

基于稀疏原子融合的 RGB-D 场景图像融合算法

刘帆, 刘鹏远, 张峻宁, 徐彬彬

陆军工程大学石家庄校区, 河北 石家庄 050003

摘要 针对当前彩色图像和深度图像(RGB-D)特征融合困难、联合识别效率不高的问题,提出了一种结合 K 奇异值分解(KSVD)和最大相关最小冗余准则(mRMR)的 RGB-D 场景图像融合算法。该算法首先采用 KSVD 稀疏图像的特征,将稀疏系数对应的字典原子作为特征融合的参数,以完整地表达图像的全部信息;之后采用互信息的 mRMR 原则求取维度最小且各维度之间相关性最小的特征原子组合;最后通过最大化原则融合特征原子对应的稀疏系数,从而完成了两种图像之间的有效信息融合。实验结果表明,该算法在信息熵、互信息和边缘保持度等方面比主成分分析-K 奇异值分解和非下采样轮廓变换-K 奇异值分解融合算法更有优势,有效提高了图像目标的识别准确率和成功率。

关键词 机器视觉; 图像融合; K 奇异值分解; 互信息; RGB-D

中图分类号 TP391

文献标识码 A

doi: 10.3788/AOS201838.0115003

RGB-D Scene Image Fusion Algorithm Based on Sparse Atom Fusion

Liu Fan, Liu Pengyuan, Zhang Junning, Xu Binbin

Shijiazhuang Campus of the Army Engineering University, Shijiazhuang, Hebei 050003, China

Abstract To solve the problems of difficulty of feature fusion and low efficiency of joint recognition in color image and depth image(RGB-D), a new algorithm of RGB-D scene image fusion is proposed based on K singular value decomposition (KSVD) and maximum correlation minimum redundancy atoms (mRMR) principle. Firstly, the features of the sparse KSVD image and the dictionary atoms corresponding to the sparse coefficients are used as the parameters of feature fusion to fully express the whole information of image. Secondly, the mRMR principle of mutual information is used to determine the characteristic atom combination which has minimum dimensions and minimum correlation among different dimensions. Finally, the sparse coefficients are fused by the maximization principle to obtain the effective information fusion between two images. Experimental results show that the proposed algorithm has advantages over principal component analysis-K singular value decomposition and non-subsampled contour transform-K singular value decomposition fusion algorithms in terms of information entropy, mutual information and edge preservation, which improves recognition accuracy and success rate of the image targets effectively.

Key words machine vision; image fusion; K singular value decomposition; mutual information; RGB-D

OCIS codes 150.0155; 150.1135

1 引 言

彩色图像和深度图像(RGB-D)是当前计算机视觉领域新的研究热点,主要由深度相机获得,包括 RGB 场景图像及其对应的深度场景图像。RGB 场景图像包含所拍摄物体的表面颜色和纹理等信息,而深度场景图像包含物体的空间形状等信息,两种图像彼此之间是很好的补充。因此,有效融

合利用 RGB 场景信息和深度场景信息,能够提高场景中物体识别的准确率。RGB-D 图像属于多模态数据,需要使用特定的融合算法进行融合。RGB-D 图像融合算法可分为像素级、特征级和决策级 3 个层次^[1]。

特征级融合是指分别提取 RGB 图像和深度图像的特征后,再采用相应的方法进行融合,其优点是减少了计算量;但由于该方法只是对图像的某种或

收稿日期: 2017-06-26; 收到修改稿日期: 2017-08-28

基金项目: 国家自然科学基金(51205405,51305454)

作者简介: 刘帆(1995—),男,硕士研究生,主要从事计算机视觉方面的研究。E-mail: 2434344286@qq.com

导师简介: 刘鹏远(1975—),男,博士,副教授,主要从事增强现实维修方面的研究。E-mail: lpy_jx@sina.com

者几种特征进行提取融合,不可避免地会丢失许多有用信息。决策级融合是指对 RGB 图像和深度图像的分类信息进行统计或逻辑推理,分别得到相应的决策后再进行综合判断,因此有用特征的损失更加严重。

像素级融合通过对图像像素进行融合,得到有利于识别的融合图像,尽可能地避免图像信息的丢失,从而有效提升机器识别检测的准确率。王田等^[2]融合图像的边缘信息,得到了有利于检测分类的融合轮廓图;向训文^[3]提出将 RGB 显著图与深度显著图进行融合,得到的融合图像对显著部分的特征进行了更全面的表述,有效提高了识别检测的效率。但以上算法对原始图像信息的融合还不全面,基于稀疏表达的信息提取融合是解决这一问题的有效方法。本文致力于对原始图像的全部有效信息进行融合。Yang 等^[4]提出了基于 K 奇异值分解(KSVD)字典的稀疏表达图像融合算法,余南南等^[5]利用 KSVD 字典对医学计算机断层扫描(CT)和磁共振成像(MR)图像进行融合。以上方法是对多模态图像信息进行融合的成功案例,但其并未对图像特征对应的字典原子进行充分的筛选,因此融合后的图像会出现较多的冗余信息。

基于稀疏表达方法对 RGB-D 场景图像进行融合,并引入互信息理论(mRMR)来解决字典原子筛选困难和字典原子冗余问题。融合算法通过 mRMR 原则找到冗余度最小且最有效表达图像特征的原子,对两种图像中相同的特征原子按最优法融合其对应的稀疏系数,之后对图像进行重构,对重合位置的像素取平均值。融合后的图像相比单一的 RGB 图像或深度图像具有更加丰富的视觉信息,在图像客观评价指标方面,与主成分分析-K 奇异值分解(PCA-KSVD)、非下采样轮廓变换-K 奇异值分解(NSCT-KSVD)稀疏融合算法相比,具有更好的性能,在相关数据库中进行目标定位检测时也取得了更高的识别准确率和成功率,从而验证了该算法融合图像特征的有效性。

2 算法原理

RGB 图像和深度图像是对同一目标不同模态的表达,两种图像之间存在着重复的冗余信息,另外 RGB 图像和深度图像本身也具有大量的空间相关冗余和结构相关冗余。空间相关冗余是指相邻位置之间的像素具有很大的相关性,而结构相关冗余是指图像中反复出现相同或相近的纹理特征。在目

标识别检测过程中,冗余信息会使图像数据量增加,从而增加提取特征向量的维度,影响匹配效率和精度。融合算法的目的在于尽可能去除两种图像之间和自身的冗余信息,以得到更有利于识别检测的融合图像。

2.1 KSVD 字典学习算法

KSVD 字典学习算法^[6]是当前较为有效的稀疏表达方法,能够迭代学习得到与训练样本高度适应的冗余字典,并且可较好地对特征进行稀疏表达。本文使用 KSVD 字典学习算法对 RGB 图像和深度图像的空间和结构冗余信息进行有效压缩降维,从而稀疏化图像特征,方便后续有效识别特征的筛选,减小计算量。

KSVD 模型可定义为

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \|\mathbf{a}\|_0 \text{ s.t. } \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 \leq \zeta, \quad (1)$$

式中 $\|\mathbf{a}\|_0$ 为伪范数,表示系数非零项的上限;训练样本矩阵 $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_n]$, 其中 \mathbf{y}_i 为第 i 个训练样本, n 为训练样本总个数;冗余字典 $\mathbf{D} \in R^{\beta \times K}$ ($K \gg n$), 其中 R 为实数集, β 为字典行数, K 为字典列数;编码系数 $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$, \mathbf{x}_i 为稀疏编码向量; \mathbf{a} 为字典原子集; a 为目标类别; ζ 为设定的系数上限值。

算法步骤如下:

1) 初始化字典 \mathbf{D} , 如定义完备的离散余弦变换字典。

2) 系数编码。固定字典 \mathbf{D} , 使用正交匹配追踪算法(OMP)^[7]对每一块得到的列向量 $\mathbf{b}_i \mathbf{Y}$ 求解稀疏编码向量 \mathbf{x}_i , $\mathbf{x}_i = \text{OMP}(\mathbf{D}_i, \mathbf{b}_i \mathbf{Y}, \zeta, \mathbf{T})$, 其中 \mathbf{b}_i 为选取的训练样本对应的系数, \mathbf{T} 为转换矩阵。

3) 字典学习。逐步更新字典原子 d_i , 该过程可表示为

$$\|\mathbf{Y} - \mathbf{D}\mathbf{x}\|_F^2 = \left\| \left(\mathbf{Y} - \sum_{j \neq l} \mathbf{d}_j \mathbf{x}_j^T \right) - \mathbf{d}_l \mathbf{x}_l^T \right\|_F^2 = \|\mathbf{E}_l - \mathbf{d}_l \mathbf{x}_l^T\|_F^2, \quad (2)$$

式中 \mathbf{d}_l 为第 l 列字典原子, \mathbf{x}_l^T 为 \mathbf{X} 中第 $1, 2, \dots, l$ 行系数, \mathbf{E}_l 代表抽取字典 \mathbf{D}_l 后的误差矩阵, F 代表取矩阵的 F 范数。设 $\delta_v = \{v \mid 1 \leq v \leq l, \mathbf{x}_l^T(v) \neq 0\}$, 定义 $\boldsymbol{\sigma}$ 为 $N \times \delta(l)$ 阶矩阵, 且在 $\boldsymbol{\sigma}$ 中第 $\delta_i(v)$ 行、第 v 列的元素全部为 1, 其余全部为 0。向量 $\mathbf{x}_l^T(v) = \mathbf{x}_l^T(v) \boldsymbol{\sigma}$ 和矩阵 $\mathbf{E}_l^k = \mathbf{E}_l \boldsymbol{\sigma}$ 分别定义为 \mathbf{x}_l^T 和 \mathbf{E}_l 去掉零元素后的收缩结果, 对 \mathbf{E}_l^k 进行奇异值分解(SVD)分解, $\mathbf{E}_l^k = \mathbf{U} \Delta \mathbf{V}^T$ (\mathbf{U} 和 $\Delta \mathbf{V}$ 为 SVD 计算的分解值矩阵), 将字典中的第 k 列更新为 \mathbf{U} 的第 1 列, \mathbf{x}_l^T 更新为 \mathbf{V} 的第 1 列与 $\Delta(1, 1)$ 的乘积。

4) 若满足停止条件(设定的迭代次数或重构图像与原图像的误差率),则得到最终的冗余字典 \mathbf{D} , 否则返回到步骤 2)。

2.2 mRMR 原则的特征选取

当使用 KSVD 字典学习算法对图像冗余信息进行稀疏表达后,致力于选取最有利于分类识别且冗余度最小的图像特征信息,以摒弃过多冗余信息对分类识别的干扰,提高分类识别的效率,因此利用 mRMR 原则对字典原子对应的稀疏系数进行优化选取。mRMR 的主要思想是利用互信息从原始特征集中选择出最优特征集,该特征集每个特征之间的相关性最小,但与目标类之间的相关性最大。这里的原始特征集是原始图像块使用 KSVD 字典学习算法得到的字典原子集 \mathbf{Z} , 目标类是含有显著识别目标的图像块利用 KSVD 字典学习算法得到的字典原子集 \mathbf{a} 。

设 $p(z)$ 和 $p(\partial)$ 分别为随机变量 z 和 ∂ 的概率密度, $p(z, \partial)$ 为二者的联合概率密度, z 和 ∂ 的互信息可定义为

$$I(z, \partial) = \iint p(z, \partial) \lg \frac{p(z, \partial)}{p(z)p(\partial)} dz d\partial. \quad (3)$$

使用互信息得到的 mRMR 准则为

$$\max C(\mathbf{L}, \mathbf{a}), C = \frac{1}{|\mathbf{L}|} \sum_{z_i \in \mathbf{S}} I(z_i, \mathbf{a}), \quad (4)$$

$$\min R(\mathbf{L}), R = \frac{1}{|\mathbf{L}|^2} \sum_{z_i, z_j \in \mathbf{S}} I(z_i, z_j), \quad (5)$$

式中 $|\mathbf{L}|$ 为特征空间维数, $I(z_i, \mathbf{a})$ 为子空间 \mathbf{L} 中特征 z_i 与目标类别 \mathbf{a} 之间的互信息, $I(z_i, z_j)$ 为特征 z_i 与特征 z_j 之间的互信息。

联立(4)、(5)式,设定 mRMR 算法的优化准则为

$$\max[\Phi(C, R), \Phi(C, R)] = C - R, \quad (6)$$

式中 $\Phi(C, R)$ 为相关度和冗余度的值域, C 为不同信息之间的相关度。该算法是一种逐步构造最优解的贪心搜索算法,采用前向 Greedy 算法^[8]对(6)式进行求解。原始特征集为 \mathbf{Z} , \mathbf{S}_{m-1} 为已选取的 $m-1$ 个特征的集合,特征选择的目标是从余下的原始特征集 $\{\mathbf{Z} - \mathbf{S}_{m-1}\}$ 中选取第 m 个特征,且该特征需满足

$$\max_{z_j \in \mathbf{Z} - \mathbf{S}_{m-1}} \left[I(z_j, \mathbf{a}) - \frac{1}{m-1} \sum_{z_i \in \mathbf{S}_{m-1}} I(z_j, z_i) \right]. \quad (7)$$

3 图像特征融合过程

3.1 特征提取

RGB-D 场景图像融合的目的是将多模态数据

中的互补信息融合到较低的维度中,并尽可能保留有利于目标识别分类的信息。因此,首先要对原始图像特征进行完整的提取表达,基于 KSVD 的稀疏算法可较好地表达稀疏信息。

设 RGB 图像为 A , 深度图像为 B , 首先采用滑动窗技术^[9], 使用特定大小和步长的滑动窗分别对两幅图像进行分块,将每个图像块字典作为列向量,分别组成矩阵 \mathbf{V}_1 和 \mathbf{V}_2 ; 将 \mathbf{V}_1 和 \mathbf{V}_2 组成联合矩阵 $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2]$, 通过 KSVD 字典学习算法训练得到联合矩阵 \mathbf{V} 的冗余字典 $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2]$, 再使用 OMP 求解得到 \mathbf{V} 的稀疏编码系数 $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$, 最后根据稀疏系数寻找对应的字典原子,并将其作为特征融合的参数。

3.2 特征融合

设 θ_r, θ_d 分别表示 RGB 图像和深度图像的特征^[10], $\theta_r = \theta_c + \theta_1, \theta_d = \theta_c + \theta_2$, 其中 θ_c 表示 RGB 图像和深度图像共同包含的图像特征, θ_1 和 θ_2 为 RGB 图像和深度图像各自包含的图像特征。图像融合就是生成一幅既包含共同特征 θ_c , 又包含各自特征 θ_1 和 θ_2 的图像。设 $\mathbf{v}_{1i}, \mathbf{v}_{2i}, \mathbf{x}_{1i}, \mathbf{x}_{2i}$ 分别为 $\mathbf{V}_1, \mathbf{V}_2, \mathbf{X}_1, \mathbf{X}_2$ 的第 i 列向量, 则 $\mathbf{v}_{1i}, \mathbf{v}_{2i}$ 分别可表示为

$$\begin{cases} \mathbf{v}_{1i} = \mathbf{D}\mathbf{x}_{1i} = \mathbf{D}(x_{1i}^0 & 0 & 0 & \cdots & x_{1i}^c & 0)^T \\ \mathbf{v}_{2i} = \mathbf{D}\mathbf{x}_{2i} = \mathbf{D}(x_{2i}^0 & 0 & 0 & \cdots & 0 & x_{2i}^c)^T \end{cases}, \quad (8)$$

式中 x_{1i}^0, x_{2i}^0 为 \mathbf{v}_{1i} 和 \mathbf{v}_{2i} 中相同位置的非零值, 表示 \mathbf{v}_{1i} 和 \mathbf{v}_{2i} 的共同特征; x_{1i}^c, x_{2i}^c 为 \mathbf{v}_{1i} 和 \mathbf{v}_{2i} 中不同位置的非零值, 表示 \mathbf{v}_{1i} 和 \mathbf{v}_{2i} 各自的特征; $x_{1i}^0, x_{2i}^0, x_{1i}^c$ 和 x_{2i}^c 的数值大小表示特征的强弱。该算法希望融合两类图像中有利于分类识别的有效特征, 摒弃冗余信息和干扰特征, 即选择最有利于分类识别的字典原子, 因此使用 mRMR 选择稀疏系数对应的字典原子, 通过最大化原则融合稀疏系数。

mRMR 准则只能选择最优字典原子集合, 即 $\mathbf{S}_1 < \mathbf{S}_2 < \cdots < \mathbf{S}_j$, 无法确定最优原子的个数, 因此提出一种最优原子数目选取原则, 根据不同的原子集合下评价指标的得分情况, 找到融合效果最好且原子数目最少的原子集。该算法可表示为

$$\max[(P_i), P_i] = \sum_{k=1}^{n_p} \tau_k \times \text{score}_k(\mathbf{S}_i), \quad (9)$$

式中 n_p 为评价指标的个数, $\text{score}_k(\mathbf{S}_i)$ 表示在原子集合 \mathbf{S}_i 下第 k 项评价指标的得分情况, τ_k 为第 k 项评价指标对应的权重。

3.3 图像融合流程

图 1 为所提 K-mRMR 融合算法流程图, 具体算法如下:

1) 稀疏表达图像特征。使用第 2.1 节的滑动窗技术分别处理 RGB 图像和深度图像,得到 $2N$ 个图像块,并组合成 $2N$ 个列向量,利用 KSVD 字典学习算法得到这 $2N$ 个列向量的联合字典 \mathbf{D} 和每个列向量对应的稀疏系数。根据稀疏系数找出每个图像块对应的字典原子,其中 \mathbf{D}^r 为 RGB 图像和深度图像共同的原子集合, \mathbf{D}^c 为各自图像对应的原子集合。

2) 基于 mRMR 算法选取最优原子集。令 \mathbf{S} 为已选择的字典原子集合, \mathbf{F} 为待选择原子集合,令 $\mathbf{F} = \{\mathbf{D}^r, \mathbf{D}^c\}$ 。依据最优原子数目原则确定原子数目 L ,从待选择原子集 \mathbf{F} 中利用 mRMR 原则选取 L 个字典原子 $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_L$,得到最终选取的字典原子集 $\mathbf{S} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_L\}$ 。

3) 融合相同特征原子的稀疏系数。由于 RGB 图像和深度图像既具有相同的特征原子集也具有不同的特征原子集,因此对于被选择的相同特征原子

集,按照最大化原则融合其对应的稀疏系数,即

$$x_i^0 = \arg \max(|x_{1i}^0|, |x_{2i}^0|), \quad (10)$$

式中 x_{1i}^0 为 RGB 图像对应的第 i 个字典原子对应的稀疏系数, x_{2i}^0 为深度图像对应的第 i 个字典原子对应的稀疏系数, x_i^0 为融合后图像的第 i 个字典原子对应的稀疏系数。根据 KSVD 稀疏化的思想, x_{1i}^0 、 x_{2i}^0 分别对应 RGB 图像和深度图像的稀疏编码特征,按照(10)式原则对 x_{1i}^0 、 x_{2i}^0 进行融合后,两种图像同类特征中最明显的一个会被更多的表达出来,如当深度图像的边缘轮廓相对于 RGB 图像更明显时,融合后的图像会更多地体现深度图像的边缘轮廓。

4) 重构图像。根据 $\mathbf{V}_f = \mathbf{S}_m \mathbf{X}_f$ (\mathbf{X}_f 为第 f 列稀疏系数矩阵, \mathbf{V}_f 为由稀疏系数重构得到的第 f 列像素值矩阵)可以计算得到融合后的特征矩阵,再按照滑动窗逆变换重构图像,将获得的每一列向量变为方阵,然后将方阵放到图像中相应的位置,重合位置的像素取其各个对应像素的平均值。

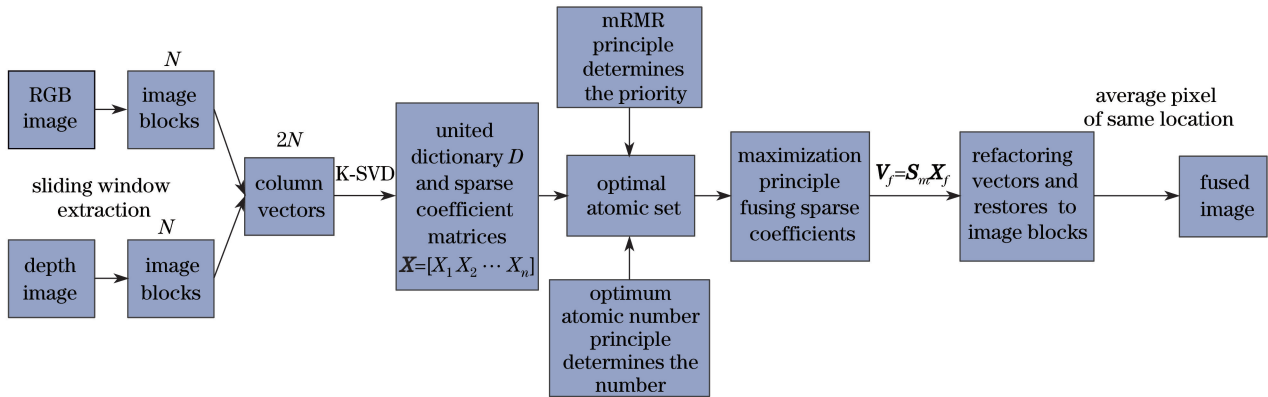


图 1 K-mRMR 融合算法流程图

Fig. 1 Flow chart of K-mRMR fusion algorithm

4 实验实现与结果分析

4.1 测试数据库

微软公司推出的 kinect 是当前比较流行的一款深度相机,基于 kinect 衍生了一些相当有价值的数据集,如华盛顿大学采集公布的 RGB-D 数据库,RGB-D 数据库包含 51 类家居用品,300 个实例,41877 张 RGB-D 图像。选用该数据集进行实验验证。

4.2 图像预处理

深度图像是根据目标距离深度摄像机距离的远近对每个像素赋予一定的数值。距离摄像头越近则像素数值越小,而距离摄像头越远则像素数值越大,将深度图像的像素值适当调整到 $0 \sim 255$ 范围内,如图 2(a)所示。

当前设备条件下采集的深度图像存在一定的边缘毛刺和无效点问题,在融合之前需要进行图像增强和优化。采用文献[11]中改进的联合双边滤波插值算法,得到处理后的图像如图 2(b)所示。处理后的图像无效点得到了插值填充,不规则的边缘毛刺则被有效滤除掉。

4.3 基于 KSVD 和 RMR 的 RGB-D 图像融合

选取^[12-13]平均梯度(A_{AVG})、信息熵(H)、互信息(M_I)、边缘保持度($Q^{AB/F}$)和结构相似度(S_{SIM}) 5 种客观评价指标。平均梯度越大,图像越清晰;信息熵越大,说明融合图像包含的信息量越大; M_I 值越大,说明融合图像从源图像中提取的信息量越大; $Q^{AB/F}$ 越接近 1,说明融合图像从源图像获得的边缘信息越多; S_{SIM} 越靠近 1,说明图像 A 与 B 之间的相似度越高。

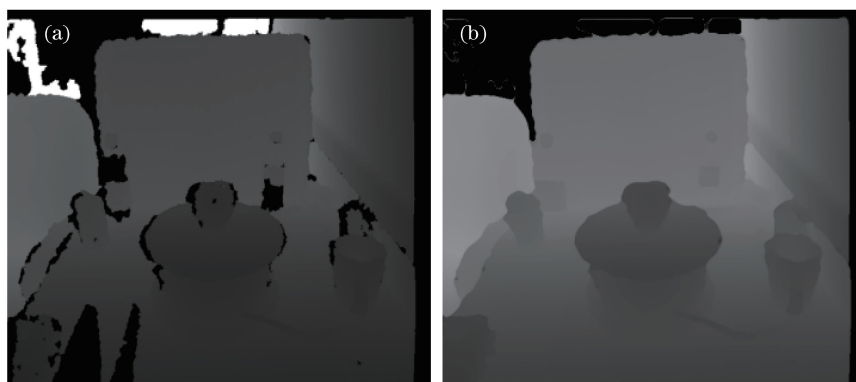


图 2 深度图像预处理。(a)原始深度图像;(b)联合双边滤波处理后图像

Fig. 2 Depth image preprocessing. (a) Original depth image; (b) image after joint bilateral filtering processing

为确定选取的最优原子数,实验选取以上 5 项指标并按(9)式对融合图像进行评分,设 5 项评价指标的权值系数分别为 W_A, W_H, W_M, W_Q, W_S ,当 $W_A = 2W_H = 2W_M = 2W_Q = 2W_S$ 时得分为 P_A ,当 $W_H = 2W_A = 2W_M = 2W_Q = 2W_S$ 时得分为 P_H ,依次类推,计算 5 种不同权重下的融合图像评分情况,得到如图 3 所示的曲线。

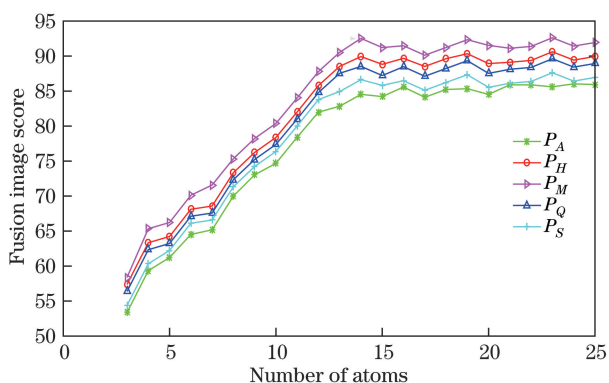


图 3 不同原子数目的融合图像评分曲线

Fig. 3 Scoring curves of fusion images with different atomic numbers

由图 3 可知,当选择的原子数为 14 时,各权重下的融合图像评分普遍达到一个极值点,若继续增加选择原子数目,则图像的评分情况逐渐趋于稳定,不再明显增加。综合以上判断,选择最优原子数为 14,以保证得到最好融合效果的同时减少计算量。优化选择前后的原子变化如图 4 所示。

由图 4 可知,原始稀疏系数对应的字典原子的原子维数较高且分布散乱,经过 mRMR 算法选择后的字典原子维数降低,且分布更加均匀有序,这体现了该算法选择原子的优势。

为了验证所提融合算法的性能,将其与另外两种稀疏表示算法(PKA-KSVD^[14]算法和 NSCT-KSVD^[15]算法)进行比较,PKA-KSVD 字典学习算法将主成分分析与 KSVD 字典学习算法结合起来,对高维的特征降维进行分解与重构;NSCT-KSVD 字典学习算法是使用 KSVD 字典对非下采样剪切波变换(NSST)分解后的低频子带系数进行学习训练。实验中使用的滑动窗大小为 8×8 ,滑动步长为 1,字典大小为 64×256 。图 5 和图 6 为 RGB 图像转化为灰度图后与深度图像进行融合的结果。

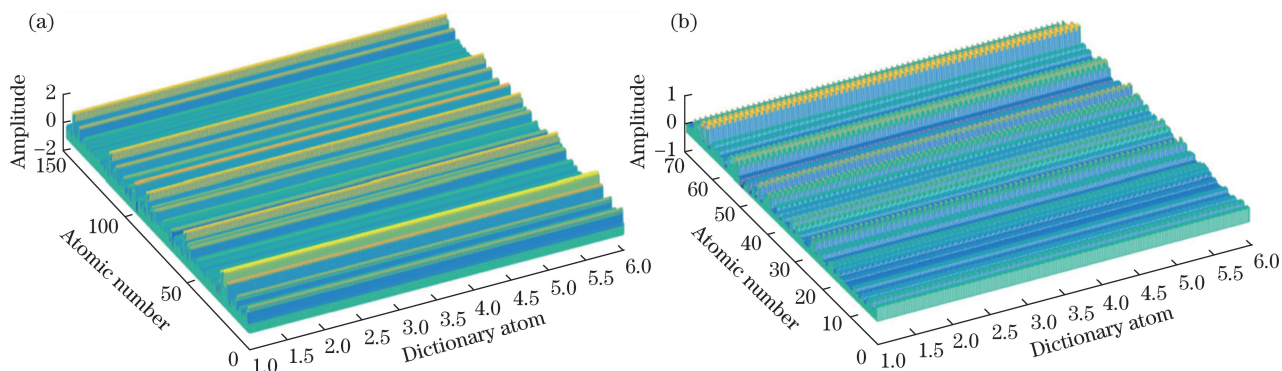


图 4 (a)原始选择的字典原子;(b)优化选择后的字典原子

Fig. 4 (a) Dictionary atoms of original selection; (b) dictionary atoms after optimized selection

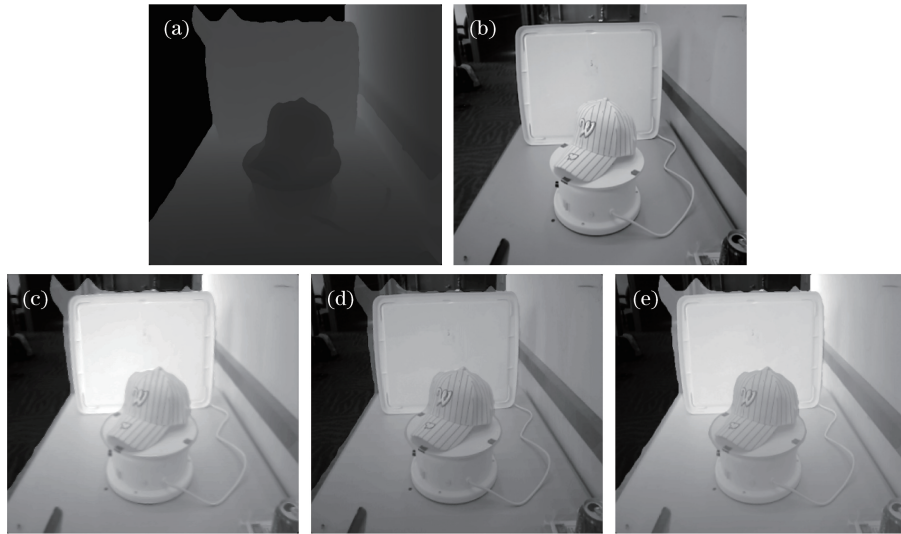


图 5 识别物体为帽子的 RGB-D 图像融合效果。(a)深度图像;(b)灰度图像;(c) PCA-KSVD 方法;
(d) NSCT-KSVD 方法;(e) K-mRMR 方法

Fig. 5 RGB-D image fusion effect of a hat used as the identification target. (a) Depth image; (b) gray image;
(c) PCA-KSVD method; (d) NSCT-KSVD method; (e) K-mRMR method

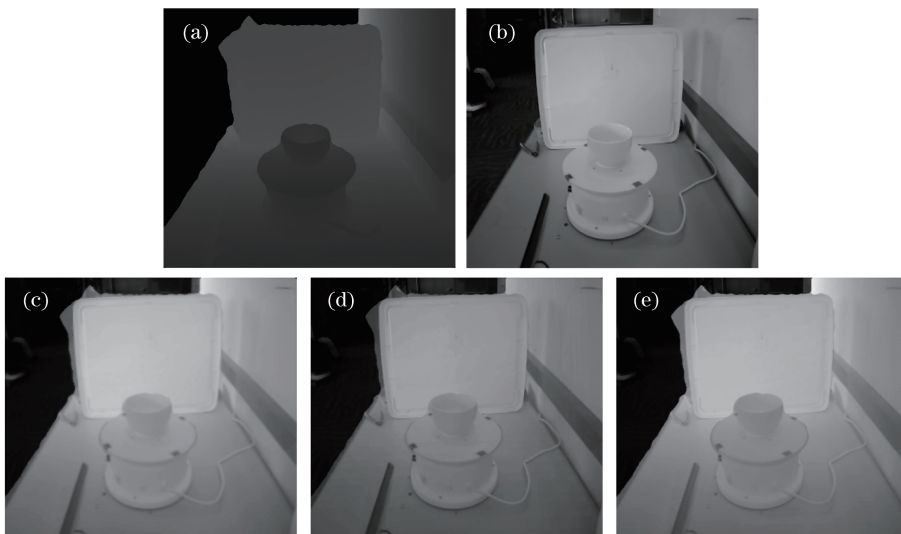


图 6 识别物体为碗的 RGB-D 图像融合效果。(a)深度图像;(b)灰度图像;(c) PCA-KSVD 方法;
(d) NSCT-KSVD 方法;(e) K-mRMR 方法

Fig. 6 RGB-D image fusion effect of a bowl used as the identification target. (a) Depth image; (b) gray image;
(c) PCA-KSVD method; (d) NSCT-KSVD method; (e) K-mRMR method

4.4 融合图像性能验证

融合后的图像对目标的边缘轮廓信息进行了强化,并在一定程度上增强了目标区域的灰度值与背景区域灰度值的区分度。图 5(b)中帽子和图 6(b)中碗的轮廓并不十分清晰,但经过与深度图像的融合后,识别物体的轮廓信息与背景区分更加明显。作为对比的 PCA-KSVD 字典学习算法得到的融合图像清晰度相对较低,NSCT-KSVD 字典学习算法得到的融合图像清晰度较高,但是由图 5(d)、图 6(d)可知,识别目标与背景的分并不是特别明显。

融合图像的评价指标如表 1 所示。通过分析表 1 可知,PCA-KSVD 字典学习算法的平均梯度最大,图像的清晰度最高;K-mRMR 算法的信息熵 H 要高于 PCA-KSVD 和 NSCT-KSVD 字典学习算法;在互信息 M_1 方面,K-mRMR 算法优势明显,说明该算法从源图像中提取融合了最大的信息量;K-mRMR 算法对物体的边缘起到了强化作用,从数据来看,边缘保持度 $Q^{AB/F}$ 也要高于另外两种算法;从结构相似度来看,3 种算法相差不多。综合分析,K-mRMR 算法得到的融合图像质量要高于

PCA-KSVD和 NSCT-KSVD 字典学习算法,相比之下具有更丰富的图像信息和更易于识别的图像特征。

表 1 融合图像评价指标

Table 1 Evaluation indexes of fusion images

Algorithm	A_{AVG}	H	M_1	$Q^{AB/F}$	S_{SIM}
PCA-KSVD	5.3784	6.3284	10.6381	0.6504	0.7338
NSCT-KSVD	6.7447	6.4763	11.7389	0.7123	0.7464
K-mRMR	6.5823	6.8841	13.5173	0.7462	0.7474

5 融合后图像的识别准确率对比

融合图像对 RGB 图像没有表达出的深度轮廓信息进行了补充,同时也强化了目标区域与背景区域灰度值的差异,因此针对轮廓、区域特征进行识别的算法在融合图像上的识别检测效果会得到改善。为了更进一步验证本文算法的融合图像特征的优势,使用尺度不变特征和支持向量机(SIFT+SVM)与卷积神经网络(CNN)^[16]两种算法对 4 种图像中的目标进行检测定位,这 4 种图像分别为单一 RGB 图像、单一深度图像、RGB 和深度联合图像,以及本文算法融合后的图像。

5.1 识别目标及评价指标

使用 RGB-D 数据库中的目标检测集进行实验,主要检测帽子、碗、手电筒和咖啡杯等 4 类物体。使用准确率和成功率来评价识别性能^[17]。准确率的定义是中心位置误差小于设定阈值的识别目标占识别总数的百分比,这里以 20 个像素点作为阈值。

成功率的定义为

$$S = \frac{|R_t \cap R_a|}{|R_t \cup R_a|}, \quad (11)$$

式中 R_t 为跟踪目标的边界区域, R_a 为目标真实的边框区域, \cap 、 \cup 分别表示对 R_t 和 R_a 做集合的交运算和并运算, $|\cdot|$ 表示集合中像素的个数。当 $S > 0.6$ 时,认为识别成功。

5.2 识别方法

使用 SIFT+SVM 和 CNN 两种算法对 4 种图像的识别检测流程如下:

1) SIFT+SVM 检测融合图像。对图像目标区和背景区域使用滑动窗提取一个个的图像块,对每个图像块使用 SIFT 算子提取特征向量,之后使用 SVM 分类器对目标区域和背景区域图像块对应的 SIFT 特征向量 V_1 和 V_2 进行分类训练,最后将训练好的分类器用于目标区域识别。单一的 RGB 图像检测和深度图像检测也使用此方法,只需要将融合图像换成各自的图像即可。

2) SIFT+SVM 检测 RGB-D 联合图像。使用 SIFT 算子分别提取 RGB 和深度图像块的特征向量 V_{RGB} 和 V_d , 然后对特征向量进行组合,即

$$V = aV_{RGB} + bV_d, \quad (12)$$

式中 a 、 b 代表两种图像特征向量对应的融合权重。对组合后的特征向量采用 SVM 分类器进行分类,最后用训练好的分类器对两种图像进行联合检测,算法流程如图 7 所示。

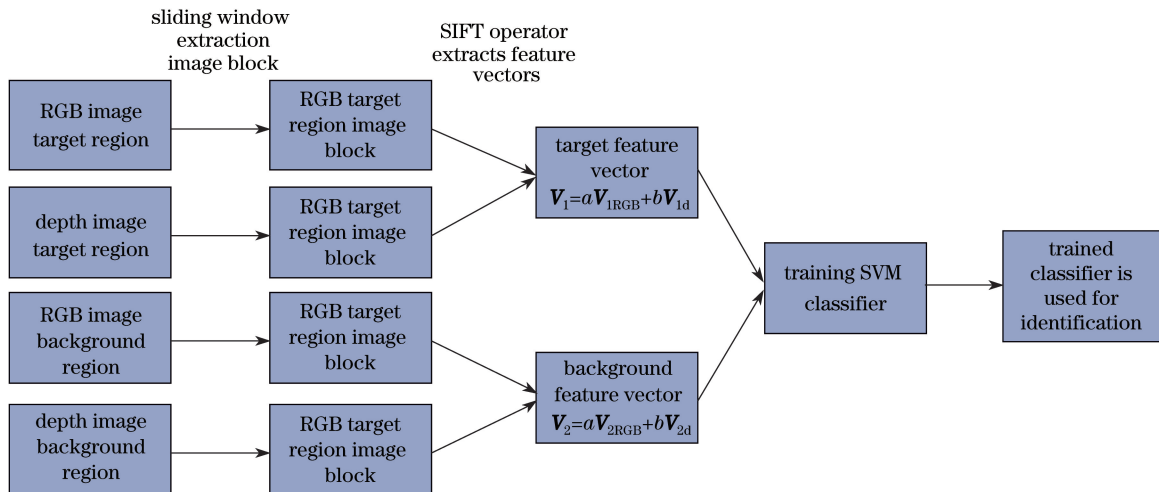


图 7 SIFT+SVM 的 RGB-D 联合检测

Fig. 7 Joint detection of RGB-D by SIFT+SVM

3) CNN 检测融合图像。将融合后的训练图像输入到 CNN 神经网络中,经过卷积网络的各层节点间运算后,将得到的输出与标签信息(目标区域中

心位置)进行对比,根据此误差对神经网络参数进行反向调节,直到误差小于一定值或达到一定训练步数后停止训练,利用训练好的模型对检测样本进行

测试,得到识别结果。使用此方法对单一的 RGB 图像和深度图像进行检测时,只需要将实验图像换成各自的图像即可。

4) CNN 检测 RGB-D 联合图像。同时将 RGB 图像和深度图像组合成一个 4 通道的信息输入

到卷积网络中,在卷积神经网络中将 4 通道的 RGB-D 信息进行融合,其示意图如图 8 所示,图中 C1~C7 代表卷积神经网络的卷积层,P1~P6 代表网络的池化层,FC1~FC3 代表网络的全连接层。

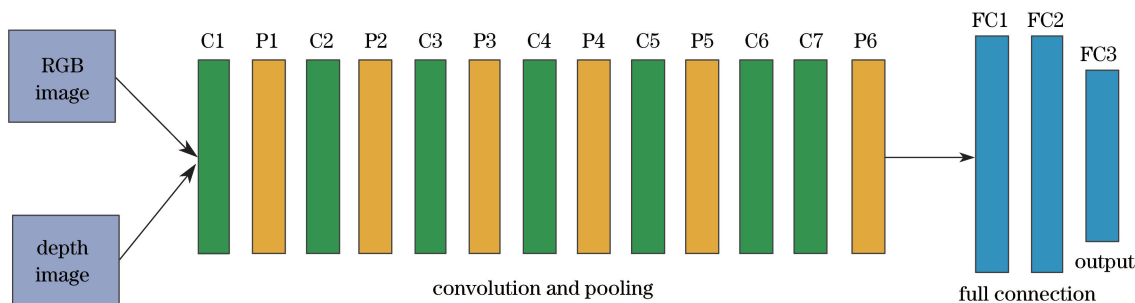


图 8 基于 CNN 的 RGB-D 联合检测

Fig. 8 Joint detection of RGB-D based on CNN

5.3 识别结果

每类识别物体选取 50 对 RGB-D 图像,4 类物体共选取 200 对图像,使用 K-mRMR 算法对这 200 对图像进行融合,得到 200 幅融合后的图像。按照

第 5.2 节中给出的识别方法并采用交叉验证法进行测试,即从每类物体图像中随机选取一半的图像用于训练,另外一半用于测试,反复进行 10 次实验,取 10 次实验结果的平均值,如表 2 和表 3 所示。

表 2 不同算法下识别准确率的对比

Table 2 Comparison of recognition accuracy under different algorithms

Algorithm	Recognition accuracy / %			
	RGB image	Depth image	RGB-D image	Proposed algorithm fused image
SIFT-SVM	74.5	71.3	78.5	81.4
CNN	80.8	77.4	85.8	88.2

表 3 不同算法下识别成功率的对比

Table 3 Comparison of recognition success rate under different algorithms

Algorithm	Recognition success rate / %			
	RGB image	Depth image	RGB-D image	Proposed algorithm fused image
SIFT-SVM	73.3	70.6	77.1	80.2
CNN	78.3	75.1	83.9	86.1

由表 2 和表 3 可以看出,无论是基于人工特征提取还是深度学习的识别检测方式,联合 RGB-D 图像进行检测的识别准确率和成功率都要高于单一的 RGB 图像或深度图像。RGB 图像和深度图像在同一成像位置所表达的特征信息既有重复又有互补。但由于 RGB-D 图像是两种不同模态的数据,因此在联合检测中,两种图像之间的重复冗余信息由于表达形式的不同而容易被判定为不同的特征,从而增加了提取特征向量的维度。另外,两种图像本身就存在着大量的空间和结构相关冗余,也会影响识别检测的精度和效率。

K-mRMR 融合算法基于 KSVD 字典稀疏图像

特征,利用 mRMR 原则求取维度最小且各维度之间相关性最小的特征原子组合,从而降低了两种图像本身的冗余度。该方法根据最大化稀疏系数的原则对重合部分的特征进行融合,使重合部分表达区分性更强的图像特征,弥补了单一图像特征不全面的缺点。因此,使用融合后的图像进行目标识别检测,有效提高了识别的准确率和成功率,从而验证了本文算法对于 RGB-D 图像联合检测的优越性。

6 结 论

提出一种应用于 RGB-D 场景图像融合的

K-mRMR算法,该算法结合了稀疏表达和 mRMR 选择准则,提取并融合 RGB 图像和深度图像中的有效信息,滤除了无关和冗余信息,得到了包含更多识别信息的融合图像,有效改善了 RGB-D 场景图像特征融合困难、联合识别效率不高的问题。将使用 KSVD 字典原子得到的稀疏系数作为特征融合参数,基于 mRMR 准则选取含有最多识别有效信息的稀疏系数,融合其对应的字典原子,能够更好地获取图像信息中最有利于识别的特征。建立了一种最优字典原子数选取原则,得到了融合后图像质量评价最高且原子数目最少的原子集合,提高了图像融合的效率。与 PCA-KSVD、NSCT-KSVD 稀疏融合算法相比,本文算法融合后的图像性能指标整体占优;利用 SIFT-SVM 和 CNN 识别多种类图像,本文算法融合的图像识别准确率和成功率更高。

参 考 文 献

- [1] Lu L F, Xie Z J, Ye H W. Object recognition algorithm based on RGB feature and depth feature fusing[J]. *Computer Engineering*, 2015, 42(5): 187-192.
卢良锋, 谢志军, 叶宏武. 基于 RGB 特征与深度特征融合的物体识别算法[J]. *计算机工程*, 2015, 42(5): 187-192.
- [2] Wang T, Zou Z L, Qiao M N. Object contour extraction based on image feature analysis[J]. *Journal of Beihang University*, 2016, 42(8): 1762-1767.
王田, 邹子龙, 乔美娜. 基于图像特征分析的物体轮廓提取[J]. *北京航空航天大学学报*, 2016, 42(8): 1762-1767.
- [3] Xiang X W. RGB-D saliency detection[D]. Guangzhou: Institutes of Technology of South China, 2015.
向训文. RGB-D 图像显著性检测研究[D]. 广州: 华南理工大学, 2015.
- [4] Yang B, Li S T. Multifocus image fusion and restoration with sparse representation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2010, 59(4): 884-892.
- [5] Yu N N, Qiu T S, Bi F, *et al.* Medical image features extraction and fusion based on K-SVD[J]. *Journal of Dalian University of Technology*, 2012, 52(4): 605-610.
余南南, 邱天爽, 毕峰, 等. 基于 K-SVD 的医学图像特征提取和融合[J]. *大连理工大学学报*, 2012, 52(4): 605-610.
- [6] Elad M, Abaron M. Image denoising via sparse and redundant representations over learned dictionaries[J]. *IEEE Transactions on Image Processing*, 2006, 15(12): 3736-3745.
- [7] Olshausen B A, Field D J. Sparse coding of sensory inputs[J]. *Current Opinion in Neurobiology*, 2004, 14(4): 481-487.
- [8] Li B, Zhang P L, Ren G Q, *et al.* Features selection for rolling bearing fault diagnosis based on mutual information[J]. *Journal of Test and Measurement Technology*, 2009, 23(2): 184-190.
李兵, 张培林, 任国全, 等. 基于互信息的滚动轴承故障特征选择方法[J]. *测试技术学报*, 2009, 23(2): 184-190.
- [9] Ohn-Bar E, Trivedi M M. Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(6): 2368-2377.
- [10] Wang W, Chen L, Liu Z, *et al.* Textured/textureless object recognition and pose estimation using RGB-D image[J]. *Journal of Real-Time Image Processing*, 2015, 10(4): 667-682.
- [11] Liu J R, Li C P, Ouyang J Q, *et al.* Depth image enhancement algorithm based on joint bilateral filtering[J]. *Computer Engineering*, 2014, 40(3): 249-253.
刘金荣, 李淳芑, 欧阳建权, 等. 基于联合双边滤波的深度图像增强算法[J]. *计算机工程*, 2014, 40(3): 249-253.
- [12] Zhang C C. Detection based on sparse representation RGB-D object[D]. Hangzhou: Zhejiang University, 2015.
张铨程. 基于稀疏表示的 RGB-D 物体检测[D]. 杭州: 浙江大学, 2015.
- [13] Yin W, Li Y X, Zhou Z M, *et al.* Remote sensing image fusion based on sparse representation[J]. *Acta Optica Sinica*, 2013, 33(4): 0428003.
尹雯, 李元祥, 周则明, 等. 基于稀疏表示的遥感图像融合方法[J]. *光学学报*, 2013, 33(4): 0428003.
- [14] Ouyang N, Zheng X Y, Yuan H. Multi-focus image fusion based on NSCT and sparse representation[J]. *Computer Engineering and Design*, 2017, 38(1): 177-182.
欧阳宁, 郑雪英, 袁华. 基于 NSCT 和稀疏表示的多聚焦图像融合[J]. *计算机工程与设计*, 2017, 38(1): 177-182.
- [15] Qin L J, Gu Y J, Jiang L L. Research on combined filtering method of PCA and K-SVD[J]. *Electro-optic Technology Application*, 2016, 31(4): 32-37.
秦丽娟, 谷雨, 蒋磊磊. PCA 与 K-SVD 联合滤波方

- 法的研究[J]. 光电技术应用, 2016, 31(4): 31-36.
- [16] Tu S Q, Xue Y J, Liang Y, *et al.* Review on RGB-D image classification[J]. Laser and Optoelectronics Progress, 2016, 53(6): 060003.
涂淑琴, 薛月菊, 梁云, 等. RGB-D 图像分类方法研究综述[J]. 激光与光电子学进展, 2016, 53(6): 060003.
- [17] Wang H Q, Cai Y N, Fu G Y, *et al.* Image recognition and tracking based on image sequences for slow moving targets in the ground[J]. Laser and Optoelectronics Progress, 2016, 53(5): 051501.
汪洪桥, 蔡艳宁, 付光远, 等. 基于图像序列的地面慢动多目标识别与跟踪[J]. 激光与光电子学进展, 2016, 53(5): 051501.