

文章编号:1004-4213(2011)07-1025-6

# 基于视觉注意的彩色图像检索方法

黄传波,金忠

(南京理工大学 计算机科学与技术学院, 南京 210094)

**摘 要:**基于视觉注意模型提取的特征能够反映图像高层语义的新特征,将视觉注意机制引入到图像分析领域能有效地减小语义鸿沟,获得高效的图像检索性能.根据视觉感知的特点,对 Itti 视觉注意模型进行了改进.采用主分量图表示亮度图,将纹理粗糙度信息融入视觉注意模型,进而提出了一种基于视觉注意空间分布特征的图像检索算法.首先由改进视觉注意模型将图像分解得到 38 个视觉特征图,然后采用网格平分法提取视觉特征图的空间分布信息,组成特征矢量来多层次地对图像特征进行描述,用于图像检索.实验结果表明,该算法利用基于改进注意力模型方法来提取图像空间分布特征进行图像检索,能获得较高的检索率.

**关键词:**视觉注意;视觉特征图;特征抽取;图像检索

**中图分类号:**TP391.41

**文献标识码:**A

**doi:**10.3788/gzxb20114007.1025

## 0 引言

基于内容图像检索(Content-Based Image Retrieval, CBIR)是利用图像自身包含的丰富视觉信息来进行检索<sup>[1]</sup>,但是 CBIR 中的图像低层特征与高层语义间存在着巨大鸿沟<sup>[2]</sup>.近年来提出了不少获取高层语义的方法,其中运用相关反馈(Relevance Feedback, RF)的图像检索和基于感兴趣区域(Region Of Interest, ROI)的图像检索是两大主流技术:RF<sup>[3]</sup>是一种查询修正技术,多数采用用户手工标记正例反例进行反馈,人机交互方式不友好,影响反馈信息的数量和质量;基于 ROI 的图像检索方法选取图像中的关键区域来描述图像内容,区分了区域的重要程度,也在一定程度上弥补了语义鸿沟.但是,通过进行图像分割得到的 ROI 并不准确,难以实现对 ROI 的客观描述.研究发现:当人们浏览图像时,通过视觉系统将图像传递到大脑,大脑则控制人眼运动来表达感兴趣的视觉对象或区域<sup>[4]</sup>.

基于人类视觉系统的研究,注意力模型(Attention Model, AM)在目标检测及基于内容的图像检索等方面,已成为一个重要研究方向.注意力模型<sup>[5]</sup>是依据人类视觉系统理论模拟人眼观察图像的过程,得到图像中最容易引起人们注意的部分,并用图像表达这种显著性的方法.视觉感知特征<sup>[6]</sup>是

图像的新特征,反映了用户的高层语义,为解决 CBIR 语义鸿沟提供了新的途径,并能有效解决 RF 图像检索和基于 RIO 图像检索中存在的问题.

在现有的视觉注意模型中,Itti 等<sup>[7]</sup>提出的基于显著性的模型最具代表性.它在运算速度上有突出的表现.文献[8-9]基于 Itti 模型获得 18 个颜色和亮度特征图,Gabor 的 4 个方向滤波选取 4 个尺度的金字塔层构成 16 个特征图.对每个图采用  $4 \times 4$  分块取每个块中像素的均值作为块特征,形成 544 维的特征向量,作为整幅图像的特征,用于自然场景识别取得了较好的效果.但 Itti 模型中,对于初始亮度频谱图的生成仅是对彩色图像的  $r, g, b$  三个频谱分量求均值获得的,没有考虑选择带有最多细节、拥有最大对比度且灰度值分布最广的频谱图,因而会影响到模型的最终效果.另外,Itti 模型仅考虑亮度、颜色和方向 3 个特征,忽略了与视觉心理感知有关的纹理粗糙度特征.本文针对以上问题进行了改进,一方面运用主分量频谱图表示亮度,另一方面嵌入了纹理粗糙度信息,不仅保证了初始亮度频谱图有较大的对比度、较广的灰度值分布,而且能够反映纹理图像的粒度特性,对图像检索更为有利.在此基础上,提出了一种基于视觉注意的图像检索算法.通过分块的均值和方差提取出能够较好地表示图像特征的向量.仿真实验结果表明,本文算法具有较好的检索性能.

**基金项目:**国家自然科学基金(No. 60973098, 60873151, 90820306)资助

**第一作者:**黄传波(1972-),男,博士研究生,主要研究方向为图像处理、人工智能. Email: huangjunfengcq@126.com

**导 师:**金忠(1961-),男,教授,博导,主要研究方向为模式识别、计算机视觉.

**收稿日期:**2010-05-31; **修回日期:**2010-11-03

### 1 Itti 注意力模型

图 1 为 Itti 模型的工作框图. 首先, 提取亮度、颜色和方向图像特征.  $r, g, b$  分别对应输入图像红、绿、蓝通道, 亮度特征表示为:  $I = (r + g + b) / 3$ . 用 4 个宽调谐的颜色通道作为颜色特征: 红色  $R = r - (g + b) / 2$ , 绿色  $G = g - (r + b) / 2$ , 蓝色  $B = b - (r + g) / 2$  和黄色  $Y = (r + g) / 2 - |r - g| / 2 - b$ . 方向特征为 Gabor 小波在  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  4 个方向的分量. 特征显著度通过计算图像  $I(x, y)$  区域中心  $c$  和周边  $s$  的高斯差分 (Difference Of Gaussian, DOG) 得到.

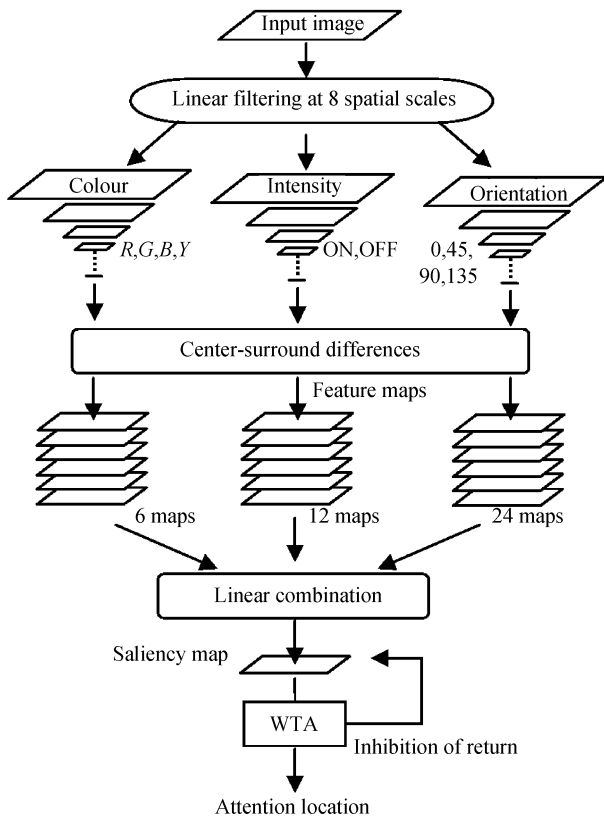


图 1 Itti 模型框图  
Fig. 1 Block diagram of the Itti model

$$DOG(x, y) = \frac{1}{2\pi\sigma_c^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_c^2}\right) - \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_s^2}\right) \quad (1)$$

式中,  $\sigma_c$  和  $\sigma_s$  分别表示中心  $c$  和周边  $s$  的尺度因子. 周边图像通过对上一级图像插值放大得到, 中心  $c$  和周边  $s$  差计算用符号  $\ominus$  表示, 其结果就是特征图. 亮度特征图

$$I(c, s) = |I(c) \ominus I(s)| \quad (2)$$

颜色特征图

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (3)$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (4)$$

方向特征图

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (5)$$

归一化后得到 3 个组合特征图  $\tilde{I}, \tilde{C}$  和  $\tilde{O}$

$$\tilde{I} = \oplus_c \oplus_s N(I(c, s)) \quad (6)$$

$$\tilde{C} = \oplus_c \oplus_s [N(RG(c, s)) + N(BY(c, s))] \quad (7)$$

$$\tilde{O} = \sum_\theta N(\oplus_c \oplus_s N(O(c, s, \theta))) \quad (8)$$

特征图线性组合为显著图

$$S = [N(\tilde{I}) + N(\tilde{C}) + N(\tilde{O})] / 3 \quad (9)$$

显著图使用胜者为王 (Winner Takes All, WTA) 竞争机制, 遵循返回抑制特点, 找到注意焦点, 然后循环下一个焦点.

### 2 改进注意力模型及特征抽取

改进的视觉注意模型的框架如图 2. 以 Itti 视觉选择性注意模型作为基础, 对其进行改进, 通过改进亮度图和嵌入粗糙度纹理特征来对图像内容进行更好的描述, 将每个特征图分成  $4 \times 4$  的网格, 以每个网格的一阶和二阶矩作为特征, 组合在一起构成一个特征向量来表示整幅图像.

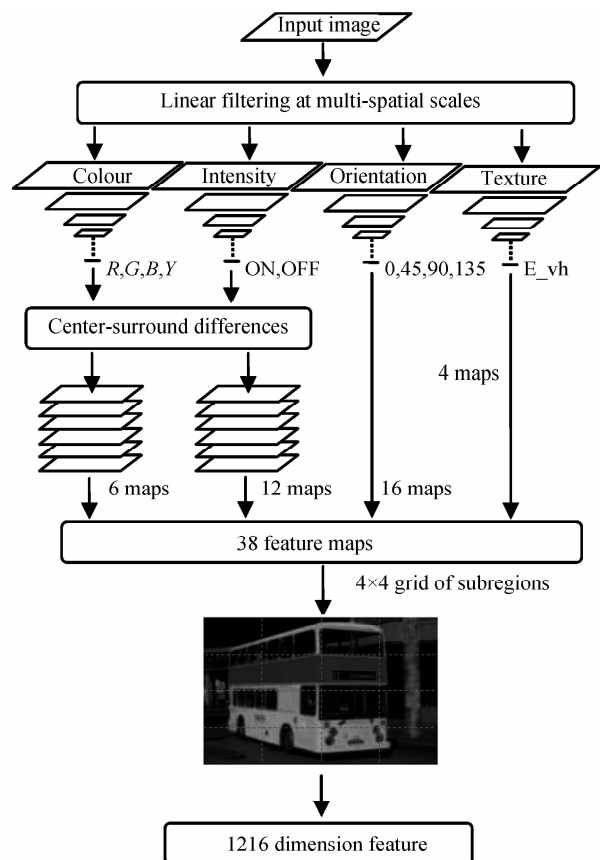


图 2 特征抽取框图  
Fig. 2 Block diagram of the extracting feature

#### 2.1 注意力模型的改进

基于 Itti 模型进行了两点改进: 一是使用主分量图改进原模型中的亮度图, 使其具有较大的对比

度和较宽的频带范围;二是分析了图像纹理视觉特征信息,在模型中引入了纹理粗糙度初级视觉特征图,且对于方向通道和纹理通道都不进行中心周边差计算。

### 2.1.1 主分量图

彩色图像是由三个相同尺寸的图像分量构成,每个图像分量都是带有特定频带的灰度图像,对应一个频谱分量,其灰度值表示该像素在这个频带上的亮度。由于在视觉注意模型中,只能用一个频谱分量来表示亮度进行处理,为了利用该模型提取更加可靠地特征描述图像进行分类或检索,应当选择带有最多细节、拥有最大对比度且灰度值分布最广的频谱分量。

彩色图像上每个像素点都是由红、绿、蓝三个频谱分量值表示,可以用一个三维空间的向量表达。显然,彩色图像的像素在彩色坐标空间形成点簇,像素的最佳延伸方向可能不是  $I=(r+g+b)/3$  的方向。为了寻找最佳延伸方向,对图像中像素在三维空间中进行主分量分析,步骤为:

1) 计算彩色空间中分布点的均值,在每个相应的频带上减去平均灰度值;

2) 算出原分布的自相关矩阵  $C(i, j)$ ;

3) 计算矩阵  $C(i, j)$  的特征值并降序排列,用最大特征值对应的特征向量作为投影向量,获得第一主分量。

用第一主分量,能保证图像有最大的对比度且用单个频带携带最多的信息。

第一主分量表示的亮度图、HSV 颜色模型亮度图  $I=(r+g+b)/3$  和 YCrCb 颜色模型亮度图  $I=0.299 * r + 0.587 * g + 0.144 * b$  分别如图 3(a)~(c)。

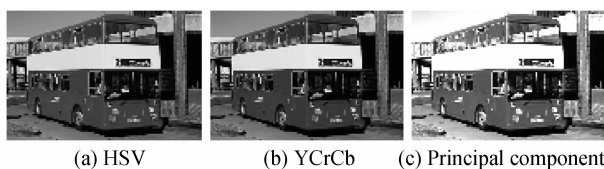


图 3 三类图的对比

Fig. 3 Comparison of three types of images

由图 3 可以看出,主分量图比 HSV 颜色模型亮度图和 YCrCb 颜色模型亮度图对比度都要强且带有更多的细节。

### 2.1.2 纹理粗糙度初级视觉特征图

粗糙度是反映纹理中粒度的一个量,也是最基本的纹理特征。两种纹理模式只有基元尺寸不同时,具有较大基元尺寸的模式给人的感觉更粗糙;对具有不同结构的纹理模式来说,基元尺寸越大或者基元重复次数越少,给人的感觉越粗糙。为了刻画纹理

的这个特性,本文采用不同大小窗口的像素滑动均值表示粗糙度,构造粗糙度初级视觉特征图的方法。主要计算方法为:

1) 计算图像中大小为  $2^n \times 2^n$  个像素的活动窗口中像素的平均强度值。

$$A_n(x, y) = \frac{1}{2^{2n}} \sum_{i=x-2^{n-1}}^{x+2^{n-1}-1} \sum_{j=y-2^{n-1}}^{y+2^{n-1}-1} g(i, j) \quad (10)$$

式中,  $n=0, 1, \dots, 5$ ;  $g(i, j)$  是位于  $(i, j)$  处像素的灰度值。

2) 对每个像素分别计算它在水平和垂直方向上互不重叠的窗口之间的平均强度差。

$$E_{n,h}(x, y) = |A_n(x+2^{n-1}, y) - A_n(x-2^{n-1}, y)| \quad (11)$$

$$E_{n,v}(x, y) = |A_n(x, y+2^{n-1}) - A_n(x, y-2^{n-1})| \quad (12)$$

3) 取水平和垂直两方向上的最大均值差值为当前像素的邻域均值差值。

$$E_n(x, y) = \max(E_{n,h}(x, y), E_{n,v}(x, y)) \quad (13)$$

经过这三步可以得到一幅新的图像,它在一定程度上反映纹理图像的粒度特性,能够表达具有不同纹理特征的图像或区域,对于图像检索非常有利。本文定义这种含有丰富纹理信息的图像为纹理粗糙度初级视觉特征图。图 4 显示了一幅图像及其纹理粗糙度初级视觉特征图。

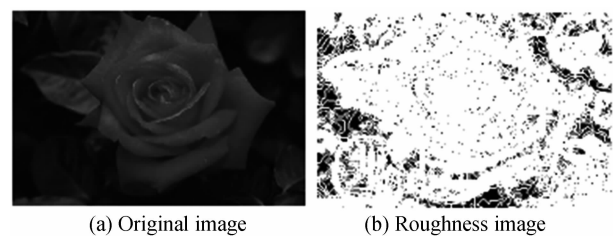


图 4 纹理粗糙度初级视觉特征图与原图像的对比

Fig. 4 Comparison between the primary visual map of texture roughness and the original image

从图 4 可知,纹理粗糙度初级视觉特征图含有较丰富的纹理粒度信息。

## 2.2 特征抽取

根据改进的 Itti 模型,采用主分量图作为亮度图,对输入图像进行多尺度表示。亮度和颜色特征采用 9 层金字塔分解(尺度 0~8),其中第 0 层为原始图像,下一层图像相对于上一层图像依次在长度和宽度上缩减一半。金字塔的  $\{1, 3, 4\}$  层是中心层  $c$ , 周边层  $s=c+\delta$ , 其中  $\delta=\{3, 4\}$ 。采用中心周边差的计算策略,抽取各个特征的对比信息,可得到 18 个特征图。对于 4 个方向特征和 1 个纹理特征,采用 4 层金字塔分解(尺度 0~3),可得 20 个特征图。初级视觉特征分解结束共生成了 38 个特征图,亮度特征 6 个,颜色特征 12 个,方向特征 16 个,纹理特征 4 个。

采用网格平分法<sup>[8-9]</sup>对生成的各特征图提取特征向量,即对每一个特征图,将其划分为固定的  $4 \times$

4 小块,然后计算每一块的平均值.均值表示的是随机变量  $X$ (像素的灰度)取值的一种位置特征,在这里作为分布的典型值,相当于图像子块中系数分布的中点位置.其定义式为

$$\mu_{kl} = \frac{1}{a * b} \sum_{i=1}^a \sum_{j=1}^b |x_{ij}| \quad (14)$$

式中,  $\mu_{kl}$  表示第  $k$  个特征图第  $l$  个子块的均值,  $a * b$  表示子块的大小.

均方差描述的是随机变量  $X$  相对于其均值的偏离程度.它反映了随机变量取值的波动性,相当于图像子块中各个系数相对于中点位置的偏离程度.若  $X$  取值比较集中,则  $\sigma(X)$  较小;反之,  $X$  取值比较分散,  $\sigma(X)$  较大.所以说  $\sigma(X)$  是衡量  $X$  取值分散程度的一个尺度,它与系数矩阵的均值共同反映了图像子块中系数的分布特征.其定义式为

$$\sigma_{kl} = \sqrt{\frac{1}{a * b} \sum_{i=1}^a \sum_{j=1}^b (x_{ij} - \mu_{kl})^2} \quad (15)$$

式中,  $\sigma_{kl}$  表示第  $k$  个特征图第  $l$  个子块的方差.

由前文分析可知,文献[8-9]仅用均值表示子块的特征是不够的,而均值和方差一起则能较好地体现各个子块的系数分布特征.本文用均值和方差对生成的 38 幅颜色、亮度、方向和纹理特征图进行分块特征抽取,共得到  $38 \times 16 \times 2 = 1216$  维向量作为下一步图像检索的特征向量.

### 3 相似性度量

由于均值和方差是不同的量纲,对于识别所起的作用是有差异的,如果简单的将各分量混在一起统一计算相似度,容易将重要特征的作用掩盖,达不到预期的效果.因此,在本文分别对  $\mu_i$ 、 $\sigma_i$  采用不同权值的欧氏距离相似度量.

$$d_{qi} = \alpha \sqrt{\sum_{r=1}^{608} (x_{qr} - x_{ir})^2} + \beta \sqrt{\sum_{r=609}^{1216} (x_{qr} - x_{ir})^2} \quad (16)$$

式中,  $\alpha$ 、 $\beta$  分别为均值和方差的权值,  $x_{qr}$  为检索图像特征向量  $x_q$  的第  $r$  个分量,  $x_{ir}$  为图像库中第  $i$  个图像特征向量  $x_i$  的第  $r$  个分量.

### 4 实验与分析

为验证本文算法的有效性,分别进行了几组不同的仿真实验.实验从国际上通用的 Corel 图像库中抽取 30 类、3 000 幅不同的彩色图像,采用“查准率(precision)”和“查全率(recall)”作为检索的评价准则.通过执行查询集合中各个查询图像来计算平均查准率和查全率(本文返回图像数选取 25),据此对系统的检索性能进行评价.

实验一:相似度量中的权值选择.实验在图像库

中选取 30 类图像,从每类图像中抽取 10 幅图像共组成 300 次查询,计算出平均查准率作为权值选取的依据.用主成分分析(Principal Component Analysis, PCA)(96%)将其由 1 216 维降至 149 维,此实验结果如表 1.

表 1 不同权值的检索查准率

Table 1 The precision to different weights

$\alpha$	$\beta$	Average precision/(%)
0.20	0.80	63.370
0.25	0.75	64.104
0.30	0.70	64.520
0.35	0.65	64.712
0.40	0.60	64.672
0.45	0.55	64.472
0.50	0.50	63.736

由表 1 可知,在  $\alpha=0.35$ 、 $\beta=0.65$  时,检索效果最好.

实验二:先在图像库中选取 10 类图像,从每类图像中抽取 50 幅图像共组成 500 次查询,对于不同类计算出平均查准率和查全率.为体现主分量图作为模型亮度图的优越性,实验不考虑加入的 4 个纹理粗糙度特征图,只用 34 个特征图,以均值和方差构成 1 088 维特征向量表示图像特征,并用 PCA 降到 144 维用于检索.将主分量图表示的亮度图与 HSV 颜色模型和 YCrCb 颜色模型的亮度特征图比较,其结果如表 2,相应查全率如图 5.

表 2 不同亮度模型的检索结果

Table 2 Retrieval results in different intensity models

Type	Principal component	HSV model	YCrCb model
1	41.10	40.08	41.58
2	41.10	40.36	40.46
3	47.18	46.16	45.78
4	69.54	69.76	66.86
5	99.86	99.86	99.89
6	61.34	60.76	58.50
7	97.14	96.76	95.58
8	89.66	89.40	88.42
9	42.54	41.84	41.30
10	54.26	53.68	52.06
Average	64.372	63.866	63.043

由表 2 和图 5 可知,即使使用同一种方法,对于不同类图像的检索效果也是有差别的.总体来看,主分量用作模型中的亮度初级视觉特征,要优于 HSV 颜色模式和 YCrCb 颜色模式,且 HSV 颜色模式优于 YCrCb 颜色模式.

实验三:先在图像库中选取 10 类图像,从每类图像中抽取 50 幅图像共组成 500 次查询,对于不同类图像计算出平均查准率和查全率.实验采用四种

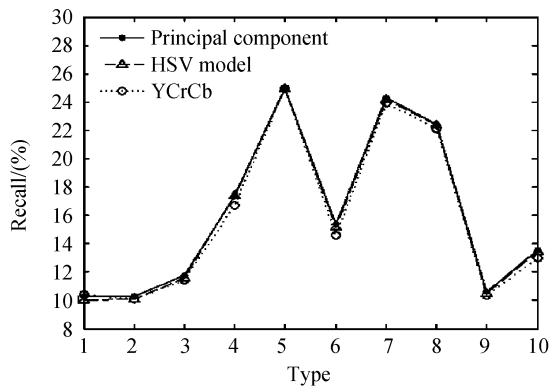


图5 不同亮度模型的10类图像查全率对比  
Fig. 5 Recall comparison of 10 types images in different intensity models

方法进行比较:方法一,在不加入纹理粗糙度特征图的情况下,仅用均值特征,采用PCA(96%)将其由544维降至86维;方法二,在不加入纹理粗糙度特征图的情况下,同时用均值和方差特征,采用PCA(96%)将其由1088维降至144维;方法三,加入纹理粗糙度特征图的情况下,同时用均值和方差特征,采用PCA(96%)将其由1216维降至149维;方法四,使用颜色直方图<sup>[10]</sup>、边缘直方图<sup>[10]</sup>、边缘方向直方图<sup>[11]</sup>和颜色矩特征<sup>[12]</sup>,共同组成一个274维特征向量,实验结果如表3和图6。

表3 不同特征的检索查准率

Table 3 The retrieval precision of different features

Type	Method 1	Method 2	Method 3	Method 4
1	45.24	41.10	43.32	43.15
2	40.48	41.10	42.96	41.56
3	43.20	47.18	47.60	50.68
4	56.16	69.54	70.56	83.24
5	98.56	99.86	99.80	97.04
6	51.36	61.34	62.67	56.20
7	89.16	97.14	97.52	71.20
8	80.60	89.66	89.64	74.72
9	39.96	42.54	42.62	43.92
10	43.16	54.26	54.44	55.62
Average	58.788	64.372	65.113	61.733

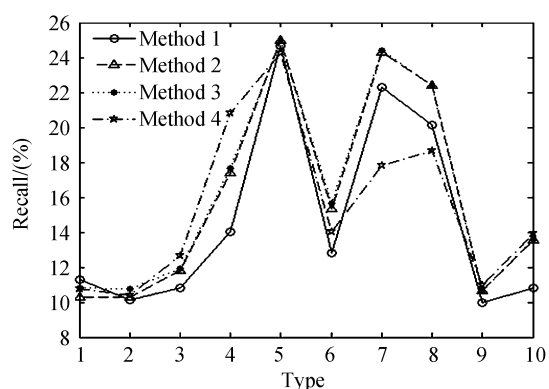


图6 不同算法10类图像查全率对比  
Fig. 6 Recall comparison of 10 types images in different methods

由表3和图6可知,方法三最优,其次是方法二,然后是方法四、方法一。方法三平均值比方法四大的多,但对于第4类,方法四的查准率和查全率明显比较高,而对于第7、8类,方法三的检索查准率和查全率明显要高。原因在于视觉注意模型特征对于目标突出的图像检索效果要好,而对那些图像中颜色分布不规则的图像,融合底层特征的检索效果要明显占优势。

## 5 结论

本文提出了一种基于视觉注意的图像检索算法,对计算模型产生的特征图通过网格平分法提取特征向量,以特征图像块的均值和方差反映图像的分布结构和统计特性,从底层特征中获取高层具有视觉特性的特征进行图像检索。实验结果表明,该方法在提高检索性能上有非常好的效果。利用视觉注意机制模拟人眼的生理视觉感知,在图像检索中估计出图像ROI,进一步融合心理视觉感知,构成完整的基于视觉感知的图像检索方法是进一步研究的方向。

### 参考文献

- [1] ZHAO Shan, ZHAI Hai-xia. Image retrieval based on bit-plane distribution feature[J]. *Acta Photonica Sinica*, 2009, **38**(8):2150-2154.  
赵珊,翟海霞. 基于位平面分布特征的图像检索算法[J]. *光子学报*, 2009, **38**(8):2150-2154.
- [2] ENSER P, SANDOM C. Towards a comprehensive survey of the semantic gap in visual image retrieval[J]. *Lecture Notes in Computer Science*, 2003, **2728**: 291-299.
- [3] SU Z, ZHANG H J, LI S, *et al.* Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning [J]. *IEEE Trans on Image Process*, 2003, **12**(8): 924-937.
- [4] GRABOWSKI T, JAIMES A, PEL Z J B, *et al.* Using human observers' eye movements in automatic image classifiers[C]. Proceedings of SPIE Human Vision and Electronic Imaging VI, San Jose, CA, USA, SPIE Press, 2001: 373-384.
- [5] CHEN I, FAN X, XIE X, *et al.* A visual attention model for adapting images on small displays [J]. *ACM Multimedia Systems Journal*, 2000, **9**(4):3353-364.
- [6] ZHANG Jing, SHEN Lan-sun, FENG D D. A survey of image retrieval based on visual perception [J]. *Acta Electronica Sinica*, 2008, **36**(3):494-499.  
张菁,沈兰荪, Dagan FENG. 基于视觉感知的图像检索的研究[J]. *电子学报*, 2008, **36**(2):494-499.
- [7] ITTI L, KOCH C. Computational modeling of visual attention [J]. *Nature Reviews Neuroscience*, 2001, **2**(3):194-203.
- [8] SIAGIAN C, ITTI L. Rapid biologically-inspired scene classification using features shared with visual attention[J]. *IEEE Trans on Pattern Anal Mach Intell*, 2007, **29**(2):300-312.
- [9] LI J, ALLINSON N M. Subspace learning-based dimensionality reduction in building recognition [J]. *Neurocomputing*, 2009, **73**(3):324-330.
- [10] MANJUNATH B S, OHM J R, VASUDEVAN V V, *et al.*

- Color and texture descriptors[J]. *IEEE Transon Circuits and Sys for Video Tech*, 2001, **11**(6): 703-715.
- [11] PINHEIRO A M G. Image description using scale-space edge pixel directions histogram[C]. *Semantic Media Adaptation and Personalization, Second International Workshop*, London, United kingdom, *Inst of Elec and Elec Eng*, 2007: 211-218.
- [12] STRICKER M A, ORENGO M. Similarity of color images [C]. *Proceedings of the SPIE*, San Jose, CA, USA, 1995: 381-392.

## A Method of Color Image Retrieval Based on Visual Attention Model

HUANG Chuan-bo, JIN Zhong

(College of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** The feature based on visual attention model is a new feature that can objectively interpret the high-level concepts. High retrieval efficiency will be achieved and the semantic gap in image retrieval can be also reduced effectively if visual attention mechanism is adopted. An improved Itti's visual attention model, inspired by human visual perception, is proposed. Primary component map is used to take place of intensity map, and the information of texture coarseness is brought in the improved Itti's model. A novel image retrieval algorithm is proposed based on visual attention distribution feature. Firstly, the image is divided into 38 feature maps by the improved Itti's visual attention model. Then, by the fixed four-by-four grid of subregions, the visual attention distribution feature to image retrieval is constructed according to statistical distribution of the 38 feature maps which contain most of the structural information of the image. The experimental results show that the proposed method has better retrieval performance.

**Key words:** Visual attention model; Visual feature map; Feature extraction; Image retrieval