



一种融合 MeanShift 聚类分析和卷积神经网络的 Vibe++ 背景分割方法

刘子豪¹, 贾小军¹, 张素兰¹, 徐志玲², 张俊³

(1. 嘉兴学院数理与信息工程学院, 浙江 嘉兴 314033;

2. 中国计量大学质量与安全工程学院, 浙江 杭州 310018;

3. 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310058)

摘要: 针对传统 Vibe+ 算法存在噪点和拖影分割错误率较高的问题, 提出了一种改进的 Vibe+ 运动目标分割算法 (Vibe++)。首先, 通过对视频帧采用传统 Vibe+ 算法处理获取二值图像, 基于区域生长算法对结果图中各连通域标记, 依据边界面积块差异获取面积筛选阈值, 将低于阈值的连通区域视为噪点并删除; 然后, 引入 5 种不同核函数优化传统 MeanShift 聚类算法, 并与卷积神经网络 (CNN) 进行顺序组合; 最后, 采用组合模型对已消除噪点图像中的拖影区、非拖影区和拖影边缘区分类, 计算拖影区中每个像素点的坐标, 定位拖影区并快速删除, 获取分割结果。所提算法用于公开数据集的实验结果表明, 其可取得 98% 以上的分割准确率, 具有较好的应用效果和较高的实用价值。

关键词: 背景分割; 聚类分析; 分割准确率; 卷积神经网络

中图分类号: TN391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021046

Vibe++ background segmentation method combining MeanShift clustering analysis and convolutional neural network

LIU Zihao¹, JIA Xiaojun¹, ZHANG Sulan¹, XU Zhiling², ZHANG Jun³

1. College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314033, China

2. College of Quality & Safety Engineering, China Jiliang University, Hangzhou 310018, China

3. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

Abstract: To solve problems of noise points and high segmentation error for image shadow brought by traditional Vibe+ algorithm, a novel background segmentation method (Vibe++) based on the improved Vibe+ was proposed. Firstly, binarization image was acquired by using traditional Vibe+ algorithm from surveillance video. The connected regions were marked based on the region-growing domain marker method. The area threshold was obtained with dif-

收稿日期: 2020-12-09; 修回日期: 2021-02-17

基金项目: 浙江省基础公益研究计划项目 (No.LGG21F030013, No.LGG20F010010, No.LGG20F030006); 嘉兴市公益计划项目 (No.2020AY10009, No.2018AY11008); 嘉兴学院科研启动基金 (No.CD70519085)

Foundation Items: Zhejiang Public Welfare Technology Research Project (No.LGG21F030013, No.LGG20F010010, No.LGG20F030006), Jiaxing City Public Welfare Technology Application Research Project of Jiaxing Science and Technology Bureau (No.2020AY10009, No.2018AY11008), Scientific Research Foundation of Jiaxing University (No.CD70519085)



ference characteristics of boundary area, the connected regions below threshold were treated as disturbing points. Secondly, five different kernel functions were introduced to improve the traditional MeanShift clustering algorithm. After improving, this algorithm was fused effectively with partitioned convolutional neural network. Finally, program of classification of trailing area, non-trailing area and trailing edge area in the resulting image was performed. Position coordinates of the trailing area were calculated and confirmed, and the trailing area was quickly deleted to obtain the final segmentation result. This segmentation accuracy was greatly improved by using the proposed method. The experimental results show that the proposed algorithm can achieve segmentation accuracy of more than 98% and has good application effect and high practical value.

Key words: background segmentation, clustering analysis, segmentation accuracy, convolutional neural network

1 引言

近年,动态目标的背景分割一直是视频处理领域的研究热点,因其面向应用场景的广泛性和多样性,吸引了大量研究者。传统动态目标的背景分割方法有背景差分法^[1]、光流法^[2]、分水岭算法^[3]、帧差法^[4]、水平集法^[5]、背景更新恢复法^[6]等,这些方法可以滤掉视频中的各种干扰背景,但也存在诸如分割目标残留、噪声分割错误以及拖影无法有效判别等问题。为了解决这些问题,有学者提出采用基于 Vibe 背景分割^[6]方法实现动态目标的精准分割。

Vibe 背景分割算法最早由 Barnich 在 2011 年提出^[7],主要思想是通过不断收集更新背景空间图样构建背景分割模型,该算法在空间背景图样的随机选择和模型更新中提出了一种空间传播机制,可将背景像素值插入邻域像素点的模型样本库中,具有较大的灵活性,但是也存在诸如抖动参数优化、背景噪声抑制不佳、填充面积不完善等问题,因此在随后的一段时间,较多相关研究者基于以上问题对传统 Vibe 算法进行了改进,例如 2013 年, Gulesir 等^[8]在原始 Vibe 方法中加入了背景像素与前景像素的函数筛选规则,探索如何最大限度地保留前景运动目标和剔除背景噪声目标;2017 年, Zhou 等^[9]通过深度融合目标像素线索和可视化背景样本图样,提出了 Vibe+算法,在一定程度上对原始 Vibe 算法中存在的问题进行了修正和优化。然而, Vibe+算法在以下两个应用中存在问题。

(1) 当监控视频中出现运动的行人、机动车辆、非机动车辆时,采用 Vibe+算法进行前景提取很容易因为树叶随风摆动或者河流流动的运动梯度模式与目标运动模式一致而导致前景目标提取失败。

(2) 在自然场景下,天气有晴天、阴天、雨天、雪天等,不同的天气条件会导致分割异常,例如,当太阳光直射在目标表面产生高光区时,常常会因为高光区域的图像信息失真,目标提取不完全,从而导致分割失败,而且阳光下运动目标的拖影较明显,拖影会随着目标运动而运动, Vibe+算法会使得拖影部分附着在前景目标中一起被分割出来,从而导致对运动目标的分割失败。

针对以上问题,本文在 Vibe+方法的基础上引入卷积神经网络(CNN)和改进 MeanShift 聚类算法,实现运动目标的精准分割。首先,采用传统的 Vibe+算法获取视频帧二值图像,利用区域生长算法对轻微扰动噪声删除;然后,基于 CNN 对运动目标拖影区域进行检测和识别,联合改进的 MeanShift 聚类分析对图像中不同目标进行分类,获取关于拖影特征的分类结果,进行拖影的消除处理;最后,整合以上两个步骤,完成前景运动目标提取。

2 改进 MeanShift 聚类分析与卷积神经网络的融合方法

2.1 噪声扰动消除算法

在采用 Vibe+算法处理后的视频中,轻微抖动

的目标例如树叶、微波扰动的湖面、被风吹动的条幅等均会在背景分割的二值化视频帧中存在相对应的噪声干扰点,这些干扰点具有断断续续且相互联接不紧密的特点,而且噪声大小和强弱随机,即无法对噪声进行建模定位和规则化处理,需要针对目标视频中每一帧图像由非目标引起的干扰点进行处理,因此,本节采用基于区域生长算法对分割图像中所有的连通区域进行标记。

2.1.1 区域生长算法

对 Vibe+算法处理后的二值图像中各个连通区域采用基于区域生长算法^[10]进行标记,其主要思想为:对其中每个像素点遍历其周围八邻域内的像素点,当检测到存在相邻像素点与当前像素点连通时,将当前像素点周围八邻域内的面积子区域与各个相邻像素点周围八邻域内的面积子区域共同确定为连通区域,因此可实现目标块区域的自动标记。

2.1.2 区域填充与阈值分割

为了避免算法过多标记出多个较小的连通域而使得连通域标记法算法执行缓慢,本节在标记连通区域后采用区域填充算法^[11]对各个二值图像中的封闭区域进行像素值为1的像素点填充,从而加快算法程序执行速度,获取完整大块的连通区域。在获取连通区域个数及其面积数值之后,

依次对连通区域的面积数值进行由大至小排序,并对相邻两个面积数值的面积差值进行统计,确定面积差值最大值对应的两个面积数值,将两个面积数值的平均值确定为面积筛选阈值,由于同一个视角拍摄的视频中运动目标尺寸相差较小,通过 Vibe+算法处理后生成的二值图像中除了有大块运动目标,还存在微小的点,这些较小的点可认为是噪声所引起的干扰因素,因此把所有小于面积筛选阈值所对应的连通区域全部删除,即可通过去噪点获取处理结果。

2.2 拖影区域检测与识别算法

在阳光充足的晴天开放式场景下,运动目标拖影是监控视频中最常见的另一类背景分割干扰项,为了消除由拖影带来的过分割问题,本节首先对传统的 MeanShift 聚类算法^[12]进行改进,然后让其与卷积神经网络^[13]进行融合,最后消除拖影。图1给出了这两种算法融合的流程。

2.2.1 改进 MeanShift 聚类算法

MeanShift 算法最早由 Fukunage 在 1975 年提出^[14],其最初表示偏移均值向量,随着理论的发展,MeanShift 的含义已经发生了诸多变化。之后,1995 年,Cheng^[15]定义了一族核函数,使得随着样本与被偏移点的远近距离不同,其偏移量对均值偏移向量的贡献也不同,作者首次提出采用单

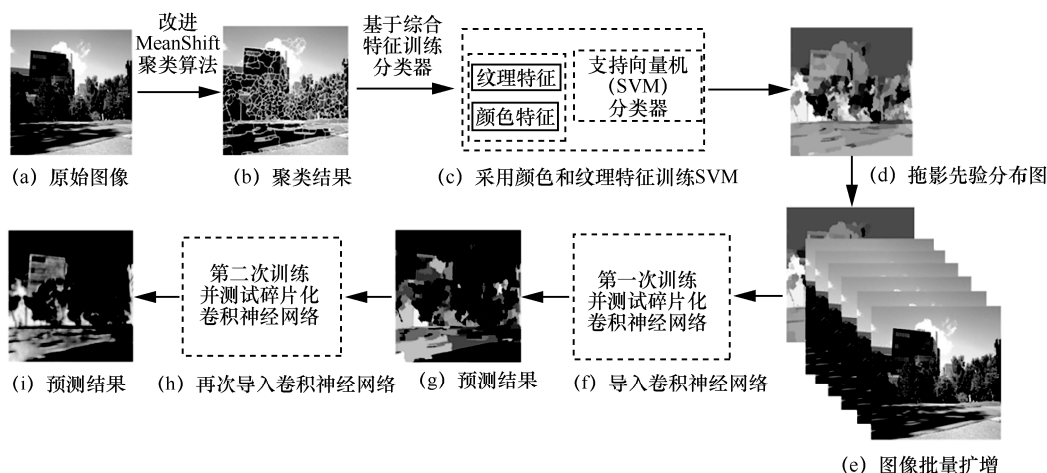


图1 融合算法的执行步骤与流程示意图



体核函数对像素点的漂移向量进行计算,获取了较好的结果,但是该方法由于把像素点由低维空间映射到高维空间中,在计算漂移向量不同空间维度的转换位置迭代时,会不断地有偏移量错误累积,存在较大计算误差,因此引入了5种核函数^[16]的均值组合模式,从感兴趣区域中心点出发寻找最佳目标像素点,最小化累积误差,同时达到最小化向量投影到高维空间中误差的目标。与传统基于单一数据密度变化不同,本节对不同维度空间漂移向量的映射加入以数据密度为基础计算漂移向量的综合均值核函数,将每个图像像素点移动到密度函数的局部极大值点处(密度梯度为0),收敛到概率密度最稠密处,获取最优的收敛结果,可降低图像像素点的漂移向量计算误差。改进 MeanShift 聚类算法的核心思想在于:引入线性核函数、多项式核函数、高斯核函数、多元二次核函数和 Sigmoid 核函数的均值组合模式,共同对视频图像进行聚类计算,得到视频图像中对应的各个区域块图像,然后依据最小化漂移向量计算误差为原则,得到最优的核函数。改进 MeanShift 所采用的5种核函数如下所示:

$$k_1(x, y) = x^T \times y + c \quad (1)$$

$$k_2(x, y) = (a \times x^T \times y + c)^d \quad (2)$$

$$k_3(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (3)$$

$$k_4(x, y) = \sqrt{\|x - y\|^2 + c^2} \quad (4)$$

$$k_5(x, y) = \tanh(a \times x^T + c) \quad (5)$$

其中,线性核函数的表达式为式(1),多项式核函数的表达式为式(2),高斯核函数的表达式为式(3),多元二次核函数的表达式为式(4),Sigmoid 核函数的表达式为式(5)。在上述表达式中,变量 x 表示视频帧提取特征, y 表示该特征被映射后的结果向量, a 、 c 表示常数项,而 σ 表示函数的惩罚系数,该系数越大意味着核函数的

值越小,反之就越大。

对5种核函数所采用的均值综合计算方法在于获取5种核函数在不同空间转换过程中漂移向量的误差累积值,在每次迭代中对误差较大的核函数乘以较大的惩罚因子,而对误差较小的核函数乘以较小的惩罚因子,而且加入常数项以弥补漂移向量空间转换的参数缺失带来的不良效果,核函数均值综合计算式如(6)所示。

$$K(x, y) = \frac{\sum_{i=1}^N \sum_{j=1}^N \sigma_i k_j(x, y)}{N^2} \quad (6)$$

其中, $K(x, y)$ 表示核函数的均值综合项, i 和 j 分别表示惩罚因子和不同核函数的下标表示, σ_i 表示不同核函数的惩罚因子项, $k_j(x, y)$ 表示不同核函数公式, N 表示所采用核函数的数量。

通过以上理论分析,本文对传统 MeanShift 聚类分析算法改进的步骤如下。

步骤 1 对于原始图像,对其中未被标记的图像像素点随机选择一个点作为起始中心点 Center。

步骤 2 将以 Center 为中心点、半径为 Radius 的区域范围内出现的所有像素点类别确定为类别 C ,同时在该聚类中记录数据点出现的次数自增 1。

步骤 3 以 Center 为中心点,计算从 Center 开始到集合 M 中每个元素的向量,分别计算向量利用不同核函数从低维空间到高维空间的映射结果,取5种核函数计算结果的误差最小项作为高维空间漂移向量 Shift 取值。

步骤 4 $\text{Center} = \text{Center} + \text{Shift}$,即 Center 沿着 Shift 的方向移动,移动距离采用欧氏距离计算公式: $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$,移动方向采用 Center 的移动方向即可。

步骤 5 重复执行步骤 2~步骤 4,重复迭代直到 Shift 收敛,取该算法迭代到收敛时的 Center 作为最后的执行结果,此迭代过程中遇到的点都

归类到当前簇 C 中。

步骤 6 如果收敛, 当前簇 C 的 Center 与其他已经存在的簇 $C2$ 中心的距离小于阈值, 那么把簇 $C2$ 和簇 C 合并, 数据点出现次数也对应合并, 否则, 把簇 C 作为新的聚类。

步骤 7 重复执行步骤 1~步骤 5, 直到二值图像中所有像素点都被标记为已访问。

步骤 8 对每个像素点进行分类, 根据每个类对每个像素点的访问频率, 取访问频率最大的那个类作为当前点集的所属类。

改进 MeanShift 聚类算法的核心步骤如图 2 所示。

在一幅视频获取的图像中, 存在 3 种需要检测的拖影区域: 拖影、非拖影、拖影边缘, 其中拖影边缘包含了拖影的轮廓信息, 而拖影区域即该轮廓所包含的内部阴影区域, 因此拖影边缘的检测是关键, 接下来在基于改进 MeanShift 聚类的结果基础上, 采用卷积神经网络对拖影边缘区域进行检测和识别。

2.2.2 卷积神经网络算法

本节所采用的拖影特征区域提取方法选用 3

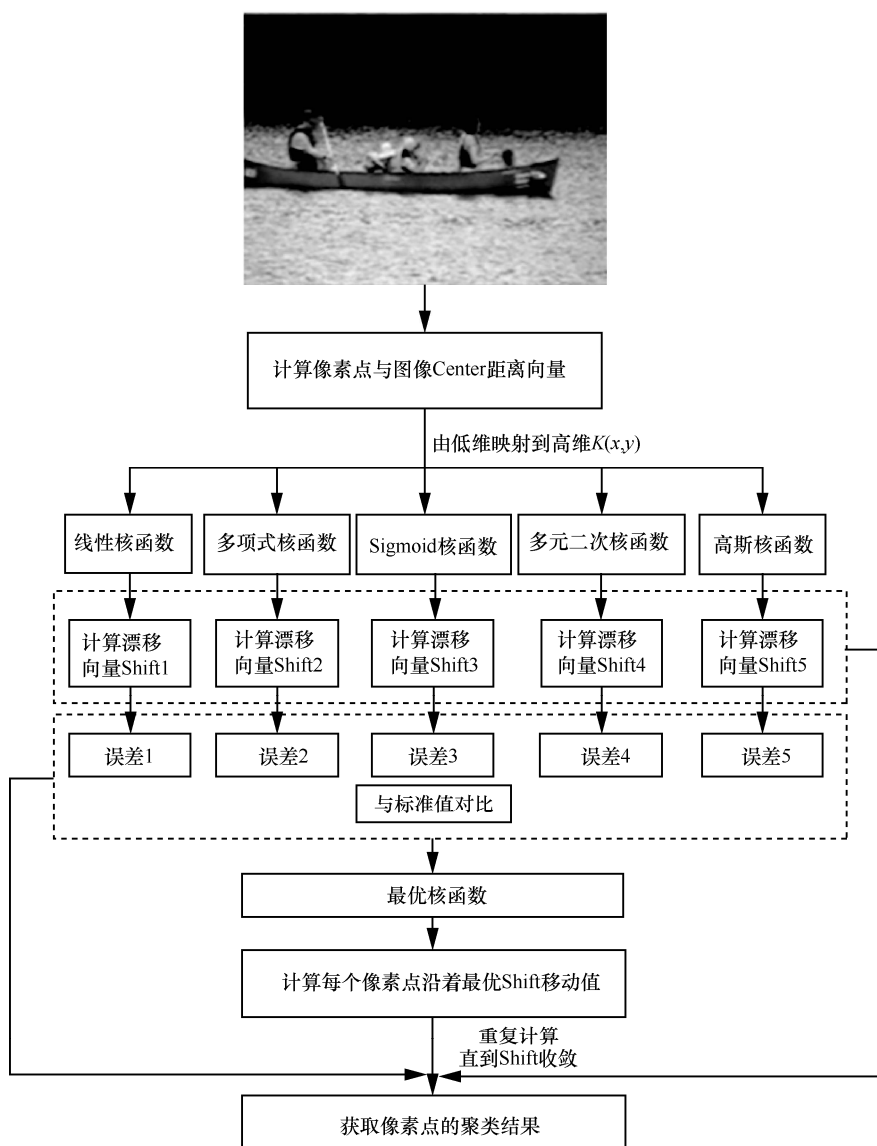


图 2 改进 MeanShift 聚类算法的核心步骤



种类型的训练图像类型：拖影区、非拖影区、拖影边缘区，训练数从 RGB 图像和阴影先验分布图融合结果中获取，所采用的 CNN 包括 6 个卷积层、两个池化层和两个全连接类别输出层，其输入数据为 224×224 图像块，输出数据为对应区域的拖影概率预测值，该网络特别在最后的全连接层 FC1 后加入基于组合分类器判别模式的新全连接层 FC2，分别进行 3 种分类器的综合分类，然后再基于 IMAJ 分类器融合规则^[17]获取最优的拖影分类结果，这样做的目的在于使得输出的概率更加接近于实际的类别标签，呈现一定的相似像素索引性。卷积神经网络结构示意图如图 3 所示。

在构建卷积神经网络后，以各个训练样本图像所对应原始 RGB 图像和阴影先验分布图作为训练数据获取源，采用拖影区域、非拖影区域、拖影边缘区域作为 3 种训练图像类型，对卷积神经网络进行训练，得到模型用于测试视频帧获取最佳的预测结果。

2.2.3 融合算法

通过以上改进 MeanShift 聚类算法执行后，可形成如图 1 (b) 所示的聚类结果，以每个区域块图像为基础在 Lab 颜色空间^[18]提取颜色直方图，再提取每一视频帧的纹理特征 Texton 直方图^[19]，训练支持向量机 (SVM) 分类器^[20]，生成阴影先验分布图，该图构建过程遵循图 1 (b) → 图 1 (c) → 图 1 (d) 顺序。接下来，考虑基于 CNN 获取精确的拖影识别结果。然而，由于训练 CNN 的准

确率极大地依赖于大批量的带拖影标注的图像或视频数据样本，现今公开的数据集中并没有大量拖影标注图像，因此考虑采用阴影先验分布图获取批量训练图像，再使用卷积神经网络作为训练图像模型，在测试阶段获取输入图像的拖影预测概率值。基于以上算法描述，将改进的 MeanShift 算法与卷积神经网络进行有效融合，具体步骤如下。

步骤 1 对于目标监控视频中的每帧图像，根据改进 MeanShift 聚类算法与第一次卷积神经网络的融合算法，识别出该帧图像中的拖影区域，并获取各个拖影区域在视频图像所对应去噪点分割处理后二值图像的拖影位置坐标，将二值图像中各个拖影位置坐标所对应像素点的像素值置为 0，得到第一次拖影分割后的二值图像；根据第一次卷积神经网络计算得到的各个区域块图像为拖影的概率预测值，计算得到视频图像中的各个拖影边缘区域，该区域由自然场景下的目标与拖影共同组成，散落在如图 1 (g) 所示的图像空间中，将拖影概率预测值大于或等于第一预设概率预测值，且小于或等于预设概率预测值的区域块图像确定为拖影边缘区域，然后将确定出的各个拖影边缘区域再一次导入卷积神经网络进行计算，采用超参数自适应调整策略中的超参数，例如学习率、样本批次、卷积核尺寸、卷积核数量和卷积步长，然后选择最优的参数，输入卷积神经网络，并向卷积神经网络输入大批量无标签的拖影边缘

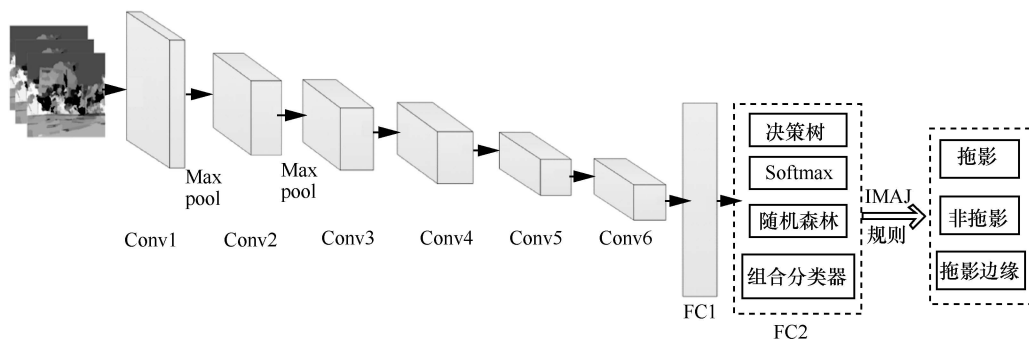


图 3 卷积神经网络结构示意图

区域进行无监督学习,由卷积神经网络输出对不同拖影边缘区域的拖影预测值。

步骤 2 采用线性判别分析法对不同拖影边缘区域的拖影预测值进行统计分析,将拖影预测值相同的各个拖影边缘区域聚合为一类,即最小化同类拖影边缘区域之间的距离,最大化不同类拖影边缘区域之间的距离,以此为基础从中寻找最大化的拖影边缘区域分类边界,求出明显的边缘分界曲线。

步骤 3 采用融合约束函数 $R(S)$ 对同类拖影边缘区域的图像进行边缘融合,融合约束函数的表达式为:

$$R(S) = \left\{ s_i \in S \mid s_i \geq \alpha \max_{1 \leq i \leq m} \max_{1 \leq i \leq m} (s_i) \right\} \quad (7)$$

其中, m 为分割区域的数量, α 为固定常数项, $R(S)$ 表示所包含的区域之间的边界像素集合。

通过上述步骤可以获取关于视频帧中所有运动目标的拖影区域坐标向量,记录并保存这些向量。本文通过 Vibe+运动目标分割算法可以提取完整的运动目标前景图像,其中可能包含有噪点和拖影,经过第 2.1 节的处理,可以剔除其中的噪点

部分,记此步产生的二值图像为 M_1 ;然后,通过第 2.2 节的处理,可检测出原始图像中是否包含拖影部分,检测后生成的拖影二值图像记为 M_2 ;最后,提取 M_2 中所有像素值为“1”的像素点坐标向量,把所有出现在 M_1 中的坐标点所对应的像素值全部置 0,即可获取最终的分割结果。

步骤 4 对于去拖影后的图像,计算该图像上下左右 4 个极值点坐标,在所对应视频图像中根据 4 个极值点坐标确定运动目标图像,实现对目标监控视频中运动目标背景的分割。

本文所提 Vibe++算法的总体流程如图 4 所示。

3 实验结果与分析

本文主要采用公开的视频数据集^[6]验证所提方法的有效性,数据集中包含 31 个视频序列,其中可分为 6 种类型:基准类(baseline)、动态背景类(dynamic background)、摄像机抖动类(camera jitter)、间断运动目标类(intermittent object motion)、拖影类(shadow)、热图类(thermal)。每个类别中的视频来源非常广泛,涵盖了自然场景、交通类、室内商场、运动员比赛、行人步行、湖

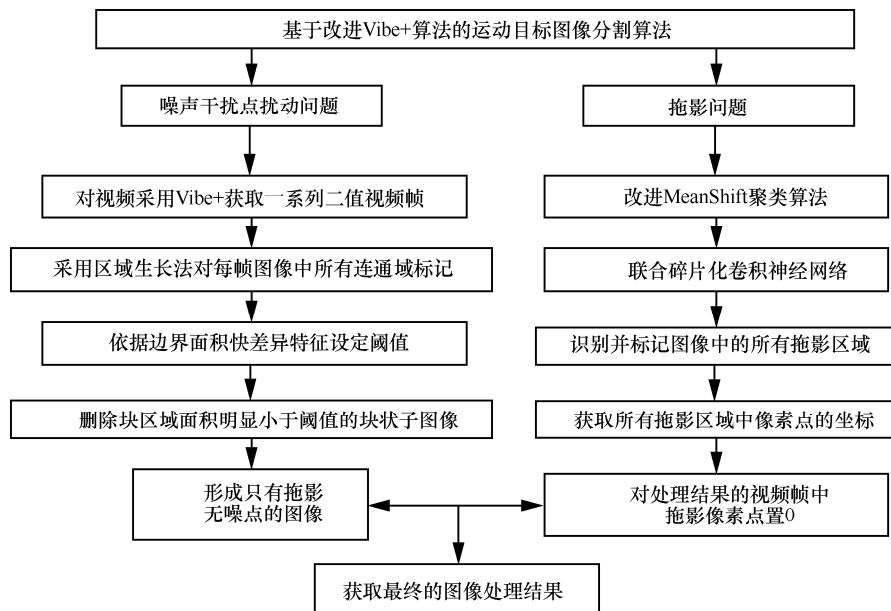


图4 本文算法总体流程



面、运动场面等视频^[21],对所提算法的对比与验证具有较好的测试性,该数据集包含的部分视频图像如图 5 所示。



图 5 公开数据集中的部分视频画面

基于 5 个场景类别,所提算法与传统背景分割算法的对比结果如图 6 所示,其中 Previous best 表示以往较优的运动背景分割算法,Baseline 表示基准对比图。

从图 6 可以看出,所提算法在拖影类、抖动类和动态背景类 3 类数据中可发挥较大优势,可一定程度上消除拖影和噪点,而且使得热图类和连续运动类得出的结果与基准图像保持一致,具有一定的适应性和鲁棒性。为了进一步验证本文所提算法有效性,且相比于其他传统经典背景分割算法的优势,本文选择了 8 种算法进行对比,实验结果见表 1。其中视频帧分割

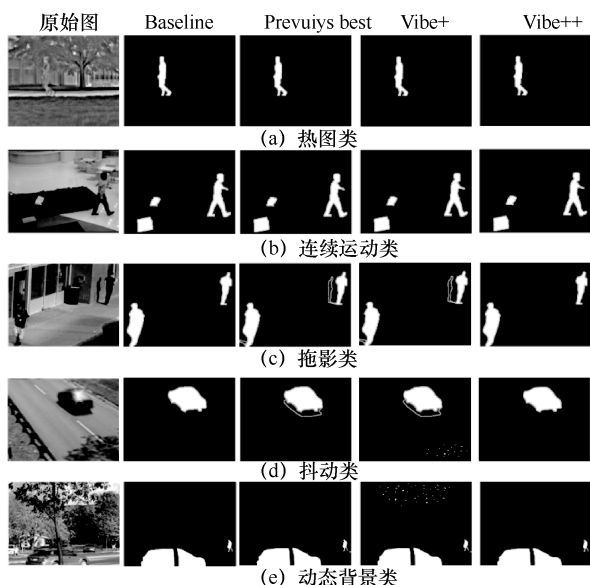


图 6 所提算法与传统背景分割算法的对比结果

时长表示分割每一帧图像所花费的时间,总体分割准确率(overall accuracy)可以用式(8)来表示:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

其中,TP(true positive)表示真阳性,即相比于基准图像,把本来正确的像素点分割正确的像素点数量;FP(false positive)表示假阳性,即相比于基准图像,把本来错误的像素点误识别为正确像素点的数量。其中较优的结果用加粗字体表示。

表 1 所提 Vibe++ 算法性能评价

分割算法	算法性能评价指标		
	视频帧 分割时长/ms	分割 准确率	算法时间 复杂度
背景差分法 ^[1]	23.5	83.31%	$O(n)$
光流法 ^[2]	86.9	93.12%	$O(n \lg n)$
分水岭算法 ^[3]	56.9	89.78%	$O(\lg n)$
帧差法 ^[4]	12.4	85.72%	$O(n)$
水平集法 ^[5]	42.1	87.93%	$O(n^2)$
Vibe ^[6]	89.4	92.13%	$O(m \times n)$
Vibe+ ^[7]	107.6	95.45%	$O(n \lg(mn))$
背景更新恢复法 ^[8]	34.5	86.05%	$O(n^3)$
Vibe++	132.7	98.58%	$O(n^2 \lg(mn^2))$

从表 1 中可以看出,相比于其他经典的视频背景分割算法,本文所提 Vibe++在分割准确率方面是最优的,这表明 Vibe++算法对经典算法中的分割错误问题,比如抖动噪声无法实时剔除、拖影分割不完全以及背景模型更新失败等方面的问题进行了修正,在原始的 Vibe+基础上引入了区域生长算法和深度学习加聚类的融合算法,消除和避免了此类问题带来的消极影响,给监控视频中运动目标的分割带来了性能提升。然而,从每一帧视频分割的时间和算法复杂度方面来看,本算法消耗的时间最多且复杂度较高,分割一张 2 MB 图像需要 132.7 ms,相比于帧差法的 12.4 ms,足足多了 120.3 ms,但是在高性能的现代计算机的中央处理器高速发展的今天,这样的分割时间是可以接受的,尤其以后随着高效率的图像矩阵处理芯片 GPU 不断推出,图像处理速度会不断提升,图像分割时间的可容忍范围也在不断扩大。

为了详细地在公开的视频数据集中给出每一类别的分割准确率,本文采用了 3 种方法进行对比研究——传统较优分割算法、Vibe+以及本文所提出的 Vibe++。表 2 给出了采用本文所提方法结合公开的视频数据集中的 5 个类别图像分割的实验结果,其中较优的结果用加粗字体表示。

表 2 本文所提方法与先前提出的方法对比

对比的 数据集类型	所提方法相比于之前方法的分割准确率		
	Previous best	Vibe+	Vibe++
基准类	95.32% ^[22]	92.62%	94.89%
动态背景类	92.12% ^[6]	72.91%	94.32%
抖动类	95.16% ^[6]	80.64%	98.05%
间断运动目标类	96.62% ^[23]	75.12%	95.98%
拖影类	94.72% ^[6]	83.02%	97.68%
热图类	98.29% ^[24]	93.63%	98.16%
所有类 (Overall)	97.36% ^[6]	87.45%	98.58%

对于每一个单独的视频数据集来说,本文所提的算法分别对 3 个数据集(动态背景类、抖动

类和拖影类)中出现的错误分类像素进行了修正和改进,取得了一定效果,所有类的分割准确率达到了 98.58%,比之前最好的 97.36%提高了 1.22%,其中主要的提升点主要在动态背景类、抖动类和拖影类 3 个类别的错误像素的修正方面;从单独数据集来看,对于抖动类视频数据,由于本文加入了区域生长算法对噪声点和抖动像素点进行了识别和剔除,并采用面积阈值特征抑制了噪声点的形成规模,与 Vibe+可以取得 80.64%的分割准确率相比,Vibe++的分割准确率可以达到 98.05%,提升了 17.41%;而对于拖影类视频数据,由于在原始的 Vibe+算法的像素背景自动更新模块中加入了改进的 MeanShift 聚类分析和卷积神经网络,新提出的方法除了兼顾 Vibe+算法的优点,还具有拖影边缘的识别和检测功能,与原始的 Vibe+可取得的 83.02%的识别精度相比,Vibe++算法可以取得 97.68%,提升了 14.66%,所提升的分割精度部分即原始 Vibe+算法对阴影区域无法识别的错误率,因此所提算法具有更好的适应性和鲁棒性。

从以上的实验结果中可以看出,所提算法取得了较高的分割准确率,为了验证所提算法中所加入的卷积神经网络的不同结构中,哪些因素对最终的分割准确率的提升产生正向贡献,这里展开消融实验在改进算法中寻优出最佳的卷积神经网络架构,以期在保证较高分割准确率的前提下,降低算法复杂度,缩短算法运行时间。表 3 展示了改变 CNN 不同结构下所获取的分割准确率以及分割一幅图像的执行时间。

从表 3 可知,当删除全连接层 FC1 或 FC2 后,两者产生的效果截然不同,删除 FC1 之后,模型的分割准确率只下降了 0.7%,只是验证精度略有下降;而当删除全连接层 FC2 之后,模型的分割准确率下降了 20.26%,即图像背景分割错误率会提升 20.26%,这是因为在 FC2 中构建了一种包含有 3 种分类器的组合分类器,其中包含网络的大



表 3 CNN 在不同结构下获取的分割准确率以及分割时间

序号	CNN 结构变更	分割准确率	分割时间/ms
1	删除 FC1	98.51%	118.91
2	删除 FC2	78.32%	103.54
3	删除 Pooling1	97.45%	124.76
4	删除 Pooling2	97.28%	128.54
3	只保留 Conv1	68.25%	73.75
4	只保留 Conv2	69.36%	69.82
5	只保留 Conv1+Conv2	83.48%	83.24
6	删除 Conv2+Pooling2	85.31%	112.42
7	删除 Conv1+Pooling1	87.37%	108.31
8	删除 Conv5	90.23%	122.13
9	删除 Conv6	91.46%	124.54
	删除 Conv5+Conv6	89.79	101.85
10	调整 Conv1,2 卷积核数量 32,64	95.14%	98.73
11	调整 Conv1,2 卷积核数量 64,128	96.58%	114.54
12 (基准)	调整 Conv1,2 卷积核数量 128,256	98.58%	132.77
13	调整 Conv1,2 卷积核数量 256,512	97.32%	156.48

部分参数, 这些参数确实会对最终拖影目标的识别产生消极影响; 同时删除两个池化层 (Pooling1 和 Pooling2) 或删除其中一个池化层对错误率的影响也相对较小。但是, 如果只保留第一个卷积层 (Conv1) 或只保留第二个卷积层 (Conv2), 那么模型的性能就会很差, 这表明模型的整体深度对于获得良好性能很重要。然后, 对模型的卷积核数量进行了调整, 在前两个卷积层中改变了卷积滤波器的数量, 模型所表现出来的分割精度不同, 但彼此之间相差不大, 这表明增加卷积层的卷积核数量多少只能在性能和模型训练时间上得到提高, 但是同时也存在诸如全连接层过拟合风险。通过以上分析, 在改进算法中所加入的 CNN 中, 删除 FC1、删除

Pooling1 或 Pooling2 以及适当调整卷积层的滤波器数量对本文所提算法的分割准确率和分割时间产生的影响均较小。

然而, 在某几个类别中, 所提出的 Vibe++ 算法并没有达到预期的效果, 例如基准类、间断运动目标类和热图类, 这是因为本文所提的方法主要针对 Vibe+ 算法在分割不完全或者背景抖动导致的噪点和运动目标存在拖影方面进行改进。此外, 为了进一步验证所提算法的鲁棒性, 本文还把所提算法应用于带有强烈抖动和具有大面积拖影区域特征的车载监控视频数据和街边监控视频数据, 并对比了 Vibe 算法、Vibe+ 算法以及本文所提的 Vibe++ 算法对运动目标进行背景分割的方法, 实验结果如图 7 和图 8 所示。

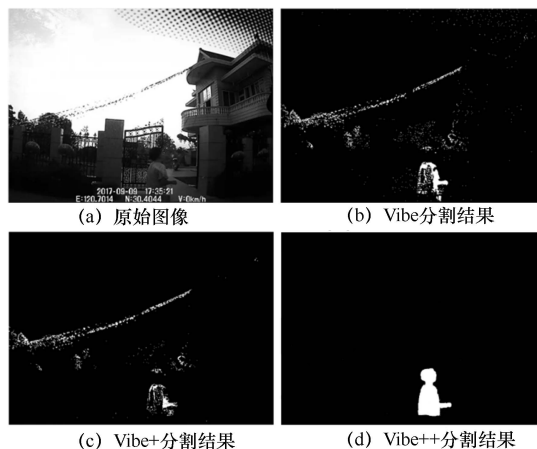


图 7 车载监控视频下的算法对比结果

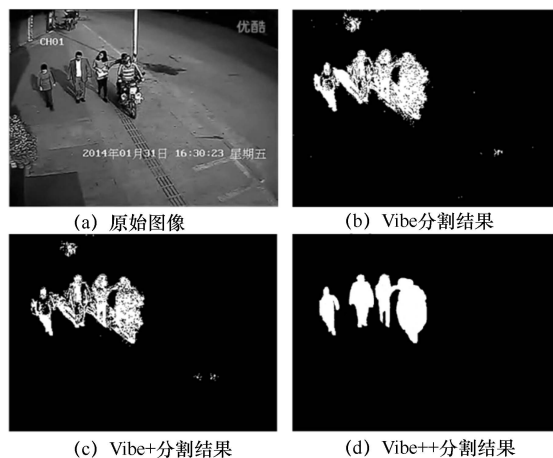


图 8 监控视频下的算法对比结果

可以看出, 经过 Vibe 算法处理后, 图 7 (b) 和图 8 (b) 依然存在较多的噪点, 且分割出的前景目标中包括高光区影响的背景目标; 而经过 Vibe+ 算法处理后得到的图 7 (c) 和图 8 (c) 虽然噪点现象略微减轻, 但噪点现象和前景目标的错误分割现象依然严重, 并且在微风吹动的条幅分割结果来看, 存在噪声扰动, 而经本文提供的 Vibe++ 算法的运动目标背景分割方法处理分割后得到的图 7 (d) 则能够准确清晰地提取到前景运动目标图像, 并可以剔除多余的噪声干扰, 填充了目标运动图像; 在强烈太阳光影响下, 运动目标 (行人和摩托车) 的拖影也可以通过本文所提出的融合改进 MeanShift 聚类算法和卷积神经网络进行拖影的检测和识别, 起到至关重要的作用。

因此, 本文提出的基于改进 Vibe+ 算法的运动目标背景分割方法与现有 Vibe 算法、Vibe+ 算法以及当今现存的运动目标分割算法相比具有较大的优势, 其对前景运动目标的分割准确率较高, 且受视频图像内容干扰的程度小, 因而具备较强的适用性。Vibe++ 算法除了在开放式场景下可以使用, 还可以应用在特定场景下的运动目标识别, 例如在工业领域中^[25], 生产线上运动的工业零部件, 可以采用 Vibe++ 实时跟踪目标, 剔除由于机械装置抖动或电压不稳定带来的分割二值图存在较多噪点干扰的情况; 而且还可以检测出因为采集视频或图像系统中的光源设计不合理导致的获取图像具有较多阴影, 对阴影的检测和剔除同样可以采用 Vibe++ 算法进行处理; 在农业领域^[26], 使用无人机进行水稻、小麦等农作物进行产量评估时, 往往需要通过无人机机载相机获取视频, 然而在白天阳光的强烈照射下会出现作物的大面积阴影和由微风所带来的抖动, 可采用本文方法进行拖影和噪点的消除; 在水产领域^[27], 在养殖鱼或虾的池塘水面上获取图像时, 获取的图像中往往会由于反光问题存在高光区, 这些区域的存

在会使得图像部分信息丢失, 而且运动目标在光线下的不断移动也会存在轻微阴影的情况, 也可采用本文提出的 Vibe++ 予以解决。因此, 本文所提算法具有在多个研究领域的应用前景和应用潜力。

4 结束语

本文针对传统 Vibe+ 算法存在的抖动运动目标和拖影区域无法准确分割的问题, 对 Vibe+ 运动目标分割算法进行了两点改进: 一是采用基于区域生长算法对分割图像中各连通区域进行标记, 依据边界面积块差异特征设定面积筛选阈值, 将不满足条件的像素点进行删除; 二是对 MeanShift 聚类算法改进, 并与卷积神经网络进行有效融合, 对结果图像中拖影区域、非拖影区域和拖影边缘区域分类识别, 计算拖影区域在图像中位置坐标, 对其中各个拖影区域进行快速删除。本文所提算法在公开的视频数据集中包含的多个子类别中与之前的最优算法相比取得较好的分割准确率, 并且通过车载视频和监控视频相关数据进一步论证了所提算法的有效性。在未来的无人驾驶汽车和视频监控研究领域中, 本研究具有广阔的应用背景和较高的实用价值。

参考文献:

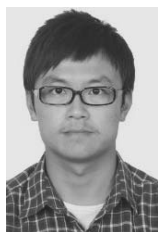
- [1] DIM M, BASHIR A, BASIT A, et al. Abandoned object detection using frame differencing and background subtraction[J]. International Journal of Applied Mathematics and Computer Science, 2020, 11 (7): 676-681.
- [2] ZHOU S T, ZHANG Q, GAO R, et al. High-accuracy atmospheric turbulence compensation based on a Wirtinger flow algorithm in an orbital angular momentum-free space optical communication system[J]. Optical Communication, 2020, 477(2): 345-353.
- [3] DERIVAUX S, FORESTIER G, WEMMERT C, et al. Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation[J]. Pattern Recognition Letters, 2010, 31(15): 2364-2374.
- [4] ALGETHAM N, REDFERN S. A robust tracking-by-detection



- algorithm using adaptive accumulated frame differencing and corner Features[J]. *Journal of Imaging*, 2020, 6(4): 1-10.
- [5] LYRAS K G, HANSON B, FAIRWEATHER M, et al. A coupled level set and volume of fluid method with a re-initialisation step suitable for unstructured meshes[J]. *Journal of Computation Physics*, 2020, 407(12): 124-139.
- [6] ALVAREZ-MEZA A M, MOLINA-GIRALDO S, CASTEL-LANOS-DOMINGUEZ G. Background modeling using object-based selective updating and correntropy adaptation[J]. *Image and Vision Computing*, 2016, 45(13): 22-36.
- [7] BARNICH O, VAN D M. ViBe: a universal background subtraction algorithm for video sequences[J]. *IEEE Transaction of Image Process*, 2011, 20(6): 1709-1724.
- [8] GULESIR G, BERGMANS L, AKSIT M, et al. Vibes: a visual language for specifying behavioral requirements of algorithms[J]. *Journal of Visual Languages & Computing*, 2013, 24(5): 350-364.
- [9] ZHOU X Q, LIU X F, JIANG A M, et al. Improving video segmentation by fusing depth cues and the visual background extractor (ViBe) algorithm[J]. *Sensors*, 2017, 17(5): 56-67.
- [10] 彭丰平, 鲍苏苏, 曾碧卿. 基于自适应区域生长算法的肝脏分割[J]. *计算机工程与应用*, 2010, 46(33): 198-200.
- PENG F P, BAO S S, ZENG B Q. Segmentation of liver based on adaptive region growing[J]. *Computer Engineering and Applications*, 2010, 46(33): 198-200.
- [11] 钮圣喆, 王盛, 杨晶晶, 等. 完全基于边缘信息的快速图像分割算法[J]. *计算机辅助设计与图形学学报*, 2012, 24(11): 1410-1419.
- NIU S X, WANG S, YANG J J, et al. A fast segmentation algorithm fully based on edge information[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2012, 24(11): 1410-1419.
- [12] CAO X F, QIU B Z, XU G D. BorderShift: toward optimal MeanShift vector for cluster boundary detection in high-dimensional data[J]. *Pattern Analysis and Applications*, 2019, 22(3): 1015-1027.
- [13] MITTAL G, KORUS P, MEMON N. FiTy: large-scale file fragment type identification using convolutional neural networks[J]. *IEEE Transactions on Information Forensics and Security*, 2021, 16(3): 28-41.
- [14] KEINOSUKE F, LARRY D H. The estimation of the gradient of a density function with application in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, 21(1): 32-40.
- [15] CHENG Y Z. Mean shift, mode seeking, and clustering[J]. *IEEE Transactions on Pattern and Intelligence*, 1995, 17(8): 790-799.
- [16] 姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述[J]. *软件学报*, 2017, 28(1): 160-183.
- JIANG F, GU Q, HAO H Z, et al. Survey on content-based image segmentation methods[J]. *Journal of Software*, 2017, 28(1): 160-183.
- [17] LIU Z, CHENG F, HONG H M. Identification of impurities in fresh shrimp using improved majority scheme-based classifier[J]. *Food Analytical Methods*, 2016, 16(3): 10-20.
- [18] 任洪娥, 白杰云. Lab 颜色空间中基于动态聚类的颜色分级[J]. *计算机工程*, 2013, 39(6): 231-235.
- REN H E, BAI J Y. Color grading based on dynamic clustering in lab color space[J]. *Computer Engineering*, 2013, 39(6): 231-235.
- [19] 彭宁嵩, 杨杰, 刘志, 等. MeanShift 跟踪算法中核函数窗宽的自动选取[J]. *软件学报*, 2005, 16(9): 1542-1550.
- PENG N S, YANG J, LIU Z, et al. Automatic selection of kernel-bandwidth for Mean-Shift object tracking[J]. *Journal of Software*, 2005, 16(9): 1542-1550.
- [20] 邓柳, 汪子杰. 基于深度卷积神经网络的车型识别研究[J]. *计算机应用研究*, 2016, 33(3): 930-932.
- DENG L, WANG Z J. Deep convolution neural networks for vehicle classification[J]. *Application Research of Computers*, 2016, 33(3): 930-932.
- [21] DROOGENBROECK M V, PAQUOT O. Background subtraction: experiments and improvements for ViBe[C]//*Proceedings: In Change Detection Workshop(CDW)*. [S.l.:s.n.], 2012.
- [22] HOFMANN M, TIEFENBACHER P, RIGOLL, G. Background segmentation with feedback: The pixel-based adaptive segmenter[C]//*Proceedings of Computer Vision and Pattern Recognition Workshops(CVPR)*. Piscataway: IEEE Press, 2012: 38-43.
- [23] GOYETTE N, JODOIN P, PORIKLI F, et al. Changedetection net: A new change detection benchmark dataset[C]//*Proceedings of Computer Vision and Pattern Recognition Workshops(CVPR)*. Piscataway: IEEE Press, 2012.
- [24] CHEN Z, ELLIS T. A self-adaptive gaussian mixture model[J]. *Computer Vision and Image Understanding*, 2014(122): 35-46.
- [25] 何利, 李伯全. 复杂背景中奶粉罐内配件数量检测研究[J]. *机电工程*, 2014, 31(8): 1098-1102.
- HE L, LI B Q. Detection of the number of accessories in the milk powder cans in complex background[J]. *Journal of Mechanical & Electrical Engineering*, 2014, 31(8): 1098-1102.
- [26] 周文静, 查志华, 吴杰. 改进圆形 Hough 变换的田间红提葡萄果穗成熟度判别[J]. *农业工程学报*, 2020, 36(9): 205-213.
- ZHOU W J, ZHA Z H, WU J. Maturity discrimination of "Red Globe" grape cluster in graper by improved circle hough transform[C]//*Proceedings of Transactions of the Chinese Society of Agriculture Engineering(Transactions of the ASABE)*. [S.l.:s.n.], 2020, 36(9): 205-213.

- [27] LIU Z H, CHENG F, ZHANG W. Recognition-based image segmentation of touching pairs of cooked shrimp (*Penaeus Orientalis*) using improved pruning algorithm for quality measurement[J]. Journal of Food Engineering, 2017, 195(6): 166-181.

[作者简介]



刘子豪（1988—），男，博士，嘉兴学院讲师、硕士生导师，主要研究方向为人工智能与图像处理、基于机器视觉的农产品无损检测。



贾小军（1974—），男，博士，嘉兴学院副教授、硕士生导师，主要研究方向为人工智能与图像处理。



张素兰（1980—），女，博士，嘉兴学院讲师，主要研究方向为三值光学计算机、系统结构、嵌入式系统。



徐志玲（1966—），女，中国计量大学教授，主要研究方向为计量检测成像技术研究 with 仪器开发、工业机器人检测系统研究与开发。



张俊（1994—），男，浙江大学生物系统工程与食品科学学院博士生，主要研究方向为基于谱图成像技术的农产品无损检测。