

# 利用颜色和熵提取感兴趣区域的感性图像检索

陆伟 倪林

(中国科技大学电子工程与信息科学系, 合肥 230027)

**摘要** 感性图像检索是一种新型的检索技术,这种检索具有较高的复杂性。一幅图像中能够使人们产生感性认识的可能只有部分区域,准确地找到感兴趣区域能有效地降低复杂度。作为图像基本特征的颜色对人的感觉有重要的影响,颜色的差异和对比使人产生了不同的情感。同时,图像的熵也反应了图像中包含信息量的大小,图像的熵也是引起人们产生感性认识的一个度量。提出了利用图像的颜色和熵提取感兴趣区域进行感性图像检索的方法,通过BP神经网络将感兴趣区域的颜色特征和熵映射到情感特征空间,具有较好的检索效果。

**关键词** 感性图像检索 感兴趣区域 熵 神经网络

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2006)04-0492-06

## Kansei Image Retrieval Based on ROI Extracted by Color and Entropy

LU Wei, NI Lin

(Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027)

**Abstract** Kansei image retrieval is a new kind of retrieval technology with high complexity. However, it's likely that only some parts of the image would attract people and produce affections. Color imposes a great impact upon the feeling as the basic feature of image, and the difference and comparison of the color would make people produce different kansei. Meanwhile, the entropy of the image also exhibits the information quantity and is a measurement of arousing people's kansei. In this paper, we present a method of kansei image retrieval utilizing the color and entropy to extract region of interest(ROI). Back propagation neural network is employed to map the color and entropy of ROI to affective feature space. Finally, we show some experimental results.

**Keywords** kansei image retrieval, region of interest (ROI), entropy, neural network

## 1 引言

近年来,数据库系统和计算机视觉的结合使得图像检索成为一个非常有前景的研究领域,人们根据自己的感知对图像进行理解,如何利用人类对图像的感知进行检索成为一个凸显出来的问题。因此,越来越多的研究开始从原来的以图像信息为核心的检索转变到以人类感觉为核心的检索。在图像信息中,颜色、形状和纹理以及一些空间信息都能让人产生一定的感觉。例如,红色、黄色和橙色会使人

感到暖意;原野会使人产生广阔的感觉<sup>[1]</sup>。利用人类对图像的感知进行检索即感性图像检索,利用图像中包含的信息来估计人们对图像产生的感觉,根据人们提供的反映情感的词汇从图像库中找到与之匹配的图像。感性图像检索技术将情感特征的组合与人们的主观情感联系起来,以用户方为核心。

然而,感性图像检索利用的大量图像特征大大增加了检索算法的复杂度,减少所需的图像特征就能够提高检索的效率。对于一幅图像来说,不一定所有的区域都可以使观众产生感觉,图像的某些区域可能就决定着人们的感受,并且不同的区域对所产生的感觉

基金项目:国家自然科学基金项目(60372059);安徽省自然科学基金项目(03042206)

收稿日期:2005-04-22;改回日期:2005-07-27

第一作者简介:陆伟(1981~),男,现为中国科技大学电子工程与信息科学系硕士研究生。从事感性图像检索等方面的研究。E-mail: greatjacklyu@yahoo.com.cn

有着不同的贡献。从图像中找到感兴趣区域并将其特征提取出来用于检索,这样就利用了图像中的有用信息而且减少了检索的底层特征。因此,利用图像的物理特性提取感兴趣区域能够大大降低映射的复杂程度<sup>[2]</sup>,提高检索速度,并使检索具有高效性。在图像的物理特性中,颜色信息是最基本也是最重要的特征,利用颜色对图像进行感兴趣区域的搜索是一种有效的方法<sup>[3]</sup>。另外,图像的熵也是衡量图像信息量的一个重要指标,采用图像的熵信息提取感兴趣区域也可以成为搜索感兴趣区域的一种解决方案。图 1 为基于感兴趣区域的感性图像检索模型。

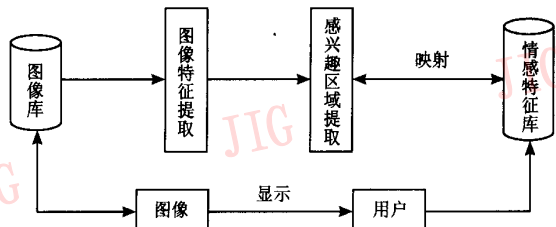


图 1 基于感兴趣区域的感性图像检索模型

Fig. 1 The model of kansei image retrieval based on ROI

## 2 感兴趣区域的提取方法

一幅图像包含了丰富的信息,颜色是图像最基本和最重要的特征。文献[2]提出了利用颜色特性的评估函数进行感兴趣区域提取的方法,这种方法提取到的感兴趣区域的准确性较低。然而,不同的颜色以及颜色的对比使人产生了复杂的印象和情感,在感性图像检索中起到非常关键的作用。除了图像的颜色信息,图像的熵是图像的另一重要特征。图像熵的大小反应了图像携带信息量的多少,它也是引起人们产生感性认识的一个度量。将图像的颜色和熵信息统一起来作为底部特征用于图像检索可以提高检索的准确率<sup>[4]</sup>,本文通过提取图像的颜色和熵进行比较筛选找到引起人们感性认识的关键区域,并与情感特征建立有效的联系,再通过映射从图像库中选取符合条件的图像,最终提供检索结果。

### 2.1 利用颜色特征提取感兴趣区域

图像的物理特征中颜色最直接影响着人们对图像产生的印象并且易于提取,用于检索时速度相对较快。颜色的对比反差对人们的感性认识有很大的影响,如绿色和黄色在一起体现出生机勃勃,红色和黑色的混合让人觉得暖和。因此,利用颜色的差异

来寻求感兴趣区域非常适合感性图像检索<sup>[5]</sup>。图像检索技术中基于颜色的检索多采用颜色直方图进行相似匹配,然而颜色直方图是颜色整体的统计特性,不同的图像可能产生相同的直方图,相似的图像之间的直方图可能差异很大。因此,采用颜色平均值的差异性来提取感兴趣区域。

颜色由 R、G、B 分量表示,使用水平和垂直两种分块方法。在这两个方向上提取颜色特征,得到 R、G、B 值在这两种分块模型中分割区域的平均分布,如图 2 所示。每幅图像按水平和垂直被平均分为  $L$  个区域,每个区域的像素数  $N_{\text{pixel}}$  为  $\frac{H \times W}{L}$ ,  $H$  和  $W$  分别表示图像的高度和宽度。设  $C$  表示像素的 R、G、B 值,  $C_{\text{RA}}$  (regional average) 表示这  $L$  个区域的颜色平均值,则

$$C_{\text{RA}} = \sum_{i=1}^{N_{\text{pixel}}} C(i) \quad (1)$$

从这  $L$  个区域中按 R、G、B 颜色特征的差异选取  $L/2$  个关键区域作为感兴趣区域,将这些感兴趣区域的颜色均值作为映射到情感空间的底层特征。

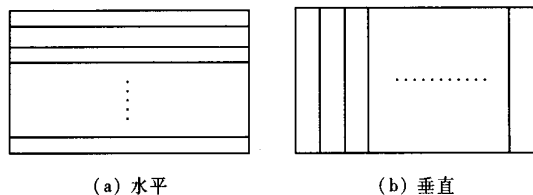


图 2 分块模型

Fig. 2 The model of divided parts

按颜色特征差异提取感兴趣区域的算法如下:首先,计算出  $L$  个区域中每个区域颜色的平均值  $C_{\text{RA}}$ ;比较各区域之间  $C_{\text{RA}}$  的差别  $D_{\text{RA}}$

$$D_{\text{RA}}(i) = \left| C_{\text{RA}}(i) - \frac{\sum_{j=1}^L C_{\text{RA}}(j)}{L} \right| \quad i = 1, \dots, L \quad (2)$$

最后,按照  $D_{\text{RA}}$  的大小选出  $D_{\text{RA}}$  最大的  $L/2$  个区域作为感兴趣区域。

$D_{\text{RA}}$  的值越大表示此区域的颜色平均值与其他区域的颜色平均值差别越大,即此区域与其他区域颜色的对比反差越大。利用这种方法选取的感兴趣区域具有较好的代表性,符合颜色差异影响感性认识的条件。因此,将感兴趣区域的  $C_{\text{RA}}$  值作为感性图像检索的一组底层特征,能够降低映射算法的复杂度,减少检索时底层图像特征的冗余度,保证映射的有效性和

准确性。这样从图像中提取出来的颜色特征总共应有  $3(C) \times L/2(C_{RA}) \times 2$  (分块模型) 个, 提取出的  $3L$  个颜色特征组成一组向量, 令其为  $v_c = \{v_{c1}, v_{c2}, \dots, v_{c3L}\}$ , 最终将这  $3L$  个图像的颜色特征按照各自的权值  $w_c = \{w_{c1}, w_{c2}, \dots, w_{c3L}\}$  映射到情感特征空间。

### 2.2 利用图像的熵提取感兴趣区域

图像的熵体现了图像灰度分布的统计特性, 它作为图像的基本信息之一, 自然影响到主观情感。图像的熵是图像包含信息量的尺度, 丰富的信息量赋予人们大量的感性认识。利用熵来描述图像的特征, 降低了提取图像特征的复杂性, 使得图像特征的维数大大降低。

图像的熵定义为图像中不同灰度的熵的累加

$$E = - \sum_{m=0}^{255} p_m \log(p_m) \quad (3)$$

其中,  $p_m$  为具有灰度  $m$  的像素在图像中出现的概率。

利用熵提取感兴趣区域的方法与利用颜色提取感兴趣区域的方法类似。仍然按照水平和垂直两种分块模型对图像进行分块, 对划分出来的  $L$  个区域分别求得它们的熵。熵越大, 则此区域的信息量越丰富, 对人产生主观感觉的影响也就越大。因此, 按照分块后各区域熵值的大小, 从这  $L$  个区域中挑选出  $L/2$  个熵最大  $E_{RA}$  的区域作为感兴趣区域, 再将感兴趣区域的熵作为映射的另一组底层特征, 得到  $3(C) \times L/2(E_{RA}) \times 2$  (分块模型) 个熵特征, 即  $3L$  个熵特征  $v_e = \{v_{e1}, v_{e2}, \dots, v_{e3L}\}$ , 同样按照各自的权值  $w_e = \{w_{e1}, w_{e2}, \dots, w_{e3L}\}$  映射到情感特征空间。

颜色和熵能够充分反应图像的物理特性, 将图像感兴趣区域的颜色特征  $v_c$  和熵特征  $v_e$  按照权值  $w_c$  和  $w_e$  的结合形成检索系统的底层特征  $v_{imagefeature} = \{v_c, v_e\}$ 。每幅图像共有  $6L$  个图像特征作为映射的底层特征, 前  $3L$  个特征为颜色特征, 后  $3L$  个特征为熵特征, 再通过神经网络的学习训练调整权值将图像特征映射到情感特征。

## 3 基于感兴趣区域的感性图像检索系统

提取感兴趣区域的图像特征后要实现的是图像特征和情感特征之间相互的映射, 由于情感特征和用户主观表述内在的差异性以及映射的不精确性, 建立映射比较复杂。神经网络有学习的能力和很强

的自适应能力, 采用神经网络能够处理复杂的映射关系<sup>[6,7]</sup>。感兴趣区域的图像特征、情感特征以及它们之间的映射组成了整个基于感兴趣区域的感性图像检索系统的基本模型。

### 3.1 神经网络的映射

从图像特征到情感特征的转换需要一个映射的过程, 这种映射是一个非线性的过程。神经网络特有的非线性适应性信息处理能力, 通过研究系统过去的记录进行训练归纳全部数据, 特别适用于这种复杂和多变量的系统。

BP 算法在神经网络中广泛用于模式识别, 将其特定的输出矢量与输入矢量联系起来, 网络结构如图 3 所示。网络的最低层是提取出来的图像特征, 最高层是情感特征。采用 BP 神经网络结构建立映射关系, 对系统进行训练学习调整权值实现图像特征到情感特征的映射。

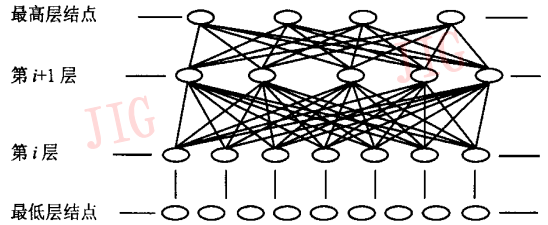


图 3 BP 神经网络结构

Fig. 3 The structure of BP neural network

### 3.2 情感特征的选取及量化

人们的情感表达多种多样, 很多表达情感的词语在很大程度上具有相关性。因此, 挑选了 18 对词语作为情感特征, 如表 1 所示, 这些词语覆盖了大多基本的感觉信息。

表 1 情感特征词语

Tab. 1 Kansei feature words

美丽的——丑陋的	潮湿的——干燥的	欢快的——阴郁的
喜欢的——讨厌的	有序的——无序的	复杂的——简单的
明亮的——阴暗的	宁静的——吵闹的	壮观的——渺小的
寒冷的——炎热的	密集的——分散的	浪漫的——不浪漫的
广阔的——狭窄的	生动的——单调的	生机勃勃的——荒凉的
清晰的——模糊的	现代的——古典的	印象深刻的——无印象的

情感特征是一个抽象的概念, 在检索时需要将其量化以便进行数值映射。为此, 以问卷调查形式获得用户对图像库中图像的评价, 再求其平均值建立情感特征数据库。用户的评价按照语义差别方法确定, 与情感特征建立度的联系。如果用户的感觉

完全符合该对词语的正面特征,则用户评价的程度为1,完全相反则为0,程度为0.5表示中性。这种方法利用用户的感受与情感特征词语的符合程度作为衡量的度,它的取值范围在0~1之间。

### 3.3 基于感兴趣区域的感性图像检索流程

基于感兴趣区域的感性图像检索是人们通过表达情感特征利用感兴趣区域的提取从图像库中找到与之匹配的图像,整个检索的流程如图4所示。

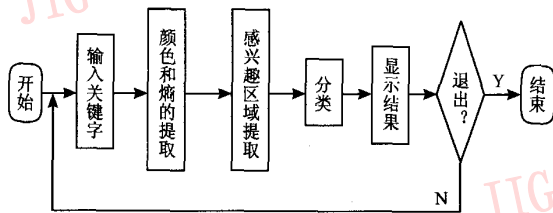


图4 基于感兴趣区域的感性图像检索流程图

Fig.4 The flowchart of kansei image retrieval based on ROI

检索的过程如下:首先由用户输入代表情感特征的关键字,对图像库中每幅图像进行感兴趣区域的提取,再通过神经网络的训练学习建立映射关系,获得与此情感特征相匹配的图像,最后由系统选择匹配程度最高的几幅图像显示给用户。由于感兴趣区域的提取减少了网络底层结点的维数,检索便能够快速有效地进行。

## 4 实验结果和分析

### 4.1 系统参数设置

实验选取300幅800×600大小的真彩色风景

图建立图像库,其中100幅图像作为样本进行训练学习,其余200幅用来测试。区域分割时取 $L=10$ ,则神经网络的最低层结点代表图像特征共有 $6L=60$ 个,最高层结点代表情感特征共有18个。本文使用3层的神经网络,输入层60个结点,输出层18个结点,隐层结点越多输出越精确。然而过度的训练学习也会使系统性能下降,因此通过实验效果寻找最佳的隐层结点,这里采用6个隐层结点。

感兴趣区域的选取采用问卷调查形式获得20名研究生对每幅图像区域分割后感兴趣区域的选择,将用户选择的结果统计后再选取具有最大值的前5个区域作为该图的感兴趣区域,在感兴趣区域提取实验中,将计算机提取出来的感兴趣区域与问卷调查获得的感兴趣区域的符合程度作为提取感兴趣区域的准确率。

情感特征数据库的建立仍然采用问卷调查形式获得20名研究生对图像库中图像的评价,再求其平均值得到神经网络训练学习的数据。在检索实验中,由神经网络学习训练获得对每幅图像情感特征度的评价,若其值与问卷调查获得的评价均处于 $[0,0.5)$ 或 $[0.5,1]$ 区间,则表明训练结果符合该情感特征的主观认识。将训练结果中符合各情感特征认识的图像个数占图像库的比例作为检索实验的准确率。

### 4.2 提取感兴趣区域的准确程度实验

利用图像的颜色和熵对图像库中作为测试的200幅图像进行感兴趣区域的提取,图5显示了分别采用颜色和熵提取感兴趣区域以及同时利用颜色和熵提取感兴趣区域的准确程度。

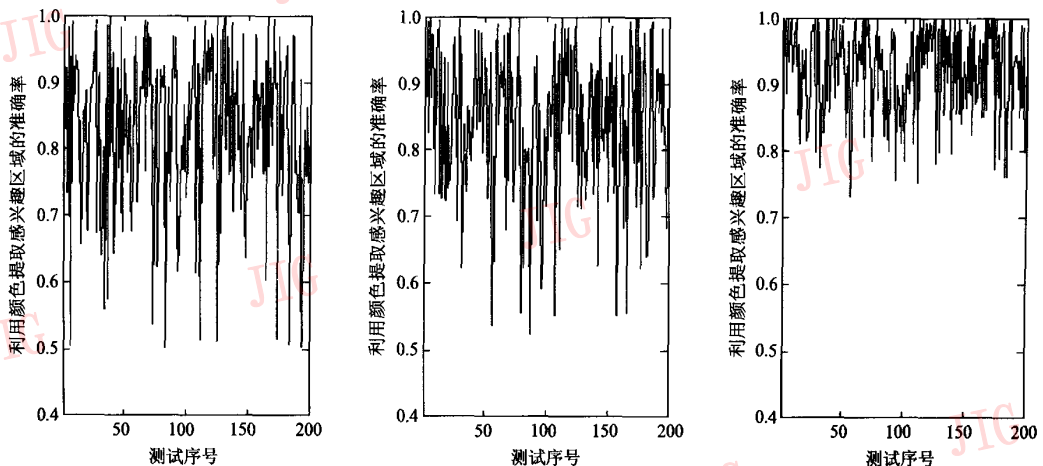


图5 感兴趣区域提取的准确率

Fig.5 Correct ROI extracted rate

由图所示,分别采用颜色和熵提取感兴趣区域的平均准确率为 83.17% 和 84.25%,而同时使用颜色和熵提取感兴趣区域的准确率达到 92.47%,提取出的感兴趣区域与人们的感性认识匹配的准确率很高。

本文方法既体现颜色的对比又考虑到了灰度分布,适合于感性图像检索。从实验结果可以看出,通过颜色和熵提取感兴趣区域具有互补作用,在利用颜色提取感兴趣区域准确率较低时,相对地利用熵提取的感兴趣区域具有较高的准确率。

#### 4.3 利用颜色和熵提取感兴趣区域的感性图像检索实验

确定模型参数后,在 Pentium IV 1.6GHz、256M SDRAM 机器上运行,通过网络学习,对每组情感特征词语进行检索。进行检索时每次网络输出层代表着图像库中每幅图对应的情感特征的符合程度,表 2 统计了每对情感特征 10 次检索结果准确率的平均值,由表中的统计数据可以看出主观性差异越

小,特征检索的结果越准确,如“浪漫的”这种感觉具有很强的主观性,不同的人感觉到的浪漫程度差异非常大,因此,该情感特征所得的检索结果相对较差。

表 3 列出了 4 种方法的检索效率。由表可见,分别采用颜色和熵提取感兴趣区域进行检索时准确率相对较低,同时利用颜色和熵提取感兴趣区域进行检索的平均准确率能够达到 77.63%。虽然仅采用颜色或熵提取感兴趣区域的检索时间较短一些,但与同时使用两种特征提取感兴趣区域检索的时间相差不大,同时利用两种特征能大大地提高检索准确率。通过 BP 神经网络映射颜色和频谱的方法<sup>[6]</sup>虽然也采用了非线性的神经网络做映射,但却没有有效地利用图像信息,没有消除图像信息的冗余,检索的时间比较长。因此,本文方法较之能取得更快的检索速度和更高的准确率,更具有优越性。

表 2 情感特征检索匹配率

Tab. 2 The matching rate of kansei features retrieval

情感特征	匹配率(%)	情感特征	匹配率(%)	情感特征	匹配率(%)
美丽的——丑陋的	88.55	潮湿的——干燥的	81.2	欢快的——阴郁的	71.55
喜欢的——讨厌的	82.6	有序的——无序的	76.45	复杂的——简单的	70.15
明亮的——阴暗的	87.8	宁静的——吵闹的	82.75	壮观的——渺小的	81.35
寒冷的——炎热的	82.15	密集的——分散的	73.05	浪漫的——不浪漫的	70.15
广阔的——狭窄的	80.15	生动的——单调的	70.85	生机勃勃的——荒凉的	79.35
清晰的——模糊的	78.25	现代的——古典的	70.65	印象深刻的——无印象的	70.4

表 3 检索效率的比较

Tab. 3 The comparison of retrieval efficiency

	颜色提取 方法	熵提取 方法	颜色和频谱 映射方法	颜色和熵 提取方法
检索时间(s)	1.0	1.1	3.2	1.5
检索准确率(%)	67.78	68.35	71.33	77.63

用户输入情感特征检索后,系统从图像库中提供最接近该情感特征的 10 幅图像给用户。图 6 给出了情感特征“宁静的”分别采用 3 种方法的检索结果,可以看出分别利用颜色和熵检索得到的结果最后两幅均不完全符合此情感特征,而同时利用颜色和熵检索的结果能够基本符合此特征。

## 5 结论

赋予计算机理解人类情感的能力成为图像检索的一项有前景的技术,感性图像检索的研究能够有效地将图像和人们的情感联系起来。由于图像内容的丰富性和主观感受的差异性,特征的提取和映射使检索结构较为复杂。本文提出了利用颜色和熵从图像中提取感兴趣区域并将其与情感特征空间建立起映射关系的方法,通过 BP 神经网络建立了映射模型。建立的这种感性图像检索系统减少了映射底层的冗余,具有较为准确的匹配程度,并在实验中取得了较好的检索效果。



图 6 基于感兴趣区域的感性图像检索结果

Fig. 6 The retrieval result of kansei image retrieval based on ROI

参 考 文 献 (References)

- 1 Kuroda Kazuhiro, Masafumi Hagiwara. An image retrieval system by impression words and specific object names-IRIS [ J ]. Neurocomputing, 2002, 43(1) : 259 ~ 276.
- 2 Shoji Tanaka, Masayuki Inoue, Michitoshi Ishiwaka, et al. A method for extracting and analyzing "Kansei" factors from pictures[A]. In: IEEE First Workshop on Multimedia Signal Processing [ C ], Princeton, New Jersey, USA, 1997:251 ~ 256.
- 3 The Color Science Association of Japan. Color Science Handbook [ M ]. Tokyo, Japan: Tokyo University Press, 1991.
- 4 Se Yoon Jeong, Kyuheon Kim, Byung Tae Chun, et al. Entropy and color correlation features for image indexing[A]. In: IEEE SMC '99 Conference Proceedings[C], Tokyo, Japan, 1999, 2: 895 ~ 899.
- 5 Toshiyuki Iwaoka, Hideyuki Kobayashi, shunji Ota. Digital safari guidebook with image retrieval [ A ]. In: Proceedings of IEEE International Conference on Multimedia Computing and Systems[C], Centro Affari, Firenze, Italy, 1999, 2: 1011 ~ 1012.
- 6 Takashi Hayashi, Masafumi Hagiwara. An image retrieval system to estimate impression words from images using a neural network[A]. In: Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation [ C ], Orlando, Florida, USA, 1997, 1: 150 ~ 155.
- 7 Nadia bianchi, Luc Berthouze, Toshikazu Kato. Supervised self-organization of user's kansei model for image retrieving[A]. In: Second International Conference on Cognitive Technology. Humanizing the Information Age Proceedings[ C ], Aizu-Wakamatsu City, Japan, 1997:185 ~ 189.