

DOI: 10.13382/j.jemi.B1902011

# 基于加权三视角运动历史图像与时序分割的动作识别算法\*

周宏宇<sup>1</sup> 严春峰<sup>2</sup> 宋旭<sup>1</sup> 刘国英<sup>1</sup>

(1. 安阳师范学院 计算机与信息工程学院 安阳 455000; 2. 桂林电子科技大学 海洋工程学院 桂林 541004)

**摘要:**针对当前人体动作识别算法中由于人体躯干遮挡而导致其检测精度不佳的问题,提出了一种基于加权三视角运动历史图像耦合时序分割的动作识别算法。首先,为了有效描述动作的形状和空间分布,从视频序列中提取运动历史图像(motion history image, MHI)。随后,应用深度相机(Kinect 相机)来提取深度图像,以获取人体目标的动作前景轮廓。为了识别由于身体部位造成的自我遮挡,动作前景轮廓被投影到3个视角(3V)平面,形成3V-MHI,增强了对动作的正确提取,利用3V-MHI构造了一个用于记录观测运动轨迹的MHI,克服了单视角MHI的信息局限性。然后,利用时序分割(temporal segmentation, TS),根据相邻的3V-MHI来计算动作的能量和方向的变化,以检测运动的开始和结束,从而输出运动结果。此外,计算MHI的梯度值作为每个平面对应的权重,从而得到加权3V-MHI。最后,将提取的每个直方图运动模板与预先建立的数据库进行比较,完成动作的分类识别。实验表明,该方法能有效地解决自遮挡问题,在复杂环境和光照变化下有较高的准确性与鲁棒性。

**关键词:**动作识别;运动历史图像;深度图像;时序分割;前景轮廓;自我遮挡

**中图分类号:** TP391.41; TN06 **文献标识码:** A **国家标准学科分类代码:** 520.2040

## Motion recognition based on weighted three-view motion history image coupled time segmentation

Zhou Hongyu<sup>1</sup> Yan Chunfeng<sup>2</sup> Song Xu<sup>1</sup> Liu Guoying<sup>1</sup>

(1. School of Computer and Information Engineering, Anyang Normal University, Anyang 455000, China;

2. College of Ocean Engineering, Guilin University of Electronic Science and Technology, Guilin 541004, China)

**Abstract:** Aiming at the problem of poor recognition accuracy caused by human trunk occlusion in current human motion recognition algorithms, an action recognition algorithm based on weighted three-view motion history image coupling time series segmentation was proposed. Firstly, in order to effectively describe the shape and spatial distribution of the action, motion history image (MHI) is extracted from the video sequence. Subsequently, the Kinect camera was used to extract the depth image to obtain the outline of the human target's action foreground. In order to recognize the self-occlusion caused by body parts, the outline of action foreground was projected to three view angles (3V) planes to form 3V-MHI, which enhances the correct extraction of action. Using 3V-MHI, a MII for recording and observing trajectories was constructed, which overcomes the information limitation of single-view MHI. Then, according to the adjacent 3V-MHI, the energy and direction of motion are calculated by using temporal segmentation to detect the beginning and end of motion and output the result of motion. In addition, the gradient value of MHI was calculated as the weight corresponding to each plane, and the weighted 3V-MHI was obtained. Finally, the extracted histogram motion template was compared with the pre-established database to complete the action classification and recognition. Experiments show that the method can effectively solve the problem of self-occlusion and has high accuracy and robustness for motion recognition under complex background and illumination changes.

**Keywords:** motion recognition; motion history image; depth image; sequence segmentation; foreground contour; self-occlusion

收稿日期: 2019-02-17 Received Date: 2019-02-17

\* 基金项目: 国家自然科学基金(U1804153)、河南省高等学校重点科研项目(18B420001)、河南省高等学校重点科研项目计划(16B520001)、河南省科技计划(132102210212)、广西高校中青年骨干教师基础能力提升项目(2019KY0223)资助

## 0 引言

人的运动和行为是计算机视觉的一个重要研究课题,不仅有重要的科学意义,同时也有非常迫切的实用价值,目前正受到各种应用的驱动,如智能视频监控、人机界面、虚拟现实、重症监护、体育辅助训练等<sup>[1]</sup>。直接利用时空模板特征分析行为,克服了传统方法的某些局限性,如光滑表面和奇点光流的计算,具有自封闭和外观变化的特征跟踪<sup>[2]</sup>。虽然模板匹配方法易于实现,并且需要较少的计算负载,但它们更容易受到噪声的影响,并且易受到运动时间间隔的影响<sup>[3]</sup>。

近年来,动作识别作为一个热点话题,受到了众多学者深入研究,提出了各种不同的方法。左国玉等<sup>[4]</sup>设计了结构优化的决策导向无环图-支持向量机(DDAG-SVM)康复训练动作识别算法。首先,将分类问题变换为二分类模式,再利用 SVM 建立各二分类器。随后利用遗传算子与特征子集区分度规则对各二分类器的核方程参数及特征集优化。通过 SVM 二分类器泛化误差来表示其分离程度,并建立泛化误差上三角矩阵。从根节点开始,分别通过各节点的泛化误差矩阵,选择其中最易被分离的 SVM 分类器形成一种 SODDAG-SVM 多分类器。该方法对于一些简单的动作能有效识别,但是当动作较复杂且相似度较高时,其准确性不太理想,并且该方法易受到外界因素的干扰。当数据较大时,SVM 的计算将耗费过多的存储与时间。Wang 等<sup>[5]</sup>定义了一种全局时空卷积神经网络(convolutional neural network, CNN)体系结构,可以用于动作特征提取。具体地说,全局时空 CNN 是由深度学习、光流和全局累积运动特征分别生成的空间流、局部时间流和全局时间流组成的,其形式为运动叠加差分图像。同时,结合高斯混合模型和 VLAD 的优点,提出了一种新的局部聚集描述符(软 VLAD)向量,根据数据的总体概率分布和聚类中心的相应差异,对数据进行编码,进一步表示提取的特征。但是, CNN 在计算过程易损失局部信息,降低了特征空间分辨率,并且对微小动作识别力较弱。Wang 等<sup>[6]</sup>设计了一种新的用于动作识别的时空表示方法(功率差分模板)。具体为根据对数域和指数域之间梯度的 Gaussian 卷积变换提取空间功率特征。利用前向-后向帧功率差分法,通过对二维水平-垂直投影的直方图进行归一化,提出了归一化投影直方图(normalized projection histogram, NPH)来表征分割后的动作空间特征。此外,从能量守恒的角度,引入运动速度(motion kinetic velocity, MKV)作为一种补充,假设能量的变化是由运动以动能的形式产生来表示功率特征的时间关系。提出的功率差分模板通过融合 NPH 和 MKV 进一步集成到一个词袋模型,用于支持向量机框架下的培

训和测试。通过对 KTH、UCF 运动、UCF101 和 HMDB 数据集的实验,测试了其性能。但是,该方法对于自我遮挡,以及运动持续时间变化等问题无法有效解决,并且在特征提取时易受到光照变化、噪声等影响。

运动历史图像(motion history image, MHI)是一个静态图像模板,其像素强度是运动历史近况的函数,其中较亮的值对应于序列中较新的运动。MHI 可以编码广泛的运动历史,但是, MHI 方法存在一些固有限制,如由于身体部位而存在的自我遮挡,以及运动持续时间变化等问题。如果一个动作(例如坐着然后站着)在其基元动作中有相反的方向,那么先前的动作信息(如坐着)将被后者的动作信息(例如站着)删除或覆盖。因此,如果一个人坐下,然后站起来,最终的 MHI 图像将在图像的上部包含更亮的像素,仅代表站起来运动。由于 MHI 方法不能明确区分运动方向,因此运动对象或人的自遮挡会覆盖先前的信息。

为了提高人体动作的识别精度,本文应用深度相机(Kinect 相机)来提取深度图像,以获取人类目标的前景轮廓。为了识别由于身体部位造成的自我遮挡,前景人体轮廓被投影到 3 个视角(3V)平面,即  $xy$ 、 $yz$  和  $xz$  平面,然后检测运动的开始和结束。利用三视角法(3V-MHI)构造了一个用于记录观测运动轨迹的 MHI,克服了单视点 MHI 的信息局限性。随后,将多分辨率运动历史直方图作为运动特征与数据库中的运动进行比较。本文算法的主要贡献包括利用运动能量和运动方向的时间分段动作来确定 MHI 中的原始运动;将深度轮廓和 2D 图像组合投影到 3V-MHI 上,增强了对动作的正确识别。

## 1 运动历史图像 MHI

由于 MHI 方法能以简单的方式表示运动序列,因此在各种应用中得到了广泛的应用<sup>[7]</sup>。但是,当前大多数方法都考虑简单或单向的动作(如举起手、坐在椅子上、站在椅子上等)来进行识别。然而,人体正常的活动通常是复杂的。这种复杂的动作或活动包含多个原始或基本的动作元素。另外,这种方法是基于人体外观行为来实现的,并且很容易通过执行时间模型来表示和识别动作。执行手臂伸展运动的人体序列的关键帧和相应的 MHI 如图 1 所示。图 1(a)为动作序列,图 1(b)为该动作形成的 MHI,从图 1(b)看出, MHI 能够描述多个时序的动作过程,动作信息显示更详细。

根据文献<sup>[7]</sup>可知, MHI 中值较高的像素表示最近的运动。因此, MHI 描述了运动的形状和空间分布,其灰度值的变化反映了运动方向。根据这种特性,可以使用一个简单的更新规则在一段时间内对图像区域分层来构造 MHI<sup>[8]</sup>:

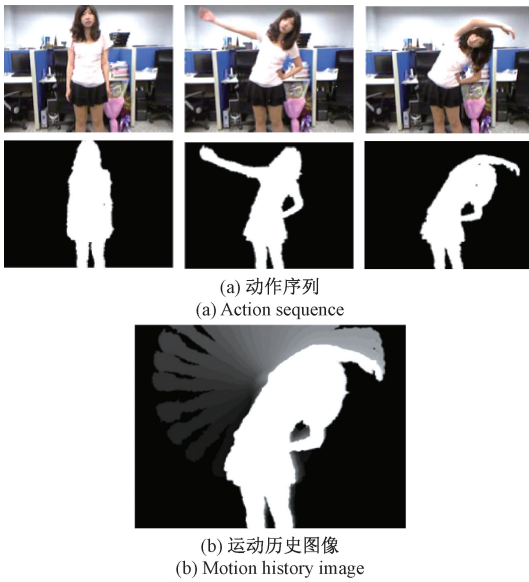


图 1 关键帧和相应的 MHI

Fig.1 Key frame and corresponding MHI

$$H_Y = \begin{cases} Y, & \varphi = 1 \\ \max(0, H_Y(x, y, t - 1) - \delta), & \text{其他} \end{cases} \quad (1)$$

其中,  $Y$  表示 MHI 的持续时间;  $\delta$  为衰减参数;  $x, y$  为像素位置;  $\varphi(x, y, t)$  代表向当前视频图像中的目标(或运动)发送信号,其更新函数为:

$$\varphi(x, y, t) = \begin{cases} 1, & \text{目标存在} \\ 0, & \text{其他} \end{cases} \quad (2)$$

更新函数的计算是进行适当运动分割的关键。定义此更新函数  $\varphi(x, y, t)$  的图像处理技术包括背景减法、帧差分 and 光流等。在所提方法中,使用深度图来定义这个更新函数。为此,采用了 Kinect 相机来提取轮廓。在深度图中,距离信息被用来计算物体像素的深度。假设 MHI 描述了与光轴正交的运动,再根据从 Kinect 传感器获取的深度值,可以在适当的深度范围内过滤出目标。因此,目标存在的更新函数重写为:

$$\varphi(x, y, t) = \begin{cases} 1, & a < Z_Y(x, y) < b \\ 0, & \text{其他} \end{cases} \quad (3)$$

式中:  $Z_Y$  为深度图中时间为  $t(x, y)$  的深度值;  $a, b$  为筛选前景的下限和上限;  $\psi(x, y, t)$  表示原图中的前景像素,而其余像素代表背景信息。

与一般二维彩色图像的背景减法相比,深度图轮廓不受复杂背景和光照变化的影响。通过高斯混合模型(Gaussian mixture model, GMM)与背景减法所提取的前景图像如图 2 所示。可以看出, GMM 很容易受到环境背景和光线的影响。相比之下,使用深度图轮廓过滤背景可以从复杂的背景环境中清晰地提取前景信息。

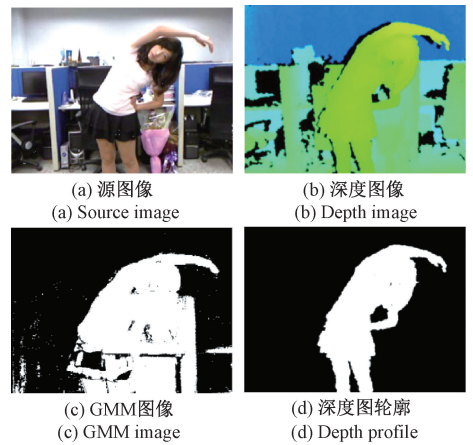


图 2 使用不同方法提取前景图像

Fig.2 Foreground image extraction using different methods

## 2 本文动作识别算法

本文的基于加权三视角运动历史图像耦合时序分割的动作识别算法过程如图 3 所示。首先,从相机中提取 MHI。通过前景提取、深度图像可获取三维前景轮廓。时序分割(temporal segmentation, TS)则是根据相邻的前景对象来计算动作的能量和方向的变化,以确定简单运动的开始和结束时间。然后将深度数据转换为 3V(前、上、右)人体轮廓,通过时间和空间归一化方法,可以输出运动分割结果。最后,将提取的每个直方图运动模板与预先建立的数据库进行比较,完成目标识别。



图 3 本文动作识别算法的过程

Fig.3 The process of action recognition proposed algorithm

### 2.1 虚拟轮廓生成

由于 Kinect 传感器中应用的光编码只检测传感器中最近的物体,实际的三维空间无法重建,因此,本文基于更新函数和深度轮廓,在空间中构建了虚拟三维轮廓



(virtual volumetric silhouette, VVS)<sup>[9]</sup>。与单个二维视图表示法相比, VVS 由于考虑了额外的深度信息, 因此其比单个二维图像信息更丰富。在特定时间的 VSS 对象  $O_i$  可通过以下公式计算:

$$O_i(x, y, z, t) = \begin{cases} 1, & \varphi(x, y, z, t) = 1 \\ 0, & \text{其他} \end{cases} \quad (4)$$

其中,  $\Theta^b = \{O_1^b, \dots, O_n^b\}$  表示包含手势的 VSS 序列 (二进制轮廓)。

## 2.2 时序分割运动历史图像

### 1) 时序分割

时序分割是将连续的运动序列分割为若干个基本片段, 包括理解、分析和聚类<sup>[10]</sup>, 如图 4 所示。运动能量是分割的主要方法之一, 其最小值可以用来检测运动之间的微小差异。因此, 分段检测只需要在 VVS 中找到最小值。当运动具有自遮挡时, 可以将一个动作分割为多个基本动作。首先, 输入三维空间中的前景三维轮廓, 然后通过计算相邻动作, 以获得相邻差分 and 人体静、运动区域。再通过静态区域和运动区域来估计运动方向的变化。最后, 通过计算运动能量和运动方向的变化来判断人体的不同状态, 即开始、移动和结束状态, 从而输出运动结果。利用 VVS 可以有效地计算出全局运动能量的近似值。

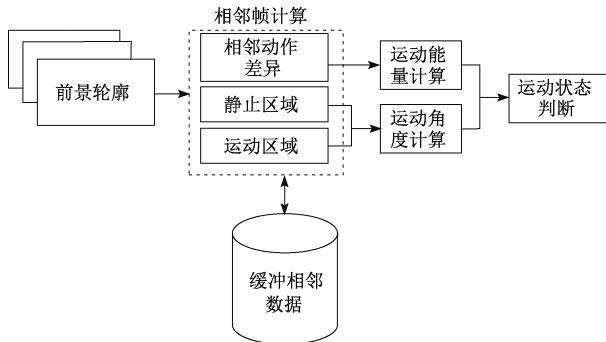


图 4 时序分割主要过程

Fig. 4 The main process of time division

### 2) 运动能量

为了估计运动区域, 计算了  $t$  时刻的两个连续动作之间的差异, 表示如下<sup>[11]</sup>:

$$\sigma_t(x, y, z) = |O_t(x, y, z) - O_{t-1}(x, y, z)| \quad (5)$$

式中:  $O_t(x, y, z)$  是当前帧的前景轮廓;  $O_{t-1}(x, y, z)$  是前一时间 ( $t - 1$ ) 帧的前景轮廓。通过直接减去两个相邻的前景轮廓, 动作差异可以表示运动的区域。在运动求和后, 可以将其视为运动能, 定义如下:

$$\mu_t = \iiint \sigma_t(x, y, z) dx dy dz \quad (6)$$

式中:  $\sigma_t$  表示时间  $t$  与前一时间  $t - 1$  之间前景轮廓, 而  $\mu_t$  代表时间  $t$  时的运动能量。物体的运动范围越大, 相应的

运动能量就越大。相反, 当运动能量较小时, 物体趋于静止。

### 3) 运动方向

在许多视频处理任务中, 大运动信息一直被认为是一个重要的特征。这一运动体现了有关事件序列的信息, 因此, 它是 TS 的一个有用特性。两个动作之间在两个不同时间的主要运动方向可以在一个时间窗口内计算。图 5 所示为运动方向估计。首先, 将每个帧分割为静态区域和运动区域, 如一个人举起双手, 身体处于静止区域, 而手臂的连续运动处于运动区域。分离运动区域后, 找到运动区域的重心, 并记录  $R_{t-w}(x, y, z)$ 。

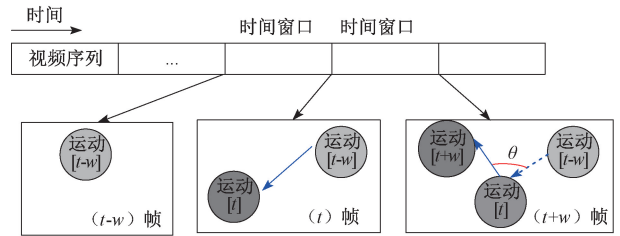


图 5 运动方向计算

Fig. 5 Motion direction calculation

在一个时间窗口  $w$  之后, 同样的计算方法也用于  $t$  时刻运动区域的重心计算, 这样就可以得到运动矢量。这样的运动矢量的方向表示对象在时间窗口的主要运动方向。时间  $t$  到时间 ( $t + w$ ) 的运动矢量可以类似地计算出来。最后, 两个向量之间的角度  $\theta$  表示运动方向的变化。当  $\theta$  大于临界值时, 可以检测到剧烈的运动方向变化, 此时序列可以分割成若干运动单元。

为了更准确计算运动方向, 将每个帧分割为静态区域和运动区域。静态区域是指物体运动时的静止部分, 在图像序列中, 流动序列中的静态区域可以通过相邻两个轮廓的交叉点来获取<sup>[12]</sup>。静态区域定义为:

$$B_t(x, y, z) = (O_t(x, y, z) \cap O_{t-\Delta t}(x, y, z)) \quad (7)$$

式中:  $O_t(x, y, z)$  为是在时间  $t$  时获取的前景对象;  $\Delta t$  为时间步长。

运动区域的估计与静态区域的估计类似, 即计算图像序列中连续的体积对象:

$$D_t(x, y, z) = (O_t(x, y, z) \cup O_{t-\Delta t}(x, y, z)) - B_t(x, y, z) \quad (8)$$

图 6 所示为分离结果, 其中深色区域和浅色区域分别是运动区域和静态区域。从上到下, 每行显示  $xz$ 、 $xy$  和  $yz$  平面上观察到的运动区域和静态区域。

由图 6 可知, 有些深色区域是不可预测的, 这些区域不是主要的运动区域, 而是受人体骨骼运动中关节连接特征的影响。而人体重心的运动区域导致了错误的运动矢量方向。为此, 本文通过加权运动区域赋予每个像素



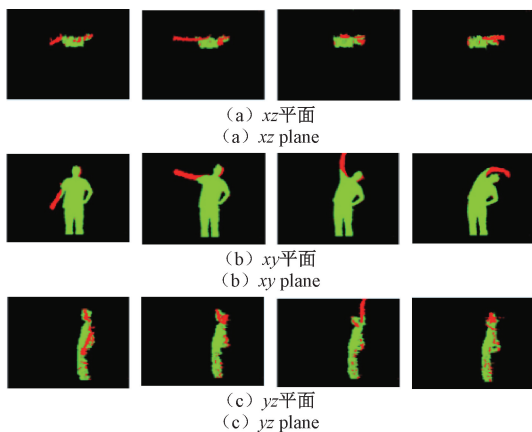


图 6 运动区域和静态区域在 3 个不同投影平面观察结果

Fig. 6 Observation results of moving and static regions on three different projection planes

相应的权重,以校正上运动区域的重心偏移

为了校正重心偏移,将运动区域中的每个像素与静态区域重心之间的距离视为加权运动区域的灰度级别。较高的权重出现在距离静态区域较远的位置;反之,将导出较低的权重。运动区域  $D_t$  到静态区域中心  $m(B_t)$  的距离定义如下:

$$dist(D_t(x,y,z), m(B_t)) = \sqrt{((x,y,z) - m(B_t))^2} \quad (9)$$

加权运动区域  $D_t^w$  定义如下:

$$D_t^w(x,y,z) = \begin{cases} dist(D_t(x,y,z), m(B_t)), & D_t(x,y,z) = 1 \\ 0, & \text{其他} \end{cases} \quad (10)$$

运动矢量  $\delta_t$  在时间  $t$  时描述如下:

$$\delta_t = m(D_t^w) - m(D_{t-1}^w) \quad (11)$$

在图 6 的“举双手”的图像中,手臂是主要的运动区域,其中计算出的主要运动方向是向上的。相反,在下手的例子中,主要运动方向是向下的。在确定主运动方向后,估计当前方向与下一时间步的夹角,得到运动方向的变化。为此,本文通过 Euclidean 空间中的点积来测量角度,点积定义为:

$$a \cdot b = |a| |b| \cos\theta \quad (12)$$

其中,  $|a|$ 、 $|b|$  是向量  $a$ 、 $b$  的长度; $\theta$  为两个向量之间的角度,其计算函数为:

$$\theta = \cos^{-1} \left( \frac{\delta_{t-w} \cdot \delta_t}{\|\delta_{t-w}\| \|\delta_t\|} \right) \quad (13)$$

以右手向上举,身体向左弯曲的动作为例,左手向上挥手时,主运动方向不显示较大的变化,因此运动方向角较小。然而,当手在某个时间向后移动时,即向右侧挥手时,主运动方向会发生剧烈变化。在这种情况下,系统将自动将序列分割成两个基本动作,从而解决了复杂动作中

出现的自遮挡问题。运动能量和运动方向的角度变化被用来检测一个没有遮挡的运动基元。当运动能量大于临界值  $\mu_{th}$  时,物体处于运动状态,否则处于静止状态。此外,当运动方向的角度变化大于临界角度时 ( $\Theta \geq \Theta_{th}$ ),运动角度突然改变,则可认为运动结束。

#### 4) 投影三视角运动历史图像

从深度摄影机获取的深度数据已转换为虚拟三维动作,以描述空间中的对象行为。在这项工作中,三维动作被映射到  $xy$ 、 $yz$  和  $xz$  平面中。这种投影类似于从正面、侧面和顶部观察物体的运动。与光轴上的平面相比,3 个平面可以从不同的角度观察目标。3 个投影矩阵(即  $P_x$ 、 $P_y$ 、 $P_z$ ) 可沿  $x$ 、 $y$  和  $z$  轴将三维动作投影到 3 个平面:

$$P_x = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, P_y = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, P_z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (14)$$

通过这些矩阵对目标轮廓进行投影,得到三维结果:

$$I_t^{3V}(x,y) = \{I_t^{xy}, I_t^{yz}, I_t^{xz}\} = \{P_z(o_t), P_x(o_t), P_y(o_t)\} \quad (15)$$

投影到 3 个平面上的动作的结果如图 7 所示。

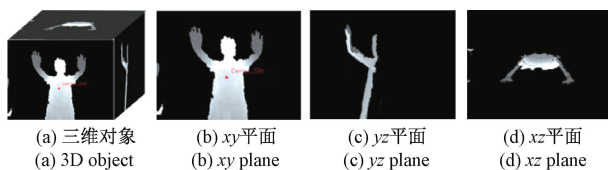


图 7 三维对象投影到 3 个平面上的结果

Fig. 7 The result of projection of three-dimensional object onto three planes

#### 5) 时空归一化

在图像配准中,数据可以是多张照片的形式,或者来自不同的传感器、时间、深度或视角的对象<sup>[13]</sup>。在这项工作中,数据是从同一深度传感器捕获的,所得到的 MHI 具有不同的时序性。因此,本文通过时间和空间归一化计算传感器到目标的不同距离和对象的大小,将 MHI 完成归一化处理:

$$MHI^{3V} = \begin{cases} \frac{H^{3V}}{\tau_{max}} \times 255, & H_r^{3V} \neq 0 \\ 0, & \text{其他} \end{cases} \quad (16)$$

### 2.3 加权运动历史图像

在得到 3V-MHI 后,需要对这些平面进行加权以改善运动特性。当 MHI 用灰度值描述运动轨迹时,具有详细灰度变化的 MHI 平面可显示该图像描述的运动。图像梯度的大小描述图像中灰度变化。随着灰度变化越大,梯度的幅度越大;相反,当运动轨迹信息很少时,梯度的幅度就越小。在观察到 MHI 梯度后,当物体垂直于光轴运动时,  $xy$  平面上的平均梯度会更高。对于平行于光

轴的运动较多的动作,  $yz$  和  $xz$  平面上会出现较高的平均梯度。根据上述特征, 利用 MHI 梯度为这些 3V-MHI 提供权重。选择 Sobel 算子生成图像梯度<sup>[14]</sup>, 表示如下:

$$G_x^{3V} = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \times MHI^{3V} \quad (17)$$

$$G_y^{3V} = \begin{bmatrix} -1 & +2 & +1 \\ -2 & 0 & +2 \\ -1 & -2 & -1 \end{bmatrix} \times MHI^{3V} \quad (18)$$

$$G^{3V} = \sqrt{(G_x^{3V})^2 + (G_y^{3V})^2} \quad (19)$$

其中,  $G_x$ 、 $G_y$  分别是  $x$ 、 $y$  方向的梯度图像的组成部分。3 个方向 (3V) 的 MHI 为  $MHI^{3V}$ , 对应的梯度图像  $G^{3V}$ 。在某个动作中, 如果  $xy$  平面上的 MHI 提供了丰富的轨迹信息, 相应的权重在转换成梯度后更高。相反, 如果动作有更多的深度分量, 那么  $yz$  和  $xz$  平面中会有相应的高梯度值。通过求和每个像素的梯度值来计算权重:

$$W^{3V} = \iint G^{3V} dx dy \quad (20)$$

每个视角对应一个权重,  $W^{3V} = \{W^{xy}, W^{yz}, W^{xz}\}$ , 通过加权, 在比较最终识别阶段的距离时, 每个视角的 MHI 乘以分配的权重。

### 2.4 多分辨率运动历史直方图

利用多分辨率运动历史直方图提取 MHI 的特征, 可以减小基于 MHI 方法的维数, 同时在识别过程中可以加速与数据库的比较。对于一个给定的带有  $b$  个 bin 的直方图  $x$ , 多分辨率结构的非均匀划分过程如下: 首先将  $b$  个 bin 划分为  $m$  个不相交的子集  $B' = \{B'_1, B'_2, \dots, B'_m, \}$ 。然后, 一个带有  $m$  个 bin 的直方图  $X'$  被定义为  $X$  的低分辨率版本。  $X'$  描述如下:

$$X'_k = \sum_{V_i, i \in B'_k} X_i \quad k = 1, \dots, m \quad (21)$$

其中,  $X'$  与  $X$  满足  $d(X, Y) \geq d(X', Y')$ ;  $d$  表示相似度度量函数 L2 范数。通过迭代使用非均匀分割方法, 可以形成若干个多分辨率直方图, 表示如下:

$$d(X, Y) = d^l((X_{b^l}, Y_{b^l})) \geq d^{l-1}((X_{b^{l-1}}, Y_{b^{l-1}})) \geq \dots \geq d^0((X_{b^0}, Y_{b^0})) \quad (22)$$

其中,  $b^i$  表示当前分辨率级别  $i$  个 bin 的大小。本文将多分辨率方法应用于 MHI, 建立了一个具有 5 个图像层次的 MHI 图像金字塔, 每个层次均被 2 倍的因子缩小。图 8 所示为一级运动历史直方图 (motion history histogram, MHH) 的示例。在这个层次上, MHI 被分成 4 个相等的部分, 其中每个像素的灰度值被求和, 然后平均。从左到右和从上到下记录产生的 bin 值。

### 2.5 比较与分类

为了完成动作的识别与分类, 在进行相似性计算的比较时, 需要找到数据库中的相应级别。在所提出的方

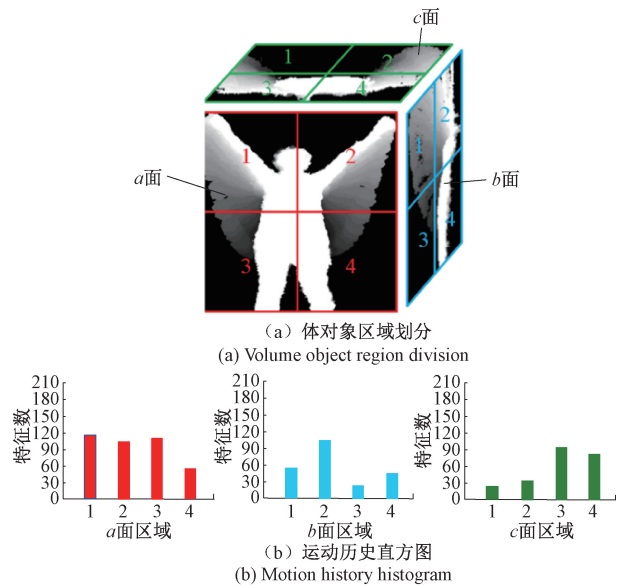


图 8 一级运动历史直方图

Fig. 8 Histogram of first-level movement history

法中, 将 3V-MHH 与相应的水平和视角进行了比较, 并获得了 3V 的相似性。通过与权重相乘, 从式 (24) 获得最终相似性  $d(x, y)$ 。用最近邻法确定运动分类, 最终识别结果如式 (25) 所示。也就是说, 测试动作序列 ( $TM$ ) 可以被归类为能够将该测试序列与所有参考动作模式 ( $RM_i, (i = 1, 2, \dots, n)$ ) 间的相似距离最小化。

$$d(TM, RM_i) = W^{xy} \cdot d^{xy}(TM, RM_i) + W^{yz} \cdot d^{yz}(TM, RM_i) + W^{xz} \cdot d^{xz}(TM, RM_i) \quad (23)$$

$$C = \underset{i}{\operatorname{argmin}} d(TM, RM_i) \quad (24)$$

## 3 实验结果与分析

为了证明本文提出的解决自遮挡和持续时间问题的方法的可行性与优势, 在 KTH、UCFSport 库中完成测试。本文使用的测试环境为 Intel (R) I3 (3.2 GHz, 32 bit)、4 GB 内存和 Windows 7 系统。借助 MATLAB 2012a 对结果分析。此外, 选取 3 组常用的算法进行比较, 分别为文献 [4-6]。所采用数据集与方法如表 1 所示。

表 1 实验数据集与方法

Table 1 Experimental data sets and methods			
数据集	动作	数量	实验方法
KTH	walking, jogging, running, boxing, hand waving 和 hand clapping	6 类别, 599 个动作	2 折交叉法
	basketball shooting, biking, diving, golf swinging, horse riding, soccer juggling, swinging, tennis swinging, trampoline jumping, volleyball spiking, walking with a dog	11 类别, 150 个动作	留一法交叉

### 3.1 KTH 数据集测试

KTH 数据集<sup>[15-17]</sup>包括 6 个动作类型,如表 1 所示,这些动作是通过 25 个表演者在 4 种不同的简单环境中拍摄,如图 9 所示。

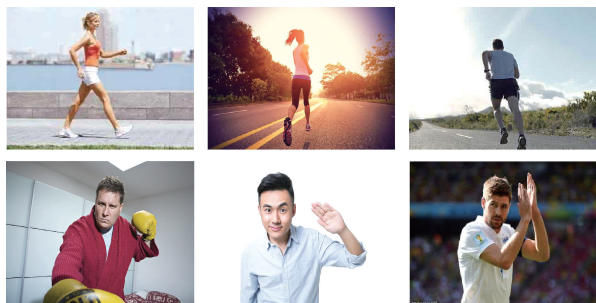


图 9 KTH 集合中的部分样例

Fig. 9 Some examples in KTH set

表 2 为通过所提算法得到的混淆矩阵。从表 2 可得出,所提算法在 KTH 中能够有效识别大部分动作。如“Hand clapping”的识别率高达 100%。“Wave”的识别率为 99%,主要是“Hand clapping”、“Wave”动作与其他 5 种行为较明显。此外,“Walk”识别率相对最低,为 92%,其主要易被误识别为“Run”。

表 2 KTH 数据集的混淆矩阵

Table 2 Confusion matrix of KTH data set

名称	本文算法					
	Walk	Jog	Run	Punch	Wave	HC
Walk	0.92	0.02	0.06			
Jog		0.95	0.03	0.02		
Run		0.02	0.96		0.02	
Punch				0.97	0.03	
Wave				0.01	0.99	
HC						1

### 3.2 UCF Sport 数据集测试

UCF Sport 数据集<sup>[18-20]</sup>含有 11 种类别(表 1),为了便于实验记录,按照表中的顺序,对这些动作进行编号,分别记为 N1、N2、N3、N4、N5、N6、N7、N8、N9、N10、N11。也就是,将“basketball shooting”标记为 N1,后面的动作以此类推。UCF Sport 数据集在视角、尺度、背景变化多样、多数情况下为手动拍摄,受到不同姿势、角度、光照变化的干扰,所以,识别该库中的各种类型具有一定的难度,部分样本如图 10 所示。



图 10 UCF Sport 集合中的部分样例

Fig. 10 Some examples in UCF Sport set

表 3 是在 UCF Sport 集合上进行实验所输出的混淆矩阵。由实验数据发现,本文算法对该集合中的所有动作都有理想的识别准确度。其中,N5、N8 的识别率为 100%;N6、N7 和 N11 的识别率为 97%。另外,将 N6 识别为 N3 的几率只有 0.05;把 N1 识别为 N7 的几率很小,仅有 0.01。

表 3 对表 1 的 11 种动作的识别统计结果

Table 3 Identification statistics results of 11 actions in table 1

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11
N1	0.94										
N2		0.94					0.06			0.03	
N3			0.94	0.03							
N4		0.03		0.95	0.03			0.07			0.02
N5					1.00						
N6			0.05			0.97					0.01
N7						0.05	0.97		0.03		
N8			0.04				0.01	1.00		0.01	
N9		0.06				0.02			0.92		
N10										0.92	
N11		0.01			0.02						0.97



### 3.3 不同算法性能测试

为进一步评估算法性能,将所提算法与对照组算法比较,表 4 与 5 分别给出了在 KTH 与 UCF Sport 中识别率。根据表 4 与 5 看出,对于不同的动作类别,其识别率存在一定差异。对于同类别的识别,本文算法取得了较大的优势。在 KTH 中,最低的正确识别率为 92.04%,最高为 99.77%,在 UCF Sport 中最低识别率为 91.26%,最高识别率为 99.56%。此外,对比表 4 与 5 发现,在 KTH 中的平均识别率高于 UCF Sport。本文算法取得的成绩主要归功于应用 Kinect 相机来提取深度图像,获取人体目标的前景轮廓。并前景人体轮廓被投影到 3V 平面,解决了身体部位造成的自我遮挡。然后,利用 TS,根据相邻的前景对象计算运动能量和方向的变化,对复杂背景和光照变化条件下同样取得较好效果,提高了准确性与鲁棒性。文献[4]易受到外界环境影响。文献[5]由于 CNN 在计算过程易损失局部信息,降低了特征空间分辨率,并且对微小动作识别力较弱。文献[6]采用的特征提取技术对自我遮挡,以及运动持续时间变化等问题无法有效解决,并且在特征提取时易受到光照变化、噪声等影响。

表 4 KTH 数据集识别率

Table 4 KTH data set recognition rate (%)

类别	文献[4]	文献[5]	文献[6]	本文算法
Walk	78.69	85.26	88.66	92.04
Jog	81.36	86.54	89.98	95.31
Run	83.62	88.21	92.81	96.14
Punch	83.98	87.91	91.67	97.33
Wave	84.52	88.66	92.24	98.65
HC	85.63	87.85	93.88	99.77

表 5 UCF Sport 数据集识别率

Table 5 UCF sport data set recognition rate (%)

类别	文献[4]	文献[5]	文献[6]	本文算法
N1	74.32	78.86	84.95	94.22
N2	75.39	81.56	85.62	94.46
N3	78.25	81.20	83.71	94.15
N4	81.56	84.65	89.20	95.56
N5	82.63	84.88	91.20	99.52
N6	81.57	83.64	88.81	96.21
N7	82.36	85.75	91.69	96.14
N8	85.65	88.48	93.29	99.54
N9	81.73	83.46	85.02	91.32
N10	74.30	78.10	83.14	91.26
N11	80.63	82.61	88.68	96.56

为描述算法的性能,绘制了不同算法在两个数据集集中的  $P-R$  曲线,如图 11 所示。图 11(a)为在 KTH 结果,图 11(b)为 UCF Sport 结果。从图 11 看出,在同一数据

集中,本文算法  $P-R$  曲线表现最优异,稳定性好,曲线下方的缓慢。并且在 KTH 的识别效果优于 UCF Sport。

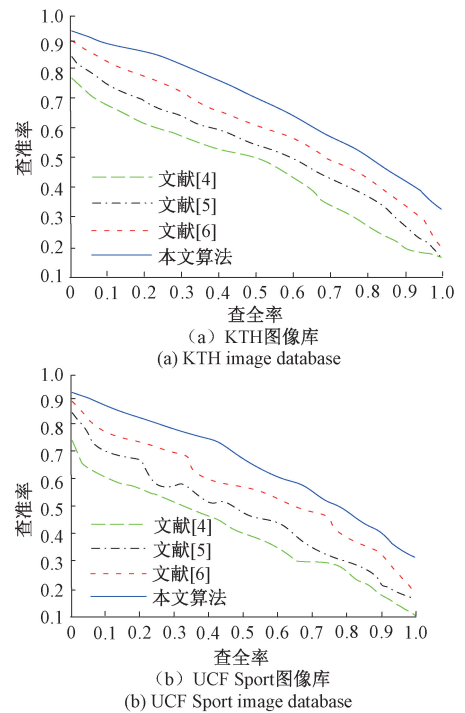


图 11 KTH 与 UCF Sport 数据集中性能测试结果

Fig. 11 Performance test results in KTH and UCF sport datasets

## 4 结论

本文提出了一种加权三视角运动历史图像耦合时序分割技术来解决人体动作识别中的动作自遮挡问题。将运动能量和运动方向角相结合作为时序分割的基础。在计算运动方向角时,提出了加权运动区域。这样就可以避免影响运动区域重心的身体摇摆运动问题。此外,通过深度相机来提取深度图像,获取人体目标的前景轮廓,并将前景人体轮廓被投影到 3 个视角平面,解决身体部位造成的自我遮挡,使得 3 个不同的视角可以观察到目标的行为。本文根据 MHI 的梯度值给出不同的权重,对 3V-MHI 进行加权。因此,该方法解决了一般相机情况下的自遮挡问题,无论运动是平行还是垂直于光轴,都可以获得较高的识别率。实验结果表明,该方法在各种行为下均能获得优异的识别结果。

## 参考文献

- [1] 李贤阳, 阳建中, 杨竣辉, 等. 深度运动图耦合正则化表示的行为识别算法[J]. 电子测量与仪器学报, 2018, 31(1): 119-128.
- LI X Y, YANG J ZH, YANG J H, et al. Behavior recognition algorithm of coupled regularization

- representation of deep motion graph [J]. *Journal of Electronic Measurement and Instrumentation*, 2018, 31(1):119-128.
- [ 2 ] 叶素芬, 赖际舟, 吕品, 等. 穿戴式钢琴弹奏手套智能感知与手势识别技术 [J]. *仪器仪表学报*, 2019, 33(5):190-197.
- YE S F, LAI J ZH, LYU P, et al. Intelligent perception and gesture recognition technology of wearable piano playing gloves [J]. *Chinese Journal of Scientific Instrument*, 2019, 33(5):190-197.
- [ 3 ] 周风余, 尹建芹, 杨阳, 等. 基于时序深度置信网络的在线人体动作识别 [J]. *自动化学报*, 2016, 42(7):1030-1039.
- ZHOU F Y, YIN J Q, YANG Y, et al. Online recognition of human actions based on temporal deep belief neural network [J]. *Acta Automatica Sinica*, 2016, 42(7):1030-1039.
- [ 4 ] 左国玉, 徐兆坤, 卢佳豪, 等. 基于结构优化的 DDAG-SVM 上肢康复训练动作识别方法 [J]. *自动化学报*, 2019, 39(28):1-13.
- ZUO G Y, XU ZH K, LU J H, et al. Action recognition method for upper limb rehabilitation training based on structure-optimized DDAG-SVM [J]. *Acta Automatica Sinica*, 2019, 39(28):1-13.
- [ 5 ] WANG L L, GE L Z, LI R F, et al. Three-stream CNNs for action recognition [J]. *Pattern Recognition Letters*, 2017, 92(6):33-40.
- [ 6 ] WANG L L, LI R F, FANG Y J. Power difference template for action recognition [J]. *Machine Vision and Applications*, 2017, 28(5):463-473.
- [ 7 ] 王满一, 宋亚玲, 李玉, 等. 结合区域光流特征的时序模板行为识别 [J]. *系统仿真学报*, 2015, 27(5):1146-1151.
- WANG M Y, SONG Y L, LI Y, et al. Behavior recognition combining regional optical flow features and temporal templates [J]. *Journal of System Simulation*, 2015, 27(5):1146-1151.
- [ 8 ] TSAI D M, CHIU W Y, LEE M H. Optical flow-motion history image (OF-MHI) for action recognition [J]. *Signal Image and Video Processing*, 2015, 9(8):1897-1906.
- [ 9 ] LIN S J, CHAO M H, LEE C Y, et al. Human action recognition using motion history image based temporal segmentation [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2016, 30(6):1-31.
- [ 10 ] BERTHOLET P, ICHIM A E, ZWICKER M. Temporally consistent motion segmentation from RGB-D video [J]. *Computer Graphics Forum*, 2018, 37(6):118-134.
- [ 11 ] TSAI T H, FANG C L, CHUANG H M. Design and implementation of efficient video stabilization engine using maximum a posteriori estimation and motion energy smoothing approach [J]. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2015, 22(6):817-830.
- [ 12 ] ROH M C, SHIN H K, LEE S W. View-independent human action recognition with volume motion template on single stereo camera [J]. *Pattern Recognition Letters*, 2009, 31(7):639-647.
- [ 13 ] 王文爽, 孙伟, 王帅. 基于改进 ORB 的警用无人机图像配准方法 [J]. *计算机工程与设计*, 2019, 40(1):190-196.
- WANG W SH, SUN W, WANG SH. Police UAV image registration based on improved ORB algorithm [J]. *Computer Engineering and Design*, 2019, 40(1):190-196.
- [ 14 ] 沈德海, 张龙昌, 鄂旭. 一种基于 Sobel 算子梯度增强的边缘检测算法 [J]. *电子设计工程*, 2015, 23(10):162-165.
- SHEN D H, ZHANG L CH, E X. A strengthening gradient edge detection algorithm based on sobel [J]. *Electronic Design Engineering*, 2015, 23(10):162-165.
- [ 15 ] 方云录, 苗茹. 时间规整耦合线性判别分析的动作识别算法 [J]. *计算机工程与设计*, 2018, 39(11):3468-3474, 3499.
- FANG Y L, MIAO R. Motion recognition algorithm based on time warping coupled linear discriminate analysis [J]. *Computer Engineering and Design*, 2018, 39(11):3468-3474, 3499.
- [ 16 ] 张国梁, 贾松敏, 张祥银, 等. 采用自适应变异粒子群优化 SVM 的行为识别 [J]. *光学精密工程*, 2017, 25(6):1669-1678.
- ZHANG G L, JIA S M, ZHANG X Y, et al. Adaptive mutation particle swarm optimization for SVM behavior recognition [J]. *Optical Precision Engineering*, 2017, 25(6):1669-1678.
- [ 17 ] 鹿天然, 于凤芹, 杨慧中, 等. 基于显著性检测和稠密轨迹的人体行为识别 [J]. *计算机工程与应用*, 2017, 54(14):163-167.
- LU T R, YU F Q, YANG H ZH, et al. Human behavior recognition based on saliency detection and dense trajectory [J]. *Computer Engineering and Application*, 2017, 54(14):163-167.
- [ 18 ] 鹿天然. 视频中人体运动行为识别方法的研究 [D]. 无锡:江南大学, 2018:30-38.
- LU T R. Research on human motion behavior recognition

- in video [D]. Wuxi: Jiangnan University, 2018: 30-38.
- [19] 冯小明, 冯乃光, 汪云云. 基运动特征与序列袋的人体动作识别[J]. 计算机工程与设计, 2018, 39(10): 228-235.
- FENG X M, FENG N G, WANG Y Y. Human motion recognition based on motion characteristics and sequence bag [J]. Computer Engineering and Design, 2018, 39(10): 228-235.
- [20] 金锋. 基于视频的人体动作识别研究[D]. 北京: 北京建筑大学, 2019: 31-43.
- JING F. Research on human motion recognition based on video [D]. Beijing: Beijing University of Architecture, 2019: 31-43.

### 作者简介



**周宏宇**, 2004 年于西安理工大学获得学士学位, 2007 年于西安理工大学获得硕士学位, 现为安阳师范学院讲师, 主要研究方向为图像处理、目标检测与模式识别。

E-mail: zhouhongyupear@163.com

**Zhou Hongyu** received B. Sc. degree from Xi'an University of Technology in 2004, M. Sc. degree from Xi'an University of Technology in 2007. Now he is a lecturer at Anyang Normal University. His main research interests include image processing, target detection and pattern recognition.



**严春峰**, 2012 年于桂林电子科技大学获得学士学位。2015 年于桂林电子科技大学获得硕士学位, 现为桂林电子科技大学讲师, 主要研究方向为图像处理、模式识别、信号检测。

E-mail: YanCfeng1988gl@126.com

**Yan Chunfeng** received B. Sc. degree from Guilin

University of Electronic Technology in 2012, and M. Sc. degrees from Guilin University of Electronic Technology in 2015. Now he is a lecturer at Guilin University of Electronic Technology. His main research interests include image processing, pattern recognition and signal detection.



**宋旭**, 2003 年于郑州轻工业学院获得学士学位, 2008 年于北京理工大学获得硕士学位, 现为安阳师范学院讲师, 主要研究方向为图像处理、信号检测、模式识别。

E-mail: SongX1980znn@yeah.net

**Song Xu** received B. Sc. degree from Zhengzhou University of Light Industry in 2003, and M. Sc. degrees from Beijing Institute of Technology in 2008. He is currently a lecturer at Anyang Normal University. His main research interests include image processing, signal detection and pattern recognition.



**刘国英**, 2001 年于长沙理工大学获得学士学位, 2004 年于长沙理工大学获得硕士学位, 2009 年于武汉大学测绘遥感信息工程国家重点实验室获得博士学位, 现为安阳师范学院计算机学院教授, 主要研究方向为计算机图像、模式识别。

E-mail: LiuGuoY1979ay@sohu.com

**Liu Guoying** received B. Sc. degree from Changsha University of Science & Technology in 2001, M. Sc. degree from Changsha University of Science & Technology in 2004, and Ph. D. degrees from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University in 2009. Now he is a professor at Anyang Normal University. His main research interests include computer image and pattern recognition.