

基于视觉的服装属性分类算法

刘 聪¹, 丁贵广²

(1 清华大学 计算机科学与技术系, 北京 100084; 2 清华大学 软件学院, 北京 100084)

摘 要: 提出了一种服装图像属性分类算法, 针对服装图像噪声多的问题, 采用人体部位检测技术定位服装关键部位并去除冗余信息, 提高了属性分类的准确率, 并提出了一种基于人体骨架与皮肤的特征提取算法, 以较少的维数表达衣型特点, 显著加快相关属性的分类速度. 针对服装属性语义复杂、需求多样化的问题, 为不同的属性构建了不同的 SVM 决策树模型, 从而提高分类效率, 并同时满足粗、细粒度的服装分类需求. 实验结果验证了该方法在多种服装属性分类任务上的有效性.

关键词: 模式识别; 图像属性分类; 服装领域; SVM 决策

中图分类号: TP391

文献标识码: A

文章编号: 1000—7180(2016)01—0028—06

Visual-Based Clothing Attribute Classification Algorithm

LIU Cong¹, DING Gui-guang²

(1 Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;

2 School of Software, Tsinghua University, Beijing 100084, China)

Abstract: We propose an algorithm for classifying clothing image attributes. To handle the noise in clothing images, key parts of clothing are located by a well-trained human part detector, and redundant information is eliminated, by which means the accuracy of clothing attribute classification is improved. Additionally, a novel feature descriptor based on human skeleton and skin is also proposed. This descriptor describes clothing feature with fewer dimensions, which significantly speeds up classifiers of related attributes. To deal with the complex semantic of clothing attributes, different SVM Decision Tree models are built for different attributes, which improves the efficiency of classification and achieves the objective of both coarse-grained and fine-grained classification. Experiments demonstrate the effectiveness of the proposed algorithm on multiple clothing attribute classification tasks.

Key words: pattern recognition; image attribute classification; clothing field; SVM decision tree

1 引言

基于视觉的服装属性分类的目的是从图像中识别和提取与服装相关的高层语义信息. 其结果可被应用于电子商务领域(服装风格检测、服装搭配推荐)乃至公共安全领域(基于衣着的人物跟踪), 因而相关研究越来越受到重视. 服装属性分类的任务是: 给定复杂背景下的人物图片, 判断图片中人物衣着的视觉属性, 包括颜色、花型、材质、袖长等.

现有的基于视觉的图像属性分类的研究大多面向物体属性^[1]和人体属性^[2], 而对服装属性的研究还比较少. 与传统的图像属性分类问题相比, 服装图片噪声较多, 拍摄的背景、拍摄的角度(如正面、侧面等)以及服装模特的姿势(如坐立、站立等), 都会对判断服装属性造成影响. 文献[3]采用文献[4]提出的人体部位检测方法, 圈出人体的关键部位, 将所有关键部位的多种图像特征拼接到一起作为服装的特征. 然而这种特征维度太高, 无效信息多, 导致分类

器的准确率和执行效率不理想. 此外, 皮肤特征^[5]对于一些服装属性(如袖长、裤长等)的判断具有重要的影响, 现有的研究未能充分利用皮肤信息. 在分类模型方面, 多分类 SVM 模型常用于解决图像分类问题, 效果良好. 但服装属性语义复杂, 具有一定的

层次结构及不同的粗细粒度. 传统的多分类 SVM 模型没有考虑服装属性的语义层次及粒度. 针对上述问题, 本文提出了一种服装图像属性分类方法(见图 1).

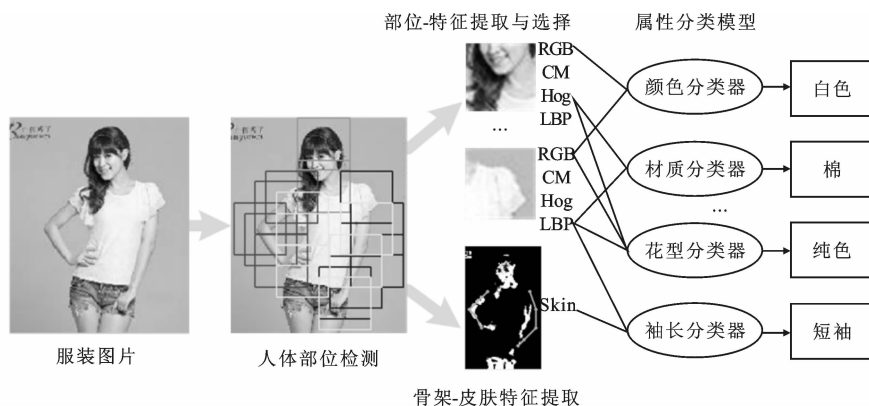


图 1 服装图像属性分类算法框架

2 本文方法

2.1 人体部位检测

由于人物姿态各异且背景环境嘈杂, 算法必须对不同姿势下的人体关键部位进行粗略定位, 来缩小工作区域, 排除噪声, 减少背景和姿势对服装属性判断的影响. 人体部位检测技术^[4]基于有人体部位关键点标注的训练集, 得到人体部位检测模型, 其中文献^[4]中的方法对上身、下身的人体部位均可适用, 具有较强的可扩展性. 图 1 的“人体部位检测”部分展示了各个关键部位检测的效果, 包含头部、颈部、胸部、腰部、胯部、肩部、大臂、小臂、手. 每一关键部位均对应一个长宽不超过 40 像素的矩形. 除头部、颈部以外, 其他部位均分为左右两个部分.

2.2 部位-特征提取与选择

完成人体部位检测之后, 对于每一个部位, 提取 4 种底层图像特征, 包括方向梯度直方图 (HOG)、局部二进制模式 (LBP)、颜色矩、颜色直方图. 文献^[3]将所有部位的所有种类的特征连接成一个长向量作为服装的特征表示, 其特征维度高达上万维, 训练效率很低, 且引入了过多无效信息, 降低模型准确率. 对此, 本文提出使用特征选择的方法进行改进.

考虑到不同部位、不同特征种类对于不同属性的影响, 除特征以外, 我们还对部位进行选择, 进一步降低特征空间维度.

综上考虑, 我们采用最优优先策略, 并避免反复验证选出的子集, 从而较快地找到较优的部位-特征

组合. 具体方法如下:

对于属性 A , 给定样本集 $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$, 其中 $y^{(i)} \in \{1, 2, \dots, |A|\}$ 为图片 $x^{(i)}$ 在属性 A 上的类别标签, 其中 $|A|$ 为属性 A 的类别数目. 人体部位集合为 $P = \{p_1, \dots, p_m\}$, 特征为 $F = \{f_1, \dots, f_k\}$.

(1) $\forall p \in P, \forall f \in F$, 取部位 p 的特征 f , 记为 $x_{p,f}^{(i)}$, 相应的样本集为 $\{(x_{p,f}^{(i)}, y^{(i)})\}_{i=1}^n$.

(2) 在样本集 $\{(x_{p,f}^{(i)}, y^{(i)})\}_{i=1}^n$ 上进行交叉验证, 训练 SVM 弱分类器: $h_{p,f}(x)$, 并采用本文 4.2 节中的正确率计算公式得到分数 $\text{score}(p, f)$.

(3) 在所有 $m \times k$ 个 (p, f) 组合中, 选出分数 $\text{score}(p, f)$ 最高的 T 个弱分类器, 所对应的 (p, f) 就是选出的部位-特征组合, 将其所对应的特征拼接起来, 即为服装的特征表示.

2.3 骨架-皮肤特征提取算法

皮肤特征对于一些服装属性的判断具有重要的影响, 比如短袖、短裤的图片可能会有较多的皮肤露出, 长袖、长裤则相对较少. 目前已经有较为成熟的模型^[5]可以将图片中的皮肤像素点标记出来. 但标记出的皮肤像素点无法向量化并用于属性分类器. 为此, 我们结合人体骨架结构, 提出了骨架-皮肤特征提取算法, 将皮肤特征向量化. 该特征维度较少, 且不受图像中人物的大小、姿势的影响.

以手臂为例, 骨架-皮肤特征的计算方法示意图见图 2. 给定一张图像 I , 具体的计算过程如下所示.

(1) 利用人体部位检测所得到的矩形及其相对

位置,可计算出人体骨架.以左臂为例,记其从肩部到手部的标记块依次为 $B_1, \dots, B_k$, 其中心点依次为 $p_1, \dots, p_k$. 将这些点依次连接,得到 $(k-1)$ 个向量,即为手臂骨架:

$$L = \overrightarrow{p_1 p_2}, \dots, \overrightarrow{p_{k-1} p_k} \quad (1)$$

(2) 获得皮肤像素点标记. 使用皮肤检测子 $S = f_{\text{skin}}(I)$, 输入图像 I , 得到皮肤像素点集合 Skin , 即如果图像 I 的 (x, y) 坐标处的像素点是皮肤, 则 $(x, y) \in \text{Skin}$.

(3) 将手臂部位的皮肤像素映射到手臂骨架上, 并将手臂骨架量化为 n 个桶, 统计每个桶是否有皮肤像素的映射, 即得到手臂骨架-皮肤向量, 记做 $\mathbf{b} = (b_1, \dots, b_n)$. 具体的做法如下所示.

① 计算手臂骨架 L 的总长度:

$$|L| = \sum_{i=1}^{k-1} |\overrightarrow{p_i p_{i+1}}| \quad (2)$$

式中, $|\overrightarrow{p_i p_{i+1}}|$ 为点 $p_i p_{i+1}$ 之间的欧式距离.

② 计算单个桶的长度: $l = |L|/n$. 那么, 对于每一段骨架 $\overrightarrow{p_i p_{i+1}}$, 桶的个数为

$$n_i = \left\lceil \frac{|\overrightarrow{p_i p_{i+1}}|}{l} \right\rceil \quad (3)$$

记该段的桶向量为 $\mathbf{b}^i = (b_1^i, \dots, b_{n_i}^i)$, 初始化 $\mathbf{b}^i = 0$.

③ 对每一段骨架 $\overrightarrow{p_i p_{i+1}}$, 更新桶向量 $\mathbf{b}^i$. 具体的做法是, 对该段骨架所连接的两个区块内的每一个皮肤像素点, 即 $\forall p \in (B_i \cup B_{i+1}) \cap \text{Skin}$, 将其映射到某一个桶 b_q^i 上, 更新 b_q^i ,

$$b_q^i := b_q^i + 1 \quad (4)$$

映射的方法是: 在直线 $\overrightarrow{p_i p_{i+1}}$ 上找到点 p_0 使得 $\overrightarrow{p p_0} \perp \overrightarrow{p_i p_{i+1}}$, 则 p_0 为映射点, p_0 所在的桶 b_q^i 即为被映射的桶, 其下标 q 为

$$q = \frac{\overrightarrow{p_i p_0} \cdot \overrightarrow{p_i p_{i+1}}}{|\overrightarrow{p_i p_{i+1}}|} \cdot \frac{1}{l} = \frac{(x_{p_0} - x_{p_i})(x_{p_{i+1}} - x_{p_i}) + (y_{p_0} - y_{p_i})(y_{p_{i+1}} - y_{p_i})}{\sqrt{(x_{p_{i+1}} - x_{p_i})^2 + (y_{p_{i+1}} - y_{p_i})^2}} \cdot \frac{1}{l} \quad (5)$$

④ 将所有 $\mathbf{b}^i$ 连接起来得到 $\mathbf{b}$, 并进行二值化处理:

$$b_j := \begin{cases} 1, & \text{if } b_j > \theta_{\text{pix}} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

式中, θ_{pix} 为经验阈值, 与单个桶长度有关, 一般可设定,

$$\theta_{\text{pix}} = \lceil \text{Height}(I)/2n \rceil \quad (7)$$

式中, $\text{Height}(I)$ 为图片高度的像素值.

⑤ 最后, 将两个手臂向量 $\mathbf{b}$ 连接在一起, 即为最终的骨架-皮肤向量.

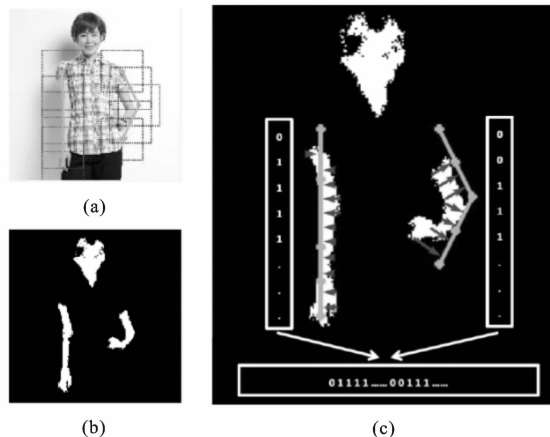


图2 手臂的骨架-皮肤特征的计算

2.4 属性分类模型

支撑向量机(SVM)模型常用于解决图像分类问题, 效果良好. 经典的 SVM 只能用于二分类^[6], 对于多分类问题, 一般的策略是 OvO (one-against-one) 和 OvA (one-against-all)^[7]. 前者被学术界认为准确率最优, 但需要 $k(k-1)/2$ 个子分类器的参与投票, 算法执行效率低. 一种较新的多分类策略是二叉决策树策略 (SVM-BDT)^[8], 即递归地对类别进行二叉划分形成一棵二叉决策树, 判别阶段只需要 $\log_2 k$ 个子分类器参与. SVM-BDT 采用数据驱动的方式自动构建决策树, 然而这种方法容易产生层数过深、平衡性差等问题.

基于以上考虑, 我们改进了 SVM 决策树策略, 提出混合粒度双层 SVM 决策树模型: 首先, 根据属性的特点, 从实际的分类需求出发, 将分类分为粗粒度分类和细粒度分类两层, 并构造两层属性树; 然后, 在进行属性分类时, 先对第一层进行粗分类, 满足粗粒度分类需求, 再在粗分类的基础上, 对第二层进行细粒度分类, 得到精细的分类结果.

基于此策略, 我们构造了不同属性的 SVM 决策树. 图3展示了袖长属性和颜色属性的 SVM 决策树. 袖长属性树的第一层对袖子长短进行粗分类, 第二层精细区分四种袖长类别; 颜色属性树是多叉树, 第一层对色系进行粗分类, 第二层确定具体颜色. 以袖长属性为例, 具体步骤如下所示.

(1) 袖长有四个类别: 长袖, 中袖, 短袖, 无袖. 这四个类别的输入将一起作为根节点 SVM 的输入.

(2) 确定粗粒度分类需求. 在实际需求中, 判断长袖还是短袖的场景比较常见, 因此, 不妨将长袖归为正例, 短袖归为负例. 而中袖和长袖比较接近, 无袖和短袖比较接近, 因此将中袖与长袖作为正例, 短

袖与无袖作为负例。

(3)进行细粒度分类,区分长袖/中袖,以及短袖/无袖。

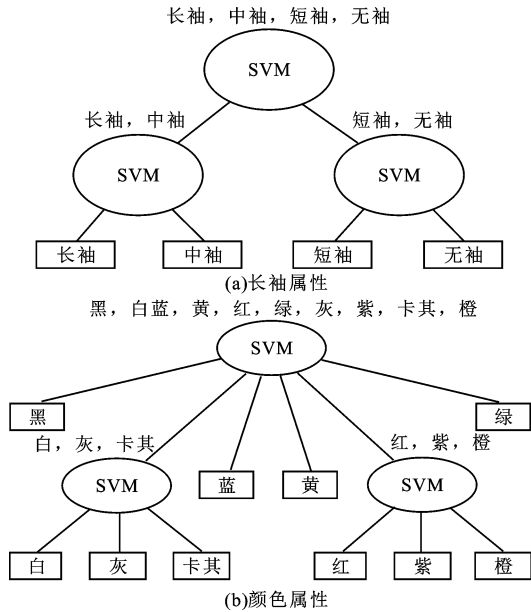


图 3 SVM 决策树

3 实验

3.1 实验数据集

目前服装领域缺乏较大规模的、有标注的、公开的数据集,文献[3]也没有公开其数据集.因此我们自行构建了规模较大、属性标注较全的数据集.该数据集包括从电商网站上爬取 11 760 张服装图片,其中上装 6 514 张,下装 1 658 张,连身装 3 588 张,包括西服、T 恤、衬衫、裤子、裙子等 22 种不同类型的服装.属性标注来自于电商网站上商家标明的信息,包括通用属性、上装属性、下装属性等共 14 种属性标注,每个属性包含 210 个类别标注,总共有 55 种类别,具体属性设置见表 1.

3.2 评价指标

本文使用常见的分类评价指标正确率 (Accuracy) 来评价分类结果.假设属性的类别标签集合为 $\{a_i\}_{i=1}^{|A|}$, TP_i 为所有属于类别 a_i 且被分类器判定为 a_i 的样本数目, N 为所有参与测试的样本,则正确率定义如下:

$$Accuracy(A) = \frac{1}{N} \sum_{i=1}^{|A|} TP_i$$

(8)

3.3 实验结果

为了更好地评价分类器效果,我们对参与训练和测试的图片数目进行了调整,使得同一个属性的

每个类别下具有大致相等的图片数目.按照 4 : 1 的比例划分训练集和测试集,训练集用来交叉验证确定模型参数、训练模型,测试集用来检验模型效果.

表 1 属性设置:通用属性、上装属性、下装属性

	属性名称	类别
通用属性	颜色	黑,白,蓝,黄,红,绿,灰,紫,卡其,橙
	花型	纯色,印花,拼接,条纹,格子
	材质	棉,聚酯纤维,雪纺,牛仔,毛,皮,天鹅绒
	季节	夏,春,秋,冬
	场合	日常,运动,宴会
上装属性	袖长	长袖,中袖,短袖,无袖
	领型	低领,高领
	前襟	开衫,套头
	版型	修身,宽松,直筒
下装属性	腰型	低腰,中腰,高腰
	裤长	长,中,短
	裙长	长,中,短
	裤型	修身,喇叭,直筒,哈伦
	裙型	修身,宽松

实验 1: 特征选择的效果.

为了检验特征选择的有效性,我们比较了三种输入不同的多分类 SVM 分类器.(1) 无特征选择:输入文献[3]中使用的无特征选择的拼接特征向量.(2) 有特征选择(top1):输入最优的 1 组部位-特征向量.(3) 有特征选择(top5):即本文第 2.2 节提出的特征选择方法,输入最优的 5 组部位-特征所拼接的向量.其中 SVM 分类器都采用线性核函数.由于裙装的图片数量较少,我们没有在裙装属性上进行实验.

图 4 展示了三种分类器在 12 种属性上的准确率.可以看出,使用了最优 5 组部位-特征的分类器在绝大多数属性上都明显优于无特征选择的分类器.只选 1 组部位-特征组合的分类器在 5 个属性上效果不佳.这说明单个部位的信息太少,没有完整地刻画待分类属性的本质特征.另一方面,最优 5 组部位-特征分类器效果的提升也证明了文献[3]中所采用的传统拼接方法所得特征包含噪声.

观察各个属性的表现,发现在颜色属性上,有特征选择比无特征选择的效果提高了约一倍.分析其原因:对于颜色属性来说,最优特征是颜色直方图,最优部位是躯干部分,而将所有特征拼接起来则引入了噪声,严重影响准确率.材质属性在三个分类器

上的表现都比较差,其原因在于材质属性与视觉特征关联度小.例如,化纤合成材质可以在视觉上模拟丝绸、棉等材质,凭肉眼无法准确判断.比较特殊的是腰型属性,特征选择没有提高腰型分类的效果,这是因为腰型依赖于其在人体部位中所处位置,单独给出下装部位特征很难对这一事实进行描述.总的来说,本文提出的特征选择方法能够有效提高服装属性分类效果.

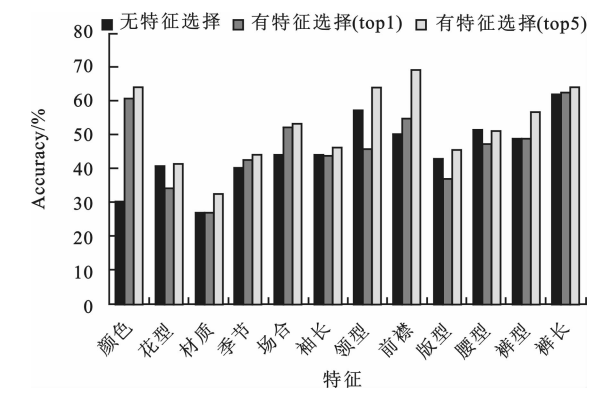


图 4 不同特征选择策略的属性分类正确率对比

实验 2:SVM 决策树的效果.

为了验证 SVM 决策树的效果,我们比较了两种策略的属性分类器:SVM-OVO 和 SVM 决策树.属性的决策树模型由 2.4 节中所述的方式来构建.以袖长属性作为二叉决策树的代表,以颜色属性作为多叉决策树的代表,比较两种策略的在测试集上的正确率和执行时间.实验结果如表 2 所示.可以发现,决策树策略与 OVO 策略相比,准确率有所提升,而且执行时间大幅度地降低.这说明我们的方法可以大幅度的提高执行效率,同时保证正确率.

表 2 SVM 决策树策略与传统 OVO 策略的对比

	颜色		袖长	
	正确率	执行时间	正确率	执行时间
OVO	63.92%	11.34 s	46.50 %	15.68 s
决策树	65.15%	6.78 s	49.25%	9.13 s
提升率	1.92%	40.21%	5.91%	41.77%

我们还以颜色属性为例,给出了使用 SVM 决策树模型进行多分类的混淆矩阵.如图 5 所示,绝大多数类别具有令人满意的正确分类比例.错分的情况大多是图像标注本身不够精确造成的.比如,有较多的样本实际类别为灰色,但被分类为黑、白、卡其色,这是因为颜色为深灰、浅灰的样本,其标注都是“灰色”,而灰色和卡其色也比较接近,因此发生混淆

也是可以理解的.

实验 3:骨架-皮肤特征的效果.

为了验证骨架-皮肤特征的效果,我们在 SVM 决策树模型的基础上,对比以下两种特征:(1) 普通特征(top5),即实验 1 中的“有特征选择(top5)”特征向量;(2) 骨架-皮肤特征,即按照第 2.3 节提出的方法构造的特征向量.表 3 展示了两种特征在袖长分类上的表现,并对比了不同的桶数 n 下的正确率.可以看出,骨架-皮肤特征的效果受参数影响较小,表现稳定.与普通特征相比,骨架-皮肤特征以极少的维数就达到了更好的正确率,且执行时间缩短了一个数量级.这说明对于皮肤相关属性来说,骨架-皮肤特征表达能力强,能够达到令人满意的正确率,并大幅提高分类效率.

	预测类别									
	黑	白	蓝	黄	红	绿	灰	紫	卡其	橙
黑	73.33	5.00	6.67	0.00	1.67	5.00	1.67	0.00	6.67	0.00
白	1.67	65.00	0.00	1.67	0.00	1.67	6.67	1.67	21.67	0.00
蓝	5.00	3.33	68.33	0.00	0.00	0.00	8.33	10.00	3.33	1.67
黄	0.00	8.33	1.67	65.00	0.00	8.33	3.33	0.00	11.67	1.67
红	0.00	10.00	0.00	0.00	70.00	1.67	1.67	3.33	5.00	8.33
绿	3.33	10.00	5.00	5.00	0.00	68.33	5.00	1.67	1.67	0.00
灰	11.67	10.00	3.33	1.67	6.67	6.67	48.33	0.00	11.67	0.00
紫	9.09	20.45	0.00	0.00	11.36	0.00	0.00	54.55	4.55	0.00
卡其	6.67	5.00	0.00	0.00	6.67	0.00	10.00	0.00	65.00	6.67
橙	0.00	0.00	0.00	2.13	12.77	0.00	0.00	2.13	10.64	72.34

图 5 颜色的多分类混淆矩阵.使用 SVM 决策树模型.

表 3 普通特征与骨架-皮肤特征在袖长分类上的对比

	普通特征	骨架-皮肤特征			
	(top5)	$n=25$	$n=50$	$n=75$	$n=100$
特征维数	4334	50	100	150	200
正确率/%	49.25	51.00	51.75	50.25	50.75
执行时间/s	9.13	0.13	0.16	0.33	0.48

4 结束语

本文提出的基于视觉的服装属性分类方法,结合人体部位检测技术进行特征选择,克服了服装特征噪声多的问题;充分利用了皮肤与人体骨架信息,大幅度减少了特征维数;从多样化的需求出发,改进了 SVM 决策树模型.该方法适用于多种属性分类问题,具有令人满意的正确率和执行效率.

下一步工作将考虑结合属性之间的关系提升属性分类效果,并基于属性分类进行服装的检索、搭配与推荐.

参考文献:

- [1] Farhadi A, Endres I, Hoiem D, et al. Describing objects by their attributes [C] // Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. USA, Miami, IEEE, 2009: 1778-1785.
- [2] Bourdev L, Maji S, Brox T, et al. Detecting people using mutually consistent poselet activations [M] // Computer Vision - ECCV 2010. Berlin, Heidelberg Springer, 2010: 168-181.
- [3] Liu S, Feng J, Song Z, et al. Hi, magic closet, tell me what to wear! [C] // Proceedings of the 20th ACM international conference on Multimedia. USA, Dallas, ACM, 2012: 619-628.
- [4] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts [C] // Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. USA, Colorado Springs, IEEE, 2011: 1385-1392.
- [5] Jones M J, Rehg J M. Statistical color models with application to skin detection [J]. International Journal of Computer Vision, 2002, 46(1): 81-96.
- [6] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000, 26(1): 32-42.
- [7] 徐勋华, 王继成. 支撑向量机的多类分类方法 [J]. 微电子学与计算机, 2005, 21(10): 149-152.
- [8] Madzarov G, Gjorgjevikj D, Chorbev I. A Multi-class SVM Classifier Utilizing Binary Decision Tree [J]. Informatica (Slovenia), 2009, 33(2): 225-233.

作者简介:

刘 聪 女, (1988-), 硕士研究生. 研究方向为计算机视觉、模式识别. E-mail: cong-liu11@mails. tsinghua. edu. cn.
丁贵广 男, (1976-), 副教授, 博士生导师. 研究方向为计算机视觉、模式识别、信息检索.

(上接第 27 页)

- [2] Niculescu D, Nath B. Ad Hoc positioning system (APS) [C] // Global Telecommunications Conference, 2001. GLOBECOM '01. San Antonio, IEEE, 2001: 2926-2931.
- [3] Brito L A, Liu Yi, Garcia Y. An improved error localization on DV-Hop scheme for wireless sensors networks [C] // Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on. Chengdu, 2010: 20-22.
- [4] Zhang Dengyi, Liu Feng, Wang Lei, et al. DV-Hop localization Algorithms based on centroid in wireless sensor networks [C] // Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on. China, Yichang, 2012: 3216-3219.
- [5] Li Mudong, Xiong Wei, Liang Qing. An improved ABC-based node localization algorithm for wireless sensor network [C] // Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference. Shanghai, 2012: 1-4.
- [6] 邱奉美, 李怀忠. 无线传感器网络 DV-Hop 定位算法的改进 [J]. 计算机工程, 2014, 40(8): 16-20.
- [7] 吴楠, 刘方爱, 王淑娟. 基于跳数修正和跳距调整的 DV-Hop 定位改进算法 [J]. 微电子学与计算机, 2015, 32(1): 91-95.

作者简介:

李长庚 男, (1970-), 博士, 教授. 研究为无线传感器网络.
刘孟思 (通讯作者) 女, (1990-), 硕士研究生. 研究为无线传感器网络定位. E-mail: 761242689@qq. com.