

基于分层分数条件随机场的行为识别

王科俊*, 吕卓纹, 孙国振, 阎涛

(哈尔滨工程大学 自动化学院, 哈尔滨 150001)

(*通信作者电子邮箱 lvlewen1988@163.com)

摘要:针对隐条件随机场(HCRF)的实时性问题和隐动态条件随机场(LDCRF)行为转换时的标记偏差问题,提出了一种基于分层分数条件随机场(SFCRF)模型的行为识别算法。该算法改进了LDCRF,并提出分数标记的概念,将人体行为的完整性和有向性具体化。实验结果表明,该算法取得了比条件随机场(CRF)、HCRF和LDCRF更好的识别效果。

关键词:隐条件随机场;隐动态条件随机场;标记偏差;行为识别;条件随机场

中图分类号: TP391.41 **文献标志码:** A

Human behavior recognition based on stratified fractal conditional random field

WANG Kejun*, LYU Zhuowen, SUN Guozhen, YAN Tao

(College of Automation, Harbin Engineering University, Harbin Heilongjiang 150001, China)

Abstract: In view of real-time issue of the Hidden Conditional Random Field (HCRF) and marked deviation problem of the Latent-Dynamic Conditional Random Field (LDCRF) during behavior transforming, this article proposed a kind of behavior recognition algorithm based on Stratified Fractal Conditional Random Field (SFCRF). The proposed algorithm improved LDCRF and put forward the concept of score mark, which made the integrity and direction of human behavior specific. The experimental results show that the proposed algorithm can obtain better recognition effect than Conditional Random Field (CRF), HCRF and LDCRF.

Key words: Hidden Conditional Random Field (HCRF); Latent-Dynamic Conditional Random Field (LDCRF); marked deviation; behavior recognition; Conditional Random Field (CRF)

0 引言

行为识别在人机交互、视觉跟踪、机器人、基于内容的视频检索等计算机视觉领域是一个非常活跃和有潜力的研究课题^[1-2]。目前研究较多的条件随机场(Conditional Random Field, CRF)模型^[3-4]实现了在整个状态空间上的全局归一化分布,将丰富的上下文特征融合到模型中,但是CRF模型不能表达序列的子结构。为了解决这一问题,文献[5]提出了隐条件随机场(Hidden CRF, HCRF)模型,实现了空间相关性的建模,但是该模型是一种脱机识别模型。为了解决HCRF模型的实时性问题,文献[6]提出了隐动态条件随机场(Latent-Dynamic CRF, LDCRF),该模型训练姿态标记间的动态特性,可以直接标记未分割序列,但LDCRF模型对于行为转换时的标记存在偏差,无法直观地显示行为的转变。

针对LDCRF标记偏差问题,本文对该模型加以改进,提出了分层分数条件随机场(Stratified Fraction CRF, SFCRF)模型,模型分为上、下两层,分别训练,共同识别,互相补充校正,将该模型应用于人体行为识别,取得了较好的识别效果。

1 分层分数条件随机场

1.1 条件随机场模型

CRF是判别式马尔可夫模型的一种,CRF能够对观测序

列的任意特征建模,能够适应重叠特征。通过最大化分类器判别权值训练CRF参数,而不训练样本的联合概率。在视频序列存在子结构的情况下,CRF不能对其建模。Wang等^[5]将隐变量加入CRF提出HCRF模型来解决子结构问题。目前,HCRF已经应用在视觉和语音领域^[7-10]。在视觉领域,HCRF已被用在目标识别中复杂图像的空间关联建模^[8],对分割的图像序列识别手臂和头部姿态^[6]。在语音识别领域,类似的模型应用在电话分类领域^[10]。HCRF采用预分割序列集训练模型,仅能获取内在结构,而不能捕获姿态标记间的动态特性。Sutton等^[11]提出了动态条件随机场(Dynamic CRF, DCRF)模型,在视频序列训练和识别过程中,需要反复修正模型的结构和参数,训练包含未观测节点的DCRF模型(隐变量)会使得该方法难以最优化。Morency等^[6]将LDCRF融合了带有显式分割的隐态变量:在训练和测试阶段采用置信传播,可以有效计算得到隐态变量的结构和参数,有效地解决HCRF未分割序列无法标记的问题。LDCRF融合隐态变量,建立姿态序列的子结构,可以训练姿态标记间的动态特性,直接应用于标记未分割序列。但是对于行为之间的转换,LDCRF无法正确标记转换过程中的行为,仅能依据概率作出标记。

图1中,给出了CRF、HCRF和LDCRF三个模型的示意图。 x_i 表示第*i*个观测(与视频序列的第*i*帧对应), h_i 是相对

收稿日期:2012-10-16;修回日期:2012-12-10。基金项目:国家863计划项目(2008AA01Z148);国家自然科学基金资助项目(60975022);博士点专项科研基金资助项目(20102304110004)。

作者简介:王科俊(1962-),男,吉林吉林人,教授,博士生导师,博士,主要研究方向:模式识别、神经网络;吕卓纹(1988-),女,吉林四平人,博士研究生,主要研究方向:模式识别、图像处理;孙国振(1985-),男,山东滕州人,硕士研究生,主要研究方向:视频监控;阎涛(1984-),男,河北衡水人,博士研究生,主要研究方向:模式识别、图像处理。

于 x_i 的一个隐态, y_i 表示 x_i 的类标记。

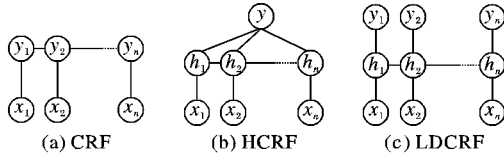


图1 CRF、HCRF与LDCRF结构示意图

1.2 分层分数条件随机场模型

针对 LDCRF 模型存在的问题,本文提出了分层分数条件随机场 (SFCRF) 模型,它改进了 LDCRF 模型,图2给出了 SFCRF 训练和识别流程。将数据库分为两部分:一部分为用整数标记的完整视频序列,另一部分为用分数标记的单个行为序列。本文提出分数标记的概念,将行为细化为几个代表性的动作,建立该行为的结构。该模型将第一层(即上层)识别得到的视频段输入第二层(即下层),进行行为识别,得到的结果是一组分数标记,进一步分析这组分数标记的方向性和完整性,验证这段视频标记是否满足一个行为的连续性,进而修正识别结果,从而提高人体行为的识别能力。

SFCRF 模型的上层识别行为的类别,下层对上层所识别的视频序列进行分数标记,并检测序列的完整性和方向性。上层类别识别先对数据库的视频序列提取轮廓特征训练模型的上层结构参数,训练结束后便可用于类别测试,将测试的输出结果作为下层的输入;从数据库中分割出单个完整行为的视频序列,训练下层结构参数,由上层输出作为测试输入,得到相应的视频段,对其进行分数标记,由所得到的标记结果,进行结果分析,进一步修正结果。测试时上层对下层的输出结果具有决定性,确定整体类别;下层对上层的输出结果进行进一步补充,使之更加完整准确。两层共同作用,有利于行为识别结果的准确、直观、连贯。

虽然模型是分层结构,但是其本身还是一体的,结构示意图如图3所示,其中 x_i 表示第 i 个观测(与视频序列的第 i 帧相对应), h_i 是相对于 x_i 的一个隐态,与 LDCRF 不同, y_i 表示 x_i 的分数类标记。

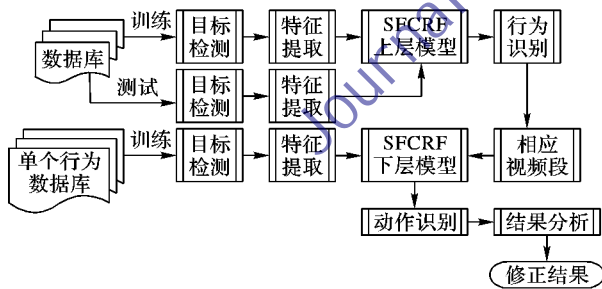


图2 SFCRF 的行为识别整体流程

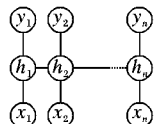


图3 SFCRF 结构示意图

2 基于分层分数条件随机场的行为建模

本文主要在联机情况下,对运动视频序列识别行为的类别,并自动检验是否符合行为的完整性和有向性。行为的完整性是指在顺序的时间序列上完成组成行为的各个动作,称之为一个行为类别。例如行走,是按照一定的速度,完成迈步前行、收腿、跨出、落地迈步等。有向性是指组成行为的动作之间的前后关系是特定的,不能随意更改其中的位置。

SFCRF 在此思想上进行建模。

观测序列 $\mathbf{X} = (x_1, x_2, \dots, x_m)$ 与标记序列 $\mathbf{Y} = (y_1, y_2, \dots, y_m)$ 间得到一组映射。每个 y_i 是视频序列中第 i 帧的一个类标记,是可能类标记集合 \mathbf{Y} 中的一个成员。每一帧观测序列 x_i 用特征向量 $\varphi(x_i) \in \mathbf{R}^d$ 表示。对每一帧序列,假设子结构变量的向量 $\mathbf{H} = (h_1, h_2, \dots, h_m)$ 。这些变量无法在训练样本中观测得到,因此会在模型中建立隐变量集。

根据以上的假设,定义一个隐条件随机模型:

$$P(y|x, \theta) = \sum_{h_i} P(y|h_i, x, \theta) P(h_i|x, \theta) \quad (1)$$

其中 θ 表示模型的参数。

为了有效地训练和学习,需要对模型加以限制,使其不含有与每个类标记关联的隐态。对于类标记 y_j , 每个 h_j 包含于所有可能隐态的集合 H_{y_j} 中。对于含有任何 $h_j \notin H_{y_j}$ 的序列, $P(y|x, \theta) = 0$, 模型表述如下:

$$P(y|x, \theta) = \sum_{h_j \in H_{y_j}} P(h_j|x, \theta) \quad (2)$$

其中 $P(h|x, \theta)$ 采用常规条件随机场^[6]:

$$P(h|x, \theta) = \exp\left(\sum_k \theta_k F_k(h, x)\right) / Z(x, \theta) \quad (3)$$

$$Z(x, \theta) = \sum_h \exp\left(\sum_k \theta_k \cdot F_k(h, x)\right) \quad (4)$$

$F_k(h, x)$ 表达式如下:

$$F_k(h, x) = \sum_{j=1}^m f_k(h_{j-1}, h_j, x, j) \quad (5)$$

其中: $Z(x, \theta)$ 为配分函数, k 表示特征数, 特征函数 $f_k(h_{j-1}, h_j, x, j)$ 为状态函数 $S_k(h_j, x, j)$ 或者转变函数 $t_k(h_{j-1}, h_j, x, j)$ 。状态函数 S_k 由模型中每个隐变量决定, 转变函数 t_k 则由隐变量对决定。

2.1 上层参数训练和行为识别

根据图2训练 SFCRF 模型。训练集分为两种类型:一种是整数标记,一种是分数标记。整数标记的训练集用于第一层的参数训练,对视频序列进行整体分类;分数标记是用于第二层的参数训练,对视频序列进行检验和校正。整数标记的训练集由 n 个标记序列 (x_i, y_i) ($i = 1, 2, \dots, n$) 组成。采用以下目标函数得到参数 θ^* ^[6]:

$$L(\theta) = \sum_{i=1}^n \ln P(y_i|x_i, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2 \quad (6)$$

其中: $\sum_{i=1}^n \ln P(y_i|x_i, \theta)$ 是训练样本的条件对数似然值, $\|\theta\| / 2\sigma^2$ 是先验方差 σ^2 的高斯对数。采用梯度下降法寻找最优参数值, 则 $\theta^* = \arg \max_{\theta} L(\theta)$ 。根据式(2)~(3), 与状态函数 S_k 关联的参数 θ_k 对应的单个训练序列 (x_i, y_i) , $\ln P(y_i|x_i, \theta)$ 的梯度为:

$$\ln P(y_i|x_i, \theta) = \sum P(h_j = a | y, x, \theta) s_k(j, a, x) - \sum P(h_j = a, y' | y, x, \theta) s_k(j, a, x) \quad (7)$$

其中梯度的边缘概率 $P(h_j = a | y, x, \theta)$ 可以由置信传播^[6]计算:

$$P(h_j = a | y, x, \theta) = \frac{\sum_{h: h_j = a \wedge h_j \in H_{y_j}} p(h|x, \theta)}{\sum_{h: h_j \in H_{y_j}} p(h|x, \theta)} \quad (8)$$

当 SFCRF 模型下层识别结果与上层识别结果出现较大偏差,即上层分类结果与其相应的结构不符,需要对比其他类

标记,通过实验发现,由式(7)得到的概率值越大,所对应的标记与检测序列结构相符程度越高,给定一段测试序列,这段序列的标记由如下公式得到:

$$y = \arg \max_y P(y | x, \theta^*) \quad (9)$$

其中 θ^* 由训练样本得到。每一类的标记都与一系列不相交的隐状态有关,式(9)还可以写成:

$$y = \arg \max_y \sum_{h: h_j \in H_{y_j}} P(h | x, \theta^*) \quad (10)$$

为了估算第 j 帧的标记 y_j , 对所有隐状态 $a \in H$ 计算边缘概率 $P(h_j = a | y, x, \theta^*)$, 用以上的方法得到的序列的标记能够使每帧误差最小。

2.2 下层参数训练和行为识别

每个行为可以分为若干连续动作,选取某几个代表性的动作代表这种行为。例如跑的行为可以由图4的几个动作关键帧代表。对运动行为序列进行标记,将跑的行为标记为1,则动作标记从右至左记为:1.1,1.2,1.3,1.4,1.5,1.6。若行为较复杂,由10个以上动作组成,可以设 i 为标记, m 为动作数,则每个动作可以标记为由 $0, 1/m, 2/m, \dots, (m-1)/m$ 加上相应标记 i 。每个训练集是由 m 个标记对序列 (x_j, y_j) , 其中 $y_j = i + (k-1)/m$ ($k = 1, 2, \dots, m$), 利用式(6)得到最优参数值 θ_i^* 。



图4 跑步中的关键序列

对于每一帧的测试序列标记为 \hat{y}_i , 并测试 \hat{y}_i 是否是递增或者递减序列, 选用最优参数值 θ_j^* , 由式(9)反复测试得到最优标记。

3 实验结果分析

实验使用 Morency 等开发的 HCRF 工具包, 它包含 CRF、HCRF、LDCRF 测试与训练的程序, 程序由三种计算机语言: VC++, Matlab 和 Python 编写而成。本文算法在奔腾双核 3.17 GHz 的 CPU, 2 GB 内存的 Windows XP 操作系统下改进了 HCRF 软件包, 用 VC++ 提取特征, Matlab 完成训练和测试过程。实验视频数据来自 Weizmann Human Action Dataset^[12], 该数据库由 9 个人的行走 (Walk)、跳 (jump) 和跑 (Run)、挥手 (Wave)、蹦 (Skip)、弯腰 (Bend) 等行为组成。本文算法加入分数标记过程, 将训练分为两部分, 模型上层完成整数标记训练, 模型下层完成分数标记训练; 对完成整数标记的测试序列利用下层模型进行分数标记。实验采用 K-折交叉^[13] 验证模型的效果, 即每次将数据库的 $K-1$ 个序列作为训练样本, 其他的作为测试样本。在不同尺寸的窗口下进行实验, $w = 0$ 表示当前帧, $w = 1$ 表示包括当前帧以及前后各一帧总共 3 帧的图像数据, 以此类推。实验中, 对比 CRF、HCRF、LDCRF 建模, 验证 SFCRF 的识别能力。

图5给出了4个模型在窗口为0时, 行走和跑的 ROC 曲线。由图可知 SFCRF 模型的识别能力优于 CRF、HCRF、LDCRF 模型。

图6对4个模型在不同尺寸窗口下的等错误率 (Equal Error Rate, EER) 识别精度进行了统计, 可以看出识别效果并不随着窗口的增大而变得更加理想, 如 LDCRF、HCRF 模型在

窗口为10的前后会发生改变, 而 SFCRF 模型在窗口较小的情况下, 识别效果优于其他模型, 随着窗口变大, 这种优势便会慢慢消失。这是由于随着序列跨度的变大, 所涉及的类别数会变多, 使整个模型的计算速度下降, 结果输出复杂, 经过实验测试 SFCRF 模型窗口在3到5个之间, 实验结果优于其他模型。

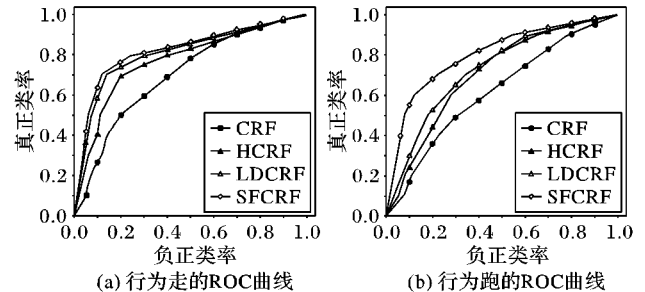


图5 行为走和跑的 ROC 曲线

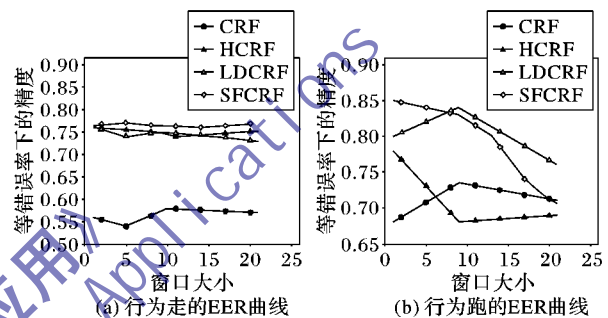


图6 行为走和跑的 EER 曲线

综合图5~6可以看出, 含有隐态变量的模型 SFCRF、LDCRF、HCRF 明显优于不含隐态的 CRF 模型。在窗口较小的情况下, SFCRF 的识别效果较好, 这是因为 SFCRF 在 LDCRF 的基础上, 增加了验证机制, 有效地减少了识别错误率。

4 结语

本文提出了基于分层分数的隐判别随机场行为模型, 获取行为外在的动态特性和隐动态特性, 使其能够同时处理视频的分段和标记工作; 引入了基于帧的隐状态变量, 对子结构行为特征进行建模, 获取行为类标记之间的动态特性。通过对比 CRF、HCRF、LDCRF 模型对分段的行为序列进行行为识别, 验证了本文提出的 SFCRF 模型具有更好的建模能力和联机识别能力。虽然这种方法具有较好的识别能力, 然而还有很多挑战性的问题没有解决, 比如, 高效、准确地提取同一行为不同角度的特征, 提高模型训练的速度等, 这些问题将会成为今后工作的重点。

参考文献:

- [1] 袁和金, 王翠茹. 人体行为识别的 Markov 随机游走半监督学习方法[J]. 计算机辅助设计与图形学学报, 2011, 23(10): 1749-1757.
- [2] 谷军霞, 丁晓青, 王生进. 行为分析算法综述[J]. 中国图象图形学报, 2009, 14(3): 377-387.
- [3] ZHAO L Y, WANG X, SUKTHANKAR G, et al. Motif discovery and feature selection for CRF-based activity recognition[C]// 20th International Conference on Pattern Recognition. Piscataway, NJ: IEEE Press, 2010: 3826-3829.
- [4] 韩磊, 李君峰, 贾云得. 基于时空单词的两人交互行为识别方法[J]. 计算机学报, 2010, 33(4): 776-784.

表6 Fisher 核的混合核与其他混合核特征提取时间 t 比较

混合核	t/s	与 Fisher 核 + 多项式核 比较时间增加/%
Fisher 核 + 多项式核	198.36	
Fisher 核 + Sigmoid 核 ^[19]	227.40	12.77
线性核 + 高斯核	254.73	22.13
多项式核 + 高斯核	311.28	36.27

4 结语

KFDA 由于其良好的分类性能在近年来模式识别的各个领域受到关注,而其中的核函数才是 KFDA 及其他分类方法中最终决定分类性能的关键。传统核函数由于参数较多造成计算时间长、占用空间资源多等缺点,许多学者做了很多研究,但因核函数本身的限制导致无法获得较好的结果。Fisher 核不同于传统核函数,它直接与样本相关且不需要参数,将它与其他核函数混合之后的参数明显少于传统的混合核函数,并且通过实验证明取得较好的效果。由于 Fisher 核可适用于不同尺寸的样本数据,本文中的图像样本数据是同尺寸的,故接下来的研究工作主要针对不同尺寸的图像样本数据,探讨应用 Fisher 核进行特征提取及分类问题。尽管图像数据较语音数据所占空间大,但随着硬件水平的不断发展,计算速度的不断提高,Fisher 核在图像领域的应用会越来越广泛。

参考文献:

- [1] VAPNIK V. An overview of statistical learning theory [J]. IEEE Transactions on Neural Networks, 1999, 10(5): 988-999.
- [2] 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.
- [3] SEWELL M. Kernel methods[R]. London: University of London, Department of Computer Science, 2007.
- [4] 李晓宇, 张新峰. 一种确定径向基核函数参数的方法[J]. 电子学报, 2005, 33(21): 2460-2463.
- [5] 颜根廷, 马广富. 一种混合核函数支持向量机算法[J]. 哈尔滨工业大学学报, 2007, 39(11): 1704-1706.
- [6] 阳春华, 王觉. 一种混合核函数 SVM 建模方法及其应用[J]. 控制工程, 2010, 17(4): 524-526.
- [7] 付丽华, 李宏伟. 尺度可调的混合核 RBF 网络[J]. 电子学报, 2011, 39(1): 184-189.
- [8] LI C H, LIN C T. An automatic method for selecting the parameter of the normalized kernel function to support vector machines [C]// Proceedings of the 2010 International Conference on Technologies and Application of Artificial Intelligence. Piscataway: IEEE Press, 2010: 226-232.
- [9] ZHAO M Y, REN JIAN, JI L P, *et al.* Parameter selection of support vector machines and genetic algorithm based on change area search [J]. Neural Computing and Applications, 2012, 21(1): 1-8.
- [10] HOU Z, YANG J, LIU G H. Selection moisture forecasting model kernel function and parameter based on support vector machine [C]// The 2011 International Symposium on Water Resource and Environmental Protection. Beijing: IEEE, 2011, 1: 717-720.
- [11] RESHMA K C, JAYADEVA, SURESH C. Optimal kernel selection in twin support vector machines [J]. Optimization Letters, 2009, 3(1): 77-88.
- [12] PERRONIN F, RODRIGUEZ-SERRANO J A. Fisher kernels for handwritten word-spotting [C]// The 10th International Conference on Document Analysis and Recognition. Beijing: IEEE, 2009: 106-110.
- [13] TRAVIESO C M, BRICENO J C, FERRER M A, *et al.* Using Fisher kernel on 2D-shape identification [C]// Computer Aided Systems Theory - EUROCAST 2007, LNCS 4739. Berlin: Springer, 2007: 740-746.
- [14] DICK U, KERSTING K. Fisher kernels for relational data [C]// Machine Learning: ECML 2006, LNCS4212. Berlin: Springer, 2006: 114-125.
- [15] RYO J, SADAOKI M. Nonparametric fisher kernel using fuzzy clustering [C]// Knowledge-Based Intelligent Information and Engineering Systems, LNCS 4252, Berlin: Springer-Verlag, 2006: 78-85.
- [16] HOLUB A D, WELLING M, PERONA P. Combining generative models and Fisher kernels for object recognition [C]// The 10th IEEE International Conference on Computer Vision. Barcelona: IEEE, 2005, 1: 136-143.
- [17] WANG C Z, YIN B C, BAI X M, *et al.* Color face recognition based on 2DPCA [C]// The 19th International Conference on Pattern Recognition. Beijing: IEEE, 2008: 1-4.
- [18] GUAN Y P. Robust video foreground segmentation and face recognition [J]. Journal of Shanghai University: English Edition, 2009, 13(4): 311-315.
- [19] 刘明, 周水生, 吴慧. 一种新的混合核支持向量机[J]. 计算机应用, 2009, 29(2): 167-169.
- [5] WANG S B, QUATTONI A. Hidden conditional random fields for gesture recognition [C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2006: 1521-1527.
- [6] MORENCY L P, QUATTONI A, DARRELL T. Latent - dynamic discriminative models for continuous gesture recognition [C]// 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2007: 1-8.
- [7] ZHANG J G, GONG S G. Action categorization with modified hidden conditional random field [J]. Pattern Recognition, 2010, 43(1): 197-203.
- [8] 李玲玲, 金泰松, 李翠华. 基于局部特征和隐条件随机场的场景分类方法[J]. 北京理工大学学报, 2012, 32(7): 720-724.
- [9] HUANG H, ZHU J. Tone modeling based on hidden conditional random field and discriminative model weight training [J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2008, 25(1): 43-49.
- [10] FUJII Y, YAMANO K, NAKAGAWA S. Hidden conditional neural fields for continuous phone speech recognition [J]. IEICE Transactions on Information and Systems, 2012, E95-D(8): 2094-2104.
- [11] SUTTON C, ROHANIMANESH K, MCCALLUM A. Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data [C]// Twenty-First International Conference on Machine Learning. New York: ACM Press, 2004: 783-790.
- [12] BLANK M, GORELICK L, SHECHTMAN E, *et al.* Actions as space-time shapes [C]// 10th IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2005: 1395-1402.
- [13] WITTEN L H, FRANK E. 数据挖掘实用机器学习技术 [M]. 董琳, 邱泉, 于晓峰, 译. 北京: 机械工业出版社, 2006: 254-238.

(上接第 959 页)