

## 摘要

人体检测逐渐成为近年来计算机视觉和模式识别领域中的一个研究热点。其之所以备受关注，主要出于以下两个原因：1) 虽然人体检测属于一般对象识别的一个特例，但是由于其问题本身具有一般性，从而使得对于该问题的有效解决能够对其他的对象识别问题提供借鉴；2) 应用上的需求进一步推动了人体检测方法的发展，例如在车载辅助预警系统、智能监控系统、人机交互系统以及基于内容的视频/图像管理系统中的应用。

人体检测中的难点问题可以归结为两个方面，即低信噪比和弱配准。低信噪比是指人体数据中噪声所占比例较大而能够被用于对人体进行判别的信息相对较少。弱配准是指因人体形态上的差异而很难将人体的各个部分进行比较好的对齐。这两个难点问题综合作用结果就是人体数据具有非常大的类内散度。针对于这两个难点问题，本文分别采用了时空分析和多粒度特征表示的技术路线，并且在此技术路线的指导下，从数据预处理和特征提取两个方面提出了一系列人体检测模型和方法。

论文的创新与主要贡献总结如下：

(1) 本文提出了一种轮廓运动特征 (Contour-Motion Feature, CMF) 进行鲁棒的人体描述。该特征使用时-空轮廓作为人体的底层描述，然后利用3维的距离变换来将1维的轮廓信息扩展到3维的空间中。通过这种方式，局部轮廓之间的关系可以被隐式的进行表达。通过使用3维的Haar特征对于静态和动态的特征进行统一的封装，可以进一步得到人体的中层的表达：运动轮廓特征。最后利用Boosting的方法来选择具有最强判别能力的特征用于分类。实验结果表明，该方法可以比其他同类方法具有更好的检测性能和可扩展性。此外，尽管该方法是在人体检测的背景下提出的，该方法还进一步被用于行为分析中，并且取得了较好的结果。

(2) 本文提出了一种多粒度特征表示方法，称为粒度可变的方向划分描述子 (Granularity-tunable Gradients Partition, GGP)。针对人体数据难以进行配准的问题，本文提出了人体的多粒度特征表示方法。这里粒度这个概念表示特征对于数据的抽象能力：精细粒度特征对于数据有较低程度的抽象，具有比较好的细节描述能力，适合用于对数据进行确定性的描述；而粗糙粒度的特征对于数据有较高度度的抽象，其所体现的通常是一种统计特性。因此，多粒度特征描述意味着可

以对人体数据进行不同层次的抽象,从而得到从确定性描述到统计性描述的一系列的具有不同描述特性人体表示。

本文在霍夫空间中对于直线的定义进行了扩展,将直线对于旋转和平移的不确定性显式的体现在直线的定义当中,并称这类直线为广义直线,其旋转和平移的不确定性为粒度参数。进一步该广义直线被作为基元,对人体数据进行解析。通过调整粒度参数,描述子可以在不同描述特性之间切换。在精细粒度的一端,GGP可以变为一种确定性的描述,如Edgelet;而在粗糙粒度的一端,GGP可以变为一种具有统计特性的描述子,如梯度方向直方图(Histograms of Oriented Gradients, HOG)。同时,梯度的位置、方向、强度和分布信息也被编码到描述子的特征向量当中,这样可以进一步增强特征的表述能力。在INRIA的人体数据库上的评测结果表明,该方法可以达到与当前领先的方法相当的检测水平,但是因为该方法中的特征和弱分类器都是线性的,所以在速度和计算复杂度上较其他方法更有优势。

(3) 本文提出了一种多粒度特征表示与时空分析相结合的人体检测方法,称为时-空域粒度可变的方向划分描述子(Spatial-Temporal Granularity-tunable Gradients Partition, STGGP)。这种描述子融合了时空分析与多粒度特征表示的优势,因而具有更强的描述能力。根据时间信息与空间信息的相关性不同,提出了3种STGGP描述子的具体实现。

在第一种实现中,不考虑运动信息与外观信息之间的相关性,将运动信息用光流梯度场表示,只将其与外观信息进行简单的串接,这种描述子称为基于光流梯度场的STGGP描述子,用STGGP\_of表示;在第二种实现中,用时空体上相互正交的3个切平面来表示三个坐标轴两两之间的相关性,并在这三个平面上分别提取多粒度特征,这种描述子称为基于时-空切平面的STGGP描述子,用STGGP\_op表示;在第三种实现中,充分考虑人体运动过程中的时-空相关性,将人体及其运动看作是3维空间中的一个实体,并且在3维霍夫空间中定义广义平面对其进行解析,称这种描述子为基于3维霍夫变换的STGGP描述子,用STGGP\_3h表示。最后,将STGGP描述子用于人体检测和行为识别,实验结果表明,STGGP描述子较其他算法具有较明显的优势。

(4) 为了进一步提高人体检测的速度和精度,本文将背景建模和运动检测也作为一个研究内容,并将其作为人体检测的预处理过程,以达到缩小检测区域、减少误检的目的。提出了一种基于非参数模型的背景建模方法用于运动目标检测,以此来降低人体检测的误检率及提高检测速度。首先,引入了一个新的

模型, 影响因素描述模型 (Effect Components Description, ECD), 来对背景的变化进行建模。通过这个模型, 可以将背景模型最好的估计与其分布的众数相关联。在ECD的基础上, 进一步提出了一个有效的背景生成方法: 可靠背景模型 (Most Reliable Background Model, MRBM)。在MRBM生成的过程中, 运用mean shift来迭代找到每个像素的分布的众数。该方法的优势主要体现在三个方面: 首先, 非参数模型可以较好的处理多峰分布数据, 该方法不需要纯净的背景作为训练, 可以用于有复杂运动物体情况下的背景模型的生成; 其次, 生成的背景图像质量较高, 可以减轻图像中由于压缩导致的块效应且不致引入模糊; 最后, 背景对于短时光照变化、噪声和摄像设备的小的抖动具有鲁棒性。

**关键词:** 人体检测; 行为识别; 轮廓运动特征; 时空分析; 多粒度特征表示; 粒度可变的的方向划分描述子; 时-空域粒度可变的的方向划分描述子; 可靠背景模型

## Abstract

Human detection is one of the most challenging and active research topics in computer vision and pattern recognition. This topic is attractive for the following two main reasons: 1) Even though human detection is a special case of the general object recognition, the problems it faced are generic and can provide valuable reference for the other topics. 2) The increasing demands in the practical applications, such as smart surveillance system, on-board driving assistance system and content based image/video management system.

The main challenges for human detection come from two aspects: low SNR (Signal to Noise Ratio) and weak alignment. Low SNR means that comparing with the noise, the discriminative information is very limited in the human data. Weak alignment means that the shape of human can not be easily normalized due to the articulation and pose variation. These two challenges lead to a huge inner-class variation of human data. From methodology point of view, we use spatial-temporal analysis and multi-granularity representation to deal with these two challenges. Based on these methodologies, a series of models and feature extraction methods have been developed.

The main contributions of the paper can be summarized as follows:

(1) A contour-motion feature for robust pedestrian detection is proposed. The space-time contours are used as the low level representation of the pedestrian. Then we apply 3D distance transform to extend the 1-dimensional contour into 3-dimensional space. By this way, the relations between the local contours can be maintained implicitly. Further, by encapsulating the static and dynamic information by 3D Haar filters, we can generate the middle level pedestrian representation: contour-motion features. Then we use boosting method to select the most representative features. Our experiments demonstrate that the proposed approach can outperform Viola's well-known pedestrian detector in both detection accuracy and generalization ability. In addition, even though our approach is presented in pedestrian detection scenario, it has been extended to human activity recognition application and remarkable performance has been achieved.

(2) A multi-granularity representation method for human detection is proposed, which we refer to as granularity-tunable gradients partition (GGP). The concept of gran-

ularity is used to define the spatial and angular uncertainty of the line segments in the Hough space. Then this uncertainty is back projected into the image space by orientation-space partitioning to achieve efficient implementation. By changing the granularity parameters, the level of uncertainty can be controlled quantitatively. Therefore a family of descriptors with versatile representation property can be generated. Specifically, the finely granular GGP descriptors can represent the specific geometry information of the object (the same as Edgelet); while the coarsely granular GGP descriptors can provide the statistical representation of the object (the same as histograms of oriented gradients, HOG). Moreover, the position, orientation, strength and distribution of the gradients are embedded into a unified descriptor to further improve the GGP’s representation power. A cascade structured classifier is built by boosting the linear regression functions. Experimental results on INRIA dataset show that the proposed method achieves comparable results to the-state-of-the-art methods.

(3) A a spatial-temporal granularity-tunable gradients partition (STGGP) descriptor for human detection is proposed. This method extend the GGP feature into the spatial-temporal domain. Therefore, it has the merits of both spatial-temporal analysis and granularity space representation. In addition, we present three methods to incorporate motion information with the appearance information. Specifically, in the first method, we represent the human body by two channels: spatial gradients field and optical field, then by calculating the GGP features on these two channels, we extract the appearance and motion information of human body. In the second method, we extract the GGP features on the three orthogonal planes ( $X - Y$ ,  $Y - T$  and  $X - T$  planes) to explore the correlation between the spatial and temporal axis. In the third method, we consider the human motions as 3D entities in the spatial-temporal domain and use the generalized planes to parse these entities. The generalized plane is defined in the 3D Hough space with explicit angular and spatial uncertainties ( granularity parameters ). By varying the granularity parameters, we can generate the granularity space representation of human motion in the spatial-temporal domain. We evaluate these methods for both human detection and activity recognition on the public dataset. Experimental results show that the propose methods can yield the comparable results as the state-of-the-art methods.

(4) A nonparametric background generation method is proposed and be used as the preprocessing step for human detection. By this means, the detection speed can be in-

creased and the false alarm can be reduced effectively. We introduce a new model, named as effect components description (ECD), to model the variation of the background, by which we can relate the best estimate of the background to the modes of the underlying distribution. Based on ECD, an effective background generation method, most reliable background model (MEBM), is developed. The basic computational module of the method is an old pattern recognition procedure, the mean shift, which can be used recursively to find the nearest stationary point of the underlying density function. The advantages of this method are: first, backgrounds can be generated from image sequence with cluttered moving objects; second, backgrounds are very clear and without blur effect; third, robust to noise and small vibration. Extensive experimental results illustrate its good performance.

**Keywords:** human detection; activity recognition; contour-motion feature (CMF); spatial-temporal analysis; multi-granularity representation; granularity-tunable gradients partition (GGP); space-time granularity-tunable gradients partition (STGGP); most reliable background model (MRBM)