曲率滤波-经验模式分解的运动人体目标检测预处理

叶 华[1,2],谭冠政[1],胡长坤[3],戴正科[2]

(1. 中南大学 信息科学与工程学院,湖南 长沙 410083;
2. 湖南文理学院 电气与信息工程学院,湖南 常德 415000;
3. 东北大学 信息科学与工程学院,辽宁 沈阳 110819)

摘 要:利用曲率滤波-经验模式分解预处理检测并提取人体目标特征,以降低图像分解运算复杂度,同步增强边缘和纹理特征,提高特征区分性。表现在:(1) 在第一层经验模态分解中,以曲率滤波曲面映射原图像的连续平滑曲面,形成包络曲面及均值面,首层分解图像纹理特征显著,以下各层凸显边缘与结构特征;(2) 从低分辨率到高分辨率尺度图像匹配出相邻层强边缘区域,易于人体目标轮廓候选区域的提取;(3) 以首层分解图纹理特征筛分背景,在相邻层中匹配前景特征区域,形成人体姿态特征的轨迹图,易于判别人体姿态及行为。在人体行为识别实验中,采用曲率滤波-经验模式分解预处理提取的轮廓特征与人体行为典型数据库 ground truth 对比,在 UIUC 示例数据的轮廓提取的精度和召回率都达到 90%以上。对人体姿态及行为做识别处理,验证了预处理方法的有效性。

关键词:图像增强; 曲率滤波平滑; 经验模式分解; 强边缘; 多尺度
中图分类号:TP29 **文献标志码**:A **DOI**:10.3788/IRLA201847.0226001

Curvature filter-empirical mode decomposition on moving human target detection preprocessing

Ye Hua[1,2], Tan Guanzheng[1], Hu Changkun[3], Dai Zhengke[2]

(1. School of Information Science and Engineering, Central South University, Changsha 410083, China;
2. College of Communication and Electric Engineering, Hunan University of Arts and Science, Changde 415000, China;
3. School of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: Curvature filtering-empirical mode decomposition preprocessing was used to detect and extract human target features. It reduced the computational complexity of image decomposition, enhanced edge and texture features simultaneously, and improved feature differentiation. Its performance was in the following areas: (1) In the first layer of empirical mode decomposition, the continuous smooth surface of the original image was mapped by curvature filtering plane to form the envelope surface and the mean surface. The first layer extracted texture features, and the following layers highlighted structure features. (2) Matching regions had sharp edges from adjacent layers which varied from low-resolution to high-

收稿日期:2017-08-10; **修订日期**:2017-10-01
基金项目:国家自然科学基金(11347132);湖南省自然科学基金(12JJ3061);湖南省优秀青年项目(10B074);湖南省教育厅科研项目(16C1087)
作者简介:叶华(1977-),女,讲师,博士生,主要从事人体行为识别、视觉认知计算等方面的研究。Email:yehuawuhan@163.com
导师简介:谭冠政(1962-),男,教授,博士,主要从事智能系统与智能控制、人工智能与认知系统、生物医学图像处理等方面的研究。Email:tgz@csu.edu.cn

resolution, easy to extract the target contour candidate regions. (3) Decomposing the texture features in the first floor to screen the background, matching the foreground feature areas in the adjacent layers to form the trajectory map of the human body posture, it made easy judging the human body posture and behavior. When applying it to human behavior recognition experiments, and comparing it with the ground truth of the human behavior database, the accuracy of the contour extraction and the recall rate of the UIUC sample data were all over 90%. And experiments verify the preprocessing method is effective in human pose recognition and behavior recognition.

Key words: image enhancement; curvature filter smoothing; empirical mode decomposition; sharp edges; multi-scale

0 引 言

深度学习的广泛应用灭亡了传统特征提取技术?恰恰相反,提取特征方法为深度学习网络结构形成、模型选择和参数调整提供基础。基于视觉的人体行为分析与识别研究的基本步骤就是特征提取、分类与定位。黑箱式的深度学习与传统的机器学习都以人体姿态特征提取为基础，检测或跟踪学习内在的本质特征[1-9]。检测的特征有灰度、纹理、颜色、形状、运动信息和融合特征[4-9]。Navneet Dalal 和 Bill Triggs 利用 HOG(Histogram of Oriented Gradient)算法检测行人,线性 SVM 作分类器,成为行人检测算法的基础框架。但该算法侧重于检测纹理,可结合其他特征检测方法以取得更好的检测效果[1]。

以上传统处理灰度图像的行人检测技术往往利用对比度高的边缘特征(如梯度算子),或者描述细节的纹理特征(如 LBP(Local Binary Pattern)算子),没有一种能准确描述影像的非稳态性，使边缘和纹理特征同步增强并得以区分，不能很好地解决背景复杂或与前景类似、图像清晰度低、图像对比度低的问题。探讨曲率滤波-经验模式分解方法直接拟合多维数据,描述影像非平稳特性,提高图像对比度,同步增强边缘和纹理特征,提高图像特征区分性。二维经验模态分解是一种空域多尺度多分辨率表达方法[10-11],分层迭代去除均值影响且同时突出局部极值,适于分析非线性和非平稳数据图像。其插值获得包络的局限性使准确获取均值成为难点。探讨以曲率滤波逼近原离散图像的连续曲面，取代计算均值的包络面求解过程。通过平滑分解多分辨图像,降低因个体差异、运动变化、光照等外部环境的影响,达到预处理的效果,特征将更显著,噪声影响将进一步降低。应用于人体行为识别数据库 KTH、UIUC 和 UCF YouTube dataset[1-9,12-16],计算检测特征的精度、人体姿态及行为的识别率，实验结果验证其在增强特征判别性、提高运算效率、提升识别率方面的有效性,为无标注样本的学习、模型构造的复杂度、计算量过大等深度学习问题提供解决思路，有明确的物理意义。

1 曲率滤波平滑的经验模式分解模型方法

采用曲率滤波方法得到原离散图像的近似连续光滑曲面，可作为经验模式分解的包络曲面进行二维固有模态函数图像的分解，并建立分解树模型来提取人体姿态轮廓特征。对提取的人体轮廓特征进行不变性描述子描述以分类识别，可以达到很好的识别效果。

1.1 二维经验模态分解方法

二维经验模态分解 (Bidimensional Empirical Mode Decomposition,BEMD)是从高频到低频分量自适应分解方法[10],能描述影像非平稳特性。BEMD 分解逻辑关系如下式描述：

$$f(x,y)=\sum_{i=1}^{n}\mathrm{BIMF}^{(i)}(x,y)+\mathrm{RESIDUE}(x,y)=\sum_{i=1}^{n}\mathrm{BIMF}^{(i)}(x,y)+M^{n}(f(x,y)) \tag{1}$$

式中:$f(x,y)$为待筛分分解的原始图像;$M(f(x,y))$为均值曲面;$M^{n}(f(x,y))$为 n 次均值曲面;$\mathrm{BIMF}^{(n)}$为本征图像;$\mathrm{RESIDUE}(x,y)$为趋势图像。对图像进行经

验模式分解就是不断去除均值影响，突出局部极值的操作，这样，图像的特征更为突出，纹理和边缘等特征同时得到增强。在进行 BEMD 分解时，使用 Somayeh Bakhtiari[11]提供的 EEMD 方法，筛选过程的迭代次数设定为 10，分解层数设定为 4 层。

1.2 曲率滤波平滑的经验模式分解方法

曲率滤波-经验模式分解方法是一种直接对多维数据拟合的经验模式分解方法。首先以高斯曲率实行源图像的平滑处理，一步得到逼近原离散图像的可展曲面。以平均曲率的极大值、极小值作为提取的局部极大值、极小值，可以直接形成曲面的上下包络曲面并进行减均值计算，筛分图像过程避免了插值求解包络的计算负荷以及插值引起的偏差。具有零 GC 的曲面可以等位映射到不变形的平面-可展曲面[17]，可展曲面与平面构成等距对应。因此，Gaussian Curvature(GC)作为理想的正则器已用于多种领域(如图像配准)[18]。对任意给定像素 x，坐标为 i、j，其强度值为 $U(x)$，用坐标表示为 $U(i,j)$，取 3×3 邻域 $N(x)$，设点为 x，到邻域 $N(x)$各点切平面的投影距离为 d_l，邻域内所有的$\{d_l, l=1,\cdots,8\}$。根据欧拉定理，$d_l \approx \kappa_1\cos^2\theta_l+\kappa_2\sin^2\theta_l$，$\kappa_1$，$\kappa_2$ 为主曲率，θ_l 对应于主曲率平面的采样角度。通过下式进行数值计算得出投影距离 d_l，推导过程见参考文献[17]。

$$\begin{bmatrix} d_1\\ d_2\\ d_3\\ d_4\\ d_5\\ d_6\\ d_7\\ d_8 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0\\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0\\ -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & -1 & 1 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} U(i-1,j-1)\\ U(i-1,j)\\ U(i-1,j+1)\\ U(i,j-1)\\ U(i,j+1)\\ U(i+1,j-1)\\ U(i+1,j)\\ U(i+1,j+1) \end{bmatrix} - U(i,j) \tag{2}$$

其最小投影距离$|d_m|=\min\{|d_l|, l=1,\cdots,8\}$，用 d_m 更新 U 就减少了 $\min\{|\kappa_1|,|\kappa_2|\}$。

$$\hat{U}(i,j)=U(i,j)+d_m \tag{3}$$

则更新像素点的投影值为$\hat{U}(x)=U(x)+d_m$。得到像素点强度值的投影值$\hat{U}(x)$，通过不相交的域分解处理，在 3×3 邻域迭代求出各像素点投影值，得到可展曲面 SGC，即可组成逼近原图像的平滑连续曲面。同高斯曲率，点的投影距离与平均曲率通过下式进行数值计算得出，推导过程见参考文献[17]。

$$\begin{bmatrix} d_1\\ d_2\\ d_3\\ d_4 \end{bmatrix} = \begin{bmatrix} 0 & \frac{5}{16} & -\frac{1}{8} & 0 & \frac{5}{8} & 0 & \frac{5}{16} & -\frac{1}{8}\\ -\frac{1}{8} & \frac{5}{16} & -\frac{1}{8} & \frac{5}{8} & 0 & 0 & \frac{5}{16} & 0\\ -\frac{1}{8} & \frac{5}{8} & -\frac{1}{8} & \frac{5}{16} & \frac{5}{16} & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{5}{16} & \frac{5}{16} & -\frac{1}{8} & \frac{5}{8} & -\frac{1}{8} \end{bmatrix} \cdot \begin{bmatrix} U(i-1,j-1)\\ U(i-1,j)\\ U(i-1,j+1)\\ U(i,j-1)\\ U(i,j+1)\\ U(i+1,j-1)\\ U(i+1,j)\\ U(i+1,j+1) \end{bmatrix} - U(i,j) \tag{4}$$

其最小平均曲率投影距离$|d_k|=\min\{|d_k|, k=1,\cdots,4\}$，最大平均曲率投影距离$|d_M|=\min\{|d_l|, l=1,\cdots,4\}$，在 3×3 邻域迭代求出各像素点投影值：

$$\hat{U}_m(i,j)=U(i,j)+d_m,\ \hat{U}_M(i,j)=U(i,j)+d_M \tag{5}$$

式中：$\hat{U}_m(i,j)$，$\hat{U}_M(i,j)$为像素点 3×3 邻域的局部极小值和局部极大值，遍历图像，计算得出逼近可展曲面的极值。

曲率滤波-经验模式分解方法算法描述如下。

(1) 初始化：$R_{ini}=I$(图像信号)，$i=1$。

(2) 以公式(3)计算点的高斯曲率，更新其强度值，对 R_{ini} 曲率滤波得到平滑图像 R_0。

(3) 设定分解层数 n，$H_0=R_{j-1}$，$j=1$。

(4) 对平滑图像 H_{j-1} 计算平均曲率，以公式(5)计算所有局部极值点并组成极大值点集，映射上包络曲面 U_{j-1}，极小值点集映射下包络曲面 B_{j-1}，求取像素均值 $M_{j-1}=(U_{j-1}-B_{j-1})/2$。

(5) 减除所有均值点构成图像 H_j，更新 H_j，$H_j=H_{j-1}-M_{j-1}$，$j=j+1$。

(6) 计算是否满足停止条件，即标准偏差 SD 不大于 0.2。满足则 $IMF_{j-1}=H_{j-1}$，$S_j=H_{j-1}$，$j=j+1$；否则重复步骤(4)~(6)的操作。

(7) 更新趋势量 $R_j=R_{j-1}-S_j$。

(8) 重复步骤(3)~(7)的操作,直到规定的层数。

1.3 曲率滤波-经验模式分解的特征提取预处理

BIMF 分解图频段呈二叉树结构,由于区域约束和分辨率约束,有效保持了不同尺度下的边缘位置精度。对各个 $BIMF^{(n)}$图特征提取采用局部区域特征搜索并分割的过程,算法提取的特征是 $BIMF^{(n)}$图局部特征 MSERs,MSERs 搜索采用 Murphy 基于分水岭的算法[19]来实现。图 1 是以 UIUC dataset dariabend 为例进行特征提取的流程示意图。图像 $f(x,y)$已完成 4 层分解得到 BIMF 图 $BIMF^{(4)}$~$BIMF^{(1)}$,尺度由粗到细变化。图中,在初始尺度空间 $BIMF^{(4)}$检测得出 MSERs 区域 9 个:E_{41}~E_{49},分别计算 9 个区域的矩特征参数。将矩特征参数与 $BIMF^{(3)}$中检测对应的 MSERs 并计算矩特征,满足相似条件的区域为 E_{31}、E_{32}。类似计算各相邻层,E_{11} 为提取的特征区域。

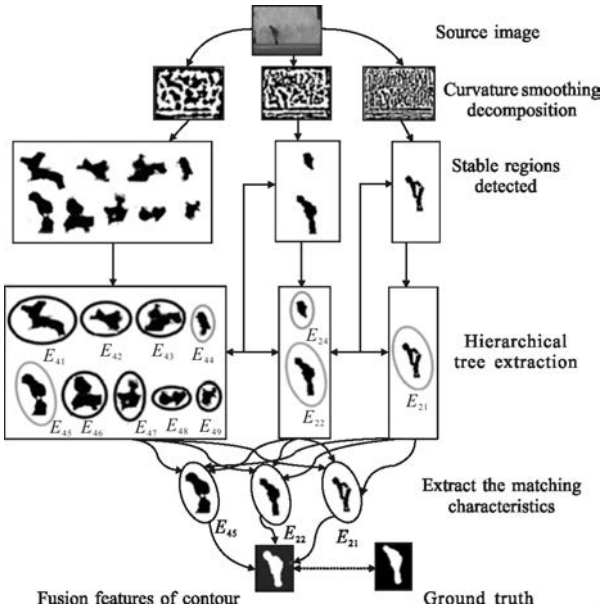

图 1 提取特征流程图:待处理的源图像来自于 UIUC

Fig.1 Extract feature flow chart: The source image to be processed comes from UIUC

2 实验与结果分析

针对人体行为识别中视觉影响问题,如光照变化、尺度不同、图像模糊、现实场景背景丰富等,选择不同数据,测试文中预处理方法提取特征的正确率与特征区分度,体现特征判别性提升给人体轮廓检测性能带来的改善。

2.1 数据库

实验选择在几个著名的公共行为识别库中进行,即 KTH 视频库,UIUC 视频库和 UCF YouTube Action Dataset 视频库。在实验中设置 EEMD 分解层数为 4,迭代次数为 10,用预处理模型提取特征,稳定区域参数阈值区域方向 d 设置为 90%,扁率比 q 设置为 50%。提取特征采用不变矩描述子,采用主成分分析(PCA)结合马氏距离(Mahalanobis distance)判别行为分类。为便于与已有方法比较形成对照结果,这里以参考文献[13]中的交叉测试设置进行实验。

2.2 提取轮廓的精度

根据生成相邻分解图形成特征轨迹图,当特征参数的变化值小于设定的阈值时提取该特征,提取的特征为 4 类:头、腰、脚、人体轮廓质心位置。特征轨迹图类似 4 条平行线,据此加速锁定待提取的轮廓特征。设 $S(A)$代表该算法提取分割的对象像素区域,$S(B)$代表数据库源图像 ground truth 的对象像素区域,则精度 P、召回率 R 的计算方法如下:

$$P=\frac{S(A\cap B)}{S(A)},R=\frac{S(A\cap B)}{S(B)}$$

UIUC 数据库提供了 ground truth 轮廓图,方便定量分析预处理方法轮廓提取的准确性,在 UIUC 示例数据的轮廓提取的精度和召回率实验结果如图2 所示,由于没有其他的实验结果作比对,仅提供该方法的实验数据。该方法的精度和召回率都达到 90%以上,其中 daria_jump 的精度与召回率均低于 daria_bend,这是由于动作本身区分度不大导致的。从实验结果可以得出,曲率滤波-经验模式分解预处理方法对图像前景与背景对比度增强、特征提取精度性能都有提升。

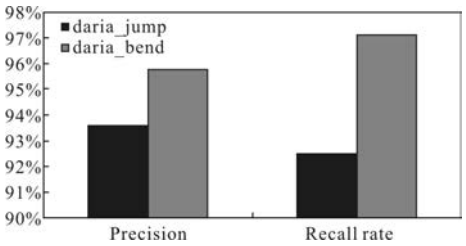

图 2 从 UIUC 示例数据库提取人体轮廓的精度与召回率

Fig.2 Extraction precision and recall rate of human contour on UIUC database

2.3 行为识别率

对比参考文献[13],在 KTH 视频序列中笔者提

取的特征使 jog 与 run 动作区分度大，在UIUC 中 stand-2-sit、jump-jack 和 jumpupward 动作判别性强，提取特征准确，无噪声污染。文中方法利用单帧或单副图像提取特征性能好，直接用于分类，识别率比较如图 3 和图 4 所示，在 KTH 数据库和 UIUC 数据库的平均识别率分别达到 95.3%和 81%。对于极其相近的动作区分，除了在 BIMF(1)上提取的特征外还加入了 BIMF(3)的特征，极大地提高了判别性。

在 UCF YouTube 中 juggle 和 tennis 视频序列的行为特征提取准确，无噪声污染，提出的方法在 UCF YouTube dataset 数据库的平均识别率达到 82.5%。少数如 swing1 序列，由于快速运动光照变化大导致图像模糊，jump 序列图像比较模糊，采用单帧多层提取特征方法，提取特征效果较差。探讨采用不精确的背景减除结合预处理提取特征以解决该问题。识别率如图 3~5 所示。

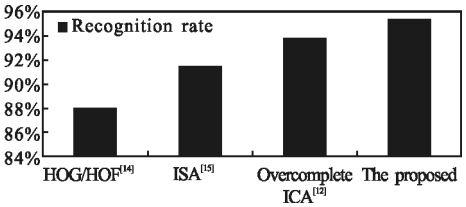

图 3 在 KTH 数据库的平均识别率

Fig.3 Average recognition rate in KTH database

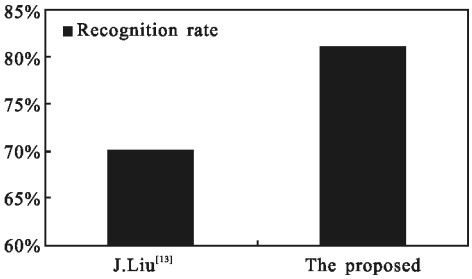

图 4 在 UIUC 数据库的平均识别率

Fig.4 Average recognition rate in UIUC database

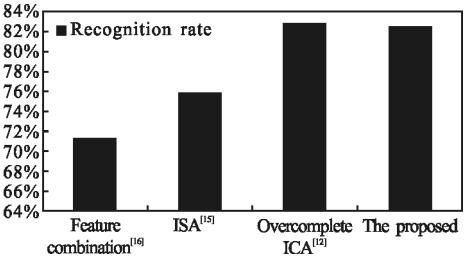

图 5 在 UCF YouTube dataset 数据库的平均识别率

Fig.5 Average recognition rate in UCF YouTube dataset database

3 结 论

以上分析采用 4 层的曲率滤波-经验模式分解对图像作分解、完成运动目标检测、特征提取等预处理。通过平滑分解多分辨图像，使得纹理和边缘特征同步增强，更易获取人体轮廓特征，达到区分背景和前景目标的目的。在 UIUC 示例数据的轮廓提取的精度和召回率均达到 90%以上，对比传统经典算法，在典型行为数据库的平均识别率较高。设图像尺寸为 $m\times n$，曲率滤波平滑是遍历整幅图像各像素点邻域的矩阵运算，设一个像素点邻域的矩阵运算量为 a，曲率滤波平滑运算量为amn。本征图像分解运算量类似，设一个像素点平均曲率运算量为 b，一层分解可近似为 bmn，M 层为 $Mbmn$。分层树提取特征基于形态学处理，可近似表示为 mn 的 k 倍，k 为检测的区域数。特征匹配与融合在相邻层完成，运算量以检测的区域数为基准，不超过 mn。因此该算法的运算量为 $amn+Mbmn+kmn+mn$，时间复杂度为 $O(Mmn)$，算法时间复杂度较低。但仍然存在检测定位查准率不高、漏检误检等问题。探讨优化预处理，结合多源数据如深度图像，其他知觉信息如声音数据、触觉数据，融入高层语义分析，检测识别率将更高。

参考文献：

[1] Luo Haibo, Xu Lingyun, Hui Bin, et al. Status and prospect of target tracking based on deep learning[J]. ***Infrared and Laser Engineering***, 2017, 46(5): 0502002. (in Chinese)

[2] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. ***IEEE Transactions on Pattern Analysis & Machine Intelligence***, 2017, 39(6): 1137-1149.

[3] Wang Xiaofei, Wang Xiaoyi, Shi Xiangyu, et al. Target detection algorithm based on spatial-contextual image one class classification [J]. ***Infrared and Laser Engineering***, 2015, 44(S1): 236-240. (in Chinese)

[4] Ke Hongchang, Sun Hongbin. A saliency target area detection method of image sequence [J]. ***Chinese Optics***, 2015, 8(5): 768-774. (in Chinese)

[5] Liu Hongbin, Chang Faliang. Moving object detection by optical flow method based on adaptive weight coefficient[J]. ***Optics and Precision Engineering***, 2016, 24(2): 460-468. (in Chinese)

[6] Lu Mu, Zhu Ming, Gao Yang, et al. Moving target detection based on dynamic background of cellular automaton [J]. ***Optics and Precision Engineering***, 2017, 25(7): 1934-1940. (in Chinese)

[7] Wang Kejia, Ping Ziliang, Sheng Yunlong. Development of image invariant moments -a short overview [J]. ***Chinese Optics Letters***, 2016, 14(9): 091001.

[8] You Ruirong, Wang Xinwei, Ren Pengdao, et al. Target observation performance evaluation method for video surveillance based on Johnson criteria[J]. ***Infrared and Laser Engineering***, 2016, 45(12): 1217003. (in Chinese)

[9] Subudhi B N, Ghosh S, Nanda P K, et al. Moving object detection using spatio-temporal multilayer compound Markov random field and histogram thresholding based change detection [J]. ***Multimedia Tools and Applications***, 2017, 76 (11): 13511-13543.

[10] Nunes J C, Bouaoune Y, Delechelle E, et al. Image analysis by bidimensional empirical mode decomposition [J]. ***Image and Vision Computing***, 2003, 21(12): 1019-1026.

[11] Bakhtiari S, Agaian S, Mo J. An enhanced empirical mode decomposition based method for image enhancement [C]// IEEE International Conference on Systems, Man, and Cybernetics, 2015: 2681-2686.

[12] Zhang Shengping, Yao Hongxun, Sun Xin, et al. Action recognition based on overcomplete independent components analysis[J].***Information Sciences***, 2014, 281(10): 635-647.

[13] Liu Jingen, Benjamin Kuipers, Silvio Savarese. Recognizing human actions by attributes[C]//Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, 2011: 3337-3344.

[14] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies [C]//Proceedings of International Conference on Pattern Recognition, 2008: 1-8.

[15] Klaser A, Marszalek M, Schmid C. A spatio-temporal descriptor based on 3d gradients [C]//Proceedings of British Machine Vision Association, 2008: 995-1004.

[16] Liu J, Luo J, Shah M. Recognizing realistic actions from videos "in the wild" [C]//Proceedings of International Conference on Pattern Recognition, 2009: 1996-2003.

[17] Gong Yuanhao, Sbalzarini Ivo F. Curvature filters efficiently reduce certain variational energies[J]. ***IEEE Transaction on Image Processing***, 2017, 26(4): 1786-1798.

[18] Brito-Loeza C, Chen K, Uc-Cetina V. Image denoising using the Gaussian curvature of the image surface [J]. ***Numerical Methods for Partial Differential Equations***, 2016, 32(3): 1066-1089.

[19] Martins P, Carvalho P, Gatta C. On the completeness of feature-driven maximally stable extremal regions [J]. ***Pattern Recognition Letters***, 2016, 74: 9-16.