基于累积边缘图像的现实人体动作识别

谌先敢 [1, 2] 刘 娟 [1] 高智勇 [2] 刘海华 [2]

摘 要 为了从现实环境下识别出人体动作, 本文研究了从无约束视频中提取特征表征人体动作的问题. 首先, 在无约束的视频上使用形态学梯度操作消除部分背景, 获得人体的轮廓形状; 其次, 提取某一段视频上每一帧形状的边缘特征, 累积到一幅图像中, 称之为累积边缘图像 (Accumulative edge image, AEI); 然后, 在该累积边缘图像上计算基于网格的方向梯度直方图 (Histograms of orientation gradients, HOG), 形成特征向量表征人体的动作, 送入分类器进行分类. YouTube 数据集上的实验结果表明, 本文的方法比其他方法更加有效.

关键词 动作识别, 累积边缘图像, 方向梯度直方图, 支持向量机

引用格式 谌先敢, 刘娟, 高智勇, 刘海华. 基于累积边缘图像的现实人体动作识别. 自动化学报, 2012, **38**(8): 1380−1384

DOI 10.3724/SP.J.1004.2012.01380

Recognizing Realistic Human Actions Using Accumulative Edge Image

CHEN Xian-Gan[1, 2] LIU Juan[1] GAO Zhi-Yong[2]
LIU Hai-Hua[2]

Abstract The problem of extracting feature from unconstrained videos for representing human actions has been investigated in order to recognize human actions in complex environment in this paper. Firstly, morphological gradient was used to eliminate most background information. Then, edge of shape was extracted and accumulated to a frame, which was named accumulative edge image (AEI). Grid-based histograms of orientation gradients (HOG) were calculated and formed a feature vector that captured the characteristic of human actions in this video sequence. Using support vector machine (SVM), the method was tested on the YouTube action dataset. The obtained impressive results showed that this method was more effective than other methods in YouTube action dataset.

Key words Action recognition, accumulative edge image (AEI), histograms of orientation gradients (HOG), support vector machine (SVM)

Citation Chen Xian-Gan, Liu Juan, Gao Zhi-Yong, Liu Hai-Hua. Recognizing realistic human actions using accumulative edge image. *Acta Automatica Sinica*, 2012, **38**(8): 1380−1384

由于在视频监控、视频检索和人机接口等领域的广泛应用, 从视频中识别出人体动作已经引起研究人员的极大兴趣. 因为现实环境是十分复杂的, 所以大部分研究工作针对的是简单环境. 其中用来测试人体动作识别算法的数据集, 如 Weizmann 和 KTH, 其数据的采集都是在简单背景和固定视角下进行. 为了进一步推进动作识别算法在现实场景中的应用, 研究人员从 YouTube、TV 广播和个人视频中收集了包含 11 类不同人体动作的数据集[1], 其中的视频包含了摄像机的运动、背景混乱、人体的外观和尺度变化, 而且部分视频还包含多个人. 因此, 在该数据集上进行人体动作的识别非常具有挑战性. 因为其中大部分视频来自 YouTube, 所以该数据集被称为 YouTube 数据集.

在该数据集上进行人体动作识别, 最大的困难在于如何从这样无约束的视频中提取可靠的特征来表征人体动作. 本文提出一种有效的特征提取方法, 其基本思想是: 人体的形状可以被边缘特征所描述, 视频中部分帧的形状边缘被累积到一幅图像上, 称之为累积边缘图像 (Accumulative edge image, AEI), 可以表征人体的动作, 用来进行动作识别. 本方法的具体步骤如下: 首先, 为了消除各种混乱背景的干扰, 采用形态学操作消除部分背景, 得到相对干净的人体轮廓; 其次, 提取某一时间窗口上每一帧的边缘特征累积到一幅图像中, 将该图像称为累积边缘图像; 接着, 在累积边缘图像上计算基于网格的方向梯度直方图 (Histograms of orientation gradients, HOG), 形成特征向量来表征人体的动作; 最后, 将该特征向量送入分类器进行分类, 识别出人体动作. 本文提出了一种从无约束的视频中提取人体动作特征的方法, 主要的贡献是提出累积边缘图像的概念, 并且在累积边缘图像上提取基于网格的 HOG 表征人体动作, 用于现实环境下的人体动作识别中.

1 相关工作

在动作识别中, 表征人体动作的方法可分为两大类: 一是全局表示法, 二是局部表示法. 前者是首先定位人体, 将感兴趣区域编码为一个整体, 形成图像描述子; 后者首先探测时空兴趣点, 在点的周围计算局部小块, 合并为一个描述子[2]. 本文的方法属于全局表示法.

全局表示法中可以通过背景相减方法获取人体侧影来定位人体. 由于提取方法的不完善, 侧影会包含一些噪声, 并且对视角变化敏感. 有许多方法对这些侧影区域编码. 早期的是运动能量图像 (Motion energy images, MEI) 和运动历史图像 (Motion history images, MHI)[3], Hu 矩被用来表征动作, 后来 MHI 被扩展到 3D 版本[4]. 此外, 视觉无关的动作识别方法[5−6] 和基于 3D 模型的动作识别方法[7] 已经被用来解决视角变化的问题.

局部表示法对噪声和部分遮挡不敏感, 并不严格需要背景相减或跟踪. 然而, 它们依靠足够相关兴趣点的提取, 有时需要预处理. 其中时空兴趣点是视频中运动突然发生变化的位置, 假设这些位置对人体动作的识别具有更多的信息. 早期的兴趣点探测器包括 Harris 角点探测器[8] 和 2D 显著点探测器[9], 分别被扩展到了 3D 空间[10−11]. 这些方法的缺点是稳定兴趣点较少, 已经通过在时空上使用 Gabor 滤波器并改变时空尺度来调整兴趣点数目的方法解决了这个问题[12]. 两人交互行为识别中也用到了兴趣点的提取[13].

全局和局部表示法都需要提取感兴趣区域的特征表征人体的动作. 作为描述特征的图像描述子之一, 方向梯度直方图首先被用于行人检测[14], 后来又发展出光流和表面的方向直方图[15]. 在 HOG 基础上加入多尺度的思想, 得到方向梯度直方图金字塔 (Pyramid histograms of orientation gradients, PHOG), 最先用于多类物体识别[16], 并已经用于人体动作识别中[17].

以上动作识别算法使用的测试数据集都是在简单环境下采集得到, 大部分算法并未扩展到复杂的现实环境中. 本文

收稿日期 2011-01-28 录用日期 2011-09-14
Manuscript received January 28, 2011; accepted September 14, 2011
国家自然科学基金 (60972158) 资助
Supported by National Natural Science Foundation of China (60972158)
本文责任编委 封举富
Recommended by Associate Editor FENG Ju-Fu
1. 武汉大学计算机学院 武汉 430072 2. 中南民族大学生物医学工程学院 武汉 430074
1. School of Computer, Wuhan University, Wuhan 430072 2. College of Biomedical Engineering, South-Central University for Nationalities, Wuhan 430074

所提出的累积边缘图像在外观上与 MHI 类似, 但本文提取的特征不是 Hu 矩, 而是基于网格的 HOG, 而且扩展到了现实环境下的人体动作识别中.

2　方法

本文的特征提取过程全貌如图 1 所示. 首先, 使用形态学梯度操作消除大部分背景, 获得人体的轮廓形状, 其作用是消除背景、减少噪声的干扰; 其次, 提取某一段视频上每一帧形状的边缘, 累积到一幅图像中, 称之为累积边缘图像; 然后, 在该累积边缘图像上计算基于网格的 HOG, 形成特征向量, 该特征向量包含了视频序列中的人体动作信息; 最后, 用该特征向量表征人体的动作, 送入支持向量机 (Support vector machine, SVM) 分类器来进行分类.

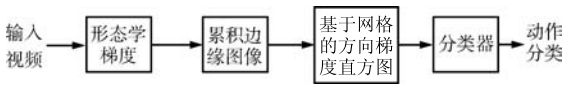

图 1　特征提取过程

Fig. 1　An overview of feature extraction chain

2.1　形态学操作

在视频图像上组合运用形态学操作, 可以消除部分背景, 保持形状特征, 得到人体的侧影轮廓, 其作用类似于背景相减技术. 组合形态学操作的公式如下:

$$G(x,y) = F(x,y) \cdot B(x,y) - F(x,y) \tag{1}$$

其中, $F(x,y)$ 是原始视频中的一帧, $B(x,y)$ 是结构元素, "$\cdot$" 代表闭合操作, $G(x,y)$ 代表经过组合形态学操作处理之后的图像. 其中闭合操作可以将原始图像中比背景暗且比结构元素尺寸小的区域除去, 选取合适的结构元素进行闭合操作可以使图像中仅剩下对背景的估计, 与原始图像相减之后可将目标提取出来. 由于 YouTube 数据集中视频图像的背景过于复杂, 利用该方法并不能将背景完全消除, 但是可以去掉部分背景. 经过在 YouTube 数据集上的反复测试, 结构元素取半径为 9 像素、高为 5 像素的半圆球结构.

图 2 的第 1 行和第 2 行分别显示了来自 YouTube 数据集中的 6 类不同动作的样本帧及其对应的形态学梯度图像. 可以清楚地看到, 通过形态学梯度操作, 原始视频中的图像得到了简化, 侧影轮廓得到了增强, 本文认为这些轮廓形状足够用来进行动作识别.

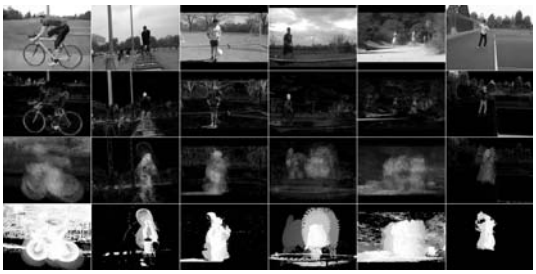

图 2　来自 YouTube 人体动作数据集的样本帧 (第 1 行) 及其对应的形态学梯度图像 (第 2 行)、累积边缘图像 (第 3 行) 和运动历史图像 (第 4 行)

Fig. 2　Sample frames from the YouTube dataset (the first row), the corresponding morphological gradient images (the second row), accumulative edge images (the third row), and motion history images (the fourth row)

2.2　累积边缘图像

包含人体动作的视频一般都有许多帧图像, 大多数时候仅仅一帧图像并不足以代表一个动作, 通常是提取多帧图像的特征来表征人体动作. YouTube 数据集中视频的帧数在 58 至 300 之间, 每个视频的帧数是不一样的, 即使是相同的动作, 也有快有慢, 即同一动作的速率可能是不一样的, 而且在不同情况下采集视频的速率也可能是不一样的. 为了处理这两种速率的变化, 本文将某一时间窗口上每一帧的边缘图像的灰度特征累积到一幅图像中, 构建累积边缘图像, 提取累积边缘图像的特征来表征人体的动作.

计算累积边缘图像的详细算法流程如下. 其中 $G(x,y)$ 表示视频中经过形态学梯度操作处理之后的一帧图像; $E(x,y)$ 表示在 $G(x,y)$ 上使用 Canny 算子得到的边缘图像, 为二值图; $I(x,y)$ 是 $G(x,y)$ 与 $E(x,y)$ 在每个像素点上相乘得到的边缘图像, $I(x,y)$ 是灰度图像, 在边缘点上包含灰度信息, 边缘之外的其他像素点的灰度值为 0; $H(x,y,t)$ 表示累积边缘图像, 尺寸与 $G(x,y)$ 大小相等, 生成 $H(x,y,t)$ 的思路是将视频中某一时间窗口上的全部 $I(x,y)$ 累积到一幅图像上.

步骤 1. 初始化 $H(x,y,t)$, 全部像素置为 0, 此时时间 t 为 0;

步骤 2. 在视频时间窗口的第一帧形态学梯度图像 $G(x,y)$ 上使用 Canny 算子得到边缘图像 $E(x,y)$;

步骤 3. $G(x,y)$ 与 $E(x,y)$ 相乘得到 $I(x,y)$;

步骤 4. $I(x,y)$ 与当前帧以前得到的 $H(x,y,t-1)$ 在每一个像素点上进行比较, 取灰度值大的像素点的灰度值为 $H(x,y,t)$ 的新值;

步骤 5. 返回至步骤 2, 直到视频的最后一帧.

本方法的创新在于提出累积边缘图像的概念, 其基本思想是将视频序列中的信息压缩到一帧图像来表示运动, 这与运动历史图像有一定的相似之处, 但累积边缘图像包含的信息比运动历史图像多. 在点 (x,y) 处, t 时刻的累积边缘图像 $H(x,y,t)$ 可由以下数学公式表示:

$$I(x,y) = G(x,y)E(x,y) \tag{2}$$

$$H(x,y,t) = \max(H(x,y,t-1), I(x,y)) \tag{3}$$

累积边缘图像并不是将每一帧二值图像 $E(x,y)$ 累积到一幅图像中, 而是在二值图像 $E(x,y)$ 与形态学梯度图像 $G(x,y)$ 在每个像素点上相乘, 得到包含灰度信息的边缘图像 $I(x,y)$ 之后, 将视频窗口上的全部边缘图像 $I(x,y)$ 累积到一幅图像中. 二值图像 $E(x,y)$ 像素的灰度值只有 0 和 1 两个值, 而边缘图像 $I(x,y)$ 在其对应的二值图像 $E(x,y)$ 中像素值为 1 的点处具有灰度值, 比二值图像 $E(x,y)$ 具有更多的信息.

与累积边缘图像相似, 运动历史图像也是将视频序列中的信息压缩到一帧图像来表示运动, 在 (x,y) 点处, 时刻 t 的运动历史图像 $H(x,y,t)$ 由下式[3] 得到:

$$H(x,y,t) = \begin{cases} \tau, & D(x,y,t) = 1 \\ \max(0, H(x,y,t-1) - 1), & \text{其他} \end{cases} \tag{4}$$

其中, $D(x,y,t)$ 是由帧差所产生的区域. 比较式 (3) 和式 (4), 可以看出, 累积边缘图像中的有效信息是每一帧边缘图像 $I(x,y)$ 的边缘点上的灰度信息, 而运动历史图像中的有效信息是由帧差图像所表示的运动区域所产生的.

由于人体是运动的, 所以从每一帧中所提取形状的边缘

虽然相似, 但几乎都是不一样的, 通过式 (3) 得到的 $H(x, y, t)$ 中每个像素点的灰度值, 是视频中全部边缘图像 $I(x, y)$ 在该点处的灰度最大值, 最终得到的结果相当于将所有的边缘图像累加到一幅图像中. 图 2 的第 3 行和第 4 行分别显示了累积边缘图像和运动历史图像的示例, 从中可以看出, 在累积边缘图像中人体的运动区域得到增强, 累积边缘图像中的信息量超过视频序列中的单帧图像和对应的运动历史图像.

2.3 基于网格的方向梯度直方图

本文提取的特征是基于网格的 HOG, 该方法的思路与基于网格的 PHOG 比较接近. 不同之处有两点: 1) PHOG 首先提取图像中目标的边缘, 然后在边缘点上计算不同尺度的 HOG, 本文中由于累积边缘图像已经包含了多帧图像的边缘信息, 所以不用再提取边缘, 而是直接在累积边缘图像的每一点上计算 HOG. 2) PHOG 中图像被依次划分成 $2^i \times 2^i$ 个网格, 联合若干尺度的特征作为最终的特征, 而本方法中图像被划分成 $i \times i$ 个网格, 仅取其中某一个尺度的特征作为动作的特征. 基于网格的 HOG 是一个空间形状描述子, 具有目标的边缘统计特性, 通过将图像分成 $i \times i$ 个网格区域得到空间布局, 并在每个区域内计算边缘方向的分布得到局部形状. 具体过程如下: 在累积边缘图像上的每个点上计算方向梯度, HOG 向量被离散化成 K 个方向柱 (Bin), 根据每个轮廓点上的梯度值进行投票. 累积边缘图像被分成 $i \times i$ 个空间网格, 在每个网格上计算 HOG 向量. 这里 i 的取值范围为 [3, 9], 其中计算 HOG 时方向柱数目 K 的取值范围为 [16, 25]. 思路如图 3 所示, 计算基于网格的方向梯度直方图算法的流程如下:

步骤 1. $\boldsymbol{P}$ 置为空;

步骤 2. 将图像划分成 $i \times i$ 个网格区域;

步骤 3. 依次计算其中每个区域的 HOG;

步骤 4. 将每个区域的 HOG 串联成一个特征向量累计添加到 $\boldsymbol{P}$.

其中, $\boldsymbol{P}$ 为最终的基于网格的 HOG 特征, i 表示图像被划分成网格的行列数目, 取值范围为 [3, 9].

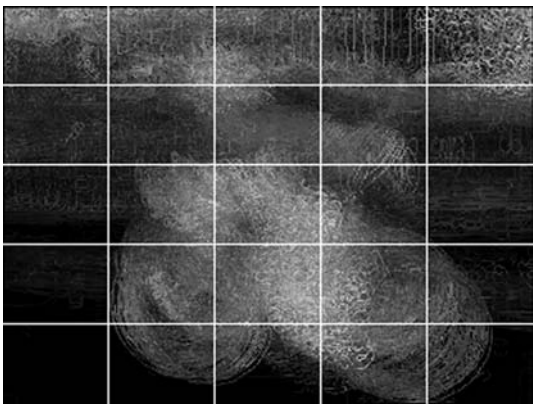

图 3 基于网格的方向梯度直方图

Fig. 3 Grid-based histograms of orientation gradients

2.4 分类器

计算出每个视频的特征之后, 本文用监督学习的方法进行动作的分类. 用支持向量机作为分类器对 YouTube 数据集中的 11 个动作进行分类, 在实验中使用 Osusvm 工具包, 经过在数据集上反复测试, 核函数采取径向基核函数, C 和 Gamma 分别取 9 和 1 000.

3 实验

本文在 YouTube 数据集上测试该方法的有效性, 关于该数据集的细节在下面给出. 本文研究了如下几个方面的问题. 首先, 测试了该方法的最佳配置: 视频序列中时间窗口的帧数、方向梯度直方图中方向柱的数目和图像被划分成网格的数目. 其次, 分析了形态学操作的作用. 接着, 与运动历史图像进行了比较. 最后, 与其他方法进行了比较分析.

3.1 数据集

YouTube 数据集有如下性质: 1) 摄像机稳定或者晃动; 2) 背景混乱; 3) 人体尺度变化; 4) 视角变化; 5) 光照变化; 6) 分辨率低; 7) 视频中包含 1 个或多个人体的动作. 这些性质导致了在该数据集上进行人体动作识别非常具有挑战性. 该数据集包含 11 类人体动作: 骑自行车、跳水、打高尔夫球、接球、跳床、骑马、投篮、扣球、荡秋千、打网球、遛狗. 每类动作被分成 25 个相对独立的组, 每组中的视频是在不同的环境下或者由不同的人拍摄, 包含 4 至 8 个视频. 本文只选择每组中的 01 至 04 号视频作为总样本.

本文实验中所有的评估都是使用 10 倍交叉验证: 数据集被随机划分成 10 份, 其中 9 份用来作为训练数据, 1 份用来作为测试数据. 对训练集和测试集进行 10 次轮换, 取这 10 次结果的平均值作为这次划分的结果. 为了消除随机划分带来的不确定性, 将这种随机划分重复做 10 次, 取这 10 次划分的实验结果的平均值作为最终的识别率. 该方法在相同参数下运行, 每次得到的识别率有少许不同, 但偏差一般不超过 0.3 %.

3.2 本方法的参数选择

本方法的性能在各种不同的配置下被评估. 有三个主要因素影响着性能, 一是视频序列中子窗口的帧数; 二是计算方向梯度直方图时方向柱的数目; 三是图像被划分成网格的数目. 帧数决定着视频中包含人体动作信息的多少, 方向柱的数目决定计算方向梯度直方图时在多少个方向统计方向梯度的分布, 网格的数目决定图像被划分的精细程度.

1) 视频的帧数. 视频序列中每一帧都包含动作的信息, 随着视频帧数的增加, 动作信息会增加, 但多少帧最合适用来进行动作识别, 这是一个值得考虑到问题. 由于 YouTube 数据集中各个视频的帧数不一样, 其中文件名为 v_shooting_24_01 的视频只有一帧, 因此用 v_shooting_24_06 替代该视频, 这样总样本中帧数最少的视频为 53 帧. 为了得到最佳的识别率, 在每个视频中选取总数为 160 至 250 帧视频序列来分别测试动作的识别率. 因此会出现某视频的帧数少于所选帧数的情况, 采取如下方法进行处理: 选取 160 帧时, 如果某视频的最大帧数不足 160, 则选择其视频中的全部帧, 其他情况也是如此.

2) 方向柱的数目. 用 HOG 统计区域内梯度的方向, 方向的范围是 0 度至 360 度, 其中每 $360/K$ 度为一个方向柱, 总共 K 个方向柱, 方向柱的数目决定着在多少个方向统计梯度的分布. 计算方向梯度直方图时, 根据每个轮廓点上的梯度值来进行投票, 方向梯度直方图的峰值代表了该区域内梯度的主方向.

3) 网格的数目. 将图像划分成 $i \times i$ 个空间网格, 在每个网格上计算 HOG, i 的取值范围是 [3, 9], 其取值决定图像被划分的精细程度. 例如, 在图 3 中, 图像被划分成 5×5 大小的网格区域. 然后, 在各个区域上计算 HOG, 将全部区域上的 HOG 特征串联在一起, 形成一个特征向量, 使用该特征向量表征人体动作.

为了得到这三个参数的最佳值，首先将网格数目固定在

某个值, 调整视频帧数和方向柱的数目. 图 4 是将网格数目固定在 5×5 大小时的识别率, 可以看出, 大多数情况下, 随着帧数的增加, 识别率升高, 增加到一个峰值后开始降低. 在不同数目的方向柱下, 峰值所对应的帧数是不同的. 视频帧数为 230, 方向柱数目为 24 时, 识别率最佳.

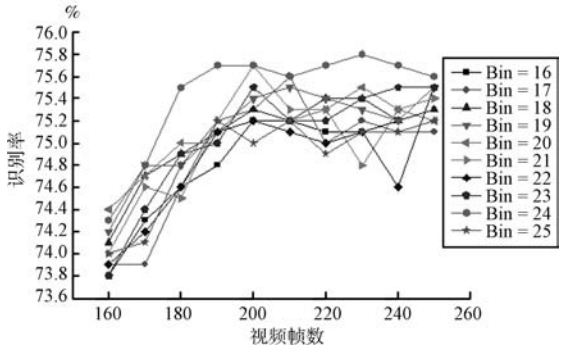

图 4　图像被划分成 5×5 大小时的识别率

Fig. 4　Classified rate while the grid is 5×5

图 5 比较了视频帧数选为 230、方向柱数目为 24 帧时不同网格数目下的识别率, 横坐标表示计算 HOG 时, 图像被划分成 3×3、4×4、5×5、6×6、7×7、8×8 和 9×9 大小的网格. 从图 5 可以看出, 网格数目在 5×5 时识别率达到峰值, 在此之前识别率随着网格数目的增加而增长, 在此之后识别率随着网格数目的增加而减少. 图像被划分成 5×5 大小的网格时, 识别率最佳.

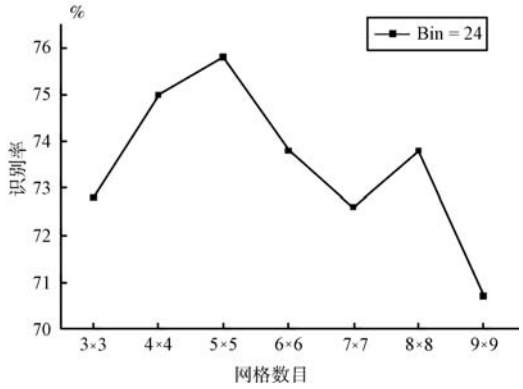

图 5　视频选取 230 帧、方向柱数目为 24 时的识别率

Fig. 5　Classified rate when the frame number is 230 of the video windows and the number of Bin is 24

3.3　形态学操作的作用

为了验证形态学梯度操作的作用, 在原始图像上计算累积边缘图像的特征, 与本方法在形态学梯度图像上计算累积边缘图像的特征进行了比较. 表 1 是在无形态学梯度操作和有形态学梯度操作情况下识别率的比较.

表 1 中的识别率 1 在帧数为 160、方向柱为 16、网格数目为 5×5 的情况下获得; 识别率 2 在帧数为 200、方向柱为 16、网格数目为 5×5 的情况下获得; 识别率 3 在帧数为 200、方向柱为 20、网格数目为 5×5 的情况下获得. 这些数据说明在不同的参数配置下, 在原始图像上计算累积边缘图像的特征表征人体动作, 可以进行人体动作识别. 加上形态学梯度操作之后, 可以去除部分噪声, 进一步提高识别率.

表 1　有形态学梯度操作和无形态学梯度操作的比较 (%)

Table 1　Comparison of the method with morphological gradient and without morphological gradient (%)

方法	识别率 1	识别率 2	识别率 3
无形态学梯度	70.5	70.6	71.6
有形态学梯度	73.8	75.2	75.5

3.4　与运动历史图像的比较

为了比较 AEI 和 MHI 这两种方法的性能, 我们设计了两种实验方案: 将 AEI + HOG 和 MHI + HOG 作为特征进行人体动作识别. 在这两种方案中, 第一种方案提取的是 AEI, 第二种方案提取的是 MHI, 除此之外, 其他条件完全相同. 视频的帧数为 230、方向柱的数目为 24、网格的数目为 5×5, 使用的分类器为 SVM. 在 YouTube 数据集上进行人体动作的识别, AEI + HOG 作为特征的识别率为 75.8 %, MHI + HOG 作为特征的识别率为 64.3 %, 这说明累积边缘图像的性能超过运动历史图像.

3.5　与其他方法的比较

YouTube 数据集是由刘金根收集得到, 他使用的识别方法是 YouTube 数据集上的经典方法, 其思路是从视频中提取静态特征和动态特征, 并对这些特征进行修剪, 得到干净的静态特征和稳定的动态特征, 联合两种特征进行动作识别, 取得了明显的效果. 但该方法依赖足够数目兴趣点的提取, 现实环境下目标区域兴趣点的提取本身就是一个困难的问题, 而且刘金根所使用的兴趣点提取方法非常耗时. 其中仅使用 Harris-Laplacian (HAR) 兴趣点探测器提取一帧图像兴趣点这一步骤的运行时间约为 36 秒, 若视频中有 100 帧图像, 则仅提取兴趣点就需 3 600 秒, 这还不包括计算特征与修剪特征的时间. 而使用本文的方法在 100 帧视频上计算特征值所用的全部时间约为 47 秒 (计算机配置: Pentium (R) Dual-Core CPU E5300 2.60 GHz, 2 GB 内存).

还有一种联合多特征的方法用于 YouTube 数据集上的人体动作识别[18]. 其思路是联合视频中的人体、物体和场景的特征进行人体动作的识别. 首先对视频进行运动补偿, 然后分别提取人体的运动和形状特征、物体的运动和形状特征、场景的形状和颜色特征. 在提取人体的特征之前, 还需使用人体探测器探测出人体[19], 而且提取人体的运动特征是光流场, 对视频中的人体进行探测和计算人体运动的光流场都是非常耗时的步骤. 该方法最终的识别率为 75.2 %. 表 2 将本文的方法与这两种方法进行了比较.

表 2　累积边缘图像和其他方法的比较

Table 2　Comparison of our method with other methods

累积边缘图像	刘的方法[1]	联合多特征的方法[18]
简单、速度快	复杂、速度慢	复杂、速度慢
无需提取兴趣点或人体检测	需要提取兴趣点	需要运动补偿和人体检测
识别率为 75.8 %	识别率为 71.2 %	识别率为 75.2 %

4　结论

本文提出了一种可以在现实场景中进行人体动作识别的方法, 无需对视频图像中的目标进行跟踪或提取兴趣点这样非常耗时的步骤. 本文首先使用形态学操作得到人体的轮廓形状, 然后提取视频中某一时间窗口上每一帧的边缘特征累积到一幅图像中, 在该累积边缘图像上计算基于网格的方向梯度直方图来表征人体的动作. 本文的主要贡献是提出累积边缘图像的概念, 并且使用了基于网格的方向梯度直方图表征人体动作, 用来进行现实环境下的人体动作识别, 比

YouTube 数据集上的其他方法的识别率高, 特征提取的速度也较快, 可应用于实际生活中.

References

1 Liu J G, Luo J B, Shah M. Recognizing realistic actions from videos "in the wild". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 1996−2003

2 Poppe R. A survey on vision-based human action recognition. *Image and Vision Computing*, 2010, **28**(6): 976−990

3 Bobick A F, Davis JW. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(3): 257−267

4 Weinland D, Ronfard R, Boyer E. Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 2006, **104**(2−3): 249−257

5 Huang Fei-Yue, Xu Guang-You. Viewpoint independent action recognition. *Journal of Software*, 2008, **19**(7): 1623−1634
(黄飞跃, 徐光祐. 视角无关的动作识别. 软件学报, 2008, **19**(7): 1623−1634)

6 Yang Yue-Dong, Hao Ai-Min, Chu Qing-Jun, Zhao Qin-Ping, Wang Li-Li. View-invariant action recognition based on action graphs. *Journal of Software*, 2009, **20**(10): 2679−2691
(杨跃东, 郝爱民, 褚庆军, 赵沁平, 王莉莉. 基于动作图的视角无关动作识别. 软件学报, 2009, **20**(10): 2679−2691)

7 Gu Jun-Xia, Ding Xiao-Qing, Wang Sheng-Jin. Human 3D model-based 2D action recognition. *Acta Automatica Sinica*, 2010, **36**(1): 46−53
(谷军霞, 丁晓青, 王生进. 基于人体行为 3D 模型的 2D 行为识别. 自动化学报, 2010, **36**(1): 46−53)

8 Harris C, Stephens M. A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference. Manchester, UK: Organising Committee AVC, 1988. 147−151

9 Kadir T, Brady M. Scale saliency: a novel approach to salient feature and scale selection. In: Proceedings of the International Conference on Visual Information Engineering. Guildford, UK: IEEE, 2003. 25−28

10 Laptev I, Lindeberg T. Space-time interest points. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 432−439

11 Oikonomopoulos A, Patras I, Pantic M. Spatiotemporal salient points for visual recognition of human actions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2006, **36**(3): 710−719

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

13 Han Lei, Li Jun-Feng, Jia Yun-De. Human interaction recognition using spatio-temporal words. *Chinese Journal of Computers*, 2010, **33**(4): 776−784
(韩磊, 李君峰, 贾云得. 基于时空单词的两人交互行为识别方法. 计算机学报, 2010, **33**(4): 776−784)

14 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 886−893

15 Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance. In: Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 428−441

16 Bosch A, Zisserman A, Munoz X. Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval. Amsterdam, Netherlands: ACM, 2007. 401−408

17 Han L, Wu X X, Liang W, Hou G M, Jia Y D. Discriminative human action recognition in the learned hierarchical manifold space. *Image and Vision Computing*, 2010, **28**(5): 836−849

18 Ikizler-Cinbis N, Sclaroff S. Object, scene and actions: combining multiple features for human action recognition. In: Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece: Springer, 2010. 494−507

19 Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. In: Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1−8

谌先敢 武汉大学计算机学院博士研究生, 中南民族大学生物医学工程学院讲师. 2002 年获武汉大学计算机学院学士学位. 主要研究方向为计算机视觉, 视觉神经计算. E-mail: chenxg@mail.scuec.edu.cn
(**CHEN Xian-Gan** Ph. D. candidate at the School of Computer, Wuhan University and lecturer in the College of Biomedical Engineering, South-Central University for Nationalities. He received his bachelor degree from Wuhan University in 2002. His research interest covers computer vision and visual neural computing.)

刘 娟 武汉大学计算机学院教授. 主要研究方向为生物信息学, 计算机视觉, 自然语言处理. 本文通信作者. E-mail: liujuan@whu.edu.cn
(**LIU Juan** Professor at the School of Computer, Wuhan University. Her research interest covers bioinformatics, computer vision, and natural language processing. Corresponding author of this paper.)

高智勇 中南民族大学生物医学工程学院副教授. 主要研究方向为计算机视觉, 视觉神经计算. E-mail: zhiyonggao@mail.scuec.edu.cn
(**GAO Zhi-Yong** Associate professor at the College of Biomedical Engineering, South-Central University for Nationalities. His research interest covers computer vision and visual neural computing.)

刘海华 中南民族大学生物医学工程学院教授. 主要研究方向为计算机视觉, 视觉神经计算. E-mail: lhh@mail.scuec.edu.cn
(**LIU Hai-Hua** Professor at the College of Biomedical Engineering, South-Central University for Nationalities. His research interest covers computer vision and visual neural computing.)