 源图像和 SPECT 源图像。从图 5(c) 中可以看出, MST-SR 方法的融合图像信息明显失真, 存在块效应。从图 5(d) 中可以看出, 相比 MST-SR 方法, CS-DCT 方法的融合图像中心上下两块“明亮”处的脑部淋巴瘤病灶区域在对比度以及清晰度上有了很大的提高, 但是颜色分辨率低, 边缘亮度失真, 但仍存在伪影。并且从表 2 可以看出, MST-SR 在 MI , Q_p , $Q^{AB/F}$, Q_w , VIF 5 项指标上的表现均不理想, 主要原因是 MST-SR 方法具有平移变化性、走样和缺乏方向性等缺点。而 CS-DCT 方法在以上 5 项指标中均明显优于 MST-SR 方法, 原因是该方法具有平移不变性。SOMP K-SVD 方法、Batch-OMP K-SVD 方法、LDPC 方法、JPC-PCA 方法、GDL 方法以及 SOMP PCA 方法与上述两种多尺度变换的方法相比, 进一步改善了图像质量, 但伪影问题依然存在。表 1 中, SOMP K-SVD, Batch-OMP K-SVD, LDPC, JPC-PCA, GDL 和 SOMP-PCA 方法在 MI 指标上优于三层分解水平的多尺度变换的方法, 在其他 4 项指标上存在一些低于多尺度变换方法的情况, 这种情况的可能原因是字典的表示能力不足以提取足够多的细节信息和融合规则导致的图像灰度不连续效应。本文提出的 JCPD 方法在图像的对比度、清晰度以及细节保留度上都优于前文提到的方法, 融合图像没有伪影, 融合图像中心上下两处的脑部淋巴瘤病灶区域清晰, 对比度高, 图像融合质量较好, 相比其他训练方法, JCPD 方法在 MI 指标上的值最大, 表明该方法在源图像中提取的信息最丰富。同时在 Q_p , $Q^{AB/F}$, Q_w 和 VIF 4 项指标上对应的值最大, 表明本文方法的图像融合质量较高, 与主观评价一致, 便于辅助医生对脑部淋巴瘤的诊断。

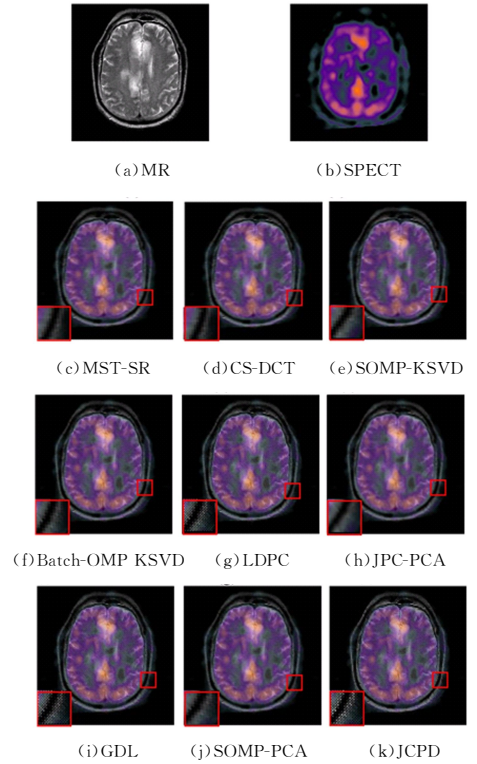


图 5 脑肿瘤的 MR/SPECT 融合结果

Fig. 5 Fusion results of brain tumors MR/SPECT

表 1 脑肿瘤 MR/SPECT 融合结果的性能比较

Table 1 Performance comparison of brain tumors MR / SPECT fusion results

图像	融合方法	评价指标				VIF
		MI	Q_p	$Q^{AB/F}$	Q_w	
图 5	MST-SR	2.2306	0.2945	0.4135	0.6847	0.1898
	CS-DCT	3.1002	0.4334	0.5407	0.7126	0.2162
	SOMP K-SVD	2.9489	0.4289	0.5897	0.7207	0.2257
	Batch-OMP K-SVD	3.6520	0.4630	0.6553	0.7525	0.2160
	LDPC	3.4802	0.3204	0.6170	0.7489	0.2345
	JPC-PCA	3.5235	0.3962	0.6256	0.7764	0.2423
	GDL	3.3030	0.4214	0.6287	0.7009	0.2144
	SOMP-PCA	3.4282	0.4489	0.6228	0.7218	0.2435
	JCPD	3.9040	0.4718	0.6908	0.7794	0.2577

为了检验所提算法的普遍适用性, 本文又做了两组实验, 分别是正常脑的 MR/PET 融合和脑部的 CT/MR 图像融合。图 6 为正常脑的 MR/PET 融合结果, 图 6(a) 和图 6(b) 分别表示 MR 源图像和 PET 源图像, 从图 6(c) 和图 6(d) 可以看出, MST-SR 方法和 CS-DCT 方法的融合图像存在块效应, 且从放大的红色框区域中可以看出, 大脑皮质区域和软组织区域细节丢失。从表 2 可以看出, MI 数值均比其他实验对比方法要小, 说明细节保证效果欠佳, 与主观观测结果一致。SOMP K-SVD, Batch-OMP K-SVD, LDPC, JPC-PCA, GDL 以及 SOMP-PCA 等方法与上述两种多尺度变换的方法相比, 保留了更多的图像细节信息, 进一步改善了图像质量, 融合图像在对比度以及清晰度上有了很大的提高, 但是颜色分辨率低, 边缘亮度失真, 软组织存在伪影。本文提出的 JCPD 方法在图像的对比度、清晰度以及细节保留度上都优于上述方法, 融合图像没有伪影, 骨骼组织、软组织清

晰,图像融合质量较好。

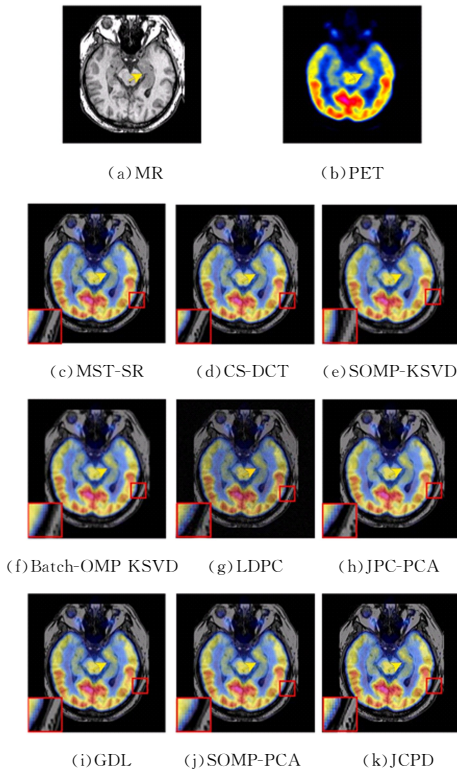


图 6 正常脑部的 MR/PET 融合结果(电子版为彩色)

Fig. 6 Fusion results of normal brain MR/PET

表 2 正常脑部 MR/PET 融合结果的性能比较

Table 2 Performance comparison of normal brain MR/PET fusion results

图像	融合方法	评价指标				
		MI	Q_p	$Q^{AB/F}$	Q_w	VIF
图 6	MST-SR	2.6115	0.2845	0.5335	0.7108	0.1981
	CS-DCT	2.6882	0.4534	0.7507	0.7328	0.2282
	SOMP K-SVD	2.7089	0.4589	0.7697	0.7587	0.2294
	Batch-OMP K-SVD	3.8524	0.4623	0.7453	0.7625	0.2196
	LDPC	3.3711	0.3130	0.6092	0.7329	0.2241
	JPC-PCA	3.4934	0.3781	0.6249	0.7460	0.2391
	GDL	3.0142	0.4512	0.6590	0.7493	0.2160
	SOMP-PCA	3.4362	0.4699	0.7128	0.7568	0.2307
	JCPD	4.0545	0.4828	0.7683	0.7798	0.2486

图 7 为脑部的 CT/MR 图像融合结果。图 7(a)和图 7(b)分别表示 CT 源图像和 MR 源图像,从图 7(c)、图 7(d)和图 7(e)可以看出, MST-SR, CS-DCT 和 SOMP K-SVD 方法的融合图像脑部右侧皮质细节丢失。从表 3 可以看出, MST-SR, CS-DCT 和 SOMP K-SVD 方法在 MI 指标上都劣于其他对比的实验方法,这种情况说明以上 3 种方法在细节信息保留上存在缺陷,与主观观察一致。Batch-OMP K-SVD, LDPC, JPC-PCA, GDL 以及 SOMP PCA 等方法保留了更多的图像细节信息,进一步改善了图像质量,但伪影问题依然存在。在表 3 中, Batch-OMP K-SVD, LDPC, JPC-PCA, GDL 以及 SOMP-PCA 等方法的 MI, Q_p , $Q^{AB/F}$, Q_w 和 VIF 5 项指标均优于前面 3 种方法,造成伪影的原因可能是在稀疏表示时对图像进行分块所造成的。本文提出的 JCPD 方法在对比度、清

晰度以及细节保留上都优于前面的所提方法,使得融合图像的左侧皮质和右侧皮质区域的血流程度更易于观测, JCPD 方法的 MI, Q_p , $Q^{AB/F}$, Q_w 和 VIF 5 项指标在上述 3 组实验中的值分别为 4.3545, 0.4806, 0.7960, 0.7498 和 0.2586, 相较于其他训练方法, JCPD 方法在 MI 指标上的值最大, 表明该方法在源图像中提取的信息最丰富。同时在 Q_p , $Q^{AB/F}$, Q_w 和 VIF 4 项指标上对应的值最大, 表明本文方法的图像融合质量较高, 与主观评价一致, 有助于对病灶区域进行诊断, 图像融合质量较好。

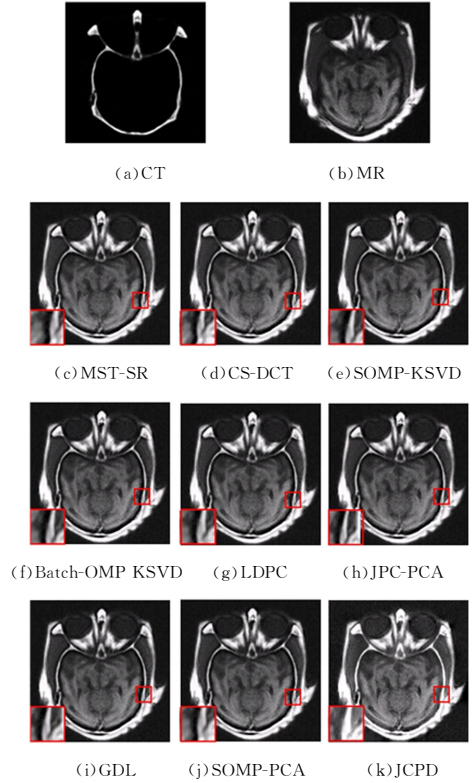


图 7 正常脑部的 CT/MR 融合结果

Fig. 7 Fusion results of normal brain CT/MR

表 3 正常脑部 CT/MR 融合结果的性能比较

Table 3 Performance comparison of normal brain CT/MR fusion results

图像	融合方法	评价指标				
		MI	Q_p	$Q^{AB/F}$	Q_w	VIF
图 7	MST-SR	3.6115	0.2745	0.4935	0.6808	0.1997
	CS-DCT	3.6882	0.4234	0.6506	0.6928	0.2372
	SOMP K-SVD	3.7089	0.4389	0.7095	0.6987	0.2304
	Batch-OMP K-SVD	3.8514	0.4423	0.7752	0.7025	0.2188
	LDPC	3.3601	0.3203	0.6198	0.7225	0.2049
	JPC-PCA	3.4945	0.3861	0.6482	0.7392	0.2382
	GDL	3.9032	0.4408	0.6703	0.7201	0.2058
	SOMP PCA	4.4362	0.4597	0.7126	0.7368	0.2487
	JCPD	4.3545	0.4806	0.7960	0.7498	0.2586

为了验证本文提出的字典构造方法的时间效率, 分别计算在相同条件下不同的图像融合方法进行 10 次实验所消耗时间的平均值, 结果如表 4 所列。由表 4 可以看出, JCPD 方法的字典训练时间、稀疏表示时间、系数融合时间和总时间都明显短于 SOMP K-SVD1, GDL 和 SOMP PCA 方法。CS-

DCT字典学习方法中的训练数据是从40幅图像中提取50000个图像块;MST-SR字典学习方法中的训练数据是在小波域中分别对每个子带提取50000个 8×8 图像块。

表4 不同方法相同条件下10次实验平均处理时间

Table 4 Average processing time of 10 experiments of different method under same conditions

	(单位:s)			
	SOMP K-SVD	GDL	SOMP PCA	JCPD
字典训练时间	45.20	42.38	38.30	20.94
稀疏表示时间	18.29	17.63	14.09	10.76
系数融合时间	12.02	13.67	8.61	4.88
总时间	95.51	93.68	76.43	42.58

结束语 针对多模态医学图像融合中过完备自适应字典存在的大量冗余信息会导致图像重建质量不佳的问题,本文提出了基于联合图像块聚类自适应字典学习的多模态医学图像融合方法。该方法分别对正常脑部、脑部肿瘤和癌细胞转移到脑部3组脑部医学图像进行了多次实验,结果表明,相比于多尺度变换的方法、基于主成分分析的稀疏表示方法、基于K-SVD的稀疏表示方法以及多尺度字典学习的方法,本文提出的JCPD方法显著提高了多模态医学图像的融合质量,能为临床医疗诊断提供有效的帮助。但本文仍存在联合图像块聚类时需要预设类簇数量的问题,如果聚类数量设定得不合适,则会影响字典学习的效率和图像融合质量,如何自适应地设定聚类的数量,是下一步的研究重点。

参考文献

- [1] ZHANG Q, LIU Y, BLUM R S, et al. Sparse Representation based Multi-sensor Image Fusion for Multi-focus and Multi-modality Images: A Review[J]. *Information Fusion*, 2017, 40(2018): 57-75.
- [2] LIU Y. Research on methods for pixel-level multi-source image fusion[D]. Hefei: University of Science and Technology of China, 2016. (in Chinese)
刘羽. 像素级多源图像融合方法研究[D]. 合肥: 中国科学技术大学, 2016.
- [3] ASIF U, BENNAMOUN M, SOHEL F. A model-free approach for the segmentation of unknown objects[C]// *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014: 4914-4921.
- [4] ZHU P, ZHU W, HU Q, et al. Subspace clustering guided unsupervised feature selection[J]. *Pattern Recognition*, 2017, 66(C): 364-374.
- [5] YIN H, LI Y, CHAI Y, et al. A novel sparse-representation-based multi-focus image fusion approach[J]. *Neurocomputing*, 2016, 216(C): 216-229.
- [6] ZONG J, QIU T. Medical image fusion based on sparse representation of classified image patches[J]. *Biomedical Signal Processing and Control*, 2017, 34(6): 195-205.
- [7] ZHU Z, QI G, CHAI Y, et al. A Geometric Dictionary Learning Based Approach for Fluorescence Spectroscopy Image Fusion[J]. *Applied Sciences*, 2017, 7(2): 161-178.
- [8] WANG K, QI G, ZHU Z, et al. A Novel Geometric Dictionary Construction Approach for Sparse Representation Based Image Fusion[J]. *Entropy*, 2017, 19(7): 306-324.
- [9] YIN F, GAO W, SONG Z. Medical Image Fusion Based on Feature Extraction and Sparse Representation[J]. *International journal of biomedical imaging*, 2017(5): 11-34.
- [10] KIM M, HAN D K, KO H. Joint patch clustering-based dictionary learning for multimodal image fusion[J]. *Information Fusion*, 2015, 27(C): 198-214.
- [11] ZHU Z, CHAI Y, YIN H, et al. A novel dictionary learning approach for multi-modality medical image fusion[J]. *Neurocomputing*, 2016, 214: 471-482.
- [12] TAKEDA H, FARSIU S, MILANFAR P. Kernel regression for image processing and reconstruction[J]. *IEEE Transactions on Image Processing*, 2007, 16(2): 349-366.
- [13] AGARWAL A, ANANDKUMAR A, NETRAPALLI P. A Clustering Approach to Learning Sparsely Used Overcomplete Dictionaries[J]. *IEEE Transactions on Information Theory*, 2016, 63(1): 575-592.
- [14] YEGANLI F, NAZZAL M, OZKARAMANLI H. Image super-resolution via sparse representation over multiple learned dictionaries based on edge sharpness and gradient phase angle[J]. *Signal Image & Video Processing*, 2015, 9(1): 285-293.
- [15] ZHANG Q, LIU Y, BLUM R S, et al. Sparse Representation based Multi-sensor Image Fusion for Multi-focus and Multi-modality Images: A Review[J]. *Information Fusion*, 2017, 9(1): 25-33.
- [16] LIU J W, CUI L P, LUO X L. Structured sparse models[J]. *Chinese Journal of Computers*, 2017, 40(6): 1309-1337. (in Chinese)
刘建伟, 崔立鹏, 罗雄麟. 结构稀疏模型[J]. *计算机学报*, 2017, 40(6): 1309-1337.
- [17] JAVED U, RIAZ M M, GHAFOR A, et al. MRI and PET image fusion using fuzzy logic and image local features[J]. *The Scientific World Journal*, 2014, 2014(2014): 8-36.
- [18] SUN J G, LIU J, ZHAO L Y. Clustering algorithms research[J]. *Journal of Software*, 2008, 19(1): 48-61. (in Chinese)
孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. *软件学报*, 2008, 19(1): 48-61.
- [19] ZHANG X L, LI X F, LI J. Validation and Correlation Analysis of Metrics for Evaluating Performance of Image Fusion[J]. *Acta automatica scinica*, 2014, 40(2): 306-315. (in Chinese)
张小利, 李雄飞, 李军. 融合图像质量评价指标的相关性分析及性能评估[J]. *自动化学报*, 2014, 40(2): 306-315.
- [20] LIU Z, BLASCH E, XUE Z, et al. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 34(1): 94-109.
- [21] JOHNSON K A, BECKER J A. The whole brain atlas [EB/OL]. [2016-10-09]. <http://www.med.harvard.edu/aanlib/home.html>.