一种新的基于语义聚类和图算法的自动图像标注方法

芮晓光 袁平波 何 芳 俞能海
(中国科学技术大学多媒体计算与通信教育部-微软重点实验室,合肥 230027)
(中国科学技术大学电子工程与信息科学系,合肥 230027)

摘 要 针对图像检索中的语义鸿沟问题,提出了一种新颖的自动图像标注方法。该方法首先采用了一种基于软约束的半监督图像聚类算法(SHMRF-Kmeans)对已标注图像的区域进行语义聚类,这种聚类方法可以同时考虑图像的视觉信息和语义信息。并利用图算法——Manifold 排序学习算法充分发掘语义概念与区域聚类中心的关系,得到两者的联合概率关系表。然后利用此概率关系表标注未知标注的图像。该方法与以前的方法相比可以更加充分地结合图像的视觉特征和高层语义。通过在通用图像集上的实验结果表明,本文提出的自动图像标注方法是有效的。

关键词 半监督聚类 软约束 图像标注 Manifold 排序学习算法
中图法分类号: TP75 **文献标识码**: A **文章编号**: 1006-8961(2007)02-0239-06

A New Approach for Automatic Image Annotation Based on Semantic Clustering and Graph Algorithm

RUI Xiao-guang, YUAN Ping-bo, HE Fang, YU Neng-hai
(*MOE-MS Key Laboratory of Multimedia Computing and Communication, University of Science and Technology of China, Hefei* 230027)
(*Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei* 230027)

Abstract A novel automatic image annotation approach is proposed to bridge the semantic gap of content-based image retrieval. Our approach first performs segmentation of images into regions, followed by the clustering of regions to blobs using a semi-supervised image clustering algorithm with soft constraints which utilizing the visual and semantic information of images. And a graph-based algorithm is used to compute the probabilistic relation between concepts and region blobs which can be used to annotate new images. Experiments conducted on standard dataset demonstrate the effectiveness and efficiency of the proposed approach for image annotation.

Keywords semi-supervised clustering, soft constraints, image annotation, Manifold ranking

1 引 言

自动图像标注是基于内容图像检索中重要而具有挑战性的工作。它可以利用已标注的图像集自动学习语义概念空间与视觉特征空间的关系模型,并用此模型标注未知语义的图像,即它试图在高层语义特征和底层视觉特征之间建立一座桥梁。因此,它可以一定程度解决大多基于内容图像检索方法存在的语义鸿沟问题。如果能实现自动图像标注,那么现有的图像检索问题实际上就可以转化成技术已经相当成熟的文本检索问题。它的潜在应用领域包括生物医学、商业、军事、教育、数字图书馆和互联网检索等。

基金项目:多媒体计算与通信教育部-微软重点实验室开放基金资助项目(05071804);国家自然科学基金项目(60672056)
收稿日期:2006-10-13;**改回日期**:2006-11-06

第一作者简介:芮晓光(1983~),男。中国科学技术大学电子工程与信息科学系硕士研究生。主要研究方向为多媒体信息检索、模式识别等。E-mail:davidrui@mail.ustc.edu.cn

基于内容的图像检索近 10 年来得到了研究者的关注,一系列的基于内容的图像检索方法和检索系统被提出来[1]。然而图像底层的视觉特征和高层语义存在不一致性,即所谓的“语义鸿沟”。自动图像标注可以缓解这一问题,所以得到了研究者的极大关注。

自动图像标注的一个方向是采用分类方法,每一个语义概念被当作一个类别进行分类。代表方法有:SVM 方法[2],贝叶斯点机方法[3]等等。这种方法当语义概念相当多时会遇到困难。

自动图像标注的另一个方向是建立图像和语义概念的统计概率模型。Duygulu 等人提出的翻译模型(TM: translation modal)[4],利用传统的语言统计翻译模型将语义概念翻译为由图像区域聚类产生的 blobs。Jeon 等人介绍了一种交叉媒体相关模型(CMRM: cross-media relevance modal)[5],将图像标注问题看作跨语言检索问题,模型通过计算 blobs 和语义概念的联合概率进行图像标注,获得了比较好的效果。但是这类概率的方法对语义和图像特征的利用比较粗糙,两者的结合不是很紧密。而且这类方法对图像区域聚类结果的好坏比较敏感。

本文提出一种新颖的自动图像标注框架,并在此基础上构建了图像检索系统。其主要思想是:首先将已标注图像集分割为图像区域,再利用提出的基于软约束的半监督图像聚类算法(SHMRF-Kmeans: soft hidden markov random field kmeans)对图像区域进行语义聚类,实现图像区域的简化表示,即图像集在视觉特征空间中的量化表示,每个子类称为 blob[6]。然后结合概率模型和 Manifold 排序学习算法[7]建立语义概念和 blobs 之间的概率关系。当有未标注的图像时,通过判断其区域所属的 blob 即可利用此概率关系进行自动标注。

2 自动图像标注框架

2.1 框架结构

自动图像标注的框架如图 1 所示,其主要基于两种技术:语义聚类(基于软约束的半监督图像聚类算法——SHMRF-Kmeans 算法)和图算法(Manifold 排序学习算法)。

基于约束的半监督聚类是在加有数据间约束的数据集上实现聚类的方法。约束的种类主要有:正关联约束(must-link constraint),它表示两数据点必须归于同一类别;负关联约束(cannot-link constraint),它表示两数据点不能归于同一类别。软约束相对于硬约束来说,它是在硬约束基础上加上了强度因子,表示约束的强度。比起硬约束,软约束可以更好地表达数据间的约束关系。

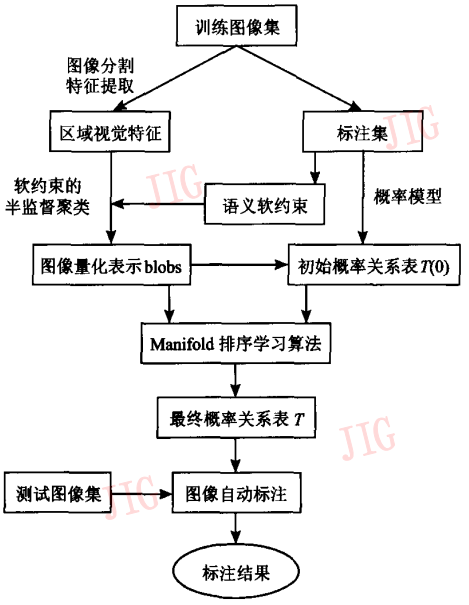

图 1 自动图像标注框架

Fig. 1 Automatic image annotation framework

本文提出的基于软约束的半监督图像聚类算法 SHMRF-Kmeans,其基本思想是利用隐马尔科夫随机场模型结合基于约束和基于距离测度函数学习的方法实现了在数据软约束条件下的聚类。算法首先利用隐马尔科夫随机场模型的先验能量建立聚类的目标函数,然后利用期望最大(EM: expectation maximization)算法计算目标函数的最小值,实现聚类。由于软约束是根据已知的图像标注确定的,所以 SHMRF-Kmeans 算法得到的图像区域聚类结果同时利用了图像的视觉信息和语义信息,改善了图像区域聚类效果,为基于区域聚类的图像标注方法奠定了良好的基础。本文算法是对 Basu 等人[8]工作的延伸。

Manifold 排序学习算法是一种基于图的算法,起初是用于发掘数据的特性和内在维数的方法。本文中由于一般的概率模型无法充分利用 blob 之间和 blob 与概念间的深层关系,所以提出利用概率模型先计算出 blob 与概念间的初始概率关系表 $T(0)$,然后利用 Manifold 排序学习算法充分发掘数据间的关系

得到最终的概率关系表。

2.2　SHMRF-Kmeans 的描述

2.2.1　软约束的描述

软约束可以表示为(A,B,S)。当$S>0$时,表示A与B存在正关联约束,强度因子为S,表明A,B属于同一类别的概率为S;同理,当$S<0$时,表示A与B存在负关联约束,强度因子为$-S$,表明A,B不属于同一类别的概率为$-S$。

软约束的强度因子可以采用下式计算:

$$S=\begin{cases}-d & \text{如果 } up\_threshold<D<1\\ \exp\left\{\dfrac{-d^2}{2\sigma^2}\right\} & \text{如果 } 0<D<down\_threshold\end{cases} \tag{1}$$

其中,d为图像区域A,B标注的语义距离,σ为调节因子。这里定义了上下两个阈值,如果D在阈值范围之外,则会形成一个软约束。LIN 方法[9]可以计算词与词之间的语义距离,它采用了信息熵理论和 WordNet 词典可以计算出两个单词间的语义距离。本文利用 LIN 方法首先计算出单词间的语义距离,进而计算出图像区域对应标注之间的语义距离。

2.2.2　目标函数

聚类的目标函数[10]可以表示为

$$J_{\text{obj}}=\sum_{x_i\in X}D(x_i,\mu_{l_i})+\sum_{(x_i,x_j)\in M}w_{ij}D(x_i,x_j)I(l_i\neq l_j)-\sum_{(x_i,x_j)\in C}\bar{w}_{ij}D(x_i,x_j)I(l_i=l_j) \tag{2}$$

其中,M表示正关联约束集合,C表示负关联约束集合。$w_{ij}/\bar{w}_{ij}$是在X_i和X_j之间违反正关联/负关联约束的代价权重,l_i是X_i所属于的类标记。μ_{l_i}是类l_i的代表元素。I是指示函数($I(\text{true})=1$,$I(\text{false})=0$)。D为距离测度。可以看出该目标函数中第 1 项是标准的K均值目标函数;第 2 项是违背正关联约束的惩罚函数;而第 3 项是违背负关联约束的惩罚函数。

但是,此目标函数不能处理软约束。这里采用了两种策略来修改目标函数:

(1) 指示函数法

在原始的目标函数中,指示函数$I(a)$是一个二值函数。为了处理软约束,指示函数被修改成一个范围从 0 到 1 的连续函数$I'(a)$。当a的值是真时,连续指示函数$I'(a)$的值与强度因子S的绝对值相等,即$I'(\text{true})=S$,$I'(\text{false})=0$。因此在修改过的目标函数中,惩罚值与强度因子S的绝对值是成比例的。

(2) MS(maximum strength)法

$MS(x_i,\mu_h)$表示当x_i在类h中时的最大惩罚值。

$$MS(x_i,\mu_h)=\frac{n_{\text{Viol}}}{n_{\text{Const}}}\times\max_{\text{violate constraints}}|S| \tag{3}$$

其中,n_{Const}是$|S|$获得最大值的次数,n_{Viol}是违背约束时$|S|$达到最大值的次数。

通过以下方式新的目标函数将 MS 与旧的目标函数结合起来:

$$J_{\text{obj}}^{MS}(x_i,\mu_h)=\frac{J_{\text{obj}}(x_i,\mu_h)}{1-MS(x_i,\mu_h)} \tag{4}$$

当MS的值接近 1 时,目标函数的值将会很大。在这种情况下,x_i将不会归入类h。因此,新的目标函数能表示软约束的强弱。

最后,将这两种策略结合起来,获得可以处理软约束的新目标函数:

$$J_{\text{obj}}^{SOFT}(x_i,\mu_h)=\left[D(x_i,\mu_h)+\sum_{S>0}w_{ij}D(x_i,x_j)I'(l_i\neq l_j)-\sum_{S<0}\bar{w}_{ij}D(x_i,x_j)I'(l_i=l_j)\right]\Big/(1-MS(x_i,\mu_h)) \tag{5}$$

2.2.3　算法步骤

SHMRF-Kmeans 算法目标是使总的惩罚值最小。这一目标是通过类似 EM 算法的方法使目标函数值到达局部最小实现的。算法的基本思想是:在 E 步,给定当前类的重心,每个数据点重新分配给使其对目标函数值最小的类;在 M 步,利用目标函数值重新估计每一类的中心。并且利用变换空间更新类距离度量D来进一步减少目标函数值。

(1) 选初始类

首先利用数据点之间的正软约束关系建立正软约束集,即此数据集的数据点之间只有正软约束关系。显然正软约束数据集中的数据点应归于一类。设这样的数据集有M个,对应的重心为μ_h,$h=1,\cdots,M$。若$M=$类别数K,则这M个数据集的重心μ_h,$h=1,\cdots,M$为初始聚类中心;若$M>N$,则从M个数据集的重心中随机选取N为初始聚类中心;若$M<N$,则还需要选取$N-M$个其他的数据点和M个数据集的重心共同作为初始聚类中心。

(2) E 步

根据初始聚类中心算法采用条件迭代模型优化算法(ICM: iterated conditional modes)计算每个中心点到聚类中心的目标函数的最小值,并将数据点加入使目标函数值最小的类别。ICM 算法的基本思想

是反复最大化局部条件概率函数，直到收敛。通过 ICM 算法可以保证目标函数可以达到局部的最小值。若 ICM 算法两次迭代的结果不变，则认为算法收敛，可以进入 M 步。

（3）M 步

M 步分为两部分：

第一，类重心的重新估计，这部分没有用到软约束条件，采用和普通 Kmeans 相同的算法。

第二，距离测度函数参数更新。数据点间距离采用 L1 距离 $D=\boldsymbol{A}|X-Y|$。$\boldsymbol{A}$ 为距离权值向量，其第 m 个分量 a_m 可以采用下式更新

$$a_m = a_m + \mu \frac{\partial J_{\text{obj}}^{SOFT}}{a_m} \tag{6}$$

E 步和 M 步反复迭代直到收敛。最后返回每个数据点对应的类别和目标函数值。

2.3 Manifold 排序学习算法描述

2.3.1 初始概率表生成

采用类似于 CMRM[5] 的概率模型来计算初始概率表。定义 $\boldsymbol{B}=\{b_1,\cdots,b_N\}$ 为区域聚类中心 blob 的集合，$\boldsymbol{W}=\{w_1,\cdots,w_c\}$ 为所有标注的集合。$\boldsymbol{\Gamma}$ 为训练图像集合，$\boldsymbol{J}\in\boldsymbol{\Gamma}$。于是可以确定任意 blob 与标注概念的联合概率为

$$\begin{aligned} P(w_i,b_j) &= \sum_{J\in\Gamma} P(\boldsymbol{J})P(w_i,b_j|\boldsymbol{J}) \\ &= \sum_{J\in\Gamma} P(\boldsymbol{J})P(w_i|\boldsymbol{J})P(b_j|\boldsymbol{J}) \end{aligned} \tag{7}$$

其中

$$P(w|\boldsymbol{J}) = (1-\alpha_J)\frac{\#(w,\boldsymbol{J})}{|\boldsymbol{J}|} + \alpha_J \frac{\#(w,\boldsymbol{\Gamma})}{|\boldsymbol{\Gamma}|} \tag{8}$$

$$P(b|\boldsymbol{J}) = (1-\beta_J)\frac{\#(b,\boldsymbol{J})}{|\boldsymbol{J}|} + \beta_J \frac{\#(b,\boldsymbol{\Gamma})}{|\boldsymbol{\Gamma}|} \tag{9}$$

其中，$\#(w,\boldsymbol{J})$ 代表图像 $\boldsymbol{J}$ 的标注中单词 w 出现的次数。$\#(w,\boldsymbol{\Gamma})$ 代表单词 w 在训练基 $\boldsymbol{\Gamma}$ 中总共出现的次数。$\#(b,\boldsymbol{J})$ 反映了 blob b 在图像 $\boldsymbol{J}$ 中出现的次数。$\#(b,\boldsymbol{\Gamma})$ 反映了 blob b 在训练基 $\boldsymbol{\Gamma}$ 中总共出现的次数。$|\boldsymbol{J}|$ 等于图像 $\boldsymbol{J}$ 中 blob 数和标注数的总和。$|\boldsymbol{\Gamma}|$ 为训练集的大小。α_J,β_J 为平滑系数。

2.3.2 Manifold 排序学习算法

上文的概率模型只考虑了标注与图像的关系和 blob 与图像的关系，而忽略了 blob 之间和 blob 与标注间的深层关系，没能紧密结合图像的视觉特征和高层语义，显然只利用概率模型不能得到比较好的结果。因此，Manifold 排序学习算法由于可以挖掘数据间的深层关系而被用来计算概率表 T，并且选取概率模型的结果为初始值。

定义 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_N\}\subset\mathbf{R}^m$（$\boldsymbol{X}$ 是 blob 的特征，m 是 $\boldsymbol{X}$ 的维数），$\boldsymbol{W}=\{w_1,\cdots,w_c\}$ 为所有标注的集合。Manifold 排序的算法过程如下：

（1）计算数据间的两两距离并按升序排列。在数据点间按距离的顺序添加边，直到数据点间形成一个连通图。

（2）计算相似度矩阵：

$$\Omega_{ij} = \begin{cases} \exp[-d^2(x_i,x_j)/2\sigma^2] & i\neq j \text{ 且顶点 } x_i,x_j \text{ 之间由边相连} \\ 0 & \text{其他} \end{cases}$$

（3）构造矩阵 $\boldsymbol{S}=\boldsymbol{U}^{-1/2}\boldsymbol{\Omega}\boldsymbol{U}^{-1/2}$，其中 $\boldsymbol{U}$ 是对角阵，对角线元素等于对应 $\boldsymbol{\Omega}$ 的行元素之和。

（4）对 $T(t+1)=\alpha\times S\times T(t)+(1-\alpha)\times T(0)$ 迭代运算到收敛，其中 t 为迭代次数，$\alpha\in[0,1]$ 为传播系数。$T(0)$ 为初始概率表。

（5）输出收敛结果 T^* 为最终的语义概念 - blob 联合概率表。

其直观的描述是：首先把每个数据点 x_i 作为顶点建立一张加权图。然后每个数据点传递自己的概率值给在加权图中的邻顶点，也就是概率值在相似度比较高的顶点间传递。这种概率值的传递可以反映所有数据之间的关系：在特征空间中，距离较远的顶点会具有不同的概率值，而距离近的顶点会分享相近的概率值。这恰好满足了自动图像标注问题的需求。

确定相似度矩阵 $\boldsymbol{\Omega}$ 是本算法的关键之一。距离度量一般采用 L2 距离。但是前人的工作[11,12] 表明，在通过图像特征比较图像时 L1 距离可以更好地表达图像间的视觉差异。因此，采用 L1 距离测度替代 L2 距离，并采用拉普拉斯核表示的相似度矩阵

$$\Omega_{ij} = \prod_{l=1}^{m} \frac{1}{2\sigma}\exp(-|x_{il}-x_{jl}|/\sigma) \tag{10}$$

3 实验结果与分析

3.1 数据集

实验图像数据集采用 Duygulu 等人[4,8] 提供的两组 Corel 图像集：

ECCV 2002：近年来此数据集已成为图像标注领域的通用图像库。图像数据集由 Corel 图像集中的5 000幅图像组成。图像分割算法采用 Normalized Cut[13] 算法，确保对每幅图像有 1 到 10 个 blobs，最终有 47 065 个图像区域。每幅图像标注 1 到 5 个关

键词。图像集一共有 374 个词,把数据集分成 2 个部分,其中 4 500 幅图像作为训练集和 500 幅图像作为测试集。测试集中包含 260 个单词。

JMLR 2003:共有 10 组 Corel 图像集,每组包括 5 000 多幅图像,共 50 000 多幅图像。每组图像集大约有 50 000 个区域,聚类后有 500 个 blob。每组图像集大约有 165 个单词。每组数据集分成 2 个部分,其中约 75% 的图像作为训练集,约 25% 的图像作为测试集。图像特征提取与 ECCV 2002 图像库相同。

3.2　自动图像标注

测试集图像也需要分割为区域,并通过相似度比较确定其属于的 blob 类别,然后利用语义概念-blob概率表 T^* 求出其与所有单词的联合概率,最后从中选出概率值最大的 5 个单词作为测试图像的标注。采用查准率、查全率和 $F1$ 值衡量实验结果。查准率定义为查询结果中正确结果的比例;查全率定义为查询结果中正确结果占数据库中正确结果的比例;$F1$ 值定义为 2 × 查准率 × 查全率/(查准率 + 查全率)。

图 2 给出了 49 个最好标注结果的平均性能[5],本文方法在与 TM、CMRM 方法相比时标注的查准率、查全率和 $F1$ 值都有所提高。

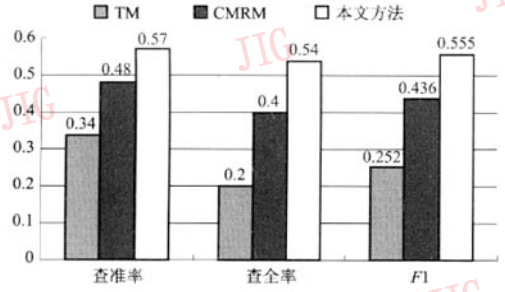

图 2　自动图像标注性能比较(TM、CMRM、本文方法)

Fig. 2　Comparison of annotation performance

为了说明本文提出的软约束半监督聚类方法(SHMRF-Kmeans)对图像标注系统的作用,分别采用 K 均值聚类方法(Kmeans)和文献[10]提出的硬约束半监督聚类方法(HMRF-Kmeans)替代 SHMRF-Kmeans 方法,得到两组新的图像标注系统。图 3 比较了这 3 组标注系统的性能。由图可见,基于语义约束聚类的标注系统性能要强于一般聚类的标注系统。并且对于语义约束聚类的标注系统,基于软约束聚类的标注系统性能要好于基于硬约束聚类的标

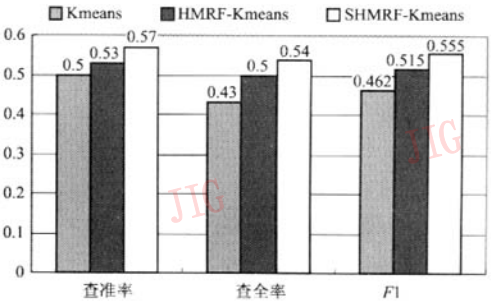

图 3　不同聚类算法标注性能比较

Fig. 3　Comparison of annotation performance based on different clustering algorithms

注系统。

表 1 给出了一些图像标注结果,并与 TM、CMRM 方法的结果做了比较。

表 1　自动注释结果(最好的 5 个)与人工注释结果比较

Tab. 1　Comparison between automatic annotation results and manual annotation results

方法	图像		
人工标注	水 草 猫 老虎	山脉 人群 路	特写 狐狸 头 雪
TM	水 树 男人 人 海滩	树 岩石 陆地 太阳 天空	天空 飞机 树 猫 男人
CMRM	水 天空 草 老虎 建筑物	树 人 建筑物 街道 游泳者	树 岩石 老虎 人 水
本文方法	老虎 猫 天空 草 地面	人 街道 路 草 山脉	雪 狐狸 树 猫 陆地

3.3　基于自动图像标注的图像检索

图 4 给出了 2 种图像检索系统在已经自动标注了的测试图像集上的检索结果。实验比较了本文方法和 CMRM 方法对应的图像检索系统在查询为 3

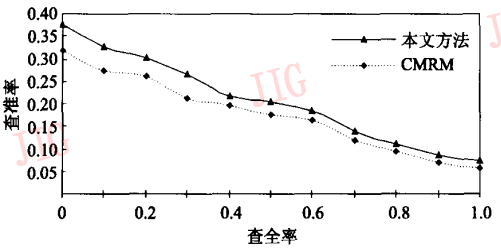

图 4　图像检索系统的查准率-查全率图

Fig. 4　Recall/Precision graph of image retrieval systems

个单词时的性能。从实验结果可见,本文的方法具有较好的检索性能。

4 结 论

针对图像检索中的语义鸿沟问题,提出了一种新颖的图像标注方法。该方法试图改变前人方法中图像视觉特征和高层语义结合不紧密的问题。提出了一种基于软约束的半监督图像聚类算法(SHMRF-Kmeans)对已标注图像的区域进行语义聚类,这种半监督的聚类方法可以同时考虑图像的视觉信息和语义特征,并提出利用图算法——Manifold排序学习算法来充分发掘语义概念与区域聚类中心的关系,得到其概率关系表。在一个包含约为55 000幅图像的数据集上进行实验,并与 TM[4]、CMRM[5]方法进行了比较,实验结果表明该方法能够获得更好的查全率和查准率。

参考文献(References)

1 Zhang Y J. Content-based Visual Information Retrieval [M]. Beijing: Science Press, 2003. [章毓晋著. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003.]

2 Claudio C, Gianluigi C, Raimondo S. Image annotation using SVM [A]. In: Proceedings of SPIE-The International Society for Optical Engineering, Internet Imaging [C], California, USA, 2004, **5304**: 330 ~ 338.

3 Edward C, Kingshy G, Gerard S, *et al*. CBSA: content-based soft annotation for multimodal image retrieval using bayes point machines [J]. IEEE Transactions on Circuits and Systems for Video Technology, January, 2003, **13**(1): 26 ~ 38.

4 Duygulu P, Barnard K, De F N, *et al*. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary[A]. In: Seventh European Conference on Computer Vision [C], Copenhagen Denmark: Springer, 2002, **4**: 97 ~ 112.

5 Jeon J, Lavernko V, Manmatha V. Automatic image annotation and retrieval using cross-media relevance models[A]. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval [C], Toronto Canada, 2003: 119 ~ 126.

6 Carson C, Thomas M, Hellerstein J, *et al*. Blobworld: A system for region-based image indexing and retrieval [A]. In: Third International Conference on Visual Information Systems [C], Amsterdam The Netherlands, 1999: 509 ~ 516.

7 Zhou D, Bousquet O. Learning with local and global consistency [A]. In: Advances in Neural Information Processing Systems[C], Vancouver, British Columbia, Canada, 2004: 321 ~ 328.

8 Barnard K, Duygulu P. Matching words and pictures[J]. Journal of Machine Learning Research, 2003, **3**(1): 1107 ~ 1135.

9 Lin D. Using syntactic dependency as local context to resolve word sense ambiguity[A]. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics [C], Madrid, Spain, 1997: 64 ~ 71.

10 Basu S, Bilenko M, Mooney R J. A probabilistic framework for semi-supervised clustering [A]. In: Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C], San Francisco, CA, USA, 2004: 59 ~ 68.

11 Kokare M, Chatterji B N, Biswas P K. Comparison of similarity metrics for texture image retrieval [A]. In: IEEE Conference on Convergent Technologies for Asia-Pacific Region [C], Bangalore, India, 2003, **1**(2): 571 ~ 575.

12 Stricker M, Orengo M. Similarity of color images[A]. In: Storage and Retrieval for Image and Video Databases[C], San Jose, CA, 1995: 381 ~ 392.

13 Shi J, Malik J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, **22**(8): 888 ~ 905.