

基于多视角融合稀疏表示的恐怖视频识别

丁昕苗^{1,2}, 李 兵², 胡卫明², 郭 文¹, 王振³

(1. 山东工商学院, 山东烟台 264005; 2. 中国科学院自动化研究所, 模式识别国家重点实验室, 北京 100190; 3. 中国矿业大学, 北京 100083)

摘 要: 现有的基于多示例学习的恐怖视频识别算法都是假设示例间是相互独立的, 而忽略了恐怖视频中存在的上下文信息和示例包的统计特性. 因此, 本文提出了一种多视角融合稀疏表示模型. 该模型分别从集合视角、上下文视角以及统计特性视角三个不同的视角来看待一个视频片段, 并利用联合稀疏表示框架将三个不同视角融合到一个分类框架中, 用来进行恐怖视频的识别. 在恐怖视频库上的实验结果验证了算法在恐怖视频识别中比现有的其它算法有更好的性能和稳定性.

关键词: 恐怖视频; 稀疏表示; 多视角; 核函数

中图分类号: TP37 **文献标识码:** A **文章编号:** 0372-2112 (2014)02-0301-05

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.02.014

Horror Video Scene Recognition Based on Multi-View Joint Sparse Coding

DING Xin-miao^{1,2}, LI Bing², HU Wei-ming², GUO Wen¹, WANG Zhen-chong³

(1. Shandong Institute of Business and Technology, Yantai, Shandong 264005 China;

2. National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing 100190 China;

3. China University of Mining and Technology, Beijing 100083 China)

Abstract: Along with the ever-growing Web, horror videos sharing in the Internet has threatened children's psychological health. It is necessary to effectively recognize and filter out these horror videos. So far, several horror video recognition methods based on Multi-Instance Learning (MIL) have been proposed. However, all these methods suppose that the instances in a bag are independent, ignoring the contextual cue and statistical cue in horror videos. In this paper, we propose a novel multi-view joint sparse coding model for horror video recognition. This model considers video from three different viewpoints including set view, contextual view and statistical view. The set view treats a video as a set of independent frames. The context view models the contextual relationship among key frames in a video using an e-graph. The statistical view represents a video as a histogram feature based on bag-of-words model. Then, three kernel functions are designed for the three viewpoints, respectively. Finally, the three kernels are integrated into a unified multi-view joint sparse coding classification framework to recognize the horror video scenes based on reconstruction residual. Experiments on a horror video dataset demonstrate that our method's performance is superior to the other existing algorithms.

Key words: horror video; sparse coding; multi-view; kernel

1 引言

当前, 互联网恐怖多媒体信息正严重影响着广大青少年的身心健康^[1]. 其中, 恐怖视频由于其信息量大, 形象生动, 相比文字类或图片类的恐怖信息对青少年的危害更大, 因此开始吸引越来越多的研究者关注专门的恐怖视频识别过滤技术.

Wang 等^[2]基于情感认知理论对每个恐怖视频的所有关键帧提取了一系列有效的音视频及情感特征, 平均化后作为整个视频的情感特征, 并使用支持向量机

(Support Vector Machine, SVM)来进行恐怖视频的分类识别. 由于恐怖视频并不是所有的帧都包含恐怖信息, 提取整个恐怖视频的平均特征必将淡化恐怖视频帧的特征信息, 从而导致识别率的下降. 基于恐怖视频中一定会出现恐怖视频帧, 而非恐怖视频中不会包含任何恐怖视频帧的特点, Wang 等^[3]和 Wu 等^[4]提出采用多示例学习 (Multi-Instance Learning, MIL) 的方法来识别恐怖视频, 并取得了较好的实验效果. 然而, 在传统多示例学习框架中, 示例之间被假定为相互独立的. 这种假设只能反映恐怖视频帧之间的一个特性. 通过对大量恐怖视

收稿日期: 2012-12-26; 修回日期: 2013-11-12; 责任编辑: 马兰英

基金项目: 国家自然科学基金 (No. 60935002, No. 61100142, No. 61174007, No. 61303086); 国家 863 高技术研究发展计划 (No. 2012AA012503, No. 2012AA012504); 山东省自然科学基金 (No. ZR2012FL09, No. ZR2011FQ039, No. ZR2011FL009); 山东省教育厅高校科研计划 (No. J11LG12)

频的观察和最新研究^[5],我们可以从三种不同的视角来分析恐怖视频的特点:(1)集合视角.即传统的独立性假设.在该视角下,视频可以看作是由一系列松散独立的帧组成的集合.每一帧可以独立地表达情感;(2)上下文视角.在该视角下,视频内各帧之间存在一定的上下文关系,反映了视频的情节变化.通过这种上下文关系以及对应的情节变化来表达情感;(3)统计直方图视角.在该视角下,每个视频帧表达一种特定的中层情感语义.而一段视频中不同语义出现的频率以及语义间的共生关系,反映了视频的情感.为了提高对各种不同类型恐怖视频识别的性能,需要兼顾三种视角的不同特性.本文针对每个视角设计了特定的核函数.基于这些核函数,提出了一种基于多视角融合稀疏表示的多示例学习框架,并应用到恐怖视频的识别中.

2 基于多视角融合稀疏表示恐怖视频识别

2.1 视频包的构建

给定一个视频,首先采用文献[2]中的方法进行结构化分析和音视频及情感特征的提取,然后将一个视频看作一个包 X_i ,而该视频基于关键帧所提取的各个音视频和情感特征看作包内的示例 x_{ij} ,如此可构建一个视频包.根据多示例学习定义^[6],假设给定 N 个训练样本 $\{(X_1, y_1), \dots, (X_i, y_i), \dots, (X_N, y_N)\}$,其中 $X_i = \{x_{i1}, \dots, x_{ij}, \dots, x_{in_i}\} \subseteq \mathcal{X}$ 表示一个包(即视频段), $y_i \in \{-1, +1\}$ 为该包的标签, $x_{ij} \in \mathcal{X}$ 为该包的第 j 个示例(关键帧), n_i 为第 i 个包 X_i 所包含的示例数.如果存在示例 $x_{ig}, g \in \{1, \dots, n_i\}$ 为正示例,则该包为正包, $y_i = +1$, 否则, $y_i = -1$.

2.2 基于集合视角核函数的构建

对于恐怖视频,视频中各帧可以独立地表达恐怖情感.比如:特别血腥的帧,包含鬼怪画面的帧等等.对于每个包,我们可以将其看成一个松散的关键帧的集合.由于无法直接对集合进行特征表示和分类,于是,本文通过一个映射函数 $\varphi^1: X \rightarrow R^d$, 将集合特征投影到另一个 d_1 维特征向量空间,即: $X \rightarrow \varphi^1(X)$. 如此得到高维投影特征 $F^1 = [\varphi^1(X_1), \dots, \varphi^1(X_i), \dots, \varphi^1(X_N)]$, $\varphi^1(X_i) \in R^d$. 在该空间定义集合核函数如下:

$$k_{set}(X_i, X_j) = [\varphi^1(X_i)]^T \varphi^1(X_j) \\ = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} Ker(x_{ia}, x_{jb})}{\sum_{l=1}^{n_i} \sum_{l'=1}^{n_i} Ker(x_{il}, x_{il'}) \sum_{s=1}^{n_j} \sum_{s'=1}^{n_j} Ker(x_{js}, x_{js'})}, \quad (1) \\ Ker(x_{ia}, x_{jb}) = \exp(-\sigma_1 \|x_{ia} - x_{jb}\|^2).$$

2.3 基于上下文视角核函数的构建

上下文视角特性是指每个视频包内帧与帧之间的

一种结构关系. ϵ -graph 是描述这种上下文关系的一种比较有效的方法^[7]. 给定一个包 X_i , $W^i \in R^{n_i \times n_i}$ 定义为该包 ϵ -graph 的相似性权重矩阵. 计算 X_i 内每两个示例之间的距离,如: x_{ik} 和 x_{il} . 如果 x_{ik} 和 x_{il} 之间的距离小于一个阈值 ϵ , 则在这两个示例节点之间建立一条边,并将该边的权重 W_{kl}^i 设为 1, 否则设为 0. 基于此,建立包的示例上下文结构图 G_i , 得到训练样本的上下文视角特征 $\{(X_1, G_1, y_1), \dots, (X_i, G_i, y_i), \dots, (X_N, G_N, y_N)\}$.

同样地,图 G_i 也无法直接在特征空间进行分类,于是,借助映射函数 $\varphi^2: G \rightarrow R^{d_2}$, 将包的特征投影到一个高维空间,即: $G \rightarrow \varphi^2(G)$. 得到高维空间特征 $F^2 = [\varphi^2(G_1), \dots, \varphi^2(G_i), \dots, \varphi^2(G_N)]$, $\varphi^2(G_i) \in R^{d_2}$. 在此空间参考[7]同样定义核函数如下:

$$k_{graph}(G_i, G_j) = [\varphi^2(G_i)]^T \varphi^2(G_j) \\ = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} \omega_{ia} \omega_{jb} Ker(x_{ia}, x_{jb})}{\sum_{a=1}^{n_i} \omega_{ia} \sum_{b=1}^{n_j} \omega_{jb}}, \quad (2) \\ Ker(x_{ia}, x_{jb}) = \exp(-\sigma_2 \|x_{ia} - x_{jb}\|^2).$$

其中, $\omega_{ia} = 1 / \sum_{u=1}^{n_i} W_{au}^i$, $\omega_{jb} = 1 / \sum_{u=1}^{n_j} W_{bu}^j$. W^i 和 W^j 分别为视频包 X_i 和 X_j 的相似性权重矩阵.

2.4 基于统计直方图视角的核函数的建立

为了烘托恐怖气氛,恐怖视频相对于普通视频会经常使用低沉的颜色情感特征,快速的镜头切换,阴森的背景音乐等.因此,本文基于词包模型^[8]建立了包内的特征直方图来表示视频的统计视角特征.

给定训练样本 $\{(X_1, y_1), \dots, (X_i, y_i), \dots, (X_N, y_N)\}$, 首先,利用 k-means 聚类在所有示例空间建立一个大小为 m_v 的词典 $D = [d_1, \dots, d_i, \dots, d_{m_v}]$. 然后,将视频包 X_i 的每个示例 x_{ij} 通过映射 $\pi(x_{ij}) = \arg \min_{1 \leq k \leq m_v} \|x_{ij} - d_k\|$ 表示为词典 D 中的一个情感单词. 统计 X_i 中情感单词频率 $h(k; X_i) = |\{x_{ij} \in X_i: \pi(x_{ij}) = k\}|$, $k = 1, \dots, m_v$, $|\cdot|$ 表示集合的势,则得到 X_i 的特征直方图 $Hist_i = [h(1; X_i), \dots, h(k; X_i), \dots, h(m_v; X_i)] / \sum_{k=1}^{m_v} h(k; X_i)$, 如此,得到训练数据的统计直方图视角特征: $\{(X_1, Hist_1, y_1), \dots, (X_i, Hist_i, y_i), \dots, (X_N, Hist_N, y_N)\}$.

为了线性可分,同样将低维的直方图特征通过映射函数 $\varphi^3: Hist \rightarrow R^{d_3}$, 投影到高维的特征空间,即: $Hist \rightarrow \varphi^3(Hist)$. 投影后,得到高维空间特征 $F^3 = [\varphi^3(Hist_1), \dots, \varphi^3(Hist_i), \dots, \varphi^3(Hist_N)]$, $\varphi^3(Hist_i) \in R^{d_3}$. 在这个空间中定义基于词包模型的核函数如下:

$$\begin{aligned}
k_{hist}(Hist_i, Hist_j) &= [\varphi^3(Hist_i)]^T \varphi^3(Hist_j) \\
&= \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} h(k_1; X_i) h(k_2; X_j) G(d_{k_1}, d_{k_2}), \quad (3) \\
G(d_{k_1}, d_{k_2}) &= \exp(-\sigma_3 \|d_{k_1} - d_{k_2}\|^2).
\end{aligned}$$

2.5 基于多视角稀疏表示的核函数融合

2.5.1 多视角稀疏表示

给定测试样本 (X_{test}, y_{test}) , 通过 2.2 ~ 2.4 中叙述的处理过程, 可以得到对应的三个不同视角的特征 $f_{test}^1 = \varphi^1(X_{test})$, $f_{test}^2 = \varphi^2(G_{test})$, $f_{test}^3 = \varphi^3(Hist_{test})$.

为了能同时兼顾三个视角的特征, 本文构建了如下的多视角稀疏表示模型, 实现对测试样本的重构:

$$\min_Q \sum_{i=1}^3 \|f_{test}^i - F^i q^i\|_2^2 + \eta \|Q\|_{2,1} \quad (4)$$

其中 $Q = [q^1, q^2, q^3]$. 公式(4)的第一项为重构误差项, 约束三个视角的总重构误差要小; 第二项为重构系数约束项, 此处采用 $\ell_{2,1}$ 范数表示尽可能使用同类样本的三种视角来重构测试样本, 达到在稀疏重构时兼顾三个视角的要求, 实现了视频多视角特征的融合.

2.5.2 优化求解算法

首先对式(4)进行一些变形. 假设训练样本一共有 M 个类, 把每一种视角特征 F^i 按样本类别分组得 $[F_1^i, \dots, F_k^i, \dots, F_M^i]$, 对应的将系数 q^i 也用相同的方法分组得 $[(q_1^i)^T, \dots, (q_k^i)^T, \dots, (q_M^i)^T]^T$, 令 $Q = [q_k^i]_{k,i}$, $i = 1, 2, 3, k = 1, \dots, M$, 令 $q_k = [q_k^1, q_k^2, q_k^3]$, 则式(4)可表示为:

$$(F^1)^T F^1 = \begin{bmatrix} k_{set}(X_1, X_1) & k_{set}(X_1, X_2) & \cdots & k_{set}(X_1, X_N) \\ k_{set}(X_2, X_1) & k_{set}(X_2, X_2) & \cdots & k_{set}(X_2, X_N) \\ \cdots & \cdots & \cdots & \cdots \\ k_{set}(X_N, X_1) & k_{set}(X_N, X_2) & \cdots & k_{set}(X_N, X_N) \end{bmatrix} = K^1, \quad (F^1)^T f_{test}^1 = \begin{bmatrix} k_{set}(X_1, X_{test}) \\ k_{set}(X_2, X_{test}) \\ \cdots \\ k_{set}(X_N, X_{test}) \end{bmatrix} = H^1 \quad (9)$$

$$(F^2)^T F^2 = \begin{bmatrix} k_{graph}(G_1, G_1) & k_{graph}(G_1, G_2) & \cdots & k_{graph}(G_1, G_N) \\ k_{graph}(G_2, G_1) & k_{graph}(G_2, G_2) & \cdots & k_{graph}(G_2, G_N) \\ \cdots & \cdots & \cdots & \cdots \\ k_{graph}(G_N, G_1) & k_{graph}(G_N, G_2) & \cdots & k_{graph}(G_N, G_N) \end{bmatrix} = K^2, \quad (F^2)^T f_{test}^2 = \begin{bmatrix} k_{graph}(G_1, G_{test}) \\ k_{graph}(G_2, G_{test}) \\ \cdots \\ k_{graph}(G_N, G_{test}) \end{bmatrix} = H^2 \quad (10)$$

$$(F^3)^T F^3 = \begin{bmatrix} k_{hist}(Hist_1, Hist_1) & k_{hist}(Hist_1, Hist_2) & \cdots & k_{hist}(Hist_1, Hist_N) \\ k_{hist}(Hist_2, Hist_1) & k_{hist}(Hist_2, Hist_2) & \cdots & k_{hist}(Hist_2, Hist_N) \\ \cdots & \cdots & \cdots & \cdots \\ k_{hist}(Hist_N, Hist_1) & k_{hist}(Hist_N, Hist_2) & \cdots & k_{hist}(Hist_N, Hist_N) \end{bmatrix} = K^3, \quad (F^3)^T f_{test}^3 = \begin{bmatrix} k_{hist}(Hist_1, Hist_{test}) \\ k_{hist}(Hist_2, Hist_{test}) \\ \cdots \\ k_{hist}(Hist_N, Hist_{test}) \end{bmatrix} = H^3 \quad (11)$$

分别把式(1), (2), (3)代入式(9), (10), (11)便可得到 K^i 和 H^i , $i = 1, 2, 3$ 的值. 然后由式(7), 即可得到

$$\nabla^{i,t} = -H^i + K^i v^{i,t}, \quad i = 1, 2, 3.$$

2.5.3 分类

得到了重构系数矩阵 Q 以后, 每类训练样本对测试样本的重构误差 $r_m(X_{test})$, $m \in \{1, \dots, M\}$ 则可计算如下:

$$\min_Q \sum_{i=1}^3 \left\| f_{test}^i - \sum_{k=1}^M F_k^i q_k^i \right\|_2^2 + \eta \sum_{k=1}^M \|q_k\|_2 \quad (5)$$

下面便可采用基于 $\ell_{2,1}$ 范式的加速近似梯度 (Accelerated Proximal Gradient, APG) 算法^[9] 来求解式(5). 该算法不断地交替更新一个系数矩阵序列 $\{\hat{Q}^t = [q_k^{i,t}]\}_{t \geq 1}$ 和一个集结矩阵序列 $\{\hat{V}^t = [v_k^{i,t}]\}_{t \geq 1}$, 其中变量的上标 t 表示第 t 次迭代时对应的结果. 每次迭代包含两个步骤:

(1) 梯度映射. 给定前次迭代的集结矩阵 \hat{V}^t , 本次迭代的系数矩阵 \hat{Q}^{t+1} 则根据下面的公式进行更新:

$$\begin{aligned}
\hat{q}^{i,t+1} &= \hat{v}^{i,t} - \mu \nabla^{i,t}, \quad i = 1, 2, 3, \\
\hat{q}_k^{t+1} &= \left[1 - \frac{\eta \mu}{\|\hat{q}_k^{t+1}\|_2} \right]_+ \hat{q}_k^{t+1}, \quad k = 1, \dots, M
\end{aligned} \quad (6)$$

其中,

$$\nabla^{i,t} = -(F^i)^T f_{test}^i + (F^i)^T F^i v^{i,t} \quad (7)$$

μ 为步长, $[\cdot]_+ = \max(\cdot, 0)$.

(2) 集结矩阵更新. \hat{V}^{t+1} 可由前次迭代的系数矩阵 \hat{Q}^t 和本次迭代的系数矩阵 \hat{Q}^{t+1} 的线性组合更新如下:

$$\hat{V}^{t+1} = \hat{Q}^{t+1} \frac{\tau_{t+1}(1-\tau_t)}{\tau_t} (\hat{Q}^{t+1} - \hat{Q}^t) \quad (8)$$

其中, 序列 $\{\tau_t\}_{t \geq 1}$ 通常被设定为 $\tau_t = 2/(t+2)$ ^[9].

在此优化算法中, 最关键的一步是根据式(7)计算 $\nabla^{i,t}$ ($i = 1, 2, 3$). 由式(7)可以看出, $\nabla^{i,t}$ 由两个内积 $(F^i)^T f_{test}^i$ 和 $(F^i)^T F^i v^{i,t}$ 决定. 其可以通过核函数矩阵表示如下:

$$\begin{aligned}
r_m(X_{test}) &= \sum_{i=1}^3 \|f_{test}^i - F_m^i q_m^i\|_2^2 \\
&= \sum_{i=1}^3 ((\delta_m(q^i))^T K^i \delta_m(q^i) - 2H^i \delta_m(q^i)) \quad (12)
\end{aligned}$$

$$[\delta_m(q^i)]_l = \begin{cases} (q^i)_l, & y_l = m \\ 0, & y_l \neq m \end{cases}$$

根据稀疏表示的假设, 同类的样本可以通过本类

的其他样本线性表示.因此,测试样本的类别 c 被判定为重构误差最小的类别:

$$c = \arg \min_m (r_m(\mathbf{X}_{test})) \quad (13)$$

3 实验结果

本节通过在恐怖视频库上的对比实验验证本文算法的有效性.

3.1 数据概述及评价标准

恐怖视频库采用了文献[3]中的数据库,图1给出了部分视频的海报示例.为了验证算法的性能,本小节中将本文提出的算法与 Wang 等[3]提出的基于多示例的恐怖视频识别算法进行了比较.此外,还将 Zhou 等[7]提出的 miGraph 方法,以及 Gartner 等[10]所提出的 MI-kernel 两种多示例学习算法应用到恐怖视频识别当中,并对实验结果进行了对比.对每种方法都采用了 10 次 10 重交叉验证的平均实验结果.

为了定量对比各种算法的性能,我们采用了视频分类中广泛使用的查准率(precision(P)),查全率(recall(R))以及 $F1$ 指标(F-measure(F_1))来评价.假设给定视频集中恐怖视频集合为 HS ,算法识别得到的恐怖视频集合为 ES ,则 P, R, F_1 的计算如下:

$$P = \frac{|HS \cap ES|}{|ES|}, R = \frac{|HS \cap ES|}{|HS|}, F_1 = \frac{2 \times P \times R}{P + R}. \quad (14)$$

3.2 实验结果及分析

各种算法的实验结果如表1所示,其中括号内表示十次重复实验的标准差.实验中,三个核函数参数 $\sigma_1, \sigma_2, \sigma_3$ 的取值范围为 $\{0.01, 0.1, 0.5, 1, 5, 10\}$,多视角融合稀疏表示的正则化系数 η 的取值范围为 $\{0.001, 0.01, 0.1, 1\}$.每次实验的参数选择则通过在训练集上进行三重交叉验证获得.

表1 恐怖视频库上的实验结果(%)

算法	查准率(P)	查全率(R)	$F1$ 指标(F_1)
Multi-view-sparse	84.8(± 0.49)	84.31(± 0.38)	84.55(± 0.33)
miGraph	81.87(± 1.95)	82.4(± 1.25)	82.14(± 1.2)
MI-kernel	80.7(± 1.42)	81.43(± 0.9)	81.05(± 0.5)
MI-SVM	79.78	78.92	79.35
CKNN	78.85	70.54	74.46
EM-DD	77.59	72.97	75.21
SI-SVM	75.41	75.41	75.41

表1中 Multi-view-sparse 为本文提出的算法. mi-Graph 和 MI-kernel 则分别为 Zhou 等[7]和 Gartner 等[10]所提出的基于图和核函数的多示例学习方法. MI-SVM, CKNN, EM-DD, SI-SVM 为文献[3]中提到的四种多示例方法,其实验结果摘自文献[3].

从表1可以看出,本文所提出的算法取得了最好的实验结果.较小的方差同时表明本文提出算法的稳定性.另外,除了本文提出的算法,miGraph 的效果也要比其他几种多示例学习方法的效果略高,说明上下文视角特征在恐怖视频识别中是一个不可忽略的特征.最后,表1中前3种方法(Multi-view-sparse, miGraph, MI-kernel)与后4种方法(MI-SVM, CKNN, EM-DD, SI-SVM)的比较表明,基于核函数的多示例学习算法要比其他多示例学习算法在恐怖视频的识别中更为有效.

4 结论

现有的基于多示例学习的恐怖视频识别算法,仅从包的集合视角进行恐怖视频的识别,忽略了其他视角特征带来的判别信息,降低了识别的准确性和鲁棒性.本文中构造了视频三种不同视角特征及对应的核函数,并通过多视角融合稀疏表示的方法将多个视角融合在一起,有效的实现了恐怖视频的识别.



(a) 恐怖视频海报

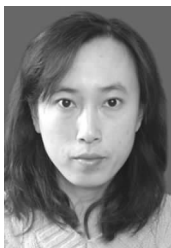
(b) 非恐怖视频海报

图1 数据库视频海报示例

参考文献

- [1] 傅丽萍, 丁芳. 恐惧信息与儿童心理发展[J]. 中国教育学报, 2004(11): 5-8.
- [2] Wang J C, Li B, Hu W M, et al. Horror movie scene recognition based on emotional perception[A]. Proceedings of IEEE International Conference on Image Processing[C]. HongKong: IEEE Computer Society, 2010. 1489-1492.
- [3] Wang J C, Li B, Hu W M, et al. Horror video scene recognition via Mutiple-Instance learning[A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Prague: Institute of Electrical and Electronics Engineers Inc., 2011. 1325-1328.
- [4] Wu B, Jiang X, Sun T, et al. A novel horror scene detection scheme on revised multiple instance learning model[A]. Proceedings of International Conference on Advances in Multimedia Modeling[C]. Taipei: Springer Verlag, 2011. 359-370.
- [5] Li B, Xiong W H, Hu W M. Web horror image recognition based on contextaware multi-instance learning[A]. Proceedings of IEEE International Conference on Data Mining[C]. Vancouver: Springer Verlag, 2011. 1158-1163.
- [6] 詹德川, 周志华. 基于流形学习的多示例回归算法[J]. 计算机学报, 2006, 29(11): 1948-1954.
- [7] Zhou Z, Sun Y, Li Y. Multi-instance learning by treating instances as non-i. i. d. samples[A]. Proceedings of International Conference on Machine Learning [C]. Montreal: Omnipress, 2009. 1249-1256.
- [8] J Winn, A Criminisi, T Minka. Object categorization by learned universal visual dictionary [A]. Proceedings of International Conference on Computer Vision[C]. Beijing: Institute of Electrical and Electronics Engineers Inc., 2005. 1800-1807.
- [9] Yuan X, Yan S. Visual Classification with Multi-Task Joint Sparse Representation[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. San Francisco: IEEE Computer Society, 2010. 3493-3500.
- [10] Gartner T, Flach P A, A Kowalczyk, Smola A J. Multi-instance kernels[A]. Proceedings of International Conference on Machine Learning[C]. Hyderabad: International Joint Conferences on Artificial Intelligence, 2002. 901-906.

作者简介



丁昕苗 女, 1979 年出生, 讲师, 研究方向为网络多媒体信息理解与识别.

E-mail: dingxinmiao@126.com



李兵 (通信作者) 男, 1983 年出生, 助理研究员, 研究方向为互联网内容安全与视觉认知计算.

E-mail: bli@nlpr.ia.ac.cn