基于体积语义局部二值模式的行为识别

吴 娴[1),2)] 赖剑煌[1)]

[1)](中山大学信息科学与技术学院 广州 510006)

[2)](南方报业传媒集团博士后科研工作站 广州 510601)

摘 要 基于视觉的行为识别是人体运动分析的重要组成,也是该领域一个富有挑战性的研究方向,因此获得广泛的关注.该文将视频中的人体行为看成由每帧轮廓图像沿时间轴堆叠而成的三维空-时体积,提出一种新颖的局部二值模式,即体积语义局部二值模式(VSLBP),用于提取空-时体积中的有效低维特征,通过计算测试序列与已标记的行为训练集特征间的最近卡方距离得到其所属行为类别.在行为库"Weizmann"上实验结果表明,该文提出方法的识别准确率略高于现有最新方法,且能容忍各种复杂的条件,如遮挡、拍摄角度、行为者外观穿着及行为方式等.

关键词 行为识别;局部二值模式;语义局部二值模式;体积局部二值模式

中图法分类号 TP391 **DOI 号**: 10.3724/SP.J.1016.2012.02652

Human Action Recognition Based on Volume Semantic Local Binary Patterns

WU Xian[1),2)] LAI Jian-Huang[1)]

[1)](*School of Information Science and Technology, Sun Yat-sen University, Guangzhou* 510006)

[2)](*Postdoctoral Research Station, Nanfang Media Group, Guangzhou* 510601)

Abstract Vision-based action recognition is an essential component of human motion analysis and has been receiving broad attention in the field of computer vision. Human action in the video sequence can be seen as a 3D space-time volume stacked by silhouette image of each frame according to temporal series. In this paper, a novel representation of local binary patterns, named volume semantic local binary patterns (VSLBP) is proposed and applied to extract low-dimensional features on the space-time volume. Action label can be assigned by computing the nearest chi-square distance of features between each probe silhouette sequence and gallery silhouette sequence sets. Experiment results on the publicly available "Weizmann" dataset demonstrate that the proposed approach outperforms the state-of-the-art methods and can also tolerate various challenging conditions, i.e., partially occluded, changes in the viewpoints, motion style, etc.

Keywords action recognition; local binary patterns; semantic local binary patterns; volume local binary patterns

1 引 言

目前,基于视觉的人体行为识别正在获得越来越多的关注,它涉及模式识别、图像处理、计算机视觉、人工智能等多门学科并具有广泛的应用前景,如视频监控、视频内容检索及人机交互等等.

近年国内外在人体行为识别方面的研究已取得

收稿日期:2008-12-31;最终修改稿收到日期:2011-06-14.本课题得到国家自然科学基金(61173084,60803083)、国家自然科学基金与广东省政府联合项目基金(U0835005)资助.**吴 娴**,女,1985 年生,博士,主要研究方向为计算机视觉和模式识别. E-mail: wuxian@nfmedia.com. **赖剑煌**(通信作者),男,1964 年生,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为人脸识别、模式识别与机器学习及图像分析与理解. E-mail: stsljh@mail.sysu.edu.cn.

一定的进展，并提出了一些解决问题的新思路. Efros 等人[1]、Dollor 等人[2]和 Ke 等人[3]直接在原始图像构成的空时体积上进行处理，避免了传统方法需要光流计算或特征跟踪等所带来的局限. 但以上大部分工作都是基于空-时体积中的梯度或强度等计算特征，因此在视频质量不佳或运动不连续等情况下，这类特征显得不够可靠. Gorelick 等人[4]、Wang 等人[5]、Weinland 等人[6]及 Ikizler 等人[7]则从人体轮廓出发，对人的行为进行分析. 在摄像机静止条件下人体轮廓较易获得，且被认为包含空间足够多的姿势及表观等信息，将人体轮廓按时间顺序接连的表示形式则同时包含了空间表观及时间运动信息，分析并度量它们之间的相似性即可对人的行为进行识别.

局部二值模式算子(Local Binary Patterns，LBP)是一种有效的纹理描述算子，近十年已广泛应用于纹理分类、图像检索及人脸分析等等. 最初的 LBP 算子是一个固定大小为 3×3 的矩形块，将周围 8 个灰度值与中心灰度值比较，大于或等于中心灰度值的点用 1 表示，反之为 0，顺时针方向得到的 8 个二进制值作为该矩形块的特征值. 最后以直方图形式统计出整个区域每个特征值数量，作为此区域的特征描述. 为了改善最初 LBP 算子无法提取大结构纹理特征的局限问题，Ojala 等人[8]对其进行了扩展，允许不同数量的邻域点 P 及不同尺寸的半径 R 即表示成 $LBP_{P,R}$. 而当邻域点 P 及半径 R 很大时，提取出的特征大部分对纹理的描述作用很有限，因此 Ojala 等人[8]又提出了 LBP 算子的另两种扩展形式，称为“均匀模式”及“旋转不变模式”. 二进制序列若至多有两个 0 与 1 之间的转换则被定义为均匀的；而通过循环移位得到的最小二进制序列被定义为旋转不变的. 这两种模式能够有效地描述出图像中大部分的纹理特征，并大大减小特征的数量. 值得注意的是，以上 LBP 特征都由二进制序列的十进制编码得到，这样可能使得语义上相似的特征经十进制编码后大相径庭，而且空间复杂度高，因此 Mu 等人[9]从几何的角度重新定义 LBP，提出了语义局部二值模式(Semantic LBP，SLBP).

另一方面，由于 LBP 出色的描述特征能力，Heikkila 等人[10]用基于 LBP 的纹理方法建立背景模型从而检测运动目标. Zhao 等人[11]将其扩展到三维形式，提出体积局部二值模式(Volume LBP，VLBP)，并发展了基于 3 个正交平面的局部二值模式(LBP from Three Orthogonal Planes，LBP-TOP)，将其用于同时包含表观及运动信息的三维体积，如面部表情、动态纹理中.

本文由 Zhao 等人[11]的工作得到启发，将人体行为看成由每帧轮廓图像沿时间轴堆叠而成的三维空-时体积，VLBP 即可用于提取其低维特征. VLBP 是 LBP 的空时扩展形式，其特征向量长度随邻域点个数增大而快速增长. 为了压缩特征向量的长度，本文采用语义局部二值模式的思想，提出体积语义局部二值模式(Volume Semantic LBP，VSLBP)，并将其用于提取空-时体积的特征. 此方法的最大优点是可直接从空-时体积中提取特征，无需降维，且特征长度仅与邻域点个数 P 有关，而与视频帧数无关，不需要时间对齐，因此可排除行为之间持续时间不同所造成的影响.

本文第 2 节介绍体积局部二值模式(VLBP)；第 3 节简述语义局部二值模式(SLBP)；第 4 节提出体积语义局部二值模式(VSLBP)；第 5 节描述基于 VSLBP 的行为识别过程；第 6 节给出本文提出方法在“Weizmann”库中的实验结果，并进行横向与纵向的比较与分析；第 7 节是对全文的总结.

2　体积局部二值模式(VLBP)

对体积 V 建模[11]，螺旋状展开，其相关符号定义及坐标计算公式列于表 1 中，用公式则可表示成：

$$V=v(g_{t_c-L,c},g_{t_c-L,0},\cdots,g_{t_c-L,P-1},g_{t_c,c},g_{t_c,0},\cdots,g_{t_c,P-1},g_{t_c+L,0},\cdots,g_{t_c+L,P-1},g_{t_c+L,c}) \quad (1)$$

表 1　VLBP 模型中符号定义及坐标计算

符号	定义
P	每帧中心点的邻域点个数
$g_{t_c,c}$	体积中心点的灰度值
$g_{t_c-L,c}$ $g_{t_c+L,c}$	体积中前与后时间间隔 L 的帧的中心点灰度值
$g_{t,p}(t=t_{c-L},t_c,t_{c+L},p=0,\cdots,P-1)$	图像 t 半径为 R 区域上的 P 个邻域点的灰度值
$g_{t_c,c}$坐标	(x_c,y_c,t_c)
$g_{t_c,p}$坐标	$(x_c+R\cos(2\pi\times p/P),y_c-R\sin(2\pi\times p/P),t_c)$
$g_{t_c\pm L,p}$坐标	$(x_c+R\cos(2\pi\times p/P),y_c-R\sin(2\pi\times p/P),t_c\pm L)$

假设 $(g_{t,p}-g_{t_c,c})$ 独立于 $g_{t_c,c}$，忽略中心点 $v(g_{t_c,c})$ 灰度值. 对 V 阈值化处理，则式(1)可近似写成 V_1：

$$V_1=v(s(g_{t_c-L,c}-g_{t_c,c}),s(g_{t_c-L,0}-g_{t_c,c}),\cdots,s(g_{t_c-L,P-1}-g_{t_c,c}),s(g_{t_c,0}-g_{t_c,c}),\cdots,$$

$$s(g_{t_c,P-1}-g_{t_c,c}),s(g_{t_c+L,0}-g_{t_c,c}),\cdots,$$
$$s(g_{t_c+L,P-1}-g_{t_c,c}),s(g_{t_c+L,c}-g_{t_c,c})) \quad (2)$$

$$s(x)=\begin{cases}1, & x\geq 0\\ 0, & x<0\end{cases} \quad (3)$$

由 V_1 可看出其为 $3P+2$ 维表示，则 V_1 的十进制编码可简化描述成：

$$VLBP_{L,P,R}=\sum_{q=0}^{3P+1}v_q 2^q \quad (4)$$

图 1 表示 $VLBP_{R,P,L}(R=1,P=4,L=1)$ 建模及形成过程，对体积中逐帧采样邻域点，将其与中间帧的中心点灰度值大小进行比较并二值化，最后乘以权重因子构成体积局部二值模式.

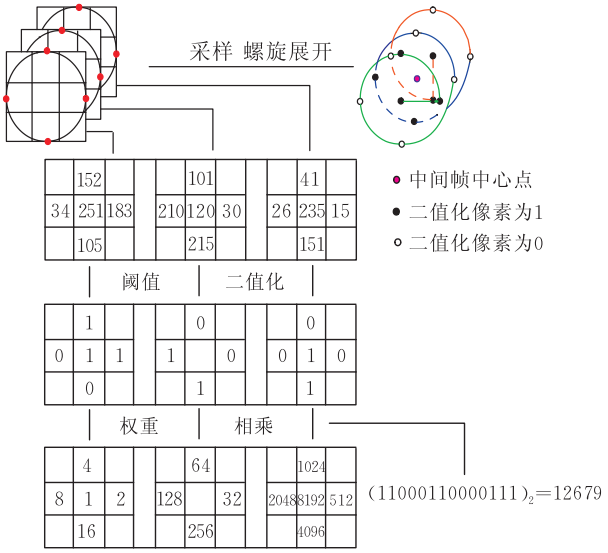


图 1 $VLBP_{1,4,1}$ 建模及形成过程

3 语义局部二值模式(SLBP)

语义局部二值模式(SLBP)最近由 Mu 等人[9]提出，它从几何的角度解释并重新定义 LBP. 基本 LBP 算子是通过对二进制序列进行十进制编码进而统计直方图特征，这么做，无法保证语义相似的特征能落入直方图的邻近的区域. 例如，$(00001111)_2$ 逆时针左移一个比特构成 $(10000111)_2$，它们之间具有很高的相似性，但经十进制编码后却相差甚远(15 vs. 135). 语义局部二值模式则在均匀 LBP 算子的基础上，将二进制序列顺时针表示成几何意义上的圆弧，用弧长和主轴角度两个特性代替基本 LBP 算子的十进制编码数. 如图 2，两个二进制序列弧长一致，主轴角度相差 45°，符合语义相似的特性. 值得注意的是，均匀 LBP 算子至多有两个 0 与 1 之间的转换，则它所对应的只可能是一段圆弧(若序列中只含单个“1”，则可视为圆弧上的一个点).

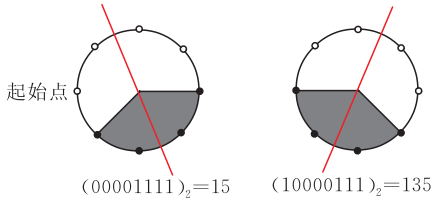


图 2 均匀局部二值模式的弧状表示(弧长及主轴角度)

4 体积语义局部二值模式(VSLBP)

由本文第 2 节可知，VLBP 特征向量长度由邻域点个数 P 决定，P 值大则特征向量长度较大，而小的 P 值则可能丢失有用信息. VLBP 的可能模式个数为 2^{3P+2}，当 P 值增大时，VLBP 所提取的特征长度呈指数级增长. 另外，VLBP 也是对二进制序列的十进制编码，这就不能保证语义相似的特征经编码后所得十进制数仍十分近似. 本节将 VLBP 的概念从灰度图扩展至二值图，为从人体轮廓出发而进行的行为识别过程作铺垫，并结合 SLBP，提出了体积语义局部二值模式(VSLBP)的思想，保留语义信息并大大地减小特征量.

本文建模的体积是由二值轮廓按时序堆叠构成. 因此，我们对第 2 节 VLBP 中的灰度图扩展至二值图，以图 3 的方式对体积建模. 首先类似地，对体积中逐帧($L=1$ 情形)采样邻域点，不同的是，我们将其与中间帧的中心点的值进行异或比较，如果相同则标记为 0，反之为 1，这么做是为了记录局部空间及时间上像素之间的变化. 将体积中如上操作所得的二进制序列组成集合，丢弃其中的全零序列(表示无变化)，且表示成均匀模式(见第 3 节解释，即舍弃圆弧个数大于 1 的二进制序列)，用弧长及主轴角度两个特性代替 VLBP 的十进制编码数，如

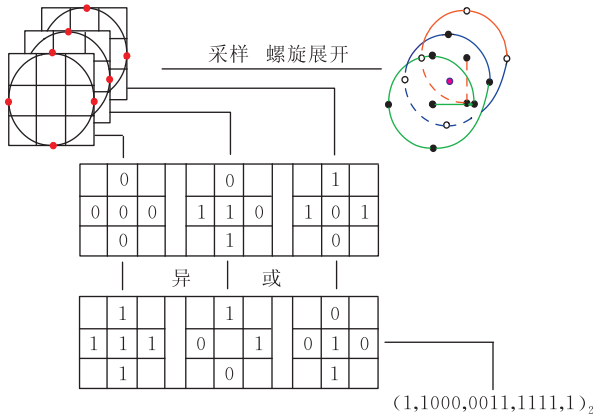


图 3 $VLBP_{1,4,1}$ 建模及二值扩展形式

图 4. 若主轴角度落入两邻域点之间，取靠近其的最小主轴角度（如图 4(b)虚线所示），而圆弧长度的不同则可以区分不同的二进制序列.

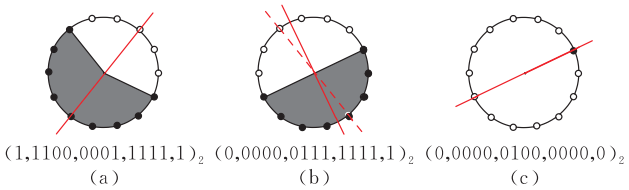


图 4　均匀体积局部二值模式的弧状表示（弧长及主轴角度）

这样弧长和主轴角度的可能值均为 $(3P+2)$ 个，可用 $(3P+2)\times(3P+2)$ 大小的矩阵统计弧长及主轴角度，在相应的弧长及主轴角度上累加计数，如图 5，这也近似于“直方图统计”思想. 最后将二维矩阵中行向量接连成单一向量，表示体积的 LBP 特征，即本文提出的体积语义局部二值模式表示.

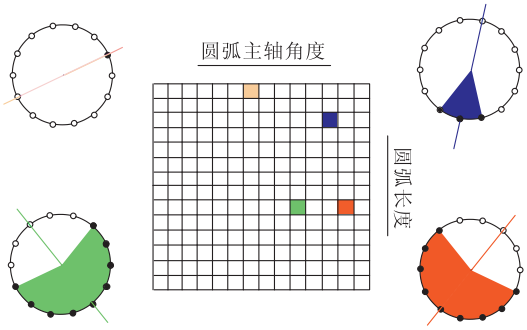


图 5　VSLBP 的统计矩阵（横轴圆弧角度，纵轴圆弧长度）

由以上的分析得出，VLBP 的可能模式个数为 2^{3P+2}；Zhao 等人[11]提出的 3 个正交平面的局部二值模式（LBP-TOP），即在 XY、XT 及 YT 平面采用均匀 LBP 算子，可能的模式个数为 $3\times P(P+1)$；VSLBP 的向量长度为 $(3P+2)^2$. 图 6 给出了其特征向量长度的对比情况，VSLBP 比 VLBP 大大压缩了特征向量的长度，但当邻域点个数 P 增大时，其特征向量长度增幅大于 LBP-TOP. 但其后从理论及实验上分析，LBP-TOP 用于基于人体轮廓的行为识别，性能不如 VSLBP. 与 LBP-TOP 及 VLBP 相比，VSLBP 旨在“简便性”及“有效性”间寻求折衷.

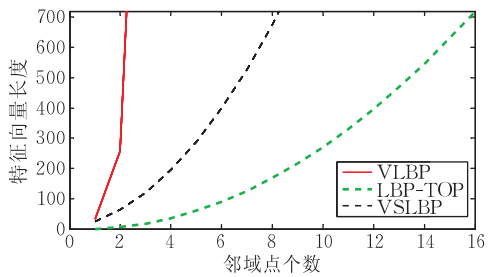


图 6　特征向量长度与邻域点个数的对应关系

5　基于 VSLBP 的行为识别

本文从视频帧的原始图像中提取人体轮廓，排除了杂乱背景的干扰，同时也避免了行为者表观及穿着不同等所带来的混淆信息. 对轮廓序列进行空间对齐并检测其周期，然后将一个周期内的轮廓接连成体积（Volume）作为输入（如图 7 左边所示）.

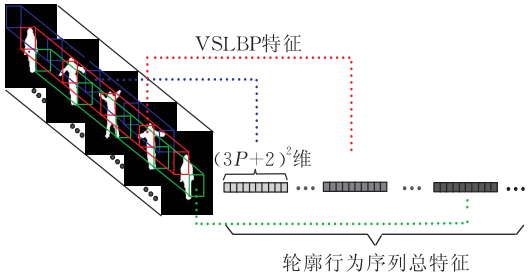


图 7　基于 VSLBP 的行为识别描述

Zhao 等人[11]提出 LBP-TOP，如图 8，在体积中心点的 XY、XT 及 YT 这 3 个正交平面上各作局部二值模式表示，并将其串接起来的直方图序列作为体积的特征，已成功用于动态纹理与面部表情识别中. 通常动态纹理与面部表情不仅在 XY 平面有空间信息，且 XT、YT 平面也包含丰富的运动信息，因此 LBP-TOP 显得可行并且有效. 但人体轮廓序列在 XT、YT 运动信息十分不明显（如图 9），若仅在体积中心点的 3 个正交平面上提取 LBP 特征，将丢失其余帧的大量细节信息，导致识别准确率不高.

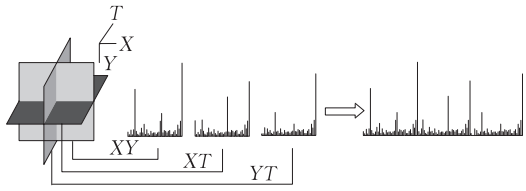


图 8　3 个正交平面局部二值模式表示（LBP-TOP）

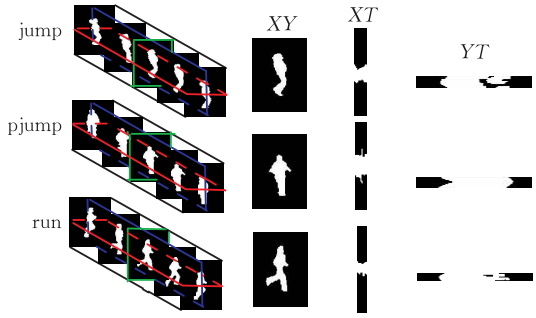


图 9　轮廓行为序列“jump、pjump”和“run”及其在中心点正交 XY、XT 及 YT 平面图

本文从另一方面考虑，提出 VSLBP 的表示方法，并描述了基于 VSLBP 的行为识别的具体过程. 首先将人体轮廓序列构成的体积分成若干个互不重叠的块体积(Block Volume)，如图 7，对每个块体积提取 VSLBP 特征向量，按照其空间结构顺序接连成单一向量表示人体轮廓行为序列的总特征.

获取所有轮廓行为序列总特征后，采用留一法(Leave-One-Group-Out)进行测试，即如果数据库中有 p 个人 q 种不同行为，假设某个人的行为未知，将其作为测试行为，而其余 $(p-1)$ 个人 q 种已知行为作为训练集，通过计算测试序列特征 $\boldsymbol{T}$ 与已知行为训练集特征 $\boldsymbol{M}$ 间的最近卡方距离，即以最小 χ^2 测度，得到此测试行为所属行为类别.

$$\chi^2(\boldsymbol{T},\boldsymbol{M}) = \sum_{i=0}^{n-1} \frac{(T_i - M_i)^2}{(T_i + M_i + \varepsilon)} \tag{5}$$

其中，n 代表行为序列总特征的向量长度，足够小的值 ε 是为了防止 $T_i = M_i = 0$ 时的溢出错误.

6 仿真试验

6.1 测试数据库

为了分析本文提出方法的性能，在公测行为库“Weizmann”库上对其进行了测试. “Weizmann”库包含 90 个视频序列，由 9 个人完成 10 种行为，如图 10，依次是：弯腰(bend)、蹦高挥手(jack)、双腿向前蹦(jump)、原地蹦(pjump)、跑(run)、侧边走(side)、跳(skip)、走(walk)、挥单手(wave1)及挥双手(wave2). 此数据库中的行为具有高度相似性，且行为者穿着、外貌及行为方式都不尽相同，比较接近于实际情形. 由于数据库中背景已知，采用背景剪除法提取所有人体行为轮廓，对其进行空间对齐并统一大小为 60×80，预处理后构成体积，分块后逐帧 $(L=1)$ 在半径 $R=1$ 区域中采样邻域点，再进行基于 VSLBP 的行为识别.

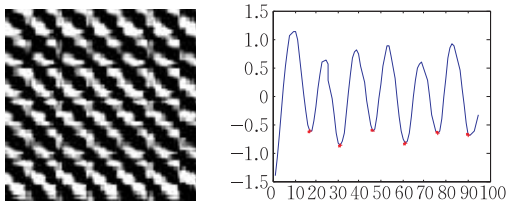
图 10 “Weizmann”库中 10 种行为

6.2 预处理

数据库中的行为一般具有周期性. 本文采用 Culter 等人[12]提出的方法检测行为的周期性，然后将一个周期内的帧接连成体积. 首先计算 t_1 和 t_2 时刻目标 O_t 的自相似性：

$$S_{t_1,t_2} = \min_{|\mathrm{d}x,\mathrm{d}y|\le r} \sum_{(x,y)\in B_{t_1}} |O_{t_1}(x+\mathrm{d}x, y+\mathrm{d}y) - O_{t_2}(x,y)| \tag{6}$$

式中，B_{t_1} 是 O_{t_1} 所占边框大小，经空间对齐后，B_{t_1} 恒定. 较小的半径 r 是为了容许分割误差. 由图 11(a)看出，周期性行为的 $\boldsymbol{S}_{t_1,t_2}$ 亦呈现周期性. 取 $\boldsymbol{S}_{t_1,t_2}$ 的列向量(即固定 t_1 取所有 t_2)进行高斯平滑滤波得图 11(b)，其中的波谷间的时间帧长度即为行为序列的周期长度. 虽然每种行为的持续时间不同(一个行为周期内的视频帧数不同)，但由于 VSLBP 的特征向量长度仅与邻域点个数 P 有关，而与帧的长度无关，所以不需要时间对齐或标准化，因此排除了行为之间持续时间不同或是同一行为不同行为者完成时间不同所带来的影响.



(a) 自相似性$\boldsymbol{S}$ (b) 高斯平滑后的$\boldsymbol{S}$列向量

图 11 “走”行为序列周期检测

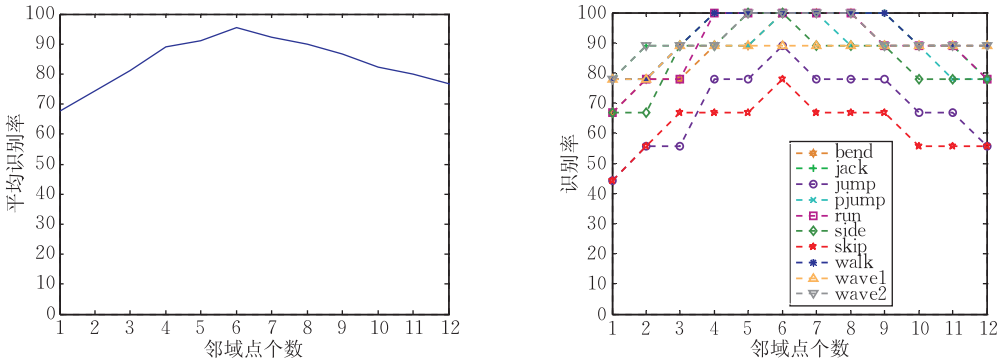
6.3 实验结果及比较

本实验中需要定义的参数是体积分块大小(按空间大小比例 3∶4 分块)及邻域点的个数 P. 表 2 给出了固定邻域点个数，平均识别率与分块大小的对应关系；图 12 分别给出分块大小固定为 12×16 时，所有行为平均识别率及单个行为识别率对应于邻域点个数的关系. 从简便及识别率考虑，取体积分块为 12×16(即分为 25 个互不重叠的块)，邻域点个数为 6. 实验结果由混淆矩阵(confusion matrix)表示. 其每行的各个元素代表此类型的行为被分类为其它行为的概率，则混淆矩阵的对角元素值(表示正确分类的概率)被期望越高越好.

表 2　平均识别率与分块大小的对应关系（固定 $P=6$，按 3∶4 比例分块）

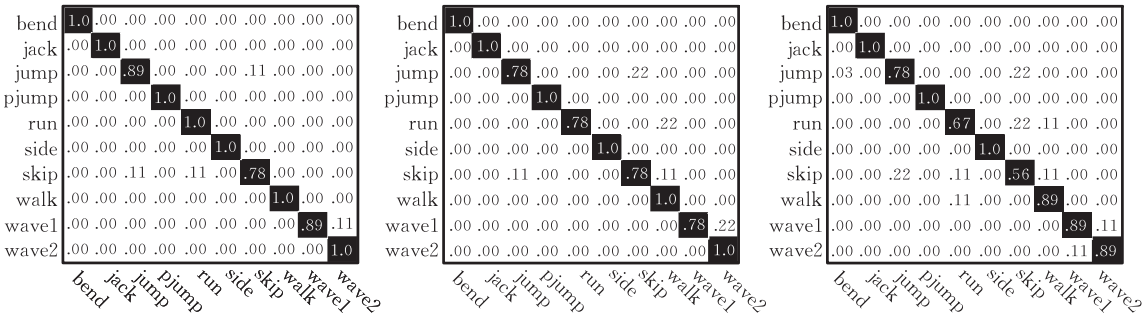
分块大小	平均识别率	分块大小	平均识别率
3×4	74.44	15×20	88.89
6×8	86.67	30×40	77.78
12×16	**95.56**	60×80	56.67

图 13 给出 VLBP、LBP-TOP 及 VSLBP 用于行为识别的纵向对比性实验结果. 相同的实验参数及设置条件下，VSLBP 的识别正确率最高，虽然 VLBP 也能正确识别大部分行为，但其所提取的特征向量长度为 2^{3P+2}，远远大于 VSLBP 的$(3P+2)^2$，



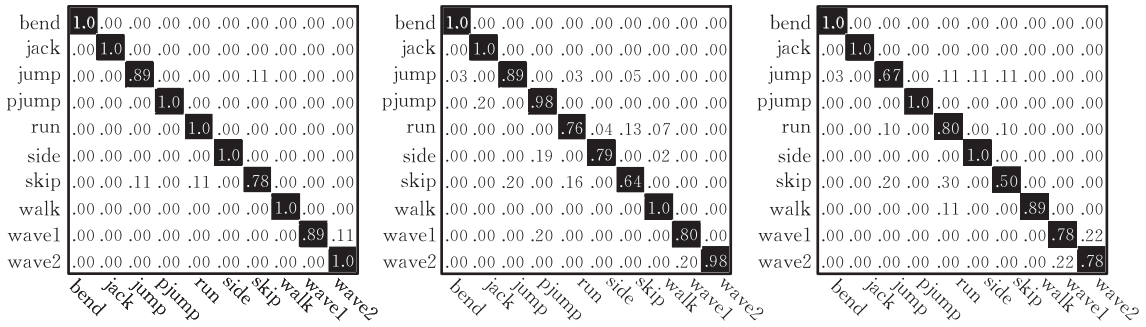
(a) 平均识别率　　(b) 每个行为的识别率

图 12　识别率与邻域点个数对应关系

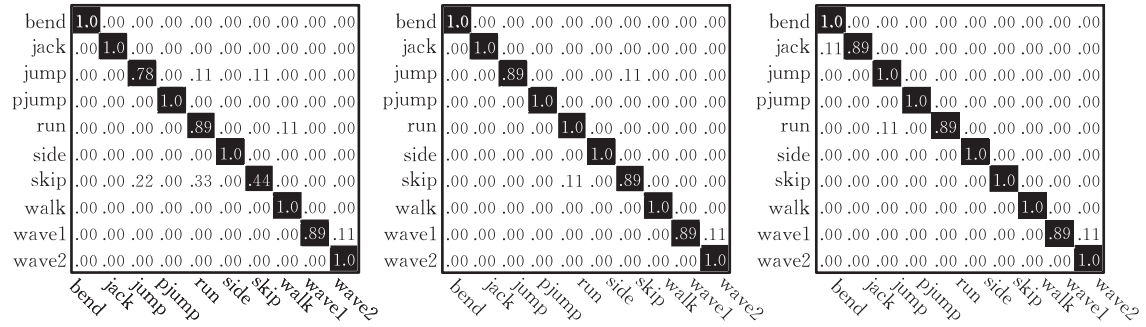


(a) VSLBP　　(b) VLBP　　(c) LBP-TOP

图 13　行为识别结果纵向比较



(a) VSLBP　　(b) 文献[13]方法　　(c) 文献[14]方法



(d) 文献[15]方法　　(e) 文献[6]方法　　(f) 文献[7]方法

图 14　行为识别结果横向比较

在本实验中即 $25\times2^{20}\gg25\times20^2$. 而由于人体轮廓行为序列中 XT 和 YT 平面所包含的时间运动信息不明显，则 LBP-TOP 的识别效果不尽如人意.

图14横向比较了本文提出的方法与一些最新方法[6-7,13-15]在 Weizmann 库上的实验结果. Niebles 等人[13]、Scovanner 等人[14]在原始视频序列上计算时空兴趣点的像素强度或梯度等特征，在 Weizmann 库上它的识别率普遍低于基于轮廓特征的方法[6-7,15]. 基于轮廓特征的方法通常假设前景已完全分割(如 Weizmann 库给出的开源数据)，将行为识别的重点转向特征的描述及识别算法的研究. 本文将 VSLBP 描绘子与基于轮廓特征的其它最新方法[6-7,15]进行比较. Jia 等人[15]在嵌入的低维流形中学习并分类行为，它包含一个降维的过程. Weinzland 等人[6]通过提取若干静态关键姿势，进而把每个行为序列看成一系列关键姿势的表达，图 14(e)列出 50 个关键姿势情形下的识别率. 这种表达虽然简便，但由于丢失时间信息，它不能区分“顺”和“倒”动作，如“起立”和“坐下”，同时它对行为的持续时间比较敏感. Ikizler 等人[7]将姿势分块表达并表示为方向矩形的直方图(HoR)，在姿势表达的基础上加入时间上的速度特征，利用分层的模型进行识别. 图 14(f)给出应用最近邻分类器的识别率. 但相对于 VSLBP，HoR 需要调整更多的参数，执行时间较长. 从图 14 可以看出，以上 3 种方法在双腿向前蹦(jump)和跳(skip)、跑(run)和跳(skip)、挥单手(wave1)和挥双手(wave2)等行为之间均有高度的混淆性. 由于轮廓提取及对齐等预处理方法等不完全相同，它们之间的精确对比较为困难. 但本文提出的方法实现相对简单，且具有较高的识别率. 它不需要降维，可以直接提取低维特征，同时对测试行为的持续时间长度等不敏感.

另外，“Weizmann”库中的 20 个行为类别均为“走”的测试视频，包括 10 个不同拍摄角度(0°～81°，每隔 9°，如图 15)和 10 个行为方式迥异的视频(如图 16)，构成鲁棒性测试库，以测试本文提出方法对拍摄角度、空间遮挡及行为方式等影响因素的稳健性. 每一个测试的“走”行为序列均与数据库中的所有行为序列匹配，找到最佳匹配的行为类型；若最佳匹配行为不是“走”，则寻找次佳匹配行为类型. 表 3 及表 4 显示了鲁棒性测试结果，可以看出，本文提出的方法能够容忍拍摄角度的小幅度变化(0°～36°)和行为方式的某些不规则性.

图 15 不同拍摄角度下的鲁棒性测试视频(图中从左向右，拍摄角度依次为 0°，9°，18°，27°，36°，45°，54°，63°，72°，81°)

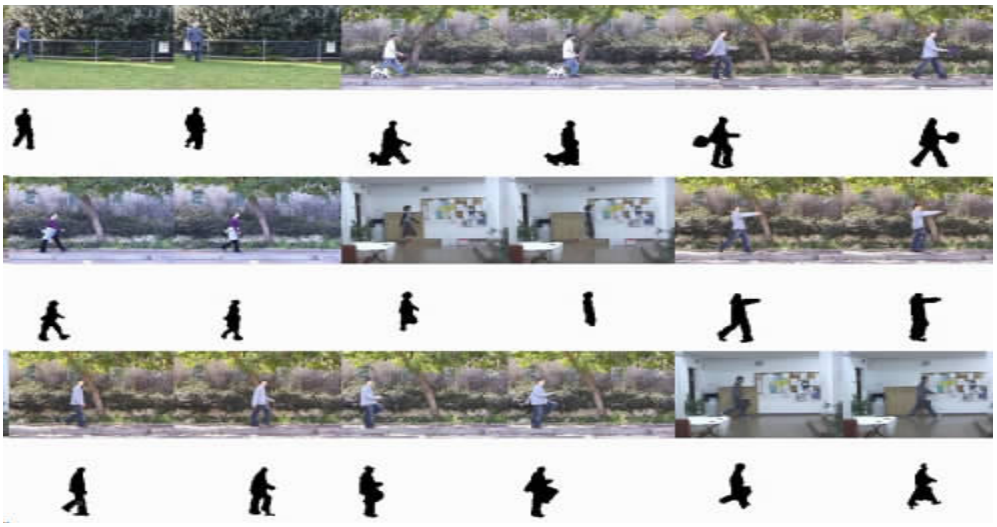


图 16 鲁棒性测试视频(从左到右自上而下依次是对角线直走(diagonal walk)，牵着狗行走(walk with a dog)，边走边甩包(walk and swinging a bag)，带着衣服走(walk in a skirt)，腿部发生遮挡(walk with the legs occluded partially)，梦游式行走(sleepwalk)，跛行(limpwalk)，膝盖抬高走(walk with knees up)，携带公文包走(walk carrying a briefcase)

表 3　鲁棒性测试结果(角度测试)

Viewpoint	Classification Results	
	1st matching	2nd matching
0	walk	/
9	walk	/
18	walk	/
27	walk	/
36	walk	/
45	jump	walk
54	side	walk
63	jump	walk
72	side	jump
81	pjump	side

表 4　鲁棒性测试结果(行为方式不规则性测试)

	Classification Results	
	1st matching	2nd matching
Diagonal walk	walk	/
Walk with dog	skip	walk
Swinging a bag	walk	/
Walk in a skirt	side	walk
Occluded legs	pjump	walk
Sleepwalk	jump	side
Limpwalk	walk	/
Walk with knees up	jack	jump
Carrying a briefcase	walk	/
Normalwalk	walk	/

7　结　论

本文提出体积语义局部二值模式(VSLBP)的表示,并描述了基于 VSLBP 的行为识别方法.此方法从一个全新的角度对行为识别问题进行探索,将刻画图像局部特征的有效算子发展为三维形式并应用于视频数据,它无需降维,能够从空-时体积直接提取有效低维特征.实验结果表明,本文提出的方法不仅能够处理持续时间不同的行为序列,而且对拍摄角度的变化、遮挡及行为方式的不规则性具有较高的鲁棒性.本文今后的研究工作将对 VSLBP 方法加入更为丰富的理论分析,以提取更具有判别性的有效特征,并应用于更自然的行为序列及更复杂的场景中.

参　考　文　献

[1] Efros A, Berg A, Mori G, Malik J. Recognizing action at a distance//Proceedings of the IEEE International Conference on Computer Vision. Nice, France, 2003: 726-733

[2] Dollar P, Rabaud V, Cottrell G, Belongie S. Behavior recognition via sparse spatio-temporal features//Proceedings of the IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China, 2005: 65-72

[3] Ke Y, Rahaud R, Cottrell G, Belongie S. Behavior recognition via spatio-temporal features//Proceedings of the IEEE International Conference on Computer Vision. Beijing, China, 2005: 166-173

[4] Gorelick G, Blank M, Shechtman E, Irani M, Basri R. Action as space-time shapes//Proceedings of the IEEE International Conference on Computer Vision. Beijing, China, 2005: 1395-1402

[5] Wang L, Sutter D. Learning and matching of dynamic shape manifolds for human action recognition. IEEE Transactions on Image Processing, 2007, 16(6): 1646-1661

[6] Weinland D, Boyer E. Action recognition using exemplar-based embedding//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8

[7] Ikizler N, Duygulu P. Histogram of oriented rectangles: A new pose descriptor for human action recognition. Image and Vision Computing, 2009, 27(10): 1515-1526

[8] Ojala T, Pietikainen M, Maenpaa T. Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987

[9] Mu Y, Yan S, Liu Y, Huang T, Zhou B. Discriminative local binary patterns for human detection in personal album//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8

[10] Heikkila M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2006, 28(4): 657-662

[11] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expression. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2007, 29(6): 915-928

[12] Culter R, Davis L. Robust real-time periodic motion detection. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2000, 22(8): 781-796

[13] Niebles JC, Wang H, Feifei L. Unsupervised learning of human action categories using spatial-temporal words. International Journal of Computer Vision, 2008, 79(3): 299-318

[14] Scovanner P, Ali S, Shah M. A 3-dimensional SIFT descriptor and its application to action recognition//Proceedings of the International Conference on Multimedia. New York, USA, 2007: 357-360

[15] Jia K, Yeung DY. Human action recognition using local spatio-temporal discriminant embedding//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008: 1-8

WU Xian, born in 1985, Ph. D.. Her current research interests include computer vision and pattern recognition.

LAI Jian-Huang, born in 1964, professor, Ph. D. supervisor. His main research interests include face recognition, pattern recognition and machine learning, image analysis and understanding.

Background

Visual analysis of human motion concerns the detection, tracking and recognition of human, and more generally, the recognition and understanding of human activities. In particular, recognition of human activities has a wide range of promising applications such as visual surveillance, perceptual interfaces, video indexing and browsing. Although there has been much work on human motion analysis over the past decade, action recognition still remains challenging, especially in complex conditions, e. g. , occlusion, illumination, camera motion and different viewpoints. In this paper, human action is treated as a volume stacked by silhouette image of each frame, and recognizing human action can be performed from the angle of local binary patterns (LBP). Volume semantics LBP (VSLBP) is proposed and applied to action recognition, which gets the high recognition accuracy in a recent action dataset named "Weizmann", and the performance demonstrate that our approach are considerably robust to viewpoint changes, partial occlusion and irregularities in the motion styles. The work in this paper is supported by the National Natural Science Foundation of China (Nos. 61173084, 60803083), and National Natural Science Foundation of China-Guangdong United Project (U0835005).