

基于 SIFT 和 HOG 特征融合的人体行为识别方法

韩欣欣,叶奇玲

(南京邮电大学 通信工程学院,江苏 南京 210003)

摘要:行为识别是视频分析的一个核心任务,而行为特征的提取与选择直接影响识别效果。针对单一特征往往受到人体外观、环境、摄像机设置等因素的影响而识别效果不佳的问题,提出一种分别提取尺度不变特征变换(SIFT)和方向梯度直方图(HOG)的特征并形成融合特征,再利用支持向量机(SVM)完成特征分类的行为识别方法。基于 Matlab 人体行为识别和检测的研究,通过采用 KTH 和 Weizmann 人体行为库来验证该算法的有效性。实验结果表明,该算法在人体行为识别中识别率可达到 90% 以上,比单独使用上述两种特征或者其他传统的描述子更高效,同时也能更好地适应光照等外部因素的变化,得到更好的识别率。

关键词:行为识别;特征融合;尺度不变特征变换;方向梯度直方图;支持向量机

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2019)06-0071-03

doi:10.3969/j.issn.1673-629X.2019.06.015

Human Action Recognition Based on Feature Fusion of SIFT and HOG

HAN Xin-xin, YE Qi-ling

(School of Communication Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Action recognition is a core task of video analysis, and the extraction and selection of behavior features directly affect the recognition effect. In view of the problem of bad recognition effect for single feature caused by human appearance, environment, camera settings and other factors, we propose a action recognition method which respectively extracts scale-invariant feature transform (SIFT) and direction gradient histogram (HOG) and fuse them, and then adopts support vector machine (SVM) for feature classification. Based on research on Matlab human action recognition and detection, the validity of the algorithm is verified by KTH and Weizmann human behavior libraries. Experiment shows that the recognition rate of the algorithm in human action recognition can reach more than 90%, which is more efficient than above two features alone or other traditional descriptors, and can better adapt to the change of external factors such as illumination and get better recognition rate.

Key words: action recognition; features fusion; SIFT; HOG; support vector machine

0 引言

在处理视频数据时,行为识别已经成为分析视频的一个关键技术,在机器学习领域和公共安全领域具有一定的研究意义。其中,在公安侦查方面,人体行为识别也应用广泛,如危险行为监控、人机交互、手势识别、智能监控和视频内容检索等^[1-3]。近年来,一些科研工作者在行为识别方面提出了运算速度较快的特征提取方法并加以改进^[4-6],如 LBP 局部二值模式^[7]、HOG 方向梯度直方图^[8]以及 3D-SIFT 三维的尺度不变特征变换^[9-10]等方法。鉴于通过单一特征描述一种动作行为存在的局限性,提出利用多种特征融合来进

行行为识别。

在图像特征提取时,处理效率可能会受到一些外界影响,像摄像机移动、物体遮挡、视角和光照明暗等,可能会导致行为识别出现误差,因此提出了一种融合基于 SIFT 特征与 HOG 特征,然后通过 SVM 完成分类识别的算法^[11]。

1 算法原理

该算法的步骤为:目标检测、单特征提取、多特征融合、SVM 分类识别。采取的人体行为数据库为小段视频,首先是预处理视频中的视频帧,从而获取图像中

收稿日期:2018-07-06

修回日期:2018-11-14

网络出版时间:2019-03-06

基金项目:江苏省自然科学基金-青年基金项目(BK20140868)

作者简介:韩欣欣(1990-),女,硕士研究生,研究方向为计算机视觉理论和系统、图像处理。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190306.0901.020.html>

的有效信息,之后将所得到的结果作为输入来分别提取 SIFT 和 HOG 的特征,然后再进一步进行多特征融合,最终通过 SVM 进行行为分类识别。算法流程如图 1 所示。

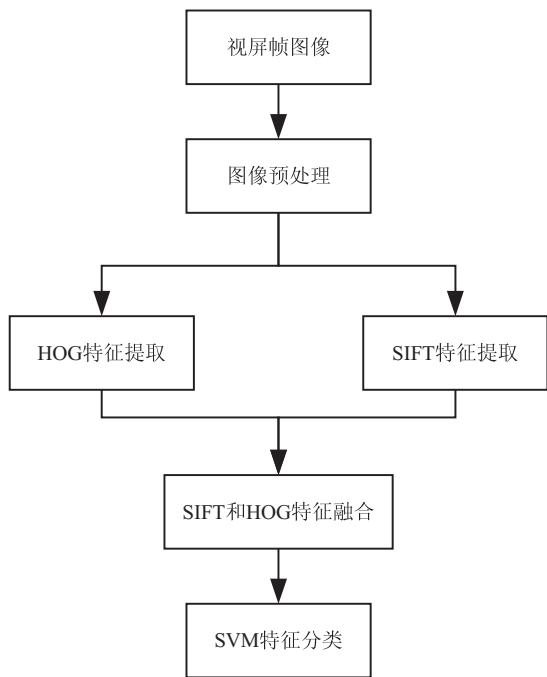


图 1 算法流程

1.1 尺度不变特征变换

在真实场景下拍摄的行为视频,一些外界因素可能会对视频中的行为识别产生影响,如不可避免的拍摄人手抖、建筑物体遮挡、背景物体变化、镜头伸缩以及光照等外在因素。SIFT 算法是可以描述图像中局部特征的算法,可以帮助识别辨认物体,此外该算法还能在空间尺度中寻找极值点数据并提取位置、关键点方向等信息。该特征描述子不会受到拍摄视角、尺度变化、亮度明暗、目标遮挡或噪声等因素的影响,因此使用 SIFT 特征可以在一定程度上解决上述问题^[12-14]。SIFT 算法在提取动作特征中主要采用以下步骤:

(1) 检测尺度空间极值点。

先建立图像金字塔,金字塔分为 O 组,每组有 S 层,其中第 $O(O \geq 2)$ 组的图像是由第 $O-1$ 组的图像经过 $1/2$ 降采样得到的,每组内的图像是由下到上进行高斯滤波得到的,在获取得到图像金字塔后,通过建立高斯差分尺度来检测稳定的特征点:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1)$$

其中, $G(x, y, k\sigma)$ 为高斯滤波器; $I(x, y)$ 为图像中点 (x, y) 的灰度值。

通过每一个采样点与其所有的邻近点(8 邻域点和上下相邻尺度对应的 18 个点)进行对比来寻找尺度

空间图像的极值点,当采样点的数据大于或者小于所有相邻点数据时,则该点为极值点数据。

(2) 精确定位极值点位置。

对于上述步骤中得到的极值点,它可能会受到一些噪音的影响,以及可能会有较强的边缘响应,关键点的尺度和位置可以使用拟合三维二次函数方法,并通过 Hessian 矩阵方法来消除边缘响应,是否为所需要的特征点由式 2 表示:

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(r+1)^2}{r} \quad (2)$$

其中, \mathbf{H} 表示点 (x, y) 处的 Hessian 矩阵。

$$\mathbf{H} = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \quad (3)$$

实验中提取了样本图像的 SIFT 特征。首先对图像进行归一化处理得到 $64 * 128$ 的图像,在高斯函数算法中,可利用其构造尺度空间,然后检测尺度空间中的关键点。文中选用的检测窗口为 $16 * 16$,在每个窗口能检测到 $16(4 * 4)$ 个关键点,然后在确定关键点的位置、尺度和方向后,通过高斯窗口进行加权计算得到具有方向的 8 个关键点。一个窗口有 128 维特征描述子,一幅 $64 * 128$ 大小的图像就会有 4 096 维 SIFT 特征。

1.2 方向梯度直方图

HOG 特征是可以用来检测物体的特征描述子,常用于计算机视觉及图像处理方面,在行人检测中也应用广泛。测试样本图像中的 HOG 特征,首先将图像分成小的单元格 (cell),然后将每个像素梯度和方向的直方图边缘的单元格组合在一起,形成测试样本图像的 HOG 特征。

图像中像素点 (x, y) 的水平方向梯度、垂直方向梯度分别为:

$$D_x(x, y) = H(x+1, y) - H(x-1, y) \quad (4)$$

$$D_y(x, y) = H(x, y+1) - H(x, y-1) \quad (5)$$

其中, $H(x, y)$ 为输入图像中像素点 (x, y) 处的像素值。

像素点 (x, y) 处的梯度幅值和梯度方向分别为:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (6)$$

$$\alpha(x, y) = \tan^{-1}(G_y(x, y)/G_x(x, y)) \quad (7)$$

实验中对样本图像进行 HOG 特征提取,与 SIFT 特征提取相同,第一步先对图像进行归一化处理,然后将图像分割为 $16 * 16$ 像素的单元格。在梯度方向上被平均划分成 9 个区间,然后统计分析单元格中每个区间像素的梯度方向直方图,从而得到一个 9 维特征向量。通过每 $2 * 2$ 个单元格组成一个块扫描样本图像,扫描步长为一个单元,最后,将所有块特性连接起来,形成完整的 HOG 特性。

1.3 分类器设计

SVM^[15]是一种有监督的学习模型,在模式识别、分类和回归的研究中有重要作用。SVM 的关键在于核函数,如线性核函数、多项式核函数和高斯核函数等都比较常用。SVM 方法是使用非线性映射算法,使非线性可分问题变成线性可分问题,其本质就是将输入变量全部映射到一个高维空间,构造出最优分类超平面来进行分类识别^[16]。其中分类函数表示为:

$$f(x) = \omega^T x + b \tag{8}$$

其中, ω 与空间维数有关,其样本点 (x_i, y_i) 到分类超平面的间距为 $y_i(\omega^T x_i + b)$,对 ω 和 b 进行归一化,用 $\omega/\|\omega\|$ 和 $b/\|b\|$ 分别替换 ω 和 b ,即几何间隔为 $f(x)/\|\omega\|$,则最大几何间隔为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s. t.} \quad & (\omega^T x_i + b) \geq 1, i = 1, 2, \dots, n \end{aligned} \tag{9}$$

式 9 为线性分类情况,如果是非线性分类,则按照式 9 进行分类则可能会降低准确率,因而加入松弛变量和惩罚因子,如下所示:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n \xi_i \\ \text{s. t.} \quad & [y_i(\omega^T x_i + b)] \geq 1 - \xi_i, \xi_i \geq 0, \\ & i = 1, 2, \dots, n \\ & \omega = x_1 y_1 + \alpha_2 x_2 y_2 + \dots + \alpha_n x_n y_n \end{aligned} \tag{10}$$

其中, ξ_i 称为松弛变量(slack variable), c 为一个参数,用于控制目标函数中两项之间的权重。

2 算法步骤

文中算法的步骤主要分为:

- (1)选取一段视频并对其提取视频帧图像,将图像转化为灰度图像,然后作归一化处理;
- (2)基于第一步分别对图像进行 SIFT 和 HOG 特征提取,从而得到两个特征集合,然后通过串行组合的方式来构造 SIFT 和 HOG 融合特征集;
- (3)通过 SVM 分类器对融合特征向量进行分类识别。

3 实验结果及分析

实验是通过 Matlab 仿真完成的。选用 KTH 和 Weizmann 人体行为库来验证该算法的有效性。在 KTH 数据库中,选出 6 个动作类:慢跑、鼓掌、画圈、拳击、散步和跑步;在 Weizmann 数据库中,也选出 6 个动作类:跳起鼓掌、跳跃、跑步、单腿跳、散步和快速双腿跳。随机在数据库的每类动作中选出 60 张图像数据,选出其中 50 张作为训练集,剩下的 10 张图像作为测试集。

将文中算法与单个特征算法进行比较,结果如表 1 所示。

表 1 各特征在 KTH、Weizmann 数据集的识别率 %

算法	KTH	Weizmann
HOG+SVM	80(48/60)	75(45/60)
SIFT+SVM	83.3(50/60)	80(48/60)
文中算法	91.7(55/60)	90(54/60)

从表 1 中可以看出,在动作的正确识别率方面,文中的融合特征算法要明显高于单特征算法,有效提高了分类器的识别率。这在基于 KTH 和 Weizmann 数据库的实验中也得到了证明,其识别率可达到 91.7%。

4 结束语

对行为识别中常用到的两种特征提取算法进行分析,利用 SIFT 和 HOG 以不同的侧重描述样本,其中 SIFT 主要用于检测尺度空间的极值点,帮助辨识物体,获取极值点位置方向等信息;HOG 主要是通过计算图像的梯度方向直方图来获取特征信息。从特征描述上来分析,SIFT 相较 HOG 复杂,因此 SIFT 的特征分类效果较好。

实验结果表明,融合特征比起两种单特征的识别效果都要好。

参考文献:

[1] 凌志刚,赵春晖,梁彦,等. 基于视觉的人行为理解综述[J]. 计算机应用研究,2008,25(9):2570-2578.

[2] 周霞,柳絮青,王宪,等. 基于特征融合的人体行为识别算法[J]. 计算机工程与应用,2013,49(7):162-166.

[3] 梁琛华,常青,胡谋法. 基于加权 Hu 矩和 HOG 特征的自适应融合人体行为识别新方法[J]. 现代电子技术,2014,37(1):14-18.

[4] 崔广才,窦风平,王春才,等. 基于傅里叶与局部特征结合的人体姿态识别方法研究[J]. 长春理工大学学报:自然科学版,2016,39(1):82-87.

[5] ZHANG Jiatao,TSOI A C,LO S L. Scale invariant feature transform flow trajectory approach with applications to human action recognition[C]//International joint conference on neural networks. Beijing,China:IEEE,2014:1197-1204.

[6] 李新德,刘苗苗,徐叶帆,等. 一种基于 2D 和 3D SIFT 特征级融合的一般物体识别算法[J]. 电子学报,2015,43(11):2277-2283.

[7] OJALA T, PIETIKAINEN M, HARWOODD. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions[C]//Proceedings of 12th international conference on pattern recognition. Jerusa-