

基于模糊神经网络的语义映射方法及其在自然图像检索中的应用

石跃祥^{1,2,3} 文华² 龚平² 莫浩澜² 金银国³

(智能计算与信息处理教育部重点实验室 湘潭 411105)¹ (湘潭大学信息工程学院 湘潭 411105)²
(湖南省轻武器研究所 益阳 413046)³

摘要 在CBIR研究中,图像底层视觉特征和高层语义特征之间存在的“语义鸿沟”成为语义图像检索的关键问题。为了避免一般映射方法把一幅图像归于一类语义图像的现象,体现自然风景图像中包含的丰富的高层语义信息和多归属类型,提出了对自然风景彩色图像中颜色较单一的目标区域,重复采用最优阈值化进行一次粗分割来提取最大目标区域,在分割区域的基础上,提取图像的局部颜色和形状特征,最后利用改进的模糊神经网络来建立底层视觉特征和高层语义特征之间的映射,实现了图像属性信息的有效传递和高层语义的自动获取。实验结果表明,该图像分割方法对自然彩色图像能够有效地提取目标物体,并对噪声图像具有一定的鲁棒性,而语义图像的部分类别的检索准确率接近90%,查全率也达到了75%,实验结果证明了该方法对自然图像检索的有效性及其先进性。

关键词 基于内容的图像检索,语义图像检索,图像分割,最优阈值化,鲁棒性,模糊神经网络
中图法分类号 TP391 文献标识码 A

Projection of Semantics and Retrieval in Natural Scenery Images Based on Fuzzy Nerve Network

SHI Yue-xiang^{1,2,3} WEN Hua² GONG Ping² MO Hao-lan² JIN Yin-guo³

(Key Laboratory of Intelligent Computing & Information Processing of Ministry of Education, Xiangtan 411105, China)¹

(College of Information Engineering, Xiangtan University, Xiangtan 411105, China)²

(The Light Weapons Institute of Hunan Province, Yiyang 413046, China)³

Abstract With the development of Content-Based Image Retrieval (CBIR), the solution of "semantic gap" which exists between the low-features and the high-level semantic features has become the key problems of the semantic image retrieval. To avoid the general method maps an image into a class of semantic image, and reflect the natural scenery image contains a wealth of high-level semantic information and multi-homing type, this paper presented a process of repeating use of the optimal threshold for a roughly extraction of the largest target area with the color image. This color target area is comparatively singleness in the natural scenery images. On the basis of the divided regions, this paper extracted the color and shape features of each region, at last, the fuzzy nerve network was used to map low-features into the high-level semantic features, so it finally realized the image attribute information transfer effectively and obtained the high-level semantic automatically. Experimental results show that the method of image segmentation of natural color image can effectively extract the target object. It also has a certain degree of robustness to the noise images. The accurate retrieval rate approaches 90% and the recall rate also achieves 75% in some class image of the nature image database. The experimental result shows the effectiveness and advancement of this method in the natural image retrieval.

Keywords Content-based image retrieval, Semantic image retrieval, Image segmentation, Optimal threshold, Robustness, Fuzzy nerve network

1 前言

基于内容的图像检索(CBIR)是通过提取图像的颜色、纹理、形状等视觉特征来实现检索的。然而,这些图像底层信息反映的只是图像的一些客观统计特性,并不能真正被人们理

解。在许多的检索识别中,人们主要是根据图像的含义来推断图像是否符合自己的需要,这些图像的含义即是图像的高层语义知识,其检索的关键词可能是蓝天、落日等语义信息,而不是图像的低层视觉特征,可见,语义检索更能满足用户的需要且有着极其广阔的应用前景。

收稿日期:2013-05-21 返修日期:2013-07-03 本文受湖南省自然科学基金(14JJ2074),“十二五”国家科技支撑计划项目(子项目),湖南省重点学科建设项目资助。

石跃祥(1964—),男,博士,教授,主要研究方向为图像处理与模式识别等,E-mail:shiyx@xtu.edu.cn;文华(1976—),女,硕士生,工程师,主要研究方向为计算机应用;龚平(1988—),女,硕士生,主要研究方向为图像处理;莫浩澜(1981—),男,硕士,主要研究方向为图像识别、智能系统;金银国(1964—),男,高级工程师,主要研究方向为自动检测与控制。

因此,如何跨越低层视觉特征和高层语义特征的鸿沟是语义图像检索中的一个难点。许多学者为此做了不少的研究和贡献^[1-18],其中 Staikopoulos A 和 Jianwei Niu 提出语义模板^[1,2]的概念建立低层视觉特征和高层语义之间的联系; Rong Zhao 等提出 LSI^[3] (latent semantics indexing) 来建立低层视觉特征和高层语义之间的联系; 鲍永生等提出语义网络^[4]用于建立低层视觉特征和高层语义之间的联系; 成洁、万华林、孙志杰等利用支持向量机^[5-7] (Support Vector Machine, SVM) 用于建立低层视觉特征和高层语义之间的联系。在利用语义视觉模板时,首先用户提供给系统初始的草图或范例图像,系统据此产生表达同样概念的其它范例,然后,用户从中提取认为满意的概念表达。在建立语义模板的过程中,人机交互起着重要的作用。

近年来, SVM 成为研究热点,其本身的众多优点被广泛应用于模式识别和机器学习领域,但 SVM 也存在一些缺点,比如只能进行两两分类,分类结果的准确性很大程度依赖于内积函数的选取,且目前得到的内积核函数形式常用的仅有 3 类,故在一定范围上限制了 SVM 的应用。更遗憾的是,一些诱人的结论(如 SVM 不像神经网络那样十分依赖对模型的选择, SVM 本身对不同的方法具有一定的不敏感性等),目前都仅仅是有限实验中观察到的现象,并没有得到正确的证明。而模糊神经网络与 SVM 类方法不同,根据模糊神经网络方法^[14,15]的性质,因引入了隶属度函数,该方法可以把一幅图像归于一类语义图像,但一幅图像归类不是很明显:例如,一幅包括山、水、蓝天的自然图像分别以某种隶属度归入几类语义图像。

本文针对基于 SVM 方法的诸多不足以及模糊神经网络的优点,引入模糊神经网络用于建立低层视觉特征和高层语义之间的联系,最后在自然风景数据库^[17]上进行了实验,并与经典的 SVM 方法进行了对比,实验结果表明了本文方法的可行性和有效性。

2 低层视觉特征提取

图像的低层特征主要包括图像的颜色、纹理、形状等特征。与文献^[16]类似,本文主要利用了图像的颜色矩^[8]和形状七阶不变矩来提取图像的低层视觉特征,下面给出其相关概念。

2.1 颜色矩特征的选择

颜色是彩色图像最显著的特征,根据其旋转、平移、尺度变化具有鲁棒性的特点,以及它的矩特征唯一的表示,选择提取颜色特征的一阶矩、二阶矩、三阶矩表示颜色分布:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (1)$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (2)$$

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (3)$$

式中, P_{ij} 是第 j 个像素的第 i 个颜色分量, N 是像素数量。事实上,一阶矩(1)定义了每个颜色分量的平均强度,二阶矩(2)和三阶矩(3)分别定义了颜色分量的方差和偏斜度。

2.2 形状七阶不变矩特征提取

形状特征通常和目标联系在一起,因而含有一定的语义信息,可以有效地对图像中感兴趣的目标进行检索。因形状

通常是由封闭的轮廓曲线包括起来的一个区域,而矩是对图像的一种统计形式,故可以通过计算不变矩对区域特征进行索引。

给定一个数字图像函数 $f(x, y)$, 则 $f(x, y)$ 的 $p+q$ 阶矩定义为:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (4)$$

m_{pq} 唯一地被 $f(x, y)$ 所确定,且 $f(x, y)$ 的 $p+q$ 阶中心矩定义为:

$$u_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (5)$$

式中, $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$, 也就是重心坐标。 $f(x, y)$ 的归一化中心矩可表示为:

$$\eta_{pq} = \frac{u_{pq}}{u_{00}^{\frac{p+q}{2}}} \quad (6)$$

$$\gamma = \frac{p+q}{2} + 1, p+q=2, 3, \dots$$

将归一化的二阶和三阶中心矩进行组合可得到 7 个对平移、旋转和尺度变化不变的矩。即有:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (7)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (8)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (9)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (10)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (11)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12}) \quad (12)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (13)$$

并且,第 7 个矩对扭曲也不变。

2.3 特征向量的内部归一化

由于不同特征的物理意义不同,取值范围也常大相径庭,这样不同的特征常不直接具有可比性,为此需要对不同特征的数值进行归一化,本文采用线性放缩法,给定一个特征取值下限 \min 和一个特征取值上限 \max , 则对特征值 x 可用下式归一化到 $[0, 1]$ 中:

$$x' = \frac{x - \min}{\max - \min} \quad (14)$$

3 低层视觉特征到高层语义的映射

3.1 改进的模糊神经网络训练算法

标准的 BP^[9] 算法存在以下不足之处:易形成局部极小而得不到全局最优;训练次数多使得学习效率低,收敛速度慢;隐层节点选取缺少理论指导;训练时学习新样本有遗忘旧样本的趋势等。

文中采用加入动量项的 BP 改进算法来克服标准 BP 算法的不足,并对学习率进行自适应调节(首先初始化学习率,若经过一批权值调整后总误差上升,则本次调整无效,且 $\Delta\eta = \beta\eta$ ($\beta < 0$));若经过一批权值调整后总误差下降,则本次调整有效,且 $\Delta\eta = \theta\eta$ ($\theta > 0$)). 在训练模糊神经网络时易出现随着训练次数的增加而总误差不再下降,这是因为神经元的输出进入了转移函数饱和区。为解决这一问题,可在原转移函

数中引入一个陡度因子 λ 。 $\sigma = \frac{1}{1 + e^{-\lambda x}}$ ，当发现总误差 E 接近于 0 而期望输出误差仍较大时，可判已进入平坦区，此时令 $\lambda > 1$ ；当退出平坦区后再令 $\lambda = 1$ 。在改进算法中采用随机抽取样本进行训练以克服同类样本太集中而使得网络训练时倾向于只建立与其匹配的映射关系。

训练算法流程如图 1 所示。

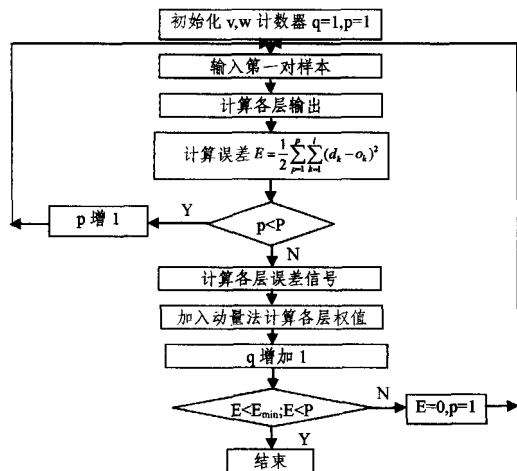


图 1 加入动量项的 BP 改进学习算法流程

训练具体步骤如下：

(1) 初始化。对权值矩阵 W, V 赋随机数 $[0, 1]$ ，将样本模式计数器 p 和训练次数计数器 q 置 1，总误差 E 置 0，学习率 η 置为 $(0, 1)$ 间的小数，网络训练精度 E_{\min} 设为一正的小数， P 表示总训练的样本数。

(2) 输入训练样本，用式(15)、式(16)计算各层输出 C, Y 。

$$C_k = f[(\omega_{1KH} * \mu_1) + (\omega_{2KH} * \mu_2) + \dots + (\omega_{jKH} * \mu_j) + \theta_{KH}] \quad (15)$$

式中， ω_{jKH} 表示隐含层第 K 个节点与第一层中低层特征向量的节点 j 相连的模糊权系数， θ_{KH} 表示隐含层第 K 个节点的模糊阈值。

$$Y_m = f[(\omega_{1mO} * C_1) + (\omega_{2mO} * C_2) + \dots + (\omega_{KmO} * C_K) + \theta_{mO}] \quad (16)$$

式中， ω_{KmO} 表示隐含层第 K 个节点与输出层第 m 个节点相连的模糊权系数， θ_{mO} 表示输出层第 m 个节点的模糊阈值。

(3) 计算误差 $E = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^l (d_k - o_k)^2$ 。

(4) 若样本模式计数器 p 超过总训练的样本数 P ，则执行(5)，否则 p 增 1 转(2)。

(5) 用 E 计算各层误差信号。公式如下：

$$\delta_k = \sqrt{E/P} * (1 - Y_k) * Y_k, k=1, 2, \dots, l \quad (17)$$

$$\delta_j^i = (\sum_{k=1}^l \delta_k \omega_{jk}) * (1 - C_j) * C_j, j=1, 2, \dots, m \quad (18)$$

式中， Y_k, C_j 分别表示输出层、隐含层各神经元的输出值， ω_{jk} 表示隐含层节点 j 与输出层节点 k 的连接权值。

(6) 加入动量项调整各层权值

$$w_{jk}(p) \leftarrow w_{jk}(p) + \eta * \delta_k^i * C_j(p) + a * \eta * \delta_k^i(p-1) * C_j(p-1) \quad (19)$$

$$v_{ij}(p) \leftarrow v_{ij}(p) + \eta * \delta_j^i * x_i + a * \eta * \delta_j^i(p-1) * x_i \quad (20)$$

其中， $k=1, 2, \dots, l; j=1, 2, \dots, m$ 。 x_i 表示输入层各结点的值， p 表示训练计数器， a 为加入的动量项 $(0, 1)$ 间的值。

(7) 训练次数计数器 q 增 1，若 $E/P < E_{\min}$ ，表示达到训练要求程序结束，否则 E 置 0， p 置 1 转(2)。

3.2 高层语义自动获取

对自然图像提取颜色和形状特征作为模糊神经网络的输入层，采用改进的 BP 学习算法训练网络，在训练好的网络中对每一幅图像进行语义分类，并根据混合模糊神经网络模型输出层所得的结果对图像库中的每一幅图像进行语义标注，从而完成图像从低层视觉特征到高层语义的映射。最后，利用字符串匹配技术^[10]对自然图像进行检索。

4 自然图像检索系统设计及实验分析

本文实验的自然图像库来源于文献[17]，实验环境为 Windows 2000，实验工具：Matlab 6.5。文中利用改进的 K 均值聚类算法对图像进行分割，在得到的图像的有效区域上提取图像的颜色和形状局部低层特征，再利用改进的模糊神经网络完成低层特征到高层语义的映射，实现了图像语义的自动提取，最后利用字符串匹配技术^[10]完成自然图像检索。

4.1 自然图像检索系统设计步骤

4.1.1 图像预处理

(1) 模糊神经网络输入结点的确定：输入层结点是根据图像的低层特征维数确定的，在本文中采用 RGB 颜色空间下的颜色三阶矩(共 9 维)和形状 7 阶不变矩(共 7 维)，则输入层结点共为 16，即为模糊神经网络模型的输入。所选的图像库大致为 6 类(日落，山，马，绿水，蓝天，花卉)自然分景图图像的语义，则输出结点为 6 个。隐层结点个数的确定根据试凑法^[11]： $m = \sqrt{n * l} = \sqrt{16 * 6} \approx 10$ (个)。

(2) 训练样本图像的选取，每个类别的样本数量大致相等，每类选取 100 幅图像左右，并对训练样本进行预处理：对每幅图像选取图像目标的最大区域及剩下的区域作为训练样本，以便扩充训练样本数并使网络训练朝目标方向进行。

4.1.2 训练模糊神经网络

采用改进的 K 均值聚类算法对图像库中的图像进行分割得到最大的目标子区域图，在这些子区域图中提取颜色三阶矩(9 维向量)和形状七阶不变矩(7 维向量)，将其内部归一化后经高斯模糊函数模糊作为模糊神经网络的输入层向量。传递函数采用 S 型函数，采用改进的 BP 学习算法训练网络，学习率 η 取 0.1 至 3 的一个随机数，动量项 a 取 0.9 至 1 的一个随机数，陡度因子 λ 取 2。6 类总训练样本为 1200 个，训练次数设为 100000 次，总误差 E 少于或等于 0.005 时停止训练。根据训练次数与总误差 E 的时空图，最佳训练次数是当总误差 E 不再随训练次数的增加而减少时确定的，同时也确定了总误差 E 。

4.1.3 语义自动提取

本文实验的自然图像库大约 4000 幅图像，分为 6 类(日落，山，马，绿水，蓝天，花卉)自然分景图。对图像库中的图像采用改进的 K 均值聚类算法^[18]进行分割得到最大的目标子区域图，在这些子区域图中提取颜色三阶矩和形状七阶不变矩作为模糊神经网络的输入层向量，通过训练好的模糊神经网络得出子区域图高层语义。

4.2 检索结果及分析

当对图像库中的每一幅图像进行了语义注释后，根据模糊神经网络的输出可知，每一幅图像都有 6 个隶属度值表明该图像属于 6 类语义(日落，山，马，绿水，蓝天，花卉)的程度，

用文本保存图像的隶属度值,文本中行数对应于图像库中的图像;再采用字符串匹配技术进行图像检索,例如要检索花卉类满足隶属度在 0.99 以上的图像,只要在保存隶属度的文本中找到文本一行中最后一个值大于 0.99 的值的行数就可以找到图像库中对应的图像,如图 2 花卉类检索结果,隶属度设置为 0.99,检索结果返回的是训练集中的花卉类图像,从而验证了当用训练集作为测试集时准确率^[12]达到了 100%,但由于网络的泛化能力较差,查全率却不足 60%。

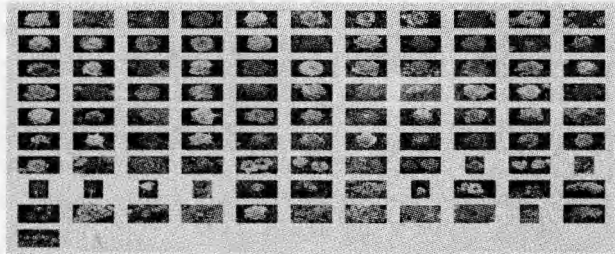


图 2 花卉类检索结果

查全率 = 正确检索结果 / (正确检索结果 + 漏检图像)

准确率 = 正确检索结果 / (正确检索结果 + 误检图像)

由于图像库较大,很难统计出漏检图像,故采用随机在图像库中抽出 120 幅作为测试集来找到最佳的隶属度设置值。图 3 是花卉类与马类图像的隶属度设置值与查全率/准确率的时空图。

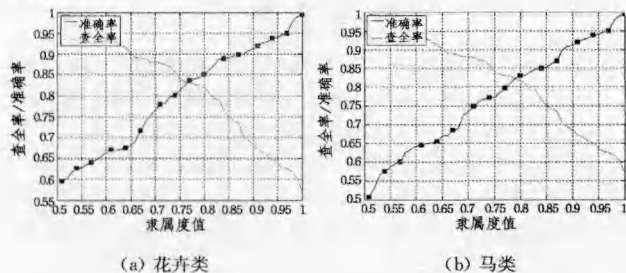


图 3 花卉类与马类图像的隶属度设置值与查全率/准确率的时空图

在设置隶属度值时,本文优先考虑准确度,因为影响查全率的因素较大的受查询用户的认识不同而不同。从图 2 中可以看出当花卉类选 0.85 时准确度接近 90% 而查全率也达 75%;当马类选 0.85 时准确度达到了 85%,查全率也达到了 75%。

当在图像库中检索满足山(0.85)、蓝天(0.75)、水(0.75)的图像时,检索结果(山、蓝天、水的检索结果图)如图 4 所示。

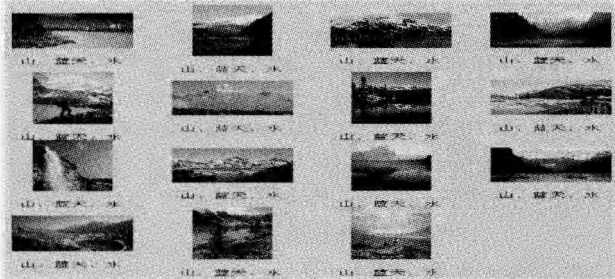


图 4 山、蓝天、水的检索结果

本文方法与 SVM 方法进行了实验结果对比,在实验库中随机抽取 6 组图像各 120 幅。表 1 和表 2 中列出了本文方法与支持向量机等方法的检索准确率与查全率(6 组图像检索的平均值)的实验结果比较,其中 SVM 方法的 1、2、3 分别

代表在 SVM 网络中所选择的不同核函数^[5];1 表示径向基核函数,2 表示多项式核函数,3 表示 Sigmoid 核函数。

表 1 本文方法与采用支持向量机等方法实验结果的准确率比较

实验方法	日落	山	绿水	马	花卉	蓝天	
SVM 方法 ^[5]	1	85.8%	73.3%	75%	75%	82.4%	72.5%
	2	78.8%	80%	68.6%	85%	76.5%	66.6%
	3	78.8%	73.3%	62.6%	75%	76.5%	66.6%
BP 神经网络 ^[14]	77.2%	48.4%	78.3%	76%	77.4%	70.5%	
本文方法	92.8%	86.7%	81.3%	85%	88.3%	72.5%	

表 2 本文方法与采用支持向量机等方法实验结果的查全率比较

实验方法	日落	山	绿水	马	花卉	蓝天	
SVM 方法 ^[5]	1	85.8%	80%	75%	75%	82.4%	83.3%
	2	92.8%	80%	81.3%	85%	76.5%	83.3%
	3	92.8%	73.3%	81.3%	75%	82.4%	77.8%
BP 神经网络	58.8%	68.9%	60.6%	62.5%	96.8%	90.7%	
本文方法	78.8%	73.3%	68.6%	75%	75%	72.5%	

实验分析:由表中可知本文方法取得了让人满意的实验结果,检索准确率超过或接近 SVM 分类方法,日落和花卉准确度超过或接近 90%,且 SVM 的检索准确率较大地受制于核函数的选取。但 SVM 方法由小样本训练 SVM 网络能获得较强的泛化能力,这是模糊神经网络无法超越的一个优点。如果查询用户对查全率要求较高,则可把隶属度的设置相应调低。

结束语 “语义鸿沟”是语义图像检索的一个难点,本文建立了一个语义图像检索模型,利用改进的模糊神经网络完成了图像的低层视觉特征与高层语义特征之间的映射。实验表明,该方法取得了较好的效果。

语义图像检索技术还处于初步发展阶段,目前的检索系统只是对图像的目标进行识别从而得到图像的场景语义,而图像的行为语义以及情感语义等更高层的语义还未涉及。如何建立一个从场景语义向更高层的语义进行逻辑推理的语义图像检索模型是下一步的工作。

参考文献

- [1] Staikopoulos A, Cliffe O, Popescu R, et al. Template-Based Adaptation of Semantic Web Services with Model-Driven Engineering[J]. IEEE Transactions on Fuzzy System, 2010, 3(2): 116-130
- [2] Niu Jian-wei, Atlee J M, Day N A. Template semantics for model-based notations[J]. IEEE Transactions on Software Engineering, 2003, 29(10): 866-882
- [3] Zhao R, Grosky W I. Narrowing the semantic gap improved text-based web document retrieval using visual features[J]. IEEE Transaction on Multimedia, 2002, 4(2): 189-200
- [4] 鲍永生, 任建锋, 郭雷. 支持语义的图像检索[J]. 南京航空航天大学学报, 2005, 37(1): 75-78
- [5] 成洁, 石跃祥. 基于 SVM 的图像低层特征与高层语义的关联[J]. 计算机应用研究, 2006, 9(12): 250-256
- [6] 万华林, Chowdhury M U. 基于支持向量机的图像语义分类[J]. 软件学报, 2003(14): 1891-1900
- [7] 孙志杰, 许宏丽. 一种图像低层视觉特征到高层语义的映射方法[J]. 计算机应用, 2004, 24(12): 22-24
- [8] 周献中. 基于 HSV 颜色空间加权 Hu 不变矩的台标识别[J]. 南京理工大学学报, 2005, 2(29): 363-367
- [9] 张智星, 孙春在. 神经-模糊和软计算[M]. 西安: 西安交通大学

[10] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003:57-130,401-421

[11] 刘宇红,刘桥,任强. 基于模糊聚类神经网络的语音识别方法[J]. 计算机学报,2006,29(10):1894-1900

[12] 章毓晋. 图像处理和分析[M]. 北京: 科学出版社, 2003(12): 180-215

[13] 韩晓微,晏磊,原忠虎,等. 基于BP神经网络的颜色模糊量化方法[J]. 系统仿真学报. 2006,18(10):3007-3010

[14] Yu W,Li X O. Fuzzy identification using fuzzy neural networks with stable learning algorithms[J]. IEEE Transactions on fuzzy

[15] Yong Y,Chi L H,Tsan M C. Intelligent Fabric Hand predication system with Fuzzy Neural Network[J]. IEEE Transactions on Au,R. System, Man, and Cybernetics, Part C: App. and Reviews,2010,6(40):619-629

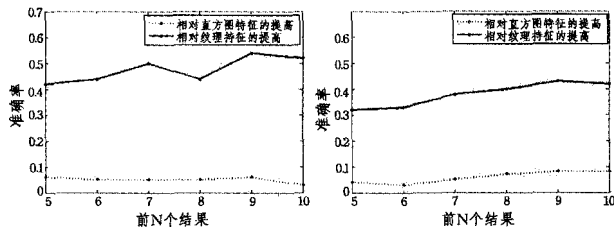
[16] 汪卫星. 基于颜色和形状特征图像检索技术研究[D]. 长沙: 长沙理工大学,2010

[17] Computer Vision Test Images[OL]. <http://www.cs.cmu.edu/~cil/v-images.html>

[18] 刘盈盈,石跃祥,莫浩瀚. 基于改进的动态聚类算法在彩色图像分割中的应用[J]. 计算机工程与应用,2008,44(29):191-192

(上接第 115 页)

准确率(查准率),最后求所有图片准确率的平均值。从以上实验结果可以看出,使用混合特征在两种情况下较单个特征准确率均有提高,其中对于纹理特征的提高比较大,对于颜色直方图特征的提高比较小。

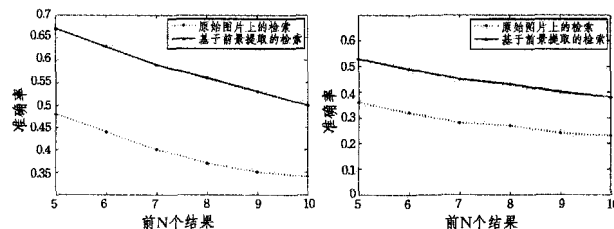


(a) 原始图片上的检索 (b) 基于前景提取的检索

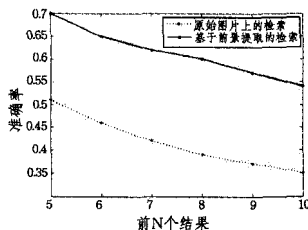
图 8 使用混合特征的提高比

5.2 算法实验结果

图 9 展示了直接用原始图片检索和基于前景提取的检索在不同的特征上准确率的对比。对于颜色直方图、LBP 纹理特征、颜色纹理混合特征,与直接使用原始图片检索的准确率相比较,本文提出的基于前景提取的复杂背景图像检索算法均有较大的提升。



(a) H-S 颜色直方图特征 (b) LBP 纹理特征



(c) 颜色纹理混合特征

图 9 使用不同特征的检索准确率

结束语 基于内容的图像检索是一种“以图搜图”的检索方式,用户往往关注的是图像中主体部分的相关性,而复杂的背景会影响检索结果的准确性。本文提出基于前景提取的图像检索算法,其能够在无人工交互的情况下提取图像的前景部分,之后针对前景部分做特征提取和相似度计算得到检索结果。通过实验证明,对于颜色直方图、LBP 纹理特征、颜色

纹理混合特征,该算法相对于直接用原始图片的检索准确率有较大的提升。

同时我们发现,对于某些图片,采用从 4 个角开始的扩散方法自动标记背景区域对某些图片并不适合,有些图片会把位于角落的前景也标记为背景,有些图片提取出的前景仍然包含大部分的背景,下一步的工作我们会考虑从图像边界向内部扩散的方法。另外,基于前景提取的图像检索算法虽然可以实现复杂背景的图像检索,但部分图像经过前景提取后仍然存在 Logo、文字等区域,这些区域大部分属于噪声信息,对图像检索的准确率会造成一定的影响,下一步的工作会将这些区域去除后再提取图像的特征向量。

参考文献

[1] Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics (TOG), ACM,2004,23(3):309-314

[2] Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images[C]// Proceedings of the International Conference on Computer Vision, ICCV 2001. 2001:105-112

[3] Kolmogorov V, Zabini R. What energy functions can be minimized via graph cuts? [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2004,26(2):147-159

[4] Rao A, Srihari R K, Zhang Z. Spatial color histograms for content-based image retrieval[C]// Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, 1999. IEEE,1999:183-186

[5] Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1996,18(8):837-842

[6] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception[J]. IEEE Transactions on Systems, Man and Cybernetics,1978,8(6):460-473

[7] 周静,郝红卫. 基于用户感兴趣区域的图像检索方法[J]. 计算机应用研究,2007,24(9):282-284

[8] Ford D R, Fulkerson D R. Flows in networks [M]. Princeton: Princeton university press,2010

[9] Goldberg A V, Tarjan R E. A new approach to the maximum-flow problem[J]. Journal of the ACM (JACM),1988,35(4): 921-940

[10] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(7):971-987