

一种基于 HOG-PCA 的高效图像分类方法^{*}

李林^{1,2}, 吴跃¹, 叶茂¹

(1. 电子科技大学 计算机科学与工程学院, 成都 611731; 2. 四川托普信息技术职业学院, 成都 611743)

摘要: 为了更有效地提高图像分类性能和准确率, 提出一种基于 HOG-PCA 的高效图像分类方法。首先通过提取方向梯度直方图(HOG)特征并作特征白化, 再随机下采样进行尺度统一, 随后采用主成分分析(PCA)进行特征映射, 最后用最小二阶范数判定进行最近邻分类。实验中, 采用 C++, 基于 OpenCV 和 Darwin 实现了提出的方法, 并在 Pascal 2012 数据集上进行测试, 比较了该方法和 BOW-SVM 方法的准确率和运行性能。实验证明, 提出的方法具有更高的准确率和更好的运行性能。

关键词: 方向梯度直方图; 主成分分析; 最小二阶范数; 图像分类; 图像特征

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1001-3695(2013)11-3476-04

doi: 10.3969/j.issn.1001-3695.2013.11.071

Effective image classification method based on HOG-PCA

LI Lin^{1,2}, WU Yue¹, YE Mao¹

(1. School of Computer Science & Engineering, University of Electronic Science & Technology of China, Chengdu 611731, China; 2. Sichuan TOP IT Vocational Institute, Chengdu 611743, China)

Abstract: This paper presented a novelty classification method based on histogram of gradient (HOG) features and principle component analysis (PCA). Firstly, it extracted HOG features of image and whitened the features. At the second stage, it subsampled the features to an uniform scale, then extracted the image HOG features and projected these features into PCA space. At the last stage, it compared each test image with the PCA projections of training images, searched the nearest image comparing to the test image, finally it decided which class the test image belonged to and its max confidence scores. It implemented this method using C++, and based on OpenCV 2.4.3 and Darwin 1.3.2 platform, tested this method on Pascal 2012 dataset with comparing to BOW-SVM method. The experiment shows this method has advantages both on running time and classification precision.

Key words: HOG (histogram of oriented gradients); PCA (principal component analysis); least squares norm; image classification; image feature

0 引言

图像分类是根据图像信息所反映的不同特征, 利用计算机对图像进行定量分析, 将图像或图像区域甚至是图像中每个像素点分为若干类别中的某一种, 从而达到分类目的。图像分类有着广泛应用, 如军事中飞机、导弹等目标识别, 工业生产中产品质量检测, 科学中卫星云图、红外图像分析, 医学中白细胞识别、超声波肝脏图像识别, 以及生活中景物识别。但受人为和自然等因素影响, 图像分类仍存在问题。到目前为止, 图像分类依然是图像处理领域中的研究难点和热点^[1]。

要获得好的分类效果, 图像特征有效表示是基础。为此, 研究人员提出了众多图像特征表示方法。其中比较有代表的是: Ojala 等人^[2]提出局部二值模式 (local binary pattern, LBP) 针对纹理特征的描述算法, 对光照有很强的鲁棒性, 但缺乏对尺度、旋转等的处理; Lowe^[3]提出的尺度不变特征变换 (scale invariant feature transform, SIFT) 能适应对图像的尺度、旋转以及仿射变换处理, 被广泛应用于场景匹配、目标识别、人脸识

别、图像拼接等^[1], 但该方法计算局部特征比较耗时; Bay 等人^[4]提出的快速健壮特征 (speed up robust features, SURF) 的计算时间优于 SIFT, 也能适应图像的尺度变化, 但牺牲了一定的精度; Dalal 等人^[5]提出的梯度直方图 (histogram of gradient, HOG) 对光照、尺度、方向有良好的适应性, 得到了广泛应用。

图像分类方法中典型的有基于 BOW-SVM^[6]的方法。BOW (bag of word, BOW) 最早来源于语言和信息处理领域, 被应用于计算机视觉领域时把图像当成文档, 特征为词组, 是一种基于独立特征的直方图表示^[7]。其通常包含特征提取、特征描述和特征词组生成三个步骤。将 BOW 用于图像分类还需要分类器, 比较典型的有支持向量机^[8]。其将数据映射到一个更高维的空间里, 在这个空间里建立一个最大间隔超平面。在分开数据的超平面两边建有两个互相平行的超平面。分隔超平面使两个平行超平面的距离最大化。该分类器具有很好分类效果, 但训练性能上没有明显的优势。总的来讲, 基于 BOW-SVM 的分类方法提供了很好的性能, 但该方法没有考虑特征点之间的空间关系, 且整个建立词组、模型训练的时间

收稿日期: 2013-01-31; **修回日期:** 2013-03-11 **基金项目:** 国家自然科学基金资助项目 (60702071); 国家“973”计划基础研究项目 (2010CB732501); 四川杰出青年基金资助项目 (09ZQ026-035)

作者简介: 李林 (1973-), 男, 四川蒲江人, 副教授, 系统分析师, 博士研究生, 主要研究方向为机器学习及其在计算机图像和视觉中的应用 (lilin200909@gmail.com); 吴跃 (1958-), 男, 四川南充人, 教授, 博导, 主要研究方向为计算机网络及数据挖掘; 叶茂 (1973-), 男, 教授, 博导, 博士, 主要研究方向为数据挖掘、计算机视觉、智能信息处理等。

较长。为此,本文提出了一种基于 HOG-PCA 的图像分类方法。首先利用 HOG 对光照、尺度、方向优越适应性很好地提取了图像特征;其次对特征白化,减小了异常点,最后利用了特征脸的计算性能优势。整个方法最大优点在于简单、计算性能好、图像类别适应度大,尤其适用于自然图像的分类。其基本框架如图 1 所示。

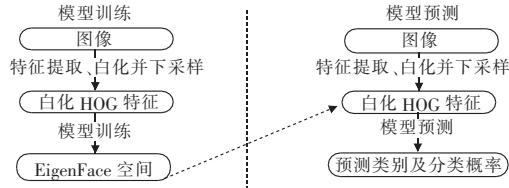


图 1 基于 HOG-PCA 图像分类方法

本文在数据集 Pascal 2012^[9] 测试了提出的方法,并与 BOW-SVM 方法进行了比较,实验验证了本文方法在准确率和运行时间方面具有明显优势。

1 基于 HOG-PCA 图像分类算法

图像分类通常包括图像特征提取与处理、分类模型建立、分类评价^[10]三个步骤。

1.1 HOG 特征提取

方向梯度直方图(histogram of oriented gradients, HOG)的核心思想是:所检测的局部物体外形能够被光强梯度或边缘方向的分布所描述。计算 HOG 共分五步:

a) 图像伽马和颜色标准化。在评价不同颜色空间(如灰度、RGB、LAB),为了更好地统一评价标准,一般统一转换到灰度图像,并且将灰度值转换到 0~1 范围。

b) 一维梯度滤波计算方法。梯度计算方法对正确描述图像特征至关重要,往往是简单方法最有效。在 HOG 计算中通常采用 Sobel 算子来计算一阶 x 或 y 轴的图像差分,以获得良好性能。Sobel 算子加上高斯模糊和差分能使结果具有更好的抗噪性能。

c) 空间和方向单元格带权投票。这是最为关键的一步,对每个单元格的方向设置累积权值,首先计算方向权值:

$$x(i) = \cos(\theta) \quad (1)$$

$$y(i) = \sin(\theta) \quad (2)$$

$$\theta = \theta + \pi / (N_{\text{direction}} + 1) \quad (3)$$

其中: i 为方向序号; θ 为角度,初值为 0; $x(i)$ 为 x 轴差分在 i 方向权值; $y(i)$ 为 y 轴差分在 i 方向权值。

其次是计算幅度和方向,幅度是 x 或 y 轴的图像差分的均方值。 $N_{\text{direction}}$ 为总的方向数,一般设置为 9。方向值取为各个方向上的带权最大值。

d) 构建块(block)并标准化。将单元格中的特征汇总组合成块,计算方法为

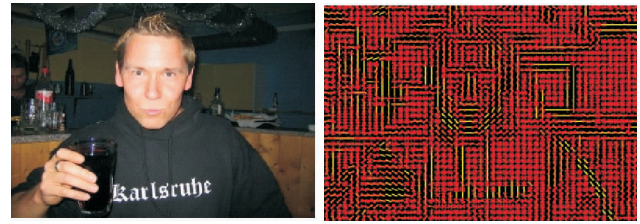
$$B(x) = (C(x) - B(\text{size}) + 1) / B(\text{step}) \quad (4)$$

其中: $B(x)$ 表示块 x 轴总数值; $C(x)$ 代表单元格 x 轴总数值; $B(\text{size})$ 为块大小; $B(\text{step})$ 是块变化的步长。同理, y 轴的块总数值计算类似。

标准化包含两步:(a) 特征剪切,根据设定最小特征和最大特征值范围进行裁剪,提出过小和过大的值,并将剩余特征值重新统一尺度到 0~1 范围;(b) 特征标准化。

e) 构建图像的方向梯度直方图。降维能减少特征数量,

降低计算复杂度,根据需要可以设置输出的维度。能量和方向加总:将不同方向和块上的特征值加总,形成图像的梯度直方图。与其他描述子相比,HOG 得到的描述子保持了几何和光学转换不变性。一个 HOG 特征的例子如图 2 所示,(a) 为图像,(b) 为图像的彩色 HOG 特征。



(a) 图像 (b) 图像的彩色 HOG 特征

图 2 图像及其 HOG 特征

1.2 HOG 特征白化和下采样

由于 HOG 特征仍然存在高度空间耦合,基于两个目的需要对 HOG 特征进行白化(whiten)处理:a) 减少特征之间的高度关联;b) 使训练特征具有相同方差。特征白化可以简述如下^[11]:假定对一组 d 维 n 行($d \times n$) 特征 X ,为了减少特征间的关联度,作如下变换

$$Y = WX \quad (5)$$

其中: Y 为变换后的特征, W 是白化矩阵。为使 W 能消除特征关联度,必须满足 YY^T 是对角线矩阵,即

$$YY^T = (n-1)I \quad (6)$$

满足式(6)要求的 W 有很多,为此作限制 $W = W^T$ 。这样可通过式(7)计算得到 W 。

$$W = \sqrt{n-1} (XX^T)^{-\frac{1}{2}} \quad (7)$$

由于待分类图像的尺度并不一致,为此还需要进行随机下采样或上采样。若白化特征数多于需要特征个数,则进行下采样:

$$Z = \text{subsample}(Y, m) \quad (8)$$

其中: Y 是白化后的特征; Z 是采样后得到的特征共有 m 个。函数 subsample 采取随机方法采样特征。

1.3 PCA 映射

主成分分析(PCA)^[12] 通过 K-L 变换抑制具有较少信息量的系数来实现数据维度的减少,但求协方差矩阵相当耗时。PCA 将多维图像信息转换成少数几个主成分,这几个成分包含了大多数的图像信息,提高了分类的效率^[13]。

假定现有 N 个样本特征 $\{x_1, x_2, \dots, x_N\}$, 每个特征是一个 n 维空间且属于类别 $\{X_1, X_2, \dots, X_c\}$ 的某一类。考虑一种线性变换,从 n 维空间映射到 m 维空间满足 $m < n$, 新特征向量 $y_k \in \mathbb{R}^m$ 定义为如下线性变换:

$$y_n = W^T x_k \quad k=1, 2, \dots, N \quad (9)$$

其中: W^T 是一个带有正交列的矩阵。假定总样本散点矩阵为

$$S_T = \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T \quad (10)$$

其中: N 为样本数量, $\mu \in \mathbb{R}^n$ 是所有样本的均值。接着可以采用线性变换:

$$W_{\text{opt}} = \arg \max_W |W^T S_T W| = [w_1, w_2, \dots, w_m] \quad (11)$$

其中: $\{w_i | i=1, 2, \dots, m\}$ 是 n 维特征空间散点集 S_T 对应的 m 个最大特征向量,也称为模型的映射特征^[14]。

1.4 分类判别

计算每个样本与训练样本的二阶范数距离,求其最小距离

的类别,将其归为该训练样本的类别。

a)将待识别图像特征投影到 PCA 空间中。

b)计算投影后的特征与 PCA 空间中训练样本之间的距离:

$$d(i) = \sqrt{(t_1(i) - p_1)^2 + (t_2(i) - p_2)^2 + \dots + (t_n(i) - p_n)^2} \quad (12)$$

其中: $d(i)$ 为与第*i*个训练样本之间的距离; $t_1(i)$ 为第*i*个训练样本的第一投影特征值; p_1 为某测试样本的第一投影特征值; $t_2(i)$ 、 p_2 、 $t_n(i)$ 、 p_n 含义类似。

c)最后选择距离最小的分类为其分类。

d)分类信度计算方法采用两者之间的距离为其信度。

2 算法实现和性能分析

2.1 Pascal VOC 2012 分类数据集简介

Pascal(pattern analysis, statistic modeling and computational learning,模式分析、统计建模和计算学习) visual object class (VOC)竞赛^[9]是欧盟信息社会技术(IST)委员会资助的一个竞赛项目。该竞赛包含两个部分:a)一个公开的数据集包含图像和对应标注,以及标准化的评估软件;b)一个年度竞赛和研讨会。Pascal VOC 2012 中对象分类数据集是最新、最完整的数据集,包含 20 个类别:人、鸟、猫、奶牛、狗、马、绵羊、飞机、自行车、船、公交车、小汽车、摩托车、火车、瓶子、椅子、餐桌、盆景、沙发和电视与显示器,用于分类的共计 11 540 个图像,用于训练的样本共计 5 717 个,测试评估样本 5 823 个。具体各分类图像个数统计如表 1 所示。

表 1 Pascal VOC 2012 实验样本统计

类别	训练	测试	类别	训练	测试
aeroplane	327	343	Dog	632	654
bicycle	268	284	horse	237	245
bird	395	370	motorbike	265	261
boat	260	248	person	1194	2093
bottle	365	341	pottedplant	269	258
bus	213	208	sheep	171	154
car	590	571	sofa	257	250
cat	539	541	train	273	271
chair	566	553	TVmonitor	290	285
cow	151	152	total	5717	5823
diningtable	269	269			

2.2 算法实现

本文基于 OpenCV2.4.3 和 Darwin 1.3.2^[15],在 VC++2012 下实现了提出的算法。具体步骤如下:

a)构建图像的 HOG 特征并白化。首先将图像转换到灰度空间,然后计算在灰度空间上的 HOG 特征。

b)下采样统一 HOG 特征大小。在 Pascal 数据集中,图像尺度差异比较大。为了便于比较并适当降低计算量,将图像下采样,抽取 100×100 共 10 000 个 HOG 特征点。

c)将训练样本进行 PCA 映射,构建训练样本的特征空间。

d)计算测试样本的 HOG 并白化,方法同训练样本计算方法一样。

e)将测试样本与训练样本的特征进行比较,选取最近邻的训练样本为其分类,并计算两样本的二阶范数距离为信度。

2.3 性能评价与分析

2.3.1 评价指标

当前,在评价分类器性能方面常用的两种方法是混淆矩阵和接收器性能曲线。混淆矩阵(confusion matrices)^[16,17]是模

式识别领域中一种常用的分类性能评价方法,它刻画了测试数据中真实属性与识别结果类型之间的关系。二分类问题的混淆矩阵如表 2 所示。

表 2 分类图像及所含对象统计

预测值	真实正值	真实负值
预测正值	TP	FP
预测负值	FN	TN

其中:TP(true positive)是正样本的正确识别个数;FP(false positive)是负样本识别为正样本的个数;FN(false negative)是负样本识别为正样本的个数;TN(true negative)是负样本的正确识别个数。

通过混淆矩阵可以计算总的准确率、平均精度、平均召回率,计算公式如下:

$$TA = (TP + TN) / (TP + FP + FN + TN) \quad (13)$$

其中:TA 代表总准确率,反映了算法总的测试精度。

查准率和查全率是图像分类效率评价的两个定量指标,不仅可以用来评价每次分类的准确性和全面性,也是在分类系统评价中衡量系统分类性能的重要指标。

查准率(precision ratio)是指分类检出的相关图像数目占分类图像总数的百分比。查准率反映分类准确性。实际应用中往往采用平均查准率来评价,对二分类问题其计算公式为

$$AP = (TP / (TP + FN) + TN / (FP + TN)) / 2 \quad (14)$$

查全率(recall ratio)是指分类检出的相关图像数占系统中相关图像总数的百分比。查全率反映分类全面性。实际应用中往往也采用平均查全率来评价,对二分类问题其计算公式为

$$AR = (TP / (TP + FP) + TN / (FN + TN)) / 2 \quad (15)$$

2.3.2 运行性能比较

用本文提出的方法测试 Pascal 2012 数据集。在 Intel i5、4 GB 内存、Windows 7(64)环境下实现了本文算法并进行了测试。每一分类的计算时间平均为 2~3 min(包括训练和测试)。而 BOW-SVM 的 BOW 特征描述生成需要 2 508 s,对每一分类的 SVM 训练和测试又需要约 3~10 min。可见,本文方法在总体性能方面明显优于 BOW-SVM 方法,如表 3 所示。

表 3 运行性能比较 /s

类别	BOW-SVM	HOG-PCA	类别	BOW-SVM	HOG-PCA
aeroplane	692.9	161.2	diningtable	183.7	110.4
bicycle	227.8	90.1	dog	330.6	110.5
bird	289.7	130.3	horse	770.8	109.0
boat	210.0	140.3	motorbike	255.6	108.8
bottle	283.2	129.0	person	794.4	108.8
bus	180.7	110.5	pottedplant	236.5	117.4
car	384.5	125.8	sheep	120.3	126.3
cat	281.2	110.0	sofa	213.3	120.1
chair	317.3	110.1	train	203.3	111.1
cow	157.2	109.0	TVmonitor	206.9	109.5

2.3.3 计算总准确率、平均查准率、平均查全率

实验结果的混淆矩阵如表 4 所示。

RP 是真实正值(存在分类对象),RN 是真实负值(不存在分类对象)。表中每一分类为两行,分别是预测正值(存在分类对象)和预测负值(不存在分类对象)。

从表 4 中可以看出,BOW-SVM 方法多数类别不能正确检出。

实验结果的总准确率、平均查准率、平均查全率如表 5 所示。

表 4 两种方法得到的混淆矩阵比较

类别	BOW-SVM		HOG-PCA	
	RP	RN	RP	RN
aeroplane	0	343	106	242
	64	5 416	189	5 286
bicycle	0	284	25	265
	0	5 539	301	5 232
bird	370	0	61	313
	5 453	0	363	5 086
boat	0	248	36	216
	0	5 575	239	5 332
bottle	0	341	52	317
	0	5 482	441	5 013
bus	0	208	51	160
	0	5 615	199	5 413
car	0	571	155	453
	0	5 252	500	4 715
cat	0	541	82	462
	0	5 282	382	4 897
chair	0	553	177	465
	6	5 264	539	4 642
cow	0	152	17	137
	0	5 671	204	5 465
diningtable	0	269	55	268
	0	5 554	314	5 186
dog	0	654	126	535
	0	5 169	562	4 600
horse	0	245	26	219
	0	5 578	228	5 350
motorbike	0	261	24	238
	0	5 562	270	5 291
person	71	2 022	1 009	1 223
	482	3 248	1 204	2 387
pottedplant	0	258	26	253
	0	5 565	268	5 276
sheep	0	154	11	144
	0	5 669	219	5 449
sofa	0	250	46	290
	0	5 573	303	5 184
train	0	271	36	239
	1	5 551	176	5 372
TVmonitor	0	285	68	227
	6	5 532	198	5 329

表 5 总准确率、平均查准率和查全率

类别	BOW-SVM			HOG-PCA		
	TA	AP	AR	TA	AP	AR
aeroplane	0.93	0.47	0.49	0.94	0.66	0.64
bicycle	0.95	- nan	0.5	0.92	0.51	0.52
bird	0.06	- nan	0.5	0.88	0.54	0.55
boat	0.96	- nan	0.5	0.92	0.55	0.55
bottle	0.94	- nan	0.5	0.87	0.52	0.53
bus	0.96	- nan	0.5	0.94	0.59	0.60
car	0.90	- nan	0.5	0.84	0.57	0.58
cat	0.91	- nan	0.5	0.86	0.55	0.54
chair	0.90	0.45	0.50	0.83	0.58	0.59
cow	0.97	- nan	0.5	0.94	0.53	0.54
diningtable	0.95	- nan	0.5	0.90	0.55	0.56
dog	0.89	- nan	0.5	0.82	0.54	0.54
horse	0.96	- nan	0.5	0.92	0.53	0.53
motorbike	0.96	- nan	0.5	0.91	0.52	0.52
person	0.57	0.37	0.45	0.58	0.56	0.56
pottedplant	0.96	- nan	0.5	0.91	0.52	0.52
sheep	0.97	- nan	0.5	0.94	0.51	0.52
sofa	0.96	- nan	0.5	0.90	0.54	0.54
train	0.95	0.48	0.50	0.56	0.56	0.55
TVmonitor	0.95	0.48	0.50	0.92	0.61	0.60

表 5 中,TA 代表总的准确率,AP 代表平均查准率,AR 代表平均查全率,-nan 为非数字。

从表 5 中可以看出,BOW 只在个别分类上检测有效,如 aeroplane、chair、person、train、TVmonitor,但其在平均查准率、平均查全率方面均低于本文的 HOG-PCA 方法。本文提出的方

法在各个类别上均有效,并具有较好的准确率。

3 结束语

本文分析了图像分类的一般原理、步骤及其评价方法,并提出了一种简单高效的图像分类方法。由于 HOG 特征固有的优点,该方法考虑了图像的旋转、光照以及尺度等因素,并能实现图像的快速分类。实验中采用 C++ 实现了提出的方法,其效率高于常用的 MATLAB 仿真方法,并与经典的 BOW-SVM 方法进行了比较。结果表明,本文方法在运行性能、总的准确率、平均查准率、平均查全率方面均具有明显的优势。图像分类具有广泛的应用价值,其分类精度的提高还有赖于图像场景、色彩、遮挡等其他更为复杂影响因素的进一步分析和研究,这一直是图像分类研究的重点和难点,也是图像分类的重要发展方向之一。

参考文献:

- [1] 向征,谭恒良,马争鸣.改进的 HOG 和 Gabor、LBP 性能比较[J]. 计算机辅助设计与图形学学报,2012,24(6):787-792.
- [2] OJALA T, PIETIKINEN M, HARWOOD D. A comparative study of texture measures with classification based on featured distributions [J]. *Pattern Recognition*, 1996, 29(1):51-59.
- [3] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- [4] BAY H, ESS A, TUYTELAARS T, *et al.* Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3):346-359.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005:886-893.
- [6] CSURKA G, DANCE C, FAN L, *et al.* Visual categorization with bags of keypoints [C]//Proc of IEEE Conference on Workshop on Statistical Learning in Computer Vision. 2004:1-22.
- [7] LI Fei-fei, FERGUS R, TORRALBA A. Recognizing and learning object categories [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2007.
- [8] CORTES C, VAPNIK V. Support-vector networks [J]. *Machine Learning*, 1995, 20(3):273-297.
- [9] EVERINGHAM M, Van Gool L, WILLIAMS C K I, *et al.* The Pascal visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010, 88(2):303-338.
- [10] LU D, WENG Q. A survey of image classification methods and techniques for improving classification performance [J]. *International Journal of Remote Sensing*, 2007, 28(5):823-870.
- [11] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images [D]. Toronto: University of Toronto, 2009.
- [12] JOLLIFFE I. Principal component analysis [M]. Toronto: Wiley Online Library, 2005.
- [13] BELHUMEUR P N, HESPAHNA J P, KRIEGMAN D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, 19(7):711-720.
- [14] SIROVICH L, KIRBY M. Low-dimensional procedure for the characterization of human faces [J]. *Journal of the Optical Society of America A*, 1987, 4(3):519-524.
- [15] GOULD S. DARWIN: a framework for machine learning and computer vision research and development [J]. *Journal of Machine Learning Research*, 2012, 13(12):3499-3503.
- [16] FAWCETT T. An introduction to ROC analysis [J]. *Pattern Recognition Letters*, 2006, 27(8):861-874.
- [17] DAVIS J, GOADRIC M. The relationship between precision-recall and ROC curves [C]//Proc of the 23rd International Conference on Machine Learning. 2006.