

## 基于支持向量机与模糊 $k$ -均值算法的部位外观模型

韩贵金\*

(西安邮电大学 自动化学院, 西安 710121)

(\* 通信作者电子邮箱 hgjin123@126.com)

**摘要:**利用梯度方向直方图(HOG)建立的部位外观模型存在两个缺陷:不同部位采用相同的细胞单元尺寸,以及线性支持向量机(SVM)分类器不能准确表征部位定位状态与外观模型的相似度。为克服这两个缺陷,建立了一种基于SVM和模糊 $k$ -均值算法的部位外观模型。部位外观模型由两个分类器构成,线性SVM分类器用于判断部位定位状态是否属于人体部位,相似度分类器由部位定位状态与利用模糊 $k$ -均值算法确定的部位聚类中心的归一化欧氏距离来构造,用于计算部位定位状态与外观模型的相似度。仿真实验结果表明,与利用SVM算法和相同细胞单元尺寸建立的基于HOG特征的部位外观模型相比,新模型建立的部位外观模型能更准确地描述真实人体部位的外观特征,用于基于树形图结构模型的人体姿态估计时准确度也更高。

**关键词:**部位外观模型;梯度方向直方图;支持向量机;模糊 $k$ -均值算法;人体姿态估计

**中图分类号:** TP391.413 **文献标志码:** A

### Part appearance model based on support vector machine and fuzzy $k$ -means algorithm

HAN Guijin\*

(School of Automation, Xi'an University of Posts and Telecommunications, Xi'an Shaanxi 710121, China)

**Abstract:** The existing part appearance models based on Histogram of Oriented Gradient (HOG) have two defects: 1) the same cell size was used for different parts; 2) the linear Support Vector Machine (SVM) classifier can not represent the similarity of the position state and appearance model accurately. For overcoming these two defects, a part appearance model based on SVM and fuzzy  $k$ -means algorithm was built. The appearance model was composed of two classifiers: the linear SVM classifier was used to determine whether a position state belonged to human part; the similarity classifier, which was built according to the normalized Euclidean distance between the position state and the clustering center determined by fuzzy  $k$ -means algorithm, was used to calculate the similarity of the position state and appearance model. The experimental results show that the proposed appearance model can represent the appearance feature of real human part more accurately than the part appearance model built by SVM algorithm and HOG of the same cell size, and can get higher estimation accuracy when it is used to human pose estimation based on the tree-like pictorial structure model.

**Key words:** part appearance model; Histogram of Oriented Gradient (HOG); Support Vector Machine (SVM); fuzzy  $k$ -means algorithm; human pose estimation

## 0 引言

随着信息技术的进步以及人类安全意识的提高,视频监控系统在银行、医院、学校、商场和住宅小区等场合获得了越来越广泛的应用。在大多数应用场合,设置视频监控系统的目的都是为了监控人类的动作和行为。人体动作与行为的识别与分析是计算机视觉中一个非常热门的研究方向,吸引了大批研究人员的关注。人体是由头部、躯干、上臂、下臂、大腿、小腿等多个部位组成的,若能确定人体各部位在图像中的具体位置,既可为人体动作与行为的识别与分析打下坚实的基础<sup>[1]</sup>。人体姿态估计即是这样一个通过分析图像特征来自动确定人体部位在图像中的定位位置的过程<sup>[2]</sup>,作为人体动作与行为的识别与分析的一个基础问题,获得了很多研究者的关注,取得了不少的研究成果<sup>[3]</sup>。

人体姿态估计算法主要包含基于整体的方法和基于模型

的方法两大类<sup>[4]</sup>。其中基于整体的方法难以覆盖人体所有可能的姿态,只适用于特定人体姿态的估计;而基于模型的方法可以覆盖人体部位所有可能的姿态,从而受到了研究者更多的关注。基于模型的人体姿态估计方法的研究主要集中在建立人体模型、建立部位外观模型、减小部位搜索空间和优化推理算法等四个方面。部位外观模型在人体姿态估计过程中用于代替真实人体部位的外观特征来计算其与部位定位状态的相似程度,部位外观模型是否准确描述了真实人体部位的外观特征对人体姿态估计的准确度起着很大的影响,得到了人体姿态估计领域最多研究者的关注<sup>[3]</sup>。

梯度方向直方图(Histogram of Oriented Gradient, HOG)特征对光照变化和局部变形具有良好的不变性<sup>[5]</sup>,其在目标检测领域获得了成功的应用。由于在建立部位外观模型时经常面临不同图像的光照条件变化以及人体体型和着装的不同,所以HOG特征在建立部位外观模型时获得了广泛的应

用<sup>[6-9]</sup>。Yang等<sup>[6]</sup>将训练图像中标注的包含人体部位的图像区域对应的HOG特征构成训练样本集,利用支持向量机(Support Vector Machine, SVM)算法训练线性SVM分类器作为部位外观模型;Wang等<sup>[7]</sup>将空间上相连的部位看作一个整体,并将训练图像对应图像区域的HOG特征构成训练样本集,利用SVM算法训练线性SVM分类器作为联合部位的外观模型;Johnson等<sup>[8]</sup>利用HOG特征和图像分割信息分别建立两种部位外观模型,并将两种部位外观模型叠加在一起建立了新的部位外观模型;Sapp等<sup>[9]</sup>将HOG、颜色直方图和形状三种图像特征合并为一个特征向量,利用SVM算法构造线性SVM分类器作为部位外观模型。

由于HOG特征对光照变化和局部变形的良好的不变性,现有利用HOG特征建立的部位外观模型取得了良好的效果,但仍存在两个缺陷:1)对不同部位建立外观模型时采用的细胞单元尺寸都相同;2)几乎都是利用线性SVM分类器建立部位外观模型,利用分类器的输出来计算部位定位状态与部位外观模型的相似度。为克服这两个缺陷,本文建立了一种基于SVM和模糊 $k$ -均值算法的部位外观模型。本文的工作主要有四个方面:1)对人体的不同部位采用不同的细胞单元尺寸;2)利用SVM算法构造线性SVM分类器,用于判断部位定位状态是否属于部位类;3)利用模糊 $k$ -均值算法确定部位聚类中心,样本初始的隶属度由线性SVM分类器的输出来确定;4)利用部位定位状态与部位聚类中心的归一化欧氏距离构造相似度分类器,用于计算部位定位状态与部位外观模型的相似度。

## 1 基于SVM和模糊 $k$ -均值的部位外观模型

HOG特征是利用细胞单元内像素点梯度的方向密度分布来描述目标的局部形状的。不同人体的同一部位的HOG特征具有类似的整体轮廓,但由于着装的不同,局部细节可能有较大的区别,所以为有效识别人体部位,需要在有效描述整体轮廓的同时抑制局部细节,为此需要采用恰当的细胞单元尺寸,细胞单元尺寸太小将过于关注局部细节,太大可能连整体轮廓都将丢失。由于同一人体的不同部位大小是不同的,所以在对不同部位建立部位外观模型时需要采用不同的细胞单元尺寸。

SVM算法在模式分类中获得了非常成功的应用,传统SVM实际上就是一个线性分类器。SVM算法通过构建分界面来实现模式分类,但分界面方程对样本的输出并不能准确表征样本与模式类的相似程度,所以线性SVM分类器并不适合用于建立部位外观模型。

为克服现有利用HOG特征建立的部位外观模型存在的两个缺陷,本文的部位外观模型由两个分类器来构造,模型如式(1)所示。线性SVM分类器用于判断部位定位状态是否属于人体部位,相似度分类器用于计算部位定位状态与部位外观模型的相似度。

$$s = s_1(x_i) \cdot s_2(x_i) \quad (1)$$

其中: $x_i$ 为第 $i$ 个部位定位状态对应的HOG特征向量, $s_1(x_i)$ 和 $s_2(x_i)$ 分别为两个分类器的输出。

### 1.1 线性SVM分类器

SVM算法<sup>[10]</sup>是一种非常有效的机器学习算法,可以较为有效地解决机器学习中广泛存在的小样本、非线性、高维数、局部极小等问题,模式分类是SVM算法的一个非常典型的应用领域。

线性SVM分类器利用SVM算法训练得到的分界面方程来构造:

$$s_1(x_i) = w \cdot x_i + b \quad (2)$$

其中: $w$ 和 $b$ 分别为利用SVM算法训练得到分界面方程的权重向量和阈值。

### 1.2 相似度分类器

模糊 $k$ -均值算法<sup>[11-12]</sup>是一种经典的软聚类方法,样本属于模式类的程度用隶属度来表示。算法的基本过程为:首先初始化模式类别数、模糊指数以及阈值或最大迭代次数的算法参数,然后根据先验知识初始化样本对各模式类的隶属度,最后通过迭代使聚类准则函数达到最小,迭代过程中模式类的聚类中心和隶属度不断地发生变化。

模糊 $k$ -均值算法可成功实现模式的聚类并确定模式类的聚类中心:

$$z_i = \left( \sum_{j=1}^{2N} u_{ji}^m x_j \right) / \left( \sum_{j=1}^{2N} u_{ji}^m \right) \quad (3)$$

其中: $2N$ 为训练样本个数; $x_j$ 为样本; $u_{ji}$ 为模糊 $k$ -均值算法收敛后第 $j$ 个样本对部位类的隶属度; $m$ 为模糊指数,取2可以得到最满意的聚类结果,故本文取2。

根据样本与聚类中心的距离即可较准确地表征样本与模式类的相似程度。相似度分类器利用部位定位状态对应HOG向量与部位聚类中心的归一化欧氏距离来构造:

$$s_2(x_i) = \frac{1}{n} \sum_{k=1}^n (1 - |x_{ik} - z_{ik}| / \max(x_{ik}, z_{ik})) \quad (4)$$

其中: $n$ 为部位定位状态对应HOG特征向量的维数; $z_i$ 为部位聚类中心。

### 1.3 部位外观模型

基于SVM和模糊 $k$ -均值算法的部位外观模型的建立过程可归纳为如下4个步骤。

步骤1 构造训练样本集。

图1以躯干为例给出了训练样本集的建立过程。首先从训练图像中剪切标注的部位区域,并在图像的其他任意区域剪切与其大小相同的一个图像块,将从所有训练图像剪切得到的图像块都化为标准大小,如图1(b)所示;然后求解所有图像块的HOG特征,如图1(c)所示;最后将所有图像块的HOG特征构成训练样本集,其中将标注的部位区域剪切得到的图像块对应的HOG特征向量属于部位类,任意剪切得到的图像块对应的HOG特征向量属于非部位类。

步骤2 构造线性SVM分类器。

对训练样本集利用线性支持向量机算法构造如式(2)所示的线性SVM分类器。

步骤3 构造相似度分类器。

将线性SVM分类器对每一个样本的输出值作为样本的初始隶属度,模糊 $k$ -均值算法收敛后即可得到部位类的聚类中心,并构造如式(4)所示的相似度分类器。



#### 步骤4 建立部位外观模型。

利用线性 SVM 分类器和相似度分类器构造如式(1)所示的分类器,即为部位外观模型。对不同部位建立部位外观模型时的最优细胞单元尺寸根据部位类中所有样本与部位外观模型相似度均值的极大化来估计:

$$d_j^* = \max_{d_j} \left( \frac{1}{N} \sum_{i=1}^N s(x_i, d_j) \right) \quad (5)$$

其中:  $d_j^*$  为部位最优细胞单元尺寸;  $N$  为样本集中部位类中样本个数;  $x_i$  为部位类的第  $i$  个样本;  $s(x_i, d_j)$  为样本  $x_i$  与细胞单元尺寸为  $d_j$  时所建立的部位外观模型的相似度。

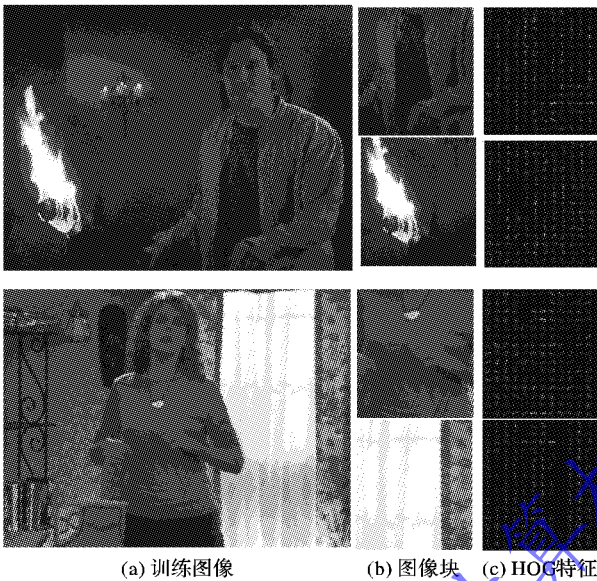


图1 样本集构造示意图

## 2 仿真实验

为便于比较,本文选择与文献[13]和[14]相同的训练图像集和测试图像集,训练图像集由从 Buffy 图像数据库中检测到人体上半身的图像中选择 513 张图像组成,测试图像集由从 Buffy 图像数据库选择的 235 张图像和 PASCAL 图像数据库中检测到人体上半身的 360 张图像组成。这两个图像数据库包含了多幅大小不一、着装风格和颜色不同的人体图片。

与文献[5]的设置相同,计算图像梯度时采用  $[-1, 0, 1]$  模板和无符号的梯度方向。本文将梯度方向分为 9 个区间,将 4 个细胞单元采用矩形块形式<sup>[5]</sup> 组成一个 HOG 块,其扫描步长与细胞单元尺寸相同。

为测试部位外观模型是否能准确描述真实人体部位的外观特征,本文采用所有测试图像标注的部位区域与部位外观模型的相似度来衡量。首先作出一个合理的假设:所有测试图像标注的部位区域与部位外观模型的相似度满足正态分布,利用极大似然估计法估计正态分布的均值和标准差。均值越大说明部位外观模型越能准确描述真实人体部位的外观特征,标准差越小说明部位外观模型的鲁棒性越好。

表 1 给出了所有测试图像与本文及利用线性 SVM 分类器所建立的部位外观模型相似度的比较结果。从表 1 可以看出,本文所建立的部位外观模型得到的相似度的正态分布的均值更大,标准差相同或更小,表明本文所建立的部位外观模

型能更准确地描述真实人体部位的外观特征,而且具有更好的鲁棒性。

将本文所建立的部位外观模型用于人体姿态估计,人体模型选择树形图结构模型<sup>[14]</sup>,图 2 给出了对几幅测试图像的人体姿态估计结果,表 2 给出了利用本文所建立的部位外观模型与其他几种同样基于树形图结构模型进行人体姿态估计的部位外观模型的比较结果。

表 1 两种算法的模型相似度比较

外观模型	躯干		头部		上臂		小臂	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差
SVM	0.75	0.11	0.80	0.10	0.69	0.14	0.67	0.15
本文模型	0.80	0.10	0.84	0.09	0.73	0.12	0.69	0.13



图2 人体姿态估计示例

表 2 人体姿态估计准确度比较

图像库	方法	躯干	头部	上臂	小臂
Buffy	文献[13]方法	90.7	95.5	79.3	41.2
	文献[15]方法	100.0	100.0	90.1	64.3
	文献[9]方法	100.0	96.2	95.3	63.0
	本文模型	100.0	100.0	91.8	64.7
PASCAL	文献[15]方法	100.0	100.0	85.0	50.4
	文献[9]方法	100.0	90.0	87.1	49.4
	本文模型	100.0	100.0	85.6	51.1

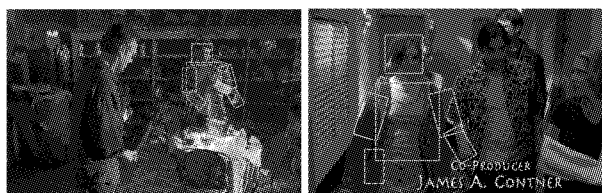
从图 2 可以看出,虽然几幅图像的光照条件、人体尺寸、体型和姿态都不相同,利用本文所建立的部位外观模型仍然获得了较为理想的人体姿态估计结果。从表 2 可以看出,与文献[13]和[15]相比,本文对人体各个部位的估计准确度都得到了提高;与文献[9]相比,除了上臂之外其他部位的估计准确度都得到了提高。本文的部位外观模型仅利用了 HOG 特征,虽然对 HOG 特征的利用方式比文献[9]更加恰当,但文献[9]中的部位外观模型利用的是颜色、HOG 和形状三种图像特征,而在相同情况下利用多种图像特征比仅利用单个图像特征能得到准确度更高的人体姿态估计结果<sup>[15]</sup>,所以上臂的估计准确度出现了下降的情况。

由于本文部位外观模型仅利用了 HOG 特征,当背景非常复杂时人体姿态估计容易出现错误,图 3 给出了利用本文部位外观模型进行人体姿态估计时两个失败的示例。图 3 中两幅测试图像对小臂的估计都出现了错误,这是由于两幅测试

图像估计的背景都比较复杂,而且第2幅测试图像中不同人体的距离很近。

### 3 结语

针对现有利用 HOG 特征建立部位外观模型时存在的两种主要缺陷,本文对如何解决这两种缺陷进行了研究,建立了一种基于 SVM 和模糊  $k$ -均值算法的部位外观模型。利用极大似然估计算法估计各部位的最优细胞单元尺寸,利用部位定位状态与模糊  $k$ -均值算法确定的部位聚类中心的归一化欧氏距离来表征相似度。仿真实验表明,所建立的部位外观模型可以更准确地描述真实人体部位的外观特征,将其用于基于树形图结构模型的人体姿态估计可以得到准确度更高的估计结果。由于同时利用多种图像特征才能更准确地表征人体部位的外观特征,所以下一步的主要工作是研究如何将 HOG 特征与颜色和形状等图像特征融合以建立更好的部位外观模型。



(a) 第1幅测试图像 (b) 第2幅测试图像

图3 本文算法估计结果

### 参考文献:

- [1] YANG W, WANG Y, MORI G. Recognizing human actions from still images with latent poses [C]// Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 2030–2037.
- [2] FELZENSZWALB P F, HUTTENLOCHER D P. Pictorial structures for object recognition [J]. International Journal of Computer Vision, 2005, 61(1): 55–79.
- [3] THOMAS B M, HILTON A, KRUGER V, et al. Visual analysis of humans [M]. Berlin: Springer, 2011: 131–138.
- [4] SAMUEL A J. Articulated human pose estimation in natural images [D]. Leeds: University of Leeds, 2012: 11–18.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2005, 1: 886–893.
- [6] YANG Y, RAMANAN D. Articulated human detection with flexible mixtures of parts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2878–2890.
- [7] WANG Y, TRAN D, LIAO Z. Learning hierarchical poselets for human parsing [C]// Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011: 1705–1712.
- [8] JOHNSON S, EVERINGHAM M. Combining discriminative appearance and segmentation cues for articulated human pose estimation [C]// Proceedings of the 12th International Conference on Computer Vision Workshops. Piscataway: IEEE, 2009: 405–412.
- [9] SAPP B, ALEXANDER T, TASKAR B. Cascaded models for articulated pose estimation [C]// Proceedings of the 11th European Conference on Computer Vision. Berlin: Springer, 2010: 406–420.
- [10] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273–297.
- [11] DUNN J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters [J]. Journal of Cybernetics, 1973, 3(3): 32–57.
- [12] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms [M]. Berlin: Springer, 1981: 65–85.
- [13] ANDRILUKA M, ROTH S, SCHIELE B. Pictorial structures revisited: people detection and articulated pose estimation [C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1014–1021.
- [14] HAN G, ZHAO Y. Human pose estimation based on tree-like picture model [J]. Journal of Xi'an University of Posts and Telecommunications, 2013, 18(3): 83–86. (韩贵金, 赵勇. 基于树形图结构模型的人体姿态估计[J]. 西安邮电大学学报, 2013, 18(3): 83–86.)
- [15] HAN G, ZHU H. Human pose estimation algorithm based on pictorial structure model [J]. Computer Engineering and Applications, 2013, 49(14): 30–33. (韩贵金, 朱虹. 一种基于图结构模型的人体姿态估计算法[J]. 计算机工程与应用, 2013, 49(14): 30–33.)
- [16] SAPP B. Pose estimation results: Buffy v2.1 [EB/OL]. [2014-12-02]. [http://www.seas.upenn.edu/~bensapp/buffy\\_all\\_tiled\\_cps.jpg](http://www.seas.upenn.edu/~bensapp/buffy_all_tiled_cps.jpg).
- [14] ALVARADO-CERVANTES R, FELIPE-RIVERON E. Improved HSI color space for color image segmentation [C]// CIARP 2012: Proceedings of the 17th Iberoamerican Congress on Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, LNCS 7441. Berlin: Springer, 2012: 348–354.
- [15] ZHANG C, WANG P. A new method of color image segmentation based on intensity and hue clustering [C]// ICPR 2000: Proceedings of the 15th International Conference on Pattern Recognition. Piscataway: IEEE, 2000: 613–616.
- [16] JU Z, CHEN J, ZHOU J. Image segmentation based on the HSI color space and an improved mean shift [C]// IETICT 2013: Proceedings of the 2013 IET International Conference on Information and Communications Technologies. Piscataway: IEEE, 2013: 135–140.
- [17] PAWLAK Z. Rough sets: Theoretical aspects of reasoning about data [M]. Norwell: Kluwer Academic Publishers, 1991: 9–16.
- [18] ZHANG Y. Image engineering (Ⅲ): image understanding [M]. Beijing: Tsinghua Publishing House, 2007: 340–343. (章毓晋. 图像工程(下册): 图像理解[M]. 北京: 清华大学出版社, 2007: 340–343.)
- [19] MUSHRIF M M, RAY A K. Color image segmentation: rough-set theoretic approach [J]. Pattern Recognition Letters, 2008, 29(4): 483–493.
- [20] CHENG H, JIANG X, WANG J. Color image segmentation based on homogram thresholding and region merging [J]. Pattern Recognition, 2002, 35(2): 373–393.

(上接第2024页)