ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 31,No. 5,Sep. 2016,pp. 958－964
DOI:10.16337/j.1004-9037.2016.05.012

http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

基于多核学习的静态图像人体行为识别方法

杨红菊[1,2]　冯进丽[1]　郭　倩[1]

(1.山西大学计算机与信息技术学院,太原,030006; 2.山西大学计算智能与中文信息处理教育部重点实验室,太原,030006)

摘　要:提出一种基于广义性多核学习的静态图像人体行为识别方法。从图像中提取基于边缘的梯度方向直方图和基于稠密采样的尺度不变特征描述子,并使用空间金字塔模型加入粗略空间信息;运用直方图内交核函数计算金字塔模型各层核矩阵,通过广义性多核学习方法求解各个核矩阵权重,以线性组合方式得到最优核矩阵;最后利用多核学习决策函数进行行为识别。Willow-actions数据集实验结果表明,本文方法比其他几种方法更加有效。

关键词:行为识别;广义性多核学习;空间金字塔模型;直方图内交核函数

中图分类号:TP391.41　　**文献标志码**:A

Action Recognition in Still Image Based on Multiple Kernel Learning

Yang Hongju[1,2], Feng Jinli[1], Guo Qian[1]

(1. School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, China; 2. Key Laboratory of Computational Intelligence & Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, China)

Abstract: A novel action recognition method based on general multiple kernel learning is proposed. Firstly, histogram of oriented gradients (HOG) based on edge of image and scale invariant feature transform (SIFT) based on dense sampling are extracted. Furthermore, spatial pyramid model is considered to obtain coarse spatial information. Then, the kernel matrix of each level in spatial model is computed by histogram intersection kernel function. With general multiple kernel learning, the weights of kernel matrixes are solved and the optimal kernel matrix is achieved by the linear combination of kernel matrixes. Finally, action recognition is realized by the decision function. The obtained impressive result shows that the proposed algorithm is more effective than some common methods in Willow-actions dataset.

Key words: action recognition; general multiple kernel learning; spatial pyramid model; histogram intersection kernel function

引　　言

静态图像人体行为识别是指计算机根据图像具有的某种视觉特征将其划分到预先设定的不同行为类别中,在大规模图像检索、视频行为标注和智能视频监控等方面拥有非常好的应用前景。但由于照相

基金项目:高等学校博士学科点专项科研基金(20121401120015)资助项目;国家自然科学基金(61201453)资助项目;山西省自然科学基金(2012011014-4)资助项目。
收稿日期:2015-05-28;**修订日期**:2015-06-14

机视角变化、遮挡现象、人体姿势多样性及人类丰富多样的衣着等客观因素导致静态图像人体行为识别变得非常困难。过去大部分有关行为识别的研究主要集中于如何从一段未知的视频[1]中识别出视频里正在发生的行为,而对于静态图像人体行为识别的研究相对较少。在现实生活中,人类视觉往往可以通过一幅图像传递出的信息识别人体在该图像中正在发生的行为,例如读书、弹吉他等,由此可见,通过静态图像识别人体行为具有可行性。在过去的研究中,大多数研究者的焦点集中于某些特定领域,例如体育运动[2-4]和弹奏乐器[5],文献[6]将这一研究推广到一般常见行为,使得该研究更加丰富且更具挑战性;文献[7]采用图片中目标人物的姿势为主要划分线索,但该方法不能有效控制姿势变化给识别带来的困难,所以部分研究者借助上下文信息进行辅助识别;文献[4]通过同时考虑交互物体和行为者姿势,捕捉上下文信息提高识别率;文献[3]通过图像里的场景、物体和行为者3个上下文信息提高识别率,并增强图像的语义性描述。视频里的"动作"为行为识别提供了足够多信息,一幅静态图像所包含的信息量却十分有限,所以静态图像人体行为识别研究更富有挑战性,需要找出可以准确测量静态图像相似度的方法。为了测量视觉相似度,许多图像特征被提出,总结得出如下结论:各种特征的区别在于识别度和不变性两方面的折中表现。没有任何一种特征在所有任务中都会呈现出最佳性能,只能根据特定识别任务的特点来选择相应的理想特征。广义性多核学习[8,9]针对特定数据库,该方法基于SVM框架对基本特征(如颜色、纹理)训练得到识别度和不变性的最佳折中特征组合,从而获得最佳视觉相似度。

本文采用的图像特征首先提取基于边缘的梯度方向直方图(Histogram of oriented gradients, HOG)[10]和基于稠密采样的尺度不变特征描述(Scale invariant feature transform, SIFT)[11,12];然后使用空间金字塔模型(Spatial pyramid model, SPM)加入粗略空间信息作为最终图像特征。提取该特征以后,使用广义性多核学习(General multiple kernel learning, GMKL)训练各层特征核矩阵,进而通过线性组合得到最优核矩阵,最后利用基于最优核矩阵决策函数进行行为识别。广义性多核学习为利于识别的特征赋予较大的权重,为不利于识别的特征赋予较低权重,进而得到折中特征组,实现最佳视觉相似度测量。此外,本文通过同时考虑整幅图像和目标对象所在矩形框特征,实现图像上下文信息结合,进而增强行为识别准确率。

1 特征提取

1.1 空间金字塔模型

特征词袋(Bag-of-feature, BoF)[13]算法借鉴文本信息处理的Bag-of-word算法思想,将图像表示为视觉关键词统计直方图。所谓视觉关键词是指图像局部区域特征(如HOG,SIFT)经过聚类形成的聚类中心。特征词袋算法的实现包括4个步骤:(1)特征点采样,包括稠密采样和稀疏采样;(2)运用描述子算法描述特征点;(3)无监督聚类算法(如K-means)对描述子进行聚类,将聚类中心视作视觉关键词;(4)将描述子量化为视觉关键词,统计视觉关键词个数,形成视觉关键词直方图作为图像特征向量。

特征词袋算法对局部特征进行整体考虑,忽略了空间分布信息,从而限制了对图像的表示能力。为了弥补这一缺陷,空间金子塔模型[14]被提出。空间金字塔在不同层次上对图像进行分块,层次越高划分图像子块数越多,在各块区域提取相应的区域特征。该特征通过大量分块方式实现了空间信息的考虑。假设一个3层空间金子塔模型(见图1),其实现方法为:第0层为原图像,第1层对图像进行2×2分块,第2层对图像进行4×4分块,共得到21个图像块区域。其他层数的空间金字塔的划分方式以此类推。由此可见,空间金字塔模型的图像块区域包含不同尺度和不同空间位置信息,因此对它们的有效组织可得到鲁棒性的图像特征。

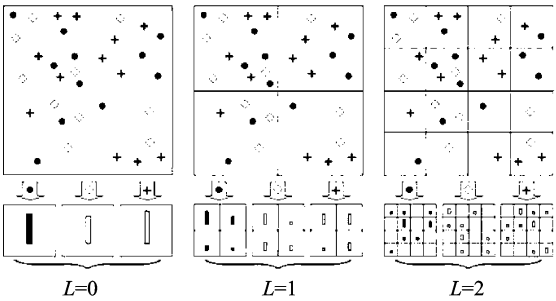


图 1 空间金字塔模型示意图

Fig. 1 Schematic diagram of spatial pyramid model

1.2 Phog 特征

Phog 是一种形状外观描述子，用来表征一幅图像整体形状外观以及该形状在空间上的布局关系。Phog 提取示意图如图 2 所示。具体算法实现如下：

(1)获取图像的边缘信息。利用 Canny 检测方法获取图像边缘像素点。图像的边缘展现出图像内容的轮廓，而轮廓能够表示图像的形状外观信息，因此 Phog 是一种形状外观描述子。

(2)计算边缘像素点的方向与幅值，并将其映射到对应的方向柄上。本文选择 8 个柄，即每 45°为一个方向区域。

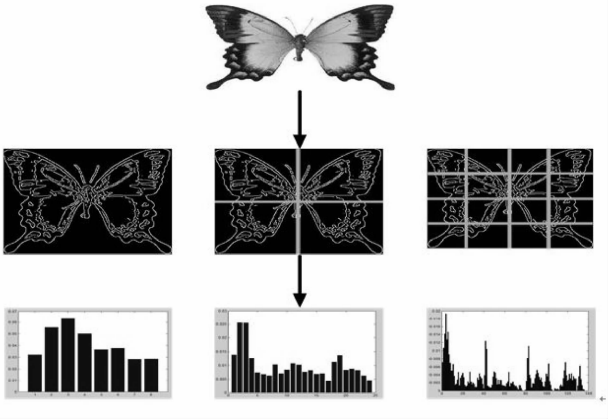


图 2 Phog 提取示意图

Fig. 2 Schematic diagram of extracting Phog

(3)利用空间金字塔模型对图像进行空间层次划分，统计每一个划分块上的 HOG，然后将各梯度直方图串联在一起便得到 Phog 特征。该特征通过在空间金字塔模型上的描述，突出了形状外观各局部部位的空间联系。在本文空间金字塔模型的层数取 3，Phog 的特征维数为：$8+8\times4+8\times16=168$。

2 基于多核学习的行为识别

2.1 广义性多核学习算法

在图像特征(如颜色、形状和纹理等)提取之后分类器使用之前，组合各种类型的特征，以求获得对图像描述能力强而全面的联合特征。联合特征一般有两种常见的方法：(1)原始特征的简单组合；(2)各种特征生成的全新特征。这种联合方式有一个很大的缺陷就是新生成的特征维数太大，容易导致溢出，处理起来较为困难。多核学习(Multiple kernel learning, MKL)是近年来图像分类、目标识别领域一种流行且经典的后期特征融合算法。该方法是在构造分类器期间对不同类型的特征进行处理和组合，最后得到决策类别。MKL 是在支持向量机(Support vector machine, SVM)[15]框架上建立的多特征融合方法。普通的 MKL 方法是对每一种特征训练出的和矩阵进行简单的线性组合，对于线性组合的权重选择具有随机性与盲目性。本文应用的是改进后的 MKL 算法，即广义性多核学习算法，该算法通过 $l1$ 正则化项选择各种特征核矩阵对应的权重，利于识别的特征赋予较高权重，不利于识别的特征赋予较低权重，从而实现了多特征的最佳折中组合。

假设 M 种基本特征,对应距离函数为 $f_1,f_2,\cdots,f_M$。这些特征和距离函数通过核技巧得到基本核矩阵 $\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_M$。所谓核技巧是指应用某种距离函数(如高斯核函数)隐式地实现特征由低维向高维转换的同时实现高维向量内积求解,是一种使复杂问题简单化算法思想。根据文献[16],直方图内交核(Histogram intersection kernel,HIK)函数优于其他核函数,所以本文基本核矩阵求解使用直方图内交核函数来实现,即

$$K_m(x,y)=\sum_{i=1}^{R}\min(x_i,y_i) \tag{1}$$

式中:x,y 为两幅图像的第 m 种基本特征;R 为第 m 种基本特征的维度。

给定基本核矩阵后,最优核矩阵可表示为 $\boldsymbol{K}_{\text{opt}}=\sum_{k=1}^{M}\boldsymbol{d}_k\boldsymbol{K}_k$,其中 d_k 表示权重,即第 k 种基本特征在识别力和不变性上的折中程度值。在 $l1$ 正则化的辅助下,利用 SVM 框架思想优化权重向量 $\boldsymbol{d}$,使训练集上的分类准确率最大。目标函数定义为

$$\min_{w,d,\xi}\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+\boldsymbol{C}\boldsymbol{l}^{\mathrm{T}}\boldsymbol{\xi}+\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{d} \tag{2}$$

式中:$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\phi}(x_i)+b)\geqslant 1-\xi_i;\xi\geqslant 0,d\geqslant 0;\boldsymbol{\phi}^{\mathrm{T}}(x_i)\boldsymbol{\phi}(x_j)=\sum_{k=1}^{M}d_k\boldsymbol{\phi}_k^{\mathrm{T}}(x_i)\boldsymbol{\phi}_k(x_j)$。与标准 SVM 目标函数唯一区别为:加入了一项有关权重向量 $\boldsymbol{d}$ 的 $l1$ 正则化项。依赖参数 $\boldsymbol{\delta}$ 向量大部分的权值 d_m 将为 0,注意参数 $\boldsymbol{\delta}$ 中可以包含优先被选特征的先验信息。在本文实验中,取 $\boldsymbol{\delta}$ 向量的所有数值均为常数 1。$\boldsymbol{\phi}^{\mathrm{T}}(x_i)\boldsymbol{\phi}(x_j)=\sum_{k=1}^{M}d_k\boldsymbol{\phi}_k^{\mathrm{T}}(x_i)\boldsymbol{\phi}_k(x_j)$ 可以等价为 $\boldsymbol{K}_{\text{opt}}=\sum_{k=1}^{M}d_k\boldsymbol{K}_k$。

该问题具有强对偶关系,转化为对偶问题进行求解,有

$$\max_{\alpha,\delta}\boldsymbol{I}^{\mathrm{T}}\alpha+\boldsymbol{\delta}\quad\delta\geqslant 0,C\geqslant\alpha\geqslant 0,\boldsymbol{1}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{\alpha}=0;\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Y}D\boldsymbol{K}_k\boldsymbol{Y}\boldsymbol{\alpha}\leqslant\boldsymbol{\delta}_k \tag{3}$$

式中:非零 α 对应特征是支持向量;$\boldsymbol{Y}$ 为对角矩阵、对角线上的元素为样本的类别标签。

该对偶问题是一个存在全局最优极值的凸优化问题。采用文献[8]提出的 minimax 优化策略进行求解。简单而言,多核学习的参数学习就是一个不断迭代过程,每次迭代包含 2 个步骤:(1)训练一个核矩阵为 $\boldsymbol{K}=\sum_{k=1}^{M}d_k\boldsymbol{K}_k$ 的标准 SVM 模型;(2)求目标函数对 d_k 的梯度。反复迭代,直到收敛或达到最大迭代次数为止,通过最后一次迭代确定权重向量 $\boldsymbol{d}$ 和向量 $\boldsymbol{\alpha}$。给定测试样本 x,其标签决策函数为

$$g(x)=\text{sign}(\sum_{i=1}^{N}\alpha_i y_i\boldsymbol{K}_{\text{opt}}(x,x_i)+b) \tag{4}$$

式中:N 为训练图像个数;x_i 为第 i 幅训练图像;$y_i\in\{1,-1\}$ 为第 i 幅训练图像的类别标签;b 为门限值。这是一个二分类决策方法,对于多分类问题,本文采用一对多策略实现。具体方法:假设一个 k 类分类问题,其中一类样本集作为一个训练类别,剩余类样本集作为另一个类别,构建二分类器,一共训练 k 个分类器;然后根据测试样本与分隔面距离最大的原则确定测试样本标签。

2.2 行为识别算法

为了得到图像上下文信息,本文分别对整幅图像和目标对象所在矩形框提取特征,然后利用广义性多核学习算法对所提取特征进行训练得到决策函数,进而对未知图像进行行为识别。具体步骤为:

(1)对整幅图像进行稠密采样,所谓稠密采样是指每隔 10 个像素点取一点作为特征点;SIFT 描述子对特征点进行描述;K-means 聚类算法对 SIFT 描述子进行聚类,构建视觉关键词(视觉关键词数量为 1 024);3 层空间金子塔模型划分图像子块区域,并在各子块区域上统计视觉关键词直方图;将 3 层视觉关键词直方图分别进行保存。对目标对象所在矩形框也做同样处理。

(2)提取整幅图像的 Phog 特征,对 3 层 Phog 特征分别进行保存。对目标对象所在矩形框也做同样处理。此时,一共得到 3×4=12 个通道特征。

(3)利用广义性多核学习求解各通道特征权重,线性组合得到基于 12 个通道特征的最优核矩阵,然后利用基于最优核矩阵的决策函数进行行为识别。

3 实验结果和分析

本文使用 Delaitre 收集的静态图像行为识别 Willow-actions 数据集进行实验,该数据集拥有 911 张图片和 7 个行为类(见图 3),图 3 中行为从上往下分别为:使用电脑、照相、演奏乐器、骑车、骑马、跑步和走路。该数据库按照 Pascal VOC 标注方式对图片里正在发生的行为对象用矩形框进行标注。需要注意的是一幅图像可能拥有多个目标行为数据。本文按照文献[6]的数据组织方式进行实验。首先取每个行为类的 70 个行为对象数据组成训练集,进行分类器模型训练;其次将剩余行为对象数据作为测试集对分类器模型的识别结果进行评价。

图 3 静态图像行为识别数据库图像示例

Fig. 3 Example images from dataset with human action in still images

3.1 各种核函数的性能比较

选择正确的核函数有助于提高识别准确率,本文通过实验验证法选取最佳核函数。本文选用 LIBSVM[14] 工具箱来比较常见的几种核函数在行为识别效果上性能的优劣。实验方法:首先在行为对象所在矩形区域提取基于 Densesift 的 3 层空间金子塔模型特征作为图像表征;然后应用 LIBSVM 工具箱进行行为识别,在工具箱中设置不同核函数进行实验,本文选取并验证了线性核函数、多项式核函数、RBF 核函数(高斯核函数)、Sigmoid 核函数以及直方图内交核函数(Histogram intersection kernel, HIK)的识别效果,基于各种核函数的识别性能比较如图 4 所示。由图可见,直方图内交核函数的识别准确率远远优于其他 4 种核函数,由此本文在广义性多核学习中,采用该核函数进行行为识别模型训练。

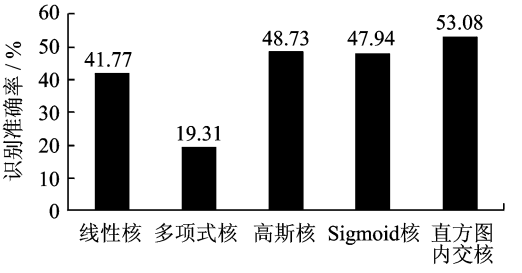


图 4 核函数性能比较图

Fig. 4 Comparison results of classification performance for different kernel functions

3.2 行为识别正确率比较

本文比较实验使用单核学习 SVM 分类器,且采用直方图内交核函数。对比算法分别是:(1)文献[11]提出的基于稠密采样的特征词袋 BoF 算法,采样间隔取 10 像素,视觉关键词数量为 1 024,分别对整幅图像和目标区域(即行为对象所在矩形框)进行实验;(2)文献[6]的方法 *A* 和方法 *B*,其中方法 *A* 对整幅图像在采用相同参数设置对比 BoF 算法基础上,应用 3 层空间金字塔模型加入粗略的空间信息;方法 *B* 将方法 *A* 的算法流程应用于目标区域。几种算法的行为识别准确率比较结果如表 1 所示。

表 1 不同算法的行为识别正确率

Tab. 1 Accuracy of action recognition for different methods %

数据库图像	BoF[14]		方法 A[6]	方法 B[6]	本文方法
	整幅图像	目标区域	整幅图像	目标区域	
使用电脑	71.79	66.67	87.18	74.36	84.26
照相	18.18	11.69	25.97	22.08	30.26
演奏乐器	55.08	34.75	64.41	40.68	54.24
骑车	70.50	62.59	66.91	64.03	70.50
骑马	50.88	45.61	57.89	57.89	75.44
跑步	34.57	43.21	34.57	51.85	39.51
走路	40.16	55.74	38.52	60.66	55.74
平均识别准确率	48.74	45.75	53.64	53.08	57.12

由表 1 可见:(1)本文算法的平均行为识别率为 57.12%,远远优于其他 4 种对比算法,证明了本文算法的有效性;(2)方法 A、方法 B 的识别准确率远远优于作用于整幅图像或目标区域的 BoF 算法,这是源于 BoF 算法没有考虑空间信息,而空间金字塔模型通过空间四叉树划分法加入了粗略的空间信息;(3)有些行为类别(如演奏乐器、使用电脑)考虑整幅图像更有利于识别,而有些行为类别(如跑步、走路)通过考虑目标区域特征对识别更为有利,所以将特征单一应用整幅图像和目标区域都不是最佳选择;(4)本文通过广义性多核学习算法将整幅图像特征和目标区域特征有效地结合在一起,通过 $l1$ 正则化为利于识别行为的特征赋予较大权重,很好地消除了特征提取于整幅图像还是目标区域的选择问题,同时有效地实现了上下文信息考虑,使得所有行为类别都取得令人更为满意的识别结果。

本文算法的识别混淆矩阵如图 5 所示,通过该混淆矩阵可以总结得出:(1)走路和跑步两个行为类别容易发生混淆,跑步类别大约有 26%被误识为走路,而走路大约有 13%被误识为跑步,这源于跑步和走路活动场所和姿势有很大的相似性,所以行为差别不是很大的行为类别很容易混淆,从而导致准确地识别行为成为一个很难的研究课题,需要找到描述性强的特征和识别率高的分类器;(2)演奏乐器和照相、使用电脑也较为容易混淆,演奏乐器大约有 20%被错误地识别为照相,大约有 14%被错误识别为使用电脑,原因在于这 3 类行为都为人类手持物体具有一定程度的相似性。

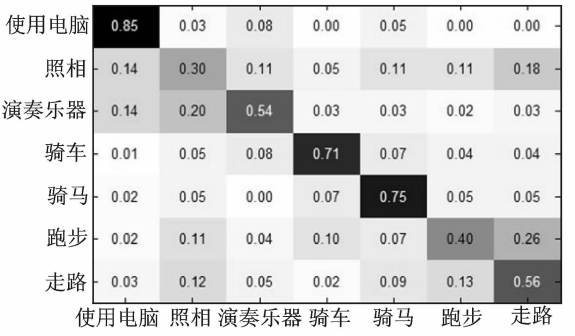


图 5 本文算法的混淆矩阵

Fig. 5 Confusion matrix obtained by method of this paper

4 结束语

为了充分利用图像信息进行静态图像行为识别,本文提出一种基于广义性多核学习的图像行为识别算法,通过实验比较可见,本文所提出的方法是一种较为有效的行为识别算法。本文的贡献可以归纳为以下几点:(1)通过同时考虑整幅图像和目标对象所在矩形框特征,实现了图像上下文信息的结合;(2)广义性多核学习算法通过 $l1$ 正则项的处理获得最佳折中特征组,实现最佳视觉相似度;(3)应用基于空间金子塔模型的 Phog 特征和稠密采样 SIFT 特征,在获取人体外观信息的同时加入了粗略的空间信息。然而,当前研究还存在一些问题,识别准确率还比较低。下一步考虑应用更多有关图像行为姿势

特征进行实验,将这些特征和广义性多核学习算法结合,期待获得一种更加准确表达图像行为的算法。

参考文献:

[1] Moeslund T B, Hilton A, Krüger V. A survey of advances in vision-based human motion capture and analysis [J]. Computer Vision and Image Understanding, 2006,104(2):90-126.

[2] Wang Y, Jiang H, Drew M S, et al. Unsupervised discovery of action classes[C]// Computer Vision and Pattern Recognition. New York,USA: IEEE Computer Society Press, 2006,2:1654-1661.

[3] Li Jiali, Li Feifei. What, where and who? Classifying events by scene and object recognition[C]// 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE Computer Society Press, 2007:1-8.

[4] Yao B, Li F F. Modeling mutual context of object and human pose in human-object interaction activities[C]// Computer Vision and Pattern Recognition (CVPR). Silvio Savarese: IEEE Computer Society Press, 2010:17-24.

[5] Yao B, Li F F. Grouplet: A structured image representation for recognizing human and object interactions[C]// Computer Vision and Pattern Recognition (CVPR). Silvio Savarese: IEEE Computer Society Press, 2010:9-16.

[6] Delaitre V, Laptev I, Sivic J. Recognizing human actions in still images: A study of bag-of-features and part-based representations[C]// BMVC 21st British Machine Vision Conference. Aberystwyth: Springer, 2010:1-11.

[7] Yang W, Wang Y, Mori G. Recognizing human actions from still images with latent poses[C]// Computer Vision and Pattern Recognition (CVPR). Silvio Savarese: IEEE Computer Society Press, 2010:2030-2037.

[8] Varma M, Ray D. Learning the discriminative power-invariance trade-off[C]// ICCV 2007. Rio de Janeiro, Brazil: IEEE Computer Society Press, 2007:1-8.

[9] Varma M, Babu B R. More generality in efficient multiple kernel learning[C]// Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada : Cambridge University Press, 2009:1065-1072.

[10] Bosch A, Zisserman A, Muoz X. Image classification using random forests and ferns[C]// 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE Computer Society Press, 2007:1-8.

[11] Lowe D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[12] 刘帅,李士进,冯钧. 多特征融合的遥感图像分类[J]. 数据采集与处理,2014,29(1):108-115.
Liu Shuai, Li Shijin, Feng Jun. Remote sensing image classification based on adaptive fushion of multiple features[J]. Journal of Data Acquisition and Processing, 2014,29(1):108-115.

[13] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[C]// 9th IEEE International Conference on Computer Vision. Nice, France: IEEE Computer Society Press , 2003:1470-1477.

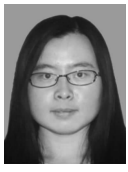
[14] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C]// Computer Vision and Pattern Recognition. New York: IEEE Computer Society Press, 2006,2:2169-2178.

[15] Chung C C, Lin C J. LIBSVM: A library for support vector machines[J]. Acm Transactions on Intelligent Systems & Technology, 2001,2(3):389-396.

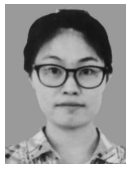
[16] Li Piji, Ma Jun. What is happening in a still picture? [C]// 2011 First Asian Conference on Pattern Recognition. Beijing, China: IEEE Computer Society Press, 2011:32-36.

作者简介:

杨红菊(1975-),女,副教授,研究方向:机器学习与计算机视觉,E-mail:yhju@sxu.edu.cn。



冯进丽(1986-),女,硕士研究生,研究方向:计算机视觉与模式识别。



郭倩(1990-),女,硕士研究生,研究方向:计算机视觉。