中图法分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1006-8961(200) - -
论文引用格式：

多视角深度运动图的人体行为识别

刘婷婷 ， 李玉鹏 ， 张良

中国民航大学智能信号与图像处理实验室, 天津 300300

摘　要：**目的** 使用运动历史点云进行人体行为识别的方法，由于点云数据量大，在提取特征时运算复杂度很高。而深度运动图进行人体行为识别的方法，提取特征简单，但是包含动作信息并不全面，限制了识别率的上限。针对上述两种方法存在的问题，提出了一种多视角深度运动图的人体行为识别算法。**方法** 首先采用深度图序列生成运动历史点云对动作进行表示，接着将运动历史点云旋转特定角度补充更多视角下的动作信息；然后将原始的和旋转后的运动历史点云投影到笛卡尔坐标平面生成多视角深度运动图，对其提取方向梯度直方图，采用串联融合生成特征向量；最后将特征向量送入到支持向量机中进行分类识别，在MSR Action3D和自建数据库上对算法进行验证。**结果** MSR Action3D数据库有两种实验设置，其中采用实验设置一时，该算法识别率达到96.8%，比APS_PHOG(axonometric projections and PHOG feature)算法高2.5%，比DMM（Depth motion maps）高1.9%，比DMM_CRC（Depth motion maps and Collaborative representation classifer）算法高1.1%。采用实验设置二时，识别率为93.82%，比DMM(Depth motion maps)算法高5.09%，比HON4D(Histogram of Oriented 4D Surface Normals)算法高4.93%。在自建数据库上该算法识别率达到97.98%，比MHPC算法高3.98%。**结论** 实验结果表明，多视角深度运动图不但解决了运动历史点云提取特征复杂的问题，而且使深度运动图包含了更多视角下的动作信息，有效的提高了人体行为识别的精度。

关键词 ： 人体行为识别; 深度图像; 深度运动图; 多视角深度运动图; 运动历史点云; 方向梯度直方图; 支持向量机

Human Action Recognition Based on Multi-perspective Depth Motion Maps

Liu Tingting, Li Yupeng, Zhang liang

Key Laboratory of Advanced Signal and Image Processing, Civil Aviation University of China, Tianjin, 300300

Abstract: Objective Due to insensitivity to illumination of depth data, action recognition based on depth data is gradually carried out. There are two main methods, one is point clouds converted from depth maps, the other is depth motion map (DMM) generated from depth maps projection. Motion history point cloud (MHPC) was proposed to represent actions, but the large amount of points in MHPC incur expensive computations when extracting features. Depth motion map is generated by stacking motion energy of depth maps sequence projected onto three orthogonal Cartesian planes. Projecting the depth maps onto a specific plane get additional body shape and motion information. However, depth motion map contains motion information inadequately, which caps the human action recognition accurate, even though it is simple to extract features from depth motion map. In other words, an action is represented by DMMs from

基金项目:国家自然科学基金（61179045）
收稿日期:　　；　**修回日期**:
Supported by: The National Natural Science Foundation of China (No. 61179045)

only three views, so the action information from other perspectives is lacking. To solve above problems, multi-perspective depth motion maps for human action recognition is proposed. **Method** In the algorithm, firstly, the motion history point cloud (MHPC) is generated from depth maps sequence to represent actions. Through rotating the motion history point cloud around axis Y a certain angle, motion information under different perspectives is supplemented. Then primary MHPC is projected onto three orthogonal Cartesian planes, and rotated MHPC is projected onto XOY planes. Multi-perspective depth motion map generated from these projected MHPC. After projection, the point clouds are distributed in plane where there are many overlapping points under the same coordinates. These points may come from the same frame of depth map, or may come from different frame.We use these overlapping points to generate DMM so as to capture the spatial energy distribution of motion. For example, the pixel in depth motion maps generated from MHPC projected onto XOY plane is the sum of absolute difference of z of the adjacent two overlapping points belonging to different frames. DMMs generation from MHPC projected onto YOZ plane and XOZ plane are similar to this, only the point of the z correspondingly is changed to the x and y. MHPC is projected onto three orthogonal Cartesian planes to generate DMM from front, side, top view respectively. The rotated MHPC is projected onto XOY plane to generate DMM under different view. Multi-perspectives depth motion maps encoding the 4D information of an action to 2D maps are utilized to represent an action, so the action information under more perspective is replenished. It should be noted that, the value of x,y,z of points in projected MHPC are normalized to fixed values as the multi-perspective depth motion maps image coordinates, which can reduce the intra-class variability due to different action performers. According to the experience, this paper normalizes the values of x and z to511, and y to1023. The histogram of oriented gradient (HOG) are extracted from each depth motion map, then they are concatenation as feature vectors of an action. Lastly, the SVM classifier is adopted to train the classifier to recognize the action. Experiments with this method on the MSR Action3D dataset and our dataset were done. **Result** The proposed algorithm exhibits improved performances on MSR Action 3D database and our dataset. There are two experimental settings for MSR Action3D. This algorithm achieves an identification rate of 96.8% in experiment setting one, which is obviously better than most algorithms. The action recognition rate of the proposed algorithm is 2.5% higher than that APS_PHOG(axonometric projections and PHOG feature) algorithm, 1.9% higher than that of DMM algorithm, 1.1% higher than that of DMM_CRC（Depth motion maps and Collaborative representation classifer）algorithm. In the second experimental setting, the recognition rate reached 93.82%, 5.09% higher than DMM algorithm, 4.93% higher than HON4D algorithm, 2.18% higher than HOPC algorithm, 1.92% higher than DMM_LBP feature fusion. In our database, the recognition rate of this algorithm is 97.98%, 3.98% higher than MHPC algorithm. **Conclusion** MHPC is used to represent the action, which supplement the action information from different perspectives by rotating certain angles. Multi-perspective depth motion maps are generated by computing the distribution of overlapping points in the projected MHPC, which captures the spatial distribution of the absolute motion energy. Coordinate normalization reduce the intra-class variability. The experimental results show that multi-perspective depth motion map not only solve the difficulty of extracting features from motion history point cloud, but also supplement motion information of traditional depth motion map. Human action recognition base on multi-perspective depth motion map outperform some existing methods. The new approach combines the method of point clouds with the method of deep motion map, which full play the advantages of both and weakens the disadvantages.

Key words: action recognition； depth maps； depth motion maps； multi-perspective depth motion maps； motion history point cloud；Histogram of Oriented Gradient；support vector machine

0 引 言

人体行为识别在智能视频监控、视频内容检索、人体运动分析、辅助医疗等领域有着广泛的应用，因此国内外对此进行了大量的研究。最初人体行为识别的研究是基于 RGB 信息的，产生了人体关键姿态[1]、剪影[2]，时空特征[3]等方法。但是由于人体动作的复杂性和

多变性，相机视角改变和抖动,光照变化,遮挡与自遮挡等因素使人体行为识别仍充满挑战。近些年深度图像获取技术逐渐成熟，深度图像受光照条件变化的影响较小，其仅与物体的空间位置有关，能直接反映物体表面的三维特性。基于以上优势，利用深度数据进行人体行为识别的研究逐渐展开。

利用骨骼关节点能够建模人体动作模型，Xia 等人[4]提出关节点位置直方图（Histogram of 3D Joint Location,HOJ3D）对人体动作进行表示，采用离散隐马尔科夫模型进行分类。冉宪宇等人[5]采用自适应骨骼中心的行为识别算法进行特征提取。Wang 等人[6]提出运用关键姿态序列（Key-pose-motif）对动作进行描述，对动作方式的差异具有鲁棒性。虽然一些利用骨骼关节点的方法能够达到较高的识别率，但是只有当人在正向面对摄像机的情况下才能准确估计骨骼关节点的位置，并且当不是直立做动作时得到的关节点会非常不稳定[7]。

通过深度相机采集的深度图像可以直接转化为 3D 点云数据，Rahmani 等人[8]提出方向主成分直方图（Histogram of Oriented Principal Components，HOPC）对动作点云数据进行描述，从局部几何形状上来刻画动作。Oreifeij 等人[9]采用直方图来捕获点云序列所构成的 4D 曲面法线的方向分布（Histogram of Oriented 4D Surface Normals，HON4D），该方法需要对动作进行时-空对齐，具有一定的局限性。Yang 等人[10]通过聚类点云序列中每个点邻域内的 4D 法线形成新的超级法向量描述子（Super Normal Vector, SNV），能够同时捕获局部运动和几何信息。以上基于点云数据的方法，提取局部特征对动作进行描述，对局部自遮挡具有鲁棒性，但是这些方法忽略了各点之间的全局约束。Liu 等人[11]提出利用运动历史点云（Motion History Point cloud，MHPC）对动作视频进行表示，其将一个动作的深度图序列看作是一个整体进行处理，完整的保留了动作的空间与时序信息，完成了对动作的全局表示。但是运动历史点云中大量的点云数据给设计高效的特征提取算法带来挑战。

深度图片能够提供动作执行者的身体形状和运动信息，Shen 等人[12]提出了一种深度差异运动历史图像（Depth Difference motion history image, DDMHI），将其轴向投影后，提取分层梯度方向直方图（Pyramid Histogram of Oriented Gradients,PHOG）对动作进行描述。Yang 等人[13]将深度图片进行投影生成深度运动图（Depth Motion Map，DMM），用来捕获时间聚集的相对运动能量，将 4 维的动作信息编码到三个视图下的 2D 图片。Chen 等人[14]对 DMM 采用 l_2 -正则化协同表示分类器实现动作识别，验证了算法的实时性。后续 Chen 等对 DMM 提取 LBP 特征，分别采用特征层融合（DMM_LBP_FF）和决策层融合（DMM_LBP_DF）两种方式对动作进行识别[15]。以上基于 DMM 的方法相对简单，计算量相对较小，但是获得的动作信息只局限在三个投影视图下，不能获得其他视角下的动作信息。

由以上分析可得，基于点云数据和 DMM 进行人体行为识别的算法各有优缺点。为了使深度运动图捕获更加全面的动作信息，本文利用运动历史点云（MHPC）将动作表示成一个三维物体，能够自由将其旋转任意角度的特性，继续采用 MHPC 对动作进行表示，然后对其进行旋转投影生成多视角深度运动图。其基本思路是：首先，由深度图序列生成 MHPC 对动作进行表示。接着，利用旋转矩阵将其绕 Y 轴旋转，增加了动作执行者在左右偏离摄像头时的动作状态信息。然后，将原始的和旋转后的 MHPC 投影到笛卡尔坐标平面上，根据投影后点云中点的分布特点生成多视角深度运动图，捕获了更多视角下的空间能量分布。堆积的能量在多视角深度运动图上表现为不同的外观与形状，利用方向梯度直方图[16]（HOG）对其进行特征提取。最后，采用支持向量机[17]（SVM）进行分类，识别出人体动作。提出的基于多视角深度运动图的人体行为识别算法将运动历史点云和深度运动图两种方法结合起来，使得深度运动图包含了更多视角下的运动信息，并且解决了运动历史点云中巨大的点云数据带来的提取特征困难的问题，很好地发挥了两者的优势，弱化了其缺点。图 1 给出了本文算法框架。

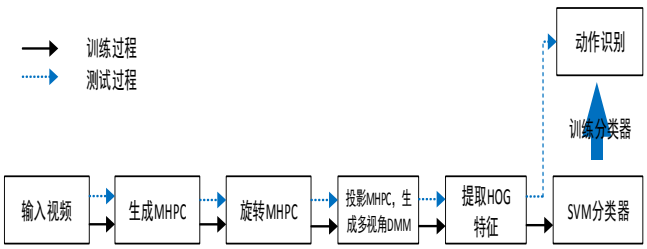


图 1　本文算法框架

Fig.1 Framework of the proposed algorithm

本文方法主要的创新点体现在以下两个方面：1）提出对 MHPC 进行旋转，使得动作信息的获取

不再局限于 1 个或 3 个视角下；2）对 DMM 进行改进，将原始的 MHPC 和经过旋转之后的 MHPC 进行投影，根据投影后点云中点的分布情况生成多视角深度运动图，生成过程中的坐标归一化降低了类内差异。

1　动作表示

本文提出基于多视角深度运动图的人体行为识别算法，首先将深度图序列生成 MHPC，接着将其进行旋转，将原始的与旋转后的 MHPC 进行投影生成多视角深度运动图，实现对动作样本的表示。

1.1 运动历史点云

运动历史点云（Motion History Point Cloud，MHPC）将一个动作样本的深度图序列压缩成一个包含空间信息与时间信息的点的集合，公式如 $M=\{P_1,P_2,\cdots P_n\}$，其中 n 表示 M 中点的个数。点云中任一点的坐标定义为 $P_j(x,y,z,h), j\in(1,n)$，其中 P_{j_x},P_{j_y},P_{j_z} 是指在相机坐标系下点的 x,y,z 值，用来记录动作发生的位置；P_{j_h} 指深度图的帧号，用来记录动作发生的时间。假如一个动作序列包含 N 帧深度图片，将每帧提取前景的深度图片，从图像坐标系映射到相机坐标系得到每帧的点云，然后将其填充到 MHPC 中，直至读取完所有深度图片,生成框架如图 2。

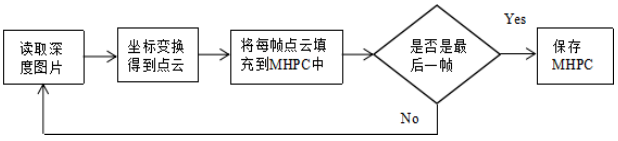


图 2　MHPC 生成流程图

Fig.2 The framework of the generation of MHPC

“高手挥舞”动作生成的 MHPC 如图 3 所示，由图可知 MHPC 是一个三维的立体，其坐标系方向以屏幕右方为 X 轴正方向，屏幕上方为 Y 轴正方向，垂直屏幕向外为 Z 轴正方向。通过旋转可以得到不同视角下的 MHPC，将 MHPC 绕 X 轴旋转 θ 度，得到上下偏离摄像头 θ 视角下的动作信息；将 MHPC 绕 Y 轴旋转 θ 度，得到左右偏离摄像头 θ 视角下的动作信息；将 MHPC 绕 Z 轴旋转一定角度，得到航偏角 θ 下的动作信息。但是在现实情况下，动作执行者多是左右偏离摄像头。所以，本文利用公式 1 将运动历史点云 M 绕 Y 轴旋转 θ 度得到 M_θ，θ 根据实验经验可以选择 $\{\pm25^\circ,\pm30^\circ,\pm45^\circ\}$ 中的一个或多个。经过旋转过后，一个动作样本由 M 和若干个旋转后的 M_θ 来表示，补充了更多视角下的动作信息，使得动作表示的更加充分。

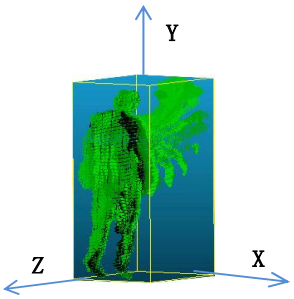


图 3　MHPC 效果图

Fig.3　MHPC of the action of high wave

$$R_y(\theta)=\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix} \tag{1}$$

1.2 多视角深度运动图

深度图片能够提供动作的形状和运动信息，将一个深度视频序列中的每一帧深度图片投影到三个正交的笛卡尔坐标平面上，具体为投影到 XOY 平面、YOZ 平面、XOZ 平面，相应的得到前视图 map_f，侧视图 map_s，俯视图 map_t，这三个投影图的像素值分别为深度图中点的 Z,X,Y 值。将上述得到的三个投影视图序列的相邻帧做差分运算，然后取绝对值与阈值判断后累加，得到深度运动图（DMM）。其计算公式如下：

$$D_v=\sum_{i=begin}^{end-1}(|map_v^{i+1}-map_v^i|>\varepsilon) \tag{2}$$

其中，i 为帧的索引，$begin$ 表示起始帧，end 表示终止帧，map_v^i 表示第 i 帧在视图 v 下的投影图，$v\in\{f,s,t\}$。$|map_v^{i+1}-map_v^i|>\varepsilon$ 是运动能量的二值图。

DMM 能够捕获运动的能量，表现为不同的外形与形状，因此能够很好的表征动作类别。但是 DMM 表示的动作只是局限在三个视图下，所以，本文利用上一节得到的多视角运动历史点云进行投影，生成多视角深度运动图。将没有旋转的 M 投影到 XOY 平面、YOZ 平面、XOZ 平面，得到前视图，侧视图，俯视图下的 $D_{f,s,t}$。将绕 Y 轴旋转 θ 度的 M_θ

投影到 XOY 平面，获得视角 θ 下的 D_θ，最终，一个动作样本由 $D_{f,s,t,\theta}$ 表示，即为多视角深度运动图，比传统的深度运动图包含更多视角下的运动信息。

将 MHPC 投影到 XOY 平面，即 MHPC 中所有点的 z 值置为 0，其他坐标值不变，MHPC 中的点全部分布在 XOY 平面。同理，投影到 XOZ 平面或者 YOZ 平面时，分别将其所有点的 y 值或 z 值置为 0。投影后得到的前视图、侧视图、俯视图下的 MHPC，如图 4 的(a)中的前三幅图。旋转后的 MHPC 投影到 XOY 平面，效果如图 4 的(a)中的后两幅图，分别表示 $\theta=-45^\circ$ 和 $\theta=45^\circ$ 的情况，获得了动作在视角 θ 下的运动信息。投影之后的点云在同一坐标下有好多重叠的点，由此来生成多视角深度运动图。此时投影时并不会把坐标值置 0，而是用来计算多视角深度运动图的像素值。以投影到 XOY 平面为例，在 MHPC 中存在一些 x,y 值相同但 z 值不同的重叠的点。这些 z 值不同的点中，若属于同一帧深度图像，其 P_h 相同，若属于不同帧的深度图像，其 P_h 不同。依次遍历这些重叠的点，如果相邻两点属于不同帧，将两点的 z 值做差取绝对值叠加，直至遍历完此坐标下所有重叠的点。将取绝对值叠加最终得到的值，作为多视角深度运动图在此坐标下的像素值，反映了动作在此坐标下的绝对空间能量。

将 XOY 平面所有坐标执行以上相同操作，得到在前视图下动作的绝对空间能量分布。类似的将 MHPC 投影到 YOZ 平面和 XOZ 平面，做差取绝对值叠加的分别是 x,y 值，得到侧视图和俯视图下动作空间能量分布。由多视角运动历史点云生成多视角深度运动图，将三维的空间信息转化到二维图片上，包含丰富的外观与形状信息，可以提取 HOG 特征对动作进行表示。相比于直接在三维的运动历史点云中提取关键点，然后对关键点特征描述后来对动作进行表示要简单的多。

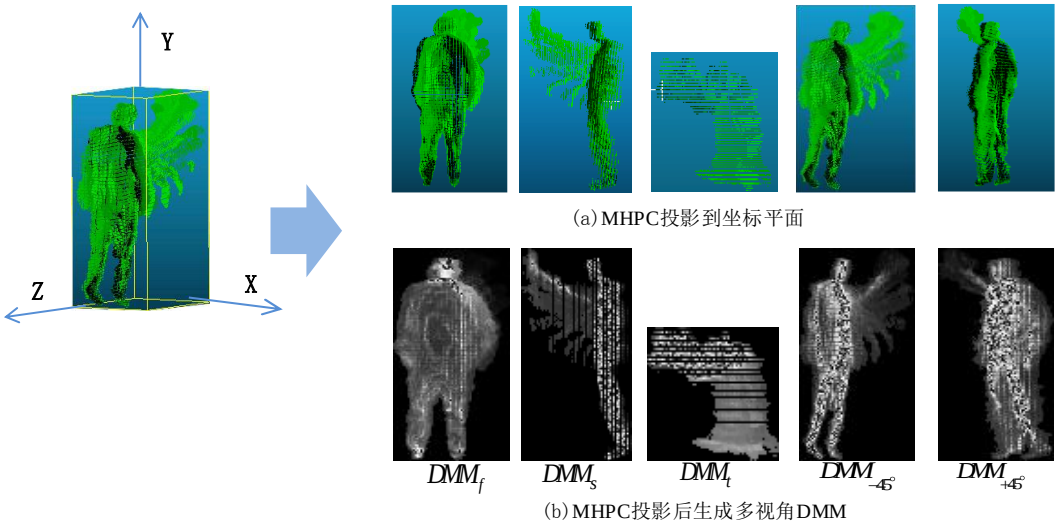
假设投影到 XOY 平面、YOZ 平面、XOZ 平面的运动历史点云中，某一坐标下有 m 个重叠的点（投影平面不同，m 不一定相同），依次遍历重叠的点，相应的将其相邻的属于不同帧的两点的 z 值、x 值、y 值进行做差取绝对值累加，作为多视角深度运动图的像素值。计算公式如下，其中 i 为重叠的点的索引：

$$D_f=\sum_{i=1}^{m-1}|P_{.z}^{i+1}-P_{.z}^{i}| \quad (if\ P_{.h}^{i+1}\neq P_{.h}^{i}) \tag{3}$$

$$D_s=\sum_{i=1}^{m-1}|P_{.x}^{i+1}-P_{.x}^{i}| \quad (if\ P_{.h}^{i+1}\neq P_{.h}^{i}) \tag{4}$$

$$D_t=\sum_{i=1}^{m-1}|P_{.y}^{i+1}-P_{.y}^{i}| \quad (if\ P_{.h}^{i+1}\neq P_{.h}^{i}) \tag{5}$$

D_θ 是由 M_θ 投影到 XOY 平面得到，计算公式如 D_f。



(a) MHPC投影到坐标平面

(b) MHPC投影后生成多视角DMM

图 4　MHPC 的投影和多视角 DMM

Fig.4 The projection of MHPC and multi-view DMM ((a) Projections of MHPC onto Cartesian planes, (b) Multi-view DMM generated from projected MHPC)

由于映射到相机坐标系的点的 x,y,z 值是实数，并且变化范围较小，不能直接作为图像的坐标，所以将其进行归一化。人有高矮胖瘦之分，当同一动作由不同人执行时，存在较大的类内差异。在生成多视角深度运动图时利用公式（6）对坐标进行归一化，能够极大的减少这种类内差异。最后生成的 $D_{f,s,\theta}$ 为 512 像素×1024 像素，D_t 为 512 像素×512 像素，

在提取特征时将尺寸分别调整到 64 像素×128 像素和 64 像素×64 像素，生成的 DMM 效果图如图 4 中的（b）。$D_{f,s,t,\theta}$ 对动作样本进行表示，旋转的次数和角度的大小决定着表示动作的丰富程度，进而影响最终的识别效果。

$$X_{norm} = C*\frac{X-X_{min}}{X_{max}-X_{min}}, \text{(C为常数, X为需要归一化的向量)} \quad (6)$$

2 特征提取与分类器

2.1 HOG 特征

梯度方向直方图（Histogram of Oriented Gradient, HOG）,采用将图像分块分单元的方法，既可以描述图像的局部形状信息，也可以表征局部像素点之间的关系。本文在对多视角深度运动图 $D_{f,s,t,\theta}$ 中的每幅图片提取 HOG 特征时，将单元大小设置为 8×8 个像素大小，块的大小为 4×4 个单元，根据生成多视角深度运动图的大小，其被划分为 2×4 个或 2×2 个互不重叠的块。每个单元内的梯度方向平均分成 12 个区间（bin）。因此得到 HOG 特征向量为 1536 维或 768 维。最后，将多视角深度运动图得到的 HOG 特征串联起来，生成该动作样本的特征向量。

2.2 SVM 分类器

支持向量机（SVM）是一种通用二分类器，假设给定线性可分数据 $=\mathrm{D}x_1 \quad \{y_1 \quad (\quad x_2, \cdot \quad y_2)$ 其中，$x_i \in X = \mathrm{R}^n$，x_i 为某一样本的特征向量。$y_i \in Y = \mathrm{R}^n = \{-1, +1\}(i = 1, 2, \cdots, m)$。在样本空间中，超平面的方程为

$$\gamma = \frac{2}{\|w\|} \quad (7)$$

γ 被称为间隔。要得到最优超平面，就要使得数据到超平面的间隔最大，即

$$\max_{w,b} \frac{2}{\|w\|} \text{ s.t. } y_i(w^{\mathrm{T}}x_i + b) \geq 1,\ i = 1, 2, \cdots, m. \quad (8)$$

此目标函数可以等价为

$$\min_{w,b} \frac{1}{2}\|w\|^2 \text{ s.t. } y_i(w^{\mathrm{T}}x_i + b) \geq 1,\ i = 1, 2, \cdots, m. \quad (9)$$

由此可以求出最优解 w, b，得到线性可分支持向量机。当训练数据近似线性可分时，通过引入松弛变量，通过软间隔最大化，学习一个线性分类器，即线性支持向量机；当训练数据线性不可分时，利用内积核函数把线性不可分的样本映射到一个更高维的空间中，通过软间隔最大化，学习非线性支持向量机。应用较广泛的内积核函数有线性核函数、多项式核函数、径向基函数（RBF）和 Sigmoid 函数。因为本文得到的样本的特征维度较高，根据实践经验，选择了线性核函数。通过网格搜索（Grid Search）方法确定参数 C（惩罚系数）的值。

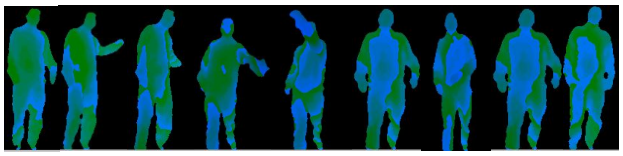
针对多分类问题，一般是将其拆分为若干个二分类问题，主要包括“一对多”（one-versus-rest, OVR）和“一对一”（one-versus-one, OVO）两种方法。假设给定 N 个类，OVR 是需要训练 N 个分类器，其中分类器 i 是将 i 类数据设置为正类，剩余 N-1 个 i 类以外的类设置为负类，因此每一个类都需要训练一个二类分类器。对于一个需要分类的数据 X，将使用投票的方式确定 X 的类别。而 OVO 方法是对 N 个类中每两个类都训练一个分类器，总共得到 $N*(N-1)/2$ 个二分类器。对于一个需要分类的数据 X，需要经过所有分类器的预测，同样使用投票的方式来决定 X 最终的类别。本文利用 LibSVM 工具包，采用“一对一”方法实现多分类。

3 实验及结果分析

3.1 数据库介绍与实验参数

为验证算法的有效性，将所提算法在 MSR Action3D 数据库[18] 与自建数据库[11]上进行实验。MSR Action3D 是目前应用较广泛的公共数据库，由 20 个动作类别构成，每个动作类别由 10 个演员执行 2~3 次，总共有 567 个动作样本。由 RGBD 摄像头进行采集，得到的深度图像的分辨率是 320 像素×240 像素。为了验证更多动作的识别效果，同时在自建数据库上进行实验。自建数据库由深度摄像头 KinectV2.0 采集，得到的深度图像的分辨率是 512 像素×424 像素。自建数据库共包含 11 个动作：*two hand-wave, run, jump, mark time, bend, squat, side, leg kick, pitch, golf-swing, side-boxing*。其中在 MSR Action3D 数据库中无 *run, jump, mark time, squat, side, side-boxing* 动作。在采集动作样本深度视频时，固定摄像机，背景不变，每一个动作由 9 个人各执行 1 次，共得到 99 个动作样本。自建数据库相似动作较少，比 MSR Action3D 数据库相对简单。

图 5 显示了上述两个数据库中动作样本的深度图像。



(a) MSR Action3D

(b) 自建数据库

图 5 实验数据库

Fig.5 Experiment datasets ((a)MAR Action3D, (b)Our dataset)

本文实验主要在惠普 Z640 工作站上进行，2.1GHzCPU,64G 内存，利用 Visual Studio 2012 和 MATLAB R2016a 进行编程。由于采用线性核函数，仅对参数 C 利用网格搜索方法寻得最优值，然后利用 LibSVM 工具包基于最优的参数 C 训练生成多分类器，进而实现动作识别。其中在 MSR Action3D 数据库中平均每个样本训练花费 7.1ms，识别需要花费 7.3ms，文献[15]采用的是基于核的极限学习机在识别时平均每个样本花费 1.21ms,相比于本算法快了 6.09ms。在自建数据库中，每个样本训练需要花费 2.3ms，识别需要花费 4.5ms。

3.2 识别结果与分析

3.2.1 MSR Action3D 数据库

数据库有两种实验设置，设置一是依据动作的复杂程度和相似性将动作集分为三个动作子集：AS1、AS2、AS3，每个动作子集中包含 8 个动作，如表 1。针对每个动作子集，有三种测试方法。方法一将每个子集的 1/3 样本用来训练，其他的样本用来测试；方法二将每个子集的 2/3 样本用来训练，其他的样本用来测试；第三种测试方法是用一半人(1,3,5,7,9)的动作样本做训练，剩下的人(2,4,6,8,10)的动作样本做测试。第二种实验设置是在整个动作集中，选取一半(1,3,5,7,9 号)演员的动作样本用来训练，剩下的(2,4,6,8,10 号)演员的动作样本用来测试。因为包含更多的样本，所以实验设置二更具有挑战性。

不断增加旋转次数，会增加不同视觉下的动作信息。因此多视角深度运动图中 DMM 的数量代表着动作信息的丰富度，对识别率产生影响。采用实验设置二进行实验，得到 DMM 数量对识别结果产生的影响如图 6 所示。图中数量 1 指 D_f，3 指 $D_{f,s,t}$，5 指 $D_{f,s,t,\pm25^o}$，7 是指 $D_{f,s,t,\pm25^o,\pm45^o}$，9 是 $D_{f,s,t,\pm25^o,\pm30^o,\pm45^o}$。当 DMM 的数量由 1 增加到 3 时，识别率增加了 7.6%，由 3 增加到 5 时，识别率增加了近 8.4%，说明随着不同视角下动作信息的增加，识别率显著提高。当由 5 增加到 7 个时，识别率增加趋于平缓，只增加了 1.8%，达到了最高识别率 93.8%。由 7 增加到 9 个时，识别率开始出现下降，可能是因为随着 DMM 的逐渐增加，提取的 HOG 特征串联融合后出现了大量的冗余，降低了识别的精度。由此可以说明，当 DMM 数量较少时获取的动作信息不充分，适当增加不同视角下的 DMM 会显著增加识别率。由本文实验可得，当多视角深度运动图包含 7 个视角时识别率达到最高，此时获得的动作信息已经足够，再增加其他视角识别率反而降低。当取 $D_{f,s,t,\pm25^o,\pm45^o}$ 对动作进行表示时，效果最

表 1 MSR Action3D 数据库的三个动作子集

Table 1 Three action subsets of MSR Action3D dataset

Action set 1(AS1)	Action set 2(AS2)	Action set 3(AS3)
Horizontal wave(2)	High wave(1)	High throw(6)
Hammer(3)	Hand catch(4)	Forward kick(14)
Forward punch(5)	Draw x(7)	Side kick(15)
High throw(6)	Draw tick(8)	Jogging(16)
Hand clap(10)	Draw circle(9)	Tennis swing(17)
Bend(13)	Two hand wave(11)	Tennis serve(18)
Tennis serve(18)	Forward kick(14)	Golf swing(19)
Pickup throw(20)	Side boxing(12)	Pickup throw(20)

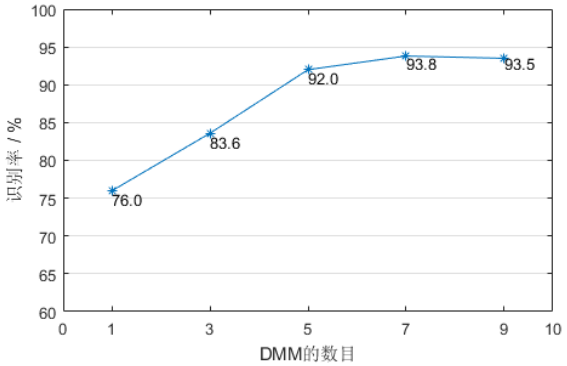


图 6 DMM 数量对识别率的影响

Fig.6 The influence of DMM number on recognition

好，所以在后面的实验中，均采用此参数设置。

采用实验设置一实验，并与现有方法进行比较，结果如表 2 所示。可以看出本文的方法在大多数情况下能取得最高的识别率。尤其在测试二中 AS1 和 AS2 两个动作序列中的识别率均为 100%，在交叉验证实验中，本文方法得到的平均识别率较高于其他方法。对三个动作子集识别率取平均，达到了 96.8%的识别率，明显优于其他算法。

表 2　在实验设置一时与现有方法比较结果(%)

Table 2　Comparison of our method with other existing methods in experiment setting 1(%)

Methods	Test one				Test two				Cross subject				Average
	AS1	AS2	AS3	Average	AS1	AS2	AS3	Average	AS1	AS2	AS3	Average	
DMM[13]	**97.3**	92.2	98.0	95.8	98.7	94.7	98.7	97.4	**96.2**	84.1	**94.6**	91.6	94.9
APS_PHOG[12]	94.8	95.2	97.9	96.0	97.8	98.8	98	98.2	90.6	81.4	**94.6**	88.8	94.3
DMM_CRC[14]	**97.3**	96.1	**98.7**	**97.4**	98.6	98.7	**100**	99.1	**96.2**	83.2	92.0	90.5	95.7
本文方法	97.0	**98.2**	96.5	97.2	**100.0**	**100.0**	99.1	**99.7**	91.7	**95.1**	93.9	**93.6**	**96.8**

注：加粗字体表示所在列最优值

采用实验设置二实验，并于现有方法进行比较，结果列在表 3 中。虽然与文献[13]一样都是采用 HOG 描述子，但是本文方法比其提高了 5.09%，说明由多视角的深度运动图比传统深度运动图表示动作更加有效。因为多视角深度运动图增加了不同视角下的动作信息，使得动作表示更加全面；同时多视角深度运动图生成过程中的坐标归一化，增加了类内差异的鲁棒性，进而提高了识别率。相比于 DMM_LBP_FF[15]，本文对由 MHPC 生成的多视角深度运动图提取 HOG 特征，然后同样采用特征融合的方法，却取得了高于其 1.92%的识别效果。说明了对多视角深度运动图提取 HOG 特征后进行特征融合是非常有效的。

表 3　在实验设置二时与现有方法比较结果(%)

Table 3　Comparison of our method with other existing methods in experiment setting 2(%)

Methods	Accuracy(%)
HOPC[8]	91.64
HON4D[9]	88.89
SNV[10]	93.45
DMM[13]	88.73
DMM_LBP_FF[15]	91.90
DMM_LBP_DF[15]	93.00
Multi_Fused Features[19]	93.30
本文方法	**93.82**

注：加粗字体表示最优值

3.2.2 自建数据库

在自建数据库上仍采用 $D_{f,s,t,\pm25^o,\pm45^o}$ 对一个动作样本进行表示。此数据库采用留一交叉验证的方法进行实验，取得了 97.98%的识别率，实验结果见表 4。文献[11]将动作样本的深度图片序列生成一个 MHPC 对动作进行表示，首先对其进行降采样，接着采用 Harris3D 检测特征点，并结合快速点特征直方图（Fast Point Feature Histogram,FPFH）对特征点进行描述，把特征描述子聚类生成单词包然后对特征点进行描述，最后采用支持向量机实现分类，在自建数据库上取得了 94%的识别率。MHPC 绕 Y 轴旋转 $\pm25^\circ$、$\pm45^\circ$ 后，与原始的 MHPC 共同表示一个动作，记为 Multi_perspective_MHPCs。针对每一个 MHPC 分别采用文献[11]的方法提取特征后进行特征融合，然后利用 SVM 分类，取得了 96.97%的识别率，相比于单一的 MHPC 识别率高出 2.97%，说明通过旋转增加的更多视角下的动作信息是非常必要的，能够有效的提高识别率。但是 MHPC 的方法提取点云特征相对繁琐，而本文算法将 Multi_perspective_MHPCs 投影生成多视角深度运动图，使得三维点云特征提取问题转化为二维图片特征提取的问题，简化了提取特征的运算复杂度的同时，比 Multi_perspective_MHPCs 识别率提高了 1.01%，达到了 97.98%的识别率，验证了将 MHPC 旋转投影到笛卡尔坐标平面生成多视角深度运动图算法的优越性。

表 4　在自建数据库中实验结果与现有方法比较(%)

Table 4　Comparison of our method with other existing methods (%) in our database

Methods	Accuracy(%)
MHPC[11]	94.00
Multi_perspective_MHPCs	96.97
本文方法	**97.98**

注：加粗字体表示最优值

4 结 论

本文提出了一种基于多视角深度运动图的人体行为识别算法，利用 MHPC 对动作进行表示，通过旋转一定的角度，补充不同视角下的动作信息。将原始的和旋转后的 MHPC 投影生成多视角深度运动图，坐标归一化操作能够增强对类内差异的鲁棒性。多视角深度运动图比传统的深度运动图拥有更多的视角，表示动作更加全面。动作的空间能量分布在多视角深度运动图上表现为不同的形状和外形，采用 HOG 特征并进行串联融合取得了良好的识别效果。实验结果验证了利用多视角深度运动图进行行为识别的有效性。在下一步的工作计划中，将尝试对多视角深度运动图提取更加有效的特征，并且在更加复杂的数据库中验证算法的有效性。

参考文献(References)

[1] Weinland D, Boyer E. Action recognition using exemplar-based embedding[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. DBLP, 2011:1-7.

[2] Guo K, Ishwar P, Konrad J. Action Recognition in Video by Sparse Representation on Covariance Manifolds of Silhouette Tunnels[M]// Recognizing Patterns in Signals, Speech, Images and Videos. Springer Berlin Heidelberg, 2010:294-305.

[3] Zhang L, Lu M M, J H. An improved scheme of visual words description and action recognition using local enhanced distribution information[J]. Journal of Electronics & Information Technology, 2016, 38(3):549-556 . [张良, 鲁梦梦, 姜华. 局部分布信息增强的视觉单词描述与动作识别[J]. 电子与信息学报, 2016, 38(3):549-556.]

[4] Xia L, Chen C C, Aggarwal J K. View invariant human action recognition using histograms of 3D joints[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2012:20-27

[5] Ran X Y, Liu K, Li G, Ding W W , Chen B. Human action recognition algorithm based on adaptive skeleton center[J] . Journal of Image and Graphics , 2018 , 23(4) : 0519-0525 .[冉宪宇, 刘凯, 李光,等. 自适应骨骼中心的人体行为识别算法[J]. 中国图象图形学报, 2018 ,23(4):0519-0525. [DOI: 10.11834 /jig.170420]]

[6] Wang C, Wang Y, Yuille A L. Mining 3D Key-Pose-Motifs for Action Recognition[C]// Computer Vision and Pattern Recognition. IEEE, 2016:2639-2647.

[7] Liu M, Liu H, Chen C. 3D Action Recognition Using Multi-scale Energy-based Global Ternary Image[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, PP(99):1-1.

[8] Rahmani H, Mahmood A, Du Q H, et al. HOPC: Histogram of Oriented Principal Components of 3D Pointclouds for Action Recognition[C]// European Conference on Computer Vision. Springer, Cham, 2014:742-757.

[9] Oreifej O, Liu Z. HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:716-723.

[10] Yang X, Tian Y. Super Normal Vector for Human Activity Recognition with Depth Cameras[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(5):1028-1039.

[11] Liu W, Jiang Y, Wang H, et al. Action description using point clouds[C]// International Workshop on Pattern Recognition. 2017:104430X.

[12] Shen X, Zhang H, Gao Z, et al. Human behavior recognition based on axonometric projections and PHOG feature[J]. Journal of Computational Information Systems, 2014, 10(8):3455-3463.

[13] Yang X, Zhang C, Tian Y L. Recognizing actions using depth motion maps-based histograms of oriented gradients[C]// ACM International Conference on Multimedia. ACM, 2012:1057-1060.

[14] Chen C, Liu K, Kehtarnavaz N. Real-time human action recognition based on depth motion maps[J]. Journal of Real-Time Image Processing, 2016, 12(1):155-163.

[15] Chen C, Jafari R, Kehtarnavaz N. Action Recognition from Depth Sequences Using Depth Motion Maps-Based Local Binary Patterns[C]// Applications of Computer Vision. IEEE, 2015:1092-1099.

[16] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005:886-893.

[17] Suykens J A K, Lukas L, Van P, et al. Least Squares Support Vector Machine Classifiers: a Large Scale Algorithm[C]// 1999:839-842.

[18] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3D points[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2010:9-14.

[19] Jalal A, Kim Y H, Kim Y J, et al. Robust Human Activity Recognition from Depth Video Using Spatiotemporal Multi-Fused Features[J]. Pattern Recognition, 2016, 61:295-308.

作者简介：

刘婷婷，1992 年生, 女, 中国民航大学在读硕士研究生，主要研究方向为计算机视觉，机器学习。
E-mail:ttliu_123@163.com

张良，男，教授，主要研究方向为计算机视觉，机器学习，模式识别。
E-mail: l-Zhang@cauc.edu.cn

李玉鹏，男，中国民航大学在读研究生，主要研究方向为机器学习，深度学习。E-mail: yupengli666@126.com