

文章编号: 1001-9081(2013)11-3176-03

doi: 10.11772/j.issn.1001-9081.2013.11.3176

# 基于视觉显著熵与 Object Bank 特征的图像记忆性模型

陈长远<sup>\*</sup>, 韩军伟, 胡新韬, 程 填, 郭 雷

(西北工业大学 自动化学院, 西安 710029)

(\*通信作者电子邮箱 [changyuanchn@gmail.com](mailto:changyuanchn@gmail.com))

**摘要:**为了提高图像的记忆性预测能力,提出了一种基于视觉显著熵与改进的 Object Bank 特征的图像记忆性自动预测方法。该方法改进了传统的 Object Bank 特征,提取图像的视觉显著熵特征,利用支持向量回归机(SVR)训练得到图像的记忆性预测模型。实验结果表明,在预测准确性方面,所提方法比现有的方法的相关系数高出 3 个百分点。所提出的模型可以应用于图像的记忆性预测、图像检索排序、广告评价分析等方向。

**关键词:**图像处理; 图像记忆性; 视觉显著熵; Object Bank; 支持向量回归机

**中图分类号:** TN911.73    **文献标志码:**A

## Image memorability model based on visual saliency entropy and Object Bank feature

CHEN Changyuan<sup>\*</sup>, HAN Junwei, HU Xintao, CHENG Gong, GUO Lei

(School of Automation, Northwestern Polytechnical University, Xi'an Shaanxi 710029, China)

**Abstract:** To improve the prediction ability of image memorability, a method for automatically predicting the memorability of an image was proposed by using visual saliency entropy and improved Object Bank feature. The proposed method improved the traditional Object Bank feature and extracted the visual saliency entropy feature. Then a prediction model of image memorability was constructed by using Support Vector Regression (SVR). The experimental results show that the correlation coefficient of the proposed method is three percentage higher than the state-of-the-art method. The proposed model can be used in image memorability prediction, image retrieval ranking and advertisement assessment analysis.

**Key words:** image processing; image memorability; visual saliency; Object Bank; Support Vector Regression (SVR)

## 0 引言

随着计算机与数字多媒体的发展,人们在生活中会碰到各种各样的图像。在对事件的阐述以及情感的表达方面,图像有着不可比拟的优势。因此,图像逐渐成为人们生活中不可缺少的一部分。然而有些图像容易被人们记住,而另一些图像则较难被记住。研究图像的记忆性是一件很有意义的事情,它有很多的应用。例如,新闻编辑们可以用容易被记住的图像作为杂志的封面,广告商可以选择容易被记住的图像作海报等。图 1 显示了一些高记忆性与低记忆性图像。针对这一问题,Isola 等<sup>[1]~[4]</sup>首先提出了图像记忆性的问题,并对其进行了初步的研究。虽然这被认为是一个很主观的问题,但是通过 Isola 等<sup>[1]~[4]</sup>的研究表明,一幅图像能否被记住是可以预测的,是有内在规律的。然而 Isola 等<sup>[1]~[4]</sup>的研究主要是研究者对图像进行标注时人工统计什么因素、什么特征会影响图像的可记性,虽然该方法提取了一些特征并试图去自动预测图像能被记住的程度,但是效果并不理想。主要原因是他们所用的特征是图像的全局特征,如尺度不变特征转换 (Scale-Invariant Feature Transform, SIFT)<sup>[2]</sup>、方向梯度直方图 (Histogram of Oriented Gradient, HOG)<sup>[3]</sup>、SSM (Self-Similarity Measures)<sup>[4]</sup>、GIST<sup>[5]</sup>等,这些特征并不能很好地预测图像的记忆性质。本文通过对图像记忆性这一问题的深入研究,提出了一种基于视觉显著熵与改进的 Object Bank 特征的图像

记忆性预测模型。通过与现有方法的对比,可以看出本文的方法明显优于现有的基于全局特征的方法;此外,通过对图像记忆性的分类研究,表明本文方法能够很好地判别一幅图像能否被人们所记住。

图 1 中,图片下面的数字是每幅图片对应的记忆性数值。该数值越高,代表图像越能被记住。

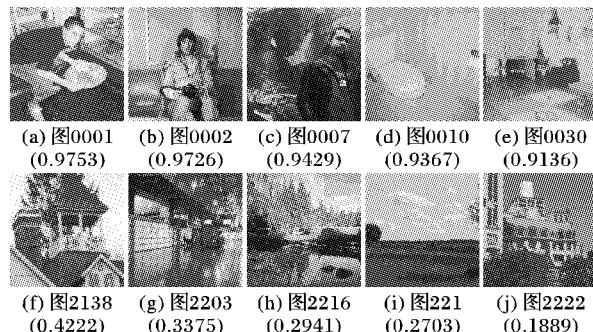


图 1 高记忆性图像与低记忆性图像示例

## 1 视觉显著熵

视觉显著性(Visual Saliency)是指一幅图像容易引起人类视觉注意的区域。通过对图像记忆性的研究,本文认为图像的视觉显著性会对一幅图像的记忆性产生影响。当一幅图像的显著性区域较分散时,图像的记忆性会较低,因此本文设

收稿日期: 2013-06-03; 修回日期: 2013-07-23。

基金项目: 国家自然科学基金资助项目(61005018); 西北工业大学基础研究基金资助项目(JC20120237)。

作者简介: 陈长远(1988-),男,山东日照人,硕士研究生,主要研究方向: 图像处理; 韩军伟(1977-),男,陕西西安人,教授,博士生导师,主要研究方向: 计算机视觉、图像与视频处理、模式识别、多媒体信息检索; 胡新韬(1980-),男,湖北随州人,副教授,主要研究方向: 脑图像处理; 程填(1984-),男,河南项城人,博士研究生,主要研究方向: 模式识别、图像处理; 郭雷(1956-),男,陕西西安人,教授,博士生导师,主要研究方向: 图像处理。

计了视觉显著熵的特征来预测一幅图像的记忆性。

对于图像视觉显著性的问题有很多的研究,如 Itti 等<sup>[6]</sup>、Judd 等<sup>[7]</sup>。本文采用 Judd 等<sup>[7]109~2112</sup>的方法来计算一幅图像的视觉显著性。具体地,先对每幅图像依次提取其底层特征(如颜色、纹理等)、中层特征(如直线等)和高层特征(如人、车等);然后对这些特征进行融合,每幅图像可得到一个 33 维的特征表示;最后将该 33 维的图像特征输入支持向量机(Support Vector Machine, SVM)分类器,便可以得到目标图像的视觉显著图。

在获得每幅图像的显著图之后,将显著图二值化,便可得到图像的显著性区域。受 Wang 等<sup>[8]</sup>启发,本文拟采用式(1)~(3)计算一幅图像的视觉显著熵。具体地,假设显著性区域的中心点的位置是  $y_i$ ,高斯分布  $G(y_i, \Sigma_i)$  用来近似估计显著区域在整个图像的位置,其中,协方差矩阵  $\Sigma_i$  由显著区域的大小和位置决定。视觉显著图的显著区域可以用下面的一个等权重混合高斯模型表示:

$$p(y) = \frac{1}{N} \sum_{i=1}^N G(y - y_i, \Sigma_i) \quad (1)$$

而显著性熵可以用式(2)计算得到:

$$H = -\log \int p(y)^2 dy \quad (2)$$

这样,将式(1)代入式(2),可以得到最终的视觉显著熵计算公式,即:

$$\begin{aligned} H = & -\log \left\{ \frac{1}{N^2} \times \int \left( \sum_{i=1}^N \sum_{j=1}^N G(y - y_i, \Sigma_i) G(y - y_j, \Sigma_j) \right) \right\} = \\ & -\log \left\{ \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(y_i - y_j, \Sigma_i + \Sigma_j) \right\} \end{aligned} \quad (3)$$

通过对式(3)的分析可以看出,当图像的显著性熵越大,图像的显著性区域就越分散,图像中吸引人们视觉注意的点就越多,这样就越会降低图像的记忆性。图 2 显示了 4 幅不同图像的视觉显著图。其中第一行为原始图像,第二行为自动预测得到的视觉显著图,第三行为二值化的视觉显著图。图 2 中 4 幅图像的视觉显著熵分别为 4.1607, 3.8947, 4.8787, 4.9148。

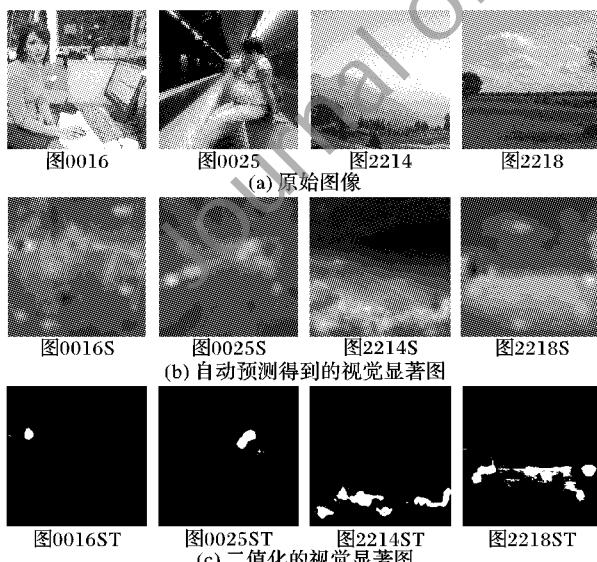


图 2 4 幅不同图像的视觉显著熵比较

## 2 改进的 Object Bank 特征

由 Isola 等<sup>[1]146</sup>的研究看出,图像中包含的物体对图像的记忆性有较大影响。因此,如果本文能够对图像中出现的物

体有一个可靠的表示,则对图像的记忆性就会有一个较好的判断。在 Belongie 等<sup>[9]</sup>、Felzenszwalb 等<sup>[10]</sup>的基础上, Li 等<sup>[11]1378~1380</sup>提出了 Object Bank 特征,并将其应用于场景分类问题,取得了不错的效果。

### 2.1 Object Bank 模型结构

Object Bank 是一种用于场景分类以及语义信息提取<sup>[11~12]</sup>的高级图像表示特征。它是由 208 个可变形模型组成的物体检测器阵列,阵列中的每一个可变形模型(检测器)对应着一类物体,如图 3 所示。这里所述的物体是指广义上的物体,共包含两大类:1)汽车、行人等具有规则外形结构的物体;2)天空、水体等不具有规则形状的物体。每个可变形模型由若干个子模型组成,而每个子模型又由两层滤波器组成:根滤波器和部件滤波器。其中:根滤波器的作用是捕获目标的整体轮廓特征,而部件滤波器能够捕获目标某个具有明显判别作用的局部特征。

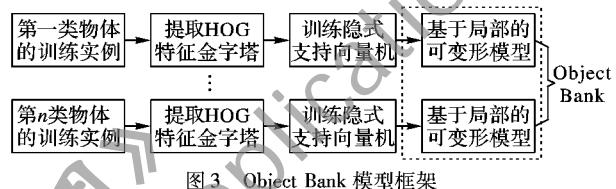


图 3 Object Bank 模型框架

### 2.2 Object Bank 模型训练

Object Bank 模型中每类物体的可变形模型的训练过程如下。

1) 标注正负训练样本。正样本只需标注包含物体的矩形框。

2) 提取训练样本的 HOG 特征金字塔。首先对训练样本通过反复的平滑操作和下采样得到  $N$  层图像金字塔,然后在每一层图像上分别计算其对应的 HOG 特征<sup>[3]887</sup>。

3) 隐式 SVM 训练。令训练样本集合为  $D = (\langle x_1, y_1 \rangle, \dots, \langle x_n, y_n \rangle)$ , 其中:  $x_i$  为正负训练样本,  $y_i \in \{-1, 1\}$ , 优化式(4)所示的目标函数:

$$\begin{cases} \boldsymbol{\beta}^*(D) = \arg \min \lambda \|\boldsymbol{\beta}\|^2 + \sum_{i=1}^n \max[0, 1 - y_i f_{\boldsymbol{\beta}}(\mathbf{x}_i)] \\ f_{\boldsymbol{\beta}}(\mathbf{x}_i) = \max_{z \in Z(\mathbf{x}_i)} \boldsymbol{\beta} \cdot \Phi(H(\mathbf{x}_i), z) \end{cases} \quad (4)$$

其中:  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_i, \dots, \boldsymbol{\beta}_m)$  为可变形模型的参数向量,  $\boldsymbol{\beta}_i$  为每个子模型的参数向量,  $H(\mathbf{x}_i)$  为样本  $\mathbf{x}_i$  的 HOG 特征金字塔,  $z$  为隐式变量,  $\Phi(H(\mathbf{x}_i), z)$  代表由隐式变量  $z$  指定的一个检测窗口,  $\boldsymbol{\beta} \cdot \Phi(H(\mathbf{x}_i), z)$  表示可变形模型在此窗口上的响应,  $\max_{z \in Z(\mathbf{x}_i)} \boldsymbol{\beta} \cdot \Phi(H(\mathbf{x}_i), z)$  代表模型在此训练样本上的最高分(对应窗口为最佳检测)。一旦隐式变量确定,也就是说每个训练样本的检测窗口确定,隐式 SVM 也就转化为普通的 SVM 优化问题。

按照上述的训练过程,分别对 208 类物体进行训练,便可得到由 208 个可变形模型组成的物体检测器阵列,即 Object Bank。

### 2.3 改进的 Object Bank 特征提取与表示

在 Li 等<sup>[11]1382</sup>的基础上,本文提出了一种改进的 Object Bank 特征提取和表示方法,用于对图像中的物体进行表示。具体地,对每类物体对应的可变形模型,本文首先计算其 12 个不同尺度的响应图,用  $responseMap(p, l)$  表示,其中,  $l$  表示尺度,本文中  $l$  取值为 1 到 12 的整数;  $p$  代表模板,取值为 1

到 208 的整数。然后,从得到的 12 个尺度上利用式(5)计算每类物体对应的最大响应图;最后,利用式(6)计算 208 个最大响应图的平均值,得到改进的 Object Bank 特征(在这里,本文仍称之为 Object Bank 特征):

$$MP(p) = \max_l(responseMap(p, l)) \quad (5)$$

$$feature = \text{aver}(MP(p)) \quad (6)$$

其中: $MP(p)$  表示每类物体对应的可变形模型在 12 个不同尺度上的最大响应; $feature$  表示改进的 Object Bank 特征,即 208 个 $MP(p)$  的平均值。图 4 给出了 Object Bank 特征的提取与表示示意图。

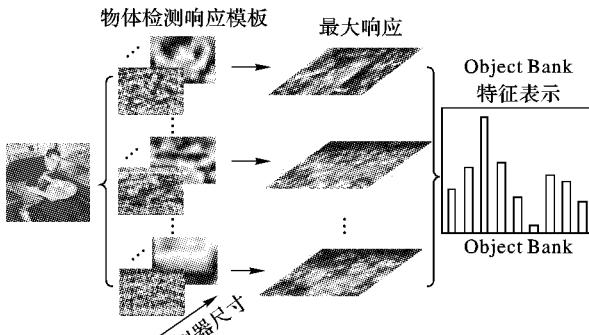


图 4 Object Bank 特征的示例图

### 3 记忆性预测模型

将视觉显著熵与改进的 Object Bank 特征相结合,便可以构建图像的记忆性预测模型,如图 5 所示,该模型的具体计算步骤如下:

- 1) 用 Judd 等<sup>[7]2109-2112</sup>的方法得到图像的视觉显著图;
- 2) 取阈值  $\alpha$  将视觉显著图二值化;
- 3) 用式(3)计算二值图像的显著性熵;
- 4) 用 Object Bank 的 208 个可变形模型与输入图像卷积,每幅图像得到  $12 \times 208$  幅响应图像;
- 5) 用式(5)求取每个可变形模型的最大响应图,然后用式(6)对最大响应图取平均,每幅图像可以得到一个 208 维的改进 Object Bank 特征;
- 6) 将视觉显著熵与改进的 Object Bank 特征进行串联融合,得到图像的特征表示;
- 7) 利用步骤 6) 中得到的图像特征表示以及图像的记忆真值,训练支持向量回归机(Support Vector Regression, SVR),得到图像的记忆性预测模型。

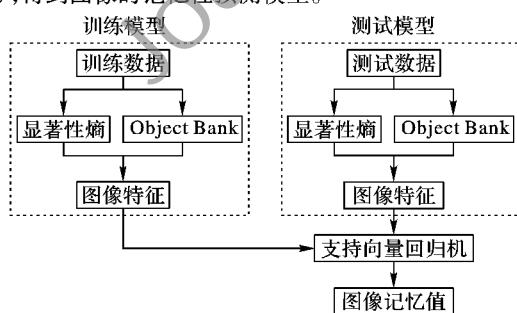


图 5 本文算法流程

## 4 实验结果

### 4.1 实验数据

本文所用的实验数据是 Isola 等<sup>[1]146</sup>的数据,数据库中共有 2222 幅图像以及每幅图像对应的记忆性真值。

### 4.2 实验设置

为了与现有方法相对比,本文与 Isola 等<sup>[1]150</sup>在相同条件下进行测试。即每次随机地把图像分为相同数量的 2 组(每组 1111 幅图像),一组用于训练,一组用于测试,应用 SVR<sup>[13]</sup>做 25 次回归分析实验。图像的记忆性数值的预测算法流程如图 5 所示。计算 25 次实验的预测值与实验真值的平均相关性系数  $\rho$ ,然后按照预测值排列图像,并计算这种排列的图像的平均记忆真值。

### 4.3 实验对比

表 1 显示了 Isola 等<sup>[1]150</sup>提出的方法以及本文方法的实验结果。由表 1 可以看出,本文提出的融合特征,即视觉显著熵与改进的 Object Bank 特征的融合与图像可记性的相关性数值可以达到 0.49,比现有的基于全局特征的方法能更好地描述图像记忆性这一问题。

表 1 基于全局特征的图像记忆性判别方法与本文方法的比较

特征方法	平均记忆真值/%				
	记忆预测值最高的 20 个图像	记忆预测值最高的 100 个图像	记忆预测值最低的 100 个图像	记忆预测值最低的 20 个图像	相关性系数
像素	74	72	61	59	0.22
GIST	82	78	58	57	0.38
SIFT	83	79	57	56	0.41
SSM	83	79	58	56	0.43
HOG2 × 2	83	80	57	55	0.43
所有全局特征	83	80	56	54	0.46
本文方法	85	81	54	53	0.49

### 4.4 分类实验

为了进一步说明本文方法在图像的记忆性问题上的判别能力,本文设计了一个分类实验。取原始数据记忆真值最高的 65 个图像及其对应的记忆性数值以及记忆真值最低的 65 个图像以及其对应的记忆性数值,实验数据共 130 幅图像,由于实验数据较少,故采用留一法进行实验,用 LIBSVM<sup>[13]</sup>进行分类。结果分类准确率为 96%。由此可见,本文方法对高记忆性图像与低记忆性图像有很好的分类判别效果。

### 4.5 实验分析

由实验结果可以看出本文的方法明显要优于 Isola 等<sup>[1]150</sup>的方法,并且通过分类实验可以看出,本文方法可以很有效地分开高记忆性与低记忆性的图像。Isola 等<sup>[1]150</sup>的方法效果不理想的原因在于其所采用的图像全局特征并不能很好地描述图像的记忆性。而由 Isola 等<sup>[1]146-149</sup>的一系列结论得知,图像中包含什么样的物体会对图像的记忆性有较大的影响,而 Object Bank 特征就是在对图像中的物体进行表示。然而由于现实世界所处的环境中有成千上万的物体,而 Object Bank 方法所用的模型仅有 208 个,因此在表示图像上必然会有较大误差,所以单独用 Object Bank 特征来表示图像的记忆性效果并不理想。视觉显著性是指图像的显著性区域,即人看图像时,视觉注意的位置。通过研究,本文得到,如果一幅图像并不能吸引人的视觉注意到某些集中区域时,这幅图像的记忆性会较低。为了对这一问题进行度量,本文引入了显著性熵的概念,用来表示一幅图像对观察者目光的吸引程度。显著性熵越小,表明人类视觉注意力越集中在某一区域。为了能让计算机自动判别图像的记忆性,本文应用了 Judd 等<sup>[7]2109-2112</sup>的方法来自动预测一幅图像的视觉显著图,然后用本文提出的方法计算一幅图像的显著性熵值。实验证明,本文方法能够取得很好的效果。

(下转第 3223 页)

响墨水扩散的两个主要因素:宣纸结构和墨水属性。给出了模拟宣纸结构的加权纤维结构,并把纤维权重和纸上剩余墨水量相结合作为扩散系数,提出了描述墨水扩散的变系数扩散方程。为了提高运行效率,本文采用预先生成宣纸纤维结构的策略,而在模拟过程中只需求解扩散方程就可获得扩散图像。用户可以通过调节初始墨水量、纤维密度和纤维权重等参数获得不同的墨水扩散效果。实验结果表明该模型能较好地模拟不同宣纸上的多种墨水扩散效果。但该方法不能真实地模拟扩散边缘的灰度变化,主要原因有两个:一是扩散图像是通过求解扩散方程得到,而本文的求解方法只有 2 阶精度;二是由于墨水扩散是一个复杂的物理过程,涉及墨水性质、宣纸结构、宣纸吸水性、液体扩散过程等诸多因素,而本文方法只考虑了墨水量和单层宣纸结构。因此未来的研究重点是提出高精度的扩散方程的数值解法和考虑宣纸空间结构的扩散方程,从而更真实地模拟各种扩散效果。

#### 参考文献:

- [1] GUO Q, KUNII T L. Modeling the diffuse paintings of Sumie [M]. Berlin: Springer, 1991.
- [2] LEE J. Diffusion rendering of black ink paintings using new paper and ink models [J]. Computers and Graphics, 2001, 25(2): 295 – 308.
- [3] 余斌, 孙济洲, 白海飞, 等. 基于纸的物理建模的水墨画扩散效果仿真[J]. 系统仿真学报, 2005, 17(9): 2305 – 2309.
- [4] WANG X, JIAO J, SUN J. Graphical simulator for Chinese ink-wash drawing [J]. Transactions of Tianjin University, 2002, 8(1): 1 – 7.
- [5] SMALL D. Simulating watercolor by modeling diffusion, pigment, and paper fibers [C]// Proceedings of SPIE 1991: Image Handling and Reproduction Systems Integration. San Jose: International Society for Optics and Photonics, 1991: 140 – 146.
- [6] CURTIS C J, ANDERSON S E, SEIMS J E, et al. Computer-generated watercolor [C]// Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1997: 421 – 430.
- [7] KUNII T L, NOSOVSKIJ G V, HAYASHI T. A diffusion model for computer animation of diffuse ink painting [C]// CA'95: Proceedings of the Computer Animation. Washington, DC: IEEE Computer Society, 1995: 98 – 102.
- [8] KUNII T L, NOSOVSKIJ G V, VECHERININ V L. Two-dimensional diffusion model for diffuse ink painting [J]. International Journal of Shape Modeling, 2001, 7(1): 45 – 58.
- [9] WAY D L, HUANG S W, SHIH Z C. Physical-based model of ink diffusion in Chinese paintings [EB/OL]. [2013-04-25]. <http://wenku.baidu.com/view/681ea52c0066f5335a8121cc.html>.
- [10] CHU N S H, TAI C L. MoXi: real-time ink dispersion in absorbent paper [J]. ACM Transactions on Graphics, 2005, 24(3): 504 – 511.
- [11] SUCCI S. Lattice Boltzmann equation [M]. Oxford: Oxford University Press, 2001.
- [12] CHU N S H, TAI C L. Real-time painting with an expressive virtual Chinese brush [J]. Computer Graphics and Applications, 2004, 24(5): 76 – 85.
- [13] CHA S, PARK J, HWANG J, et al. An efficient diffusion model for viscous fingering [J]. The Visual Computer, 2012, 28(6/7/8): 563 – 571.
- [14] WANG C M, WANG R J. Image-based color ink diffusion rendering [J]. IEEE Transactions on Visualization and Computer Graphics, 2007, 13(2): 235 – 246.
- [15] 陆金甫, 关治. 偏微分方程数值解法[M]. 北京: 清华大学出版社, 2004.

(上接第 3178 页)

## 5 结语

本文通过对图像记忆性特性的研究,提出了一种基于视觉显著熵与改进的 Object Bank 特征的记忆性预测方法,能自动地预测图像的记忆性值。与现有方法进行实验对比,表明了本文所提方法的有效性。

#### 参考文献:

- [1] ISOLA P, XIAO J, TORRALBA A, et al. What makes an image memorable [C]// Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2011: 145 – 152.
- [2] LOWE D. Distinctive image features from scale-invariant keypoint [J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [3] DALIL J, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2005: 886 – 893.
- [4] SHECHTMAN E, MICHAL I. Matching local self-similarities across images and videos [C]// Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2007: 1 – 8.
- [5] OLIVA A, TORRALBA A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. International Journal of Computer Vision, 2001, 42(3): 145 – 175.
- [6] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254 – 1259.
- [7] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look [C]// Proceedings of 2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 2106 – 2113.
- [8] WANG R, MCKENNA J, HAN J, et al. Visualizing image collections using high-entropy layout distributions [J]. IEEE Transactions on Multimedia, 2010, 12(8): 803 – 813.
- [9] BELONGIE S, MALIK J, PUZICHA J. Shape matching and object recognition using shape contexts [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509 – 522.
- [10] FELZENZWALB P, GIRSHICK R, MCALLESTER D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627 – 1645.
- [11] LI L, SU H, XING E, et al. Object Bank: a high-level image representation for scene classification and semantic feature sparsification [C]// Proceedings of the Twenty-Fourth Annual Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc, 2010: 1378 – 1386.
- [12] LI L, HAO S, LIM Y, et al. Objects as attributes for scene classification [C]// Proceedings of the 11th European Conference on Computer Vision. Berlin: Springer-Verlag, 2010: 57 – 69.
- [13] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27.