

基于内容的情感图像获取模型^{*}

王上飞 薛佳 王煦法

(中国科学技术大学计算机科学技术系 合肥230027)

摘要 随着信息技术的迅猛发展,情感信息处理已成为21世纪人工智能领域所面临的重要挑战之一。其中,情感图像获取的研究属于多学科交叉研究,本文以形容词作为处理对象,借鉴认知心理学、绘画艺术的研究成果,在图像内容的基础上,提出了一种包括公共情感和个性化情感的情感图像获取模型。在该模型中,借鉴心理学中的“维量”思想,建立情感空间;同时,抽取图像的主色调、不变矩、颜色和灰度分布,这些较容易引起情感变化的特征作为图像的视觉特征,建立图像的特征空间;采用支持向量机的方法建立图像的低层特征空间到用户的高层情感空间之间的映射,自动注释用户未曾评估的图像,实现了图像情感注释,在情感空间进行公共情感检索,快速获得用户情感信息,在此基础上,采用可视化交互式遗传算法实现因人而异的个性化情感检索,该模型应用于风景图像的情感检索,取得了较好的实验结果。

关键词 基于内容的,公共情感,个性化,图像获取

Content-Based Emotion Image Retrieval Model

WANG Shang-Fei XUE Jia WANG Xun-Fa

(Department of Computer Science and Technology of USTC, Hefei 230027)

Abstract With the rapid development and popularity of the Internet and multimedia, emotion information processing has become a great challenge faced by artificial intelligence today. The research of emotion image retrieval is an inter-sectional research. Inspired from the research products of psychology and painting, a content-based emotion image retrieval model which includes common emotion and individual emotion has been proposed in this paper. First, based on the idea of “dimension” from psychology, an emotion space is constructed. Second, dominant colors, moment invariants, color and gray compositions, which can stimulate the emotion of human very easily, have been extracted from images to construct the feature space. Then, support vector machines are used to map images from the low level feature space to the high level emotion space, and automatically annotate unevaluated images based on user’s common emotion. After that, the common emotion image retrieval has been indexed in the common emotion space and quickly grasps user’s subjectivity inherent in vision, while an interactive individual emotion image retrieval using visualized interactive genetic algorithm is presented to adapt to individual variation and improve accuracy of the retrieval results. Based on image content, an emotion scenery image retrieval system has been realized. The experimental results demonstrate the effectiveness of our approach.

Keywords Content-based, Common emotion, Individual emotion, Image retrieval

1 引言

随着信息技术的日益发展和普及,信息化社会中的用户对信息技术提出了许多更高的新要求。“以人为本”是这些众多新需求的核心。而情感信息处理,即如何处理用户的情感信息,使计算机具有识别和表达情感的能力,达到人机之间的“和谐、自然地”交流,是实现真正的“以人为本”的信息技术的一个关键,也是当前人工智能领域所面临的重要挑战之

—[1,2]。

心理学家指出人的情感具有两方面的属性^[3,4],即为公共情感(Common Emotion)和个性化的情感(Individual Emotion)。公共情感可以认为是一种“平均的”或者说“相对客观的”情感,虽然不是所有的人都一样,但对于大部分人来说是类似的,而个性化的情感则因人而异。这是人类情感信息处理的一个重要特征,然而目前的大部分情感信息管理系统没有同时考虑情感的两方面属性。

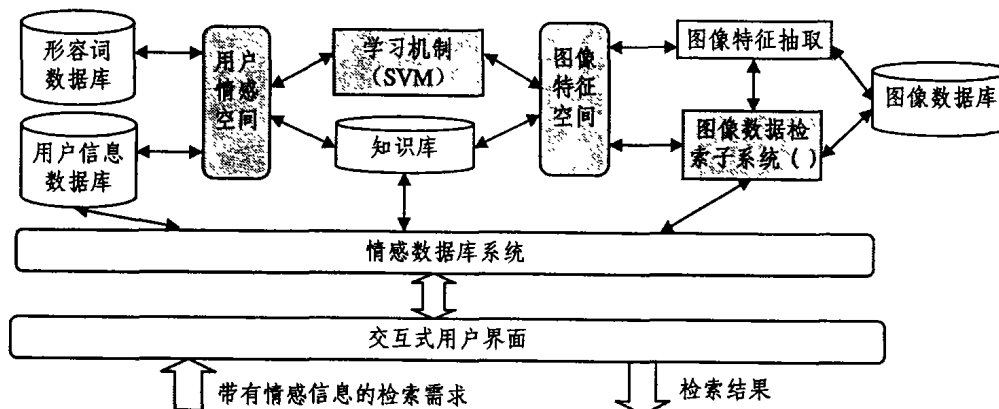


图1 基于内容的情感图像获取框图

^{*} 基金资助:973资助课题-图像、语音、自然语言理解与知识挖掘(项目编号:G1998030500);校青年基金—感性信息处理及其在多媒体中的应用研究。

针对这一问题,本文在图像内容的基础上,提出了一种包括公共情感和个性化情感的情感图像获取模型,如图1所示。首先,基于心理学中情感的“维量”思想,建立公共情感空间;然后,借鉴绘画艺术,抽取最能引起人的情绪变化的特征(如:风景图像的主色调、不变矩、颜色和灰度分布)构造图像特征空间;采用支持向量机实现图像低层特征空间到用户高层情感空间的映射,根据用户的公共情感自动注释用户未曾评估的图像;当用户提出情感检索的要求,首先在情感空间中进行公共情感图像获取,迅速捕捉用户的情感信息,在此基础上,采用了可视化交互式遗传算法实现个性化的情感检索,适应每个人的不同需求。将此模型应用于风景图像的情感检索,实验结果证明了模型的有效性。

本文第1节概述相关研究背景;第2~5节详细阐述我们提出的情感图像获取模型;第6节给出相关的实验结果;最后是结束语。

2 图像特征抽取

本文的目的是实现图像的情感检索,也即在图像内容的基础上融合用户的情感信息,建立图像的低层特征与用户的高层情感之间的联系,因此,希望提取一些较容易产生情感变化的视觉特征。对此,艺术家们在绘画艺术中已经进行了深入的研究和总结,指出^[5],绘画艺术主要是以笔、墨、色、形、位置的远近,光与影的配置、虚实与疏密的分布以及艺术神情风韵来表达它的艺术情感。联系到图片的视觉特征就是颜色(色),形状(形)以及颜色的分布和灰度分布(包含位置的远近、光与影的配置、虚实与疏密的分布),下面将提取这些特征作为图像的视觉特征。

2.1 主色调

颜色是图像内容组成的基本要素,是人识别图像的主要感知特性之一。有研究表明不同的颜色及颜色组合可以激发人们不同的情绪,比如红色象征热情,绿色充满生机,图像的颜色具有较强的感情色彩^[6]。这里采用 HSV 空间的色直方图来描述图像的整体颜色特征。一幅图像的颜色一般非常多,尤其是真彩色图像,直方图矢量的维数会非常多。如果对 HSV 空间进行适当的量化后再计算直方图,则计算量要少得多。因此根据对颜色模型的大量分析,按照人的颜色感知将 H 、 S 、 V 三个分量进行非等间隔的量化,如式(1)所示^[7]。

$$H = \begin{cases} 0 & \text{if } h \in [316, 20] \\ 1 & \text{if } h \in [21, 40] \\ 2 & \text{if } h \in [41, 75] \\ 3 & \text{if } h \in [76, 155] \\ 4 & \text{if } h \in [156, 190] \\ 5 & \text{if } h \in [191, 270] \\ 6 & \text{if } h \in [271, 295] \\ 7 & \text{if } h \in [296, 315] \end{cases}$$

$$S = \begin{cases} 0 & \text{if } s \in [0, 0.2] \\ 1 & \text{if } s \in [0.2, 0.7] \\ 2 & \text{if } s \in [0.7, 1] \end{cases}$$

$$V = \begin{cases} 0 & \text{if } v \in [0, 0.2] \\ 1 & \text{if } v \in [0.2, 0.7] \\ 2 & \text{if } v \in [0.7, 1] \end{cases} \quad (1)$$

按照以上的量化级,把3个颜色分量合成为一维特征矢量:

$$I = 9H + 3S + V \quad (2)$$

这样, H 、 S 、 V 三个分量在一维矢量上分布开来。根据(2)式, I 的取值范围为 $[0, 1, \dots, 71]$, 计算 I 获得72柄(bin)的一维直方图,由此求出每幅图片中主色调的个数和前5个主色调及其所

占的比例。

2.2 形状

不同的形状也可能引起不同的感受,比如,心形给人“浪漫,温馨的”感觉。此处采用七个不变矩来描述图像的形状特征。不变矩函数是一种被广泛用来描述形状特征的参数,它具有旋转、尺度和平移不变性,具体计算可参照文[8]。

2.3 分布特征

这里的分布特征主要指颜色和灰度分布,以 RGB 颜色在水平、垂直、辐射方向以及环行的平均分布作为图像的颜色分布特征,如图2所示。

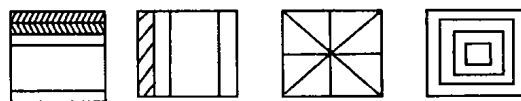


图2 颜色分布

同时,我们按照下述方法抽取灰度分布特征。由式(3)求出每个像素点的亮度,根据亮度值是否大于整幅图像的亮度均值,如果大于,记为1;否则,记为0。将彩色图像转化为二值图像。再将图像均匀地分割成 8×8 的小块,求每一块中,像素点值为1的个数,并将其归一化,作为图像的灰度分布特征。

$$H = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (3)$$

通过上面两节的特征提取,每幅图像共有196维($J=196$)个特征,分别是颜色个数(1)、五个主色调及其所占比例(10)、颜色分布特征(114)、七个不变矩(7)和灰度分布(64)。

3 情感空间

人类的感情状态向量到底有多少元素,或者说我们人类到底一共有多少种感情,是一个尚未得到充分研究的问题。中国古代有七情六欲之说,现代心理学也有不同的说法,但有一点是大多数心理学家所普遍赞同的,即人的情感分为基本情感和复合情感,基本情感是情感的基本元素,由其构成各种复合情感,这就是情绪心理学中“情绪维量”思想^[9]。维量的思想意味着可以采用适当的方法建立一个情感空间,每一个情感都可以看成该空间的一个向量。

要建立情感模型必需对表达情感的形容词(如喜欢的、浪漫的等)进行量化和分析,因此情感模型的建立主要包括三步:一是收集表达心理的形容词;二是做语义量化实验,由被调查的对象评价图像,收集数据,建立用户情感数据库;三是采用因子分析^[10]的方法对数据库中的数据分析,建立情感空间。

首先针对图像,我们精心挑选了18对形容词,如表1所示。

表1 形容词对

1. 喜欢的一不喜欢的	10. 整齐的一杂乱的
2. 美丽的一丑陋的	11. 清晰的一模糊的
3. 协调的一不协调的	12. 宁静的一不宁静的
4. 浪漫的一不浪漫的	13. 印象深刻的一平淡的
5. 舒适的一不舒适的	14. 轻松愉快的一压抑的
6. 热烈的一冰冷的	15. 富于变换的一单调的
7. 温馨的一清凉的	16. 充满生机的一荒凉的
8. 明亮的一阴暗的	17. 视野宽广的一狭窄的
9. 柔和的一不柔和的	18. 暖色调的一冷色调的

针对这18($N=18$)对形容词,由用户对样本图像进行评价,评价分五个等级,例如形容词对“喜欢的一不喜欢的”,用户的评价可以是非常喜欢,有些喜欢,中性,有点不喜欢,非常

不喜欢,分别对应0.0,0.25,0.5,0.75和1.0。这就是语义量化实验。通过实验收集数据,形成用户的情感数据库。我们的图像库中总共有1300幅风景图像,选出其中的206($M=206$)幅作为样本提供给用户评测,用户是180名大学一年级的学生,其中男生120人,女生60人。然后采用因子分析的方法来处理这些数据,具体步骤如下:

设用户 k 对图像 m 的第 n 对形容词的评价为 z_{mnk} ,按照式(4),求出平均值,并按式(5)将其标准化,得矩阵 $X(206 \times 18)$ 。

$$y_{mn} = \frac{1}{K} \sum_{k=1}^K z_{mnk} \quad (4)$$

$$x_{mn} = \frac{y_{mn} - \bar{y}_n}{s_n} \quad (5)$$

其中:

$$\bar{y}_n = \frac{1}{M} \sum_{m=1}^M y_{mn}, s_n^2 = \sum_{m=1}^M (y_{mn} - \bar{y}_n)^2$$

对矩阵 X 做因子分析,如式(6)所示:

$$X = FA' + UD \quad (6)$$

式(6)中 F 为公共因子矩阵, A 为载荷矩阵, U 为独特因子, D 是独特因子的权值。通过主元分析法求出公共因子 F 以及载荷矩阵 A ,将原先的18维的空间降维至 L ,形成 L 维的正交情感空间,则 F 矩阵的第 m 行 $f_m = (f_{m1}, f_{m2}, \dots, f_{mL})$ 对应图像 m 在情感空间的坐标, A 矩阵的第 n 行 $a_n = (a_{n1}, a_{n2}, \dots, a_{nL})$ 对应形容词 n 在 L 维空间的坐标。

4 情感注释

通过因子分析,得到了 L 维的情感空间。每一幅样本图像和每一个形容词都可以看成该空间的一个向量。而图像库中总共有1300幅图像,样本图像只有206幅,剩余的1094幅图像也希望能够映射到情感空间,这是图像的情感注释过程,主要实现的功能包括三个方面:一、建立图像低层特征空间到用户高层情感空间的映射;二、通过学习记忆用户的情感特性;三、自动对未曾评价过的图像进行情感注释。由于用户评估的样本图像不能过多,否则会引起用户疲劳,因此情感注释问题是一个小样本学习问题。

近年来兴起的支持向量机(Support Vector Machine, SVM)算法可以较好地解决小样本学习问题^[10]。这里采用函数估计的SVM来实现情感注释^[11]。图3给出了情感注释的框架,SVM的输入为经过预处理的样本图像特征,输出为样本图像在情感空间的坐标,经过学习,权值记忆了图像特征与用户情感之间的联系,当输入一幅新的图像时,网络输出该图像的在情感空间的对应值,从而实现图像的情感注释。

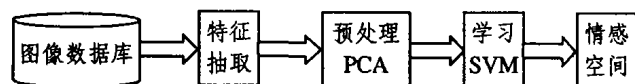


图3 情感注释框图

通过注释,图像库中每一幅图像都对应了情感空间的一个向量;而每一个形容词通过因子分析也对应着情感空间的一个向量,这些向量将用于图像的情感检索。

5 情感图像获取

一般来说,图像获取是在图像特征空间中的一个搜索过程,而情感化的图像获取要求在检索中可以反映人的主观情感。人类的主观情感具有模糊性,这种心理模糊性的量化称为

心理模糊量,心理模糊量是波动的(个性),但是这种波动也具有或表现出相对的稳定性(共性)。我国学者马谋超^[3]曾经对心理模糊量的相对稳定性和动态特征做过详细的心理实验和分析。通过实验分析,得出不同民族和不同文化水平的人对同一个模糊概念的隶属函数的曲线形状基本相同,这反映了模糊概念具有一定的稳定性。同时,实验也指出心理模糊性函数随条件的改变而发生变化。日本的印象工作组在对印象工程(Impression Engineering)的研究中,也指出情感具有两方面的属性^[4],即为公共情感(Common Emotion)和个性化的情感(Individual Emotion)。该研究组指出:公共情感可以认为是一种“平均的”或者说“相对客观的”情感,虽然不是所有的人都一样,但对于大部分人来说是类似的,而个性化的情感则因人而异。因此,本文提出了公共情感和个性化情感相结合的情感图像获取模型。

5.1 公共情感检索

本文中通过心理实验的统计分析建立了情感模型,并通过情感注释,将图像库中的所有图片映射到情感空间。对实验结果的统计分析本身包含了取平均这样一个过程,因此可以看成是对公共情感的分析,在这个情感空间中进行检索即形成了公共情感检索。检索流程如图4所示,首先用户提出检索要求,比如“检索让人感到轻松愉快的图片”,给出形容词“轻松愉快的”;系统在情感空间计算每幅图像与该形容词的相似度,找出相似度最大的12幅图像,作为结果显示给用户。如果用户对此结果已经满意,则询问用户是否退出;否则,将此结果作为初始群体,进入个性化的情感检索。图4中用向量夹角作为相似度的定义,即设 $f_m = (f_{m1}, f_{m2}, \dots, f_{mL})$ 和 $a_n = (a_{n1}, a_{n2}, \dots, a_{nL})$ 分别为图像 m 和形容词 n 在情感空间的坐标,则图像 m 到形容词 n 之间的相似度 d_{mn} 为:

$$d_{mn} = \frac{a_n \cdot f_m}{|a_n| |f_m|} \quad (7)$$

从后面的实验中,可以看出这种相似度的定义较好地反映了情感空间中向量之间的相似程度。

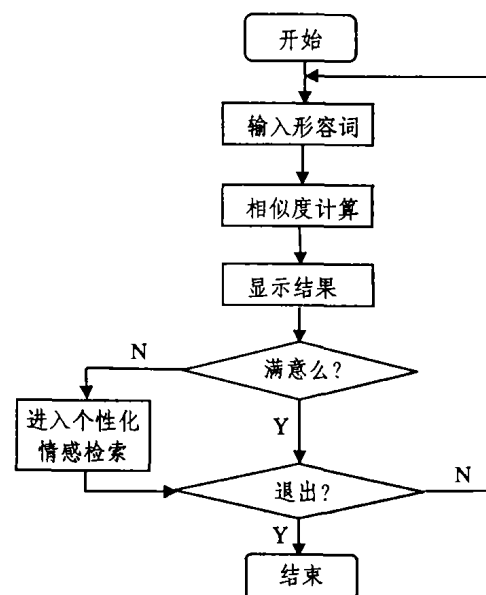


图4 公共情感图像获取流程图

公共情感检索是基于注释单元中存储的用户公共情感信息,因而检索的结果是一种“平均的”情感,要实现因人而异的个性化情感,必须采用一种动态的、交互的、自适应的检索机制提高检索的准确性。下面采用可视化交互式遗传算法(Vi-

sualized Interactive Genetic Algorithm, VIGA)来实现个性化的情感图像获取。

5.2 个性化情感检索

VIGA 是一种改进的 IGA,其采用可视化技术克服 IGA 的关键技术问题,即用户疲劳问题。该算法的基本思想^[12]是将人的2维空间上的宏观判断优势和遗传算法 J 维空间上的快速搜索结合起来。系统不仅将检索的结果图像显示给用户,同时,通过从 J 维到2维的降维,将以往搜索到的所有个体均显示给用户,个体适应度的高低以颜色的深浅来表示(如图6所示)。用户评估图像,给出适应度值,并且可以在适应度高的个体密集处选择一个新的好个体加入遗传种群,替换掉 GA 搜索结果中适应度最低的个体。这样遗传群体中好个体的数目增加了,整个算法就会更快地收敛,从而减轻了用户疲劳。

在我们的系统中,用户首先使用公共情感检索搜索图像。如果对结果不满意,则开始个性化的情感检索。公共情感检索的结果作为 VIGA 的初始群体。用户对显示的每幅图像进行评价,根据所显示图像与用户所要检索的目标图像之间的相似程度进行打分,给出适应度,同时在适应度高的个体密集处选择一个新的好个体替换掉 GA 搜索结果中适应度最低的个体。系统通过选择和交叉产生新一代群体,显示给用户,不断重复,直至用户满意。

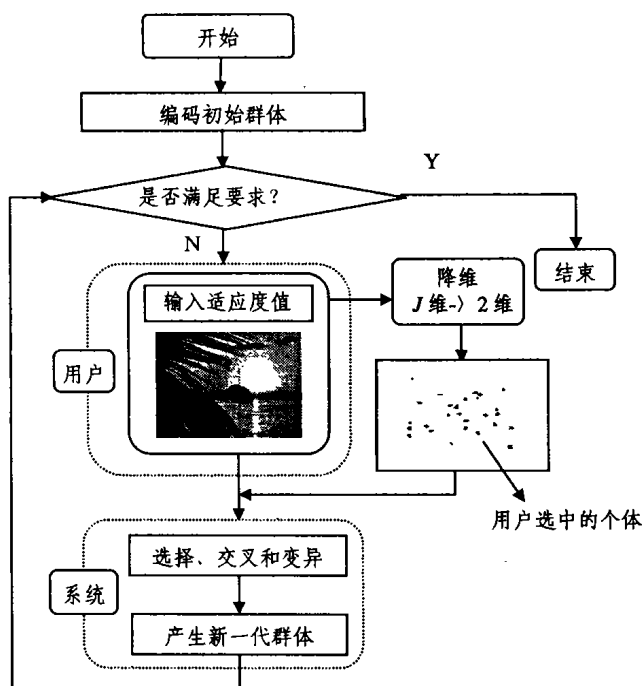


图5 个性化情感图像获取

6 实验分析

6.1 情感空间实验

用户在进入系统评估之前必须进行注册,以记录用户信息,如性别、年龄、民族和兴趣爱好等,这些信息可以用来对用户进行分类。本实验中参加实验的为大一学生,基本上处于一个年龄阶段,都是汉族。因此我们将用户按性别简单地分为两类,男性用户和女性用户。

然后,用户可以通过图样全览,对样本库中的图片有个了解,即而开始评估。用户的评估数据作为用户的情感信息存入情感数据库中,然后对这些数据进行因子分析,分析后的载荷矩阵如表2所示,累计贡献率为87.78%。

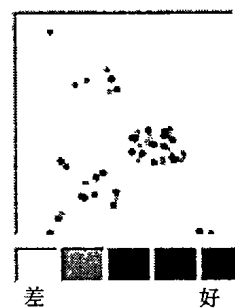


图6 数据可视化界面

表2 因子载荷矩阵

	λ_1	λ_2	λ_3	λ_4
1	0.894482	-0.362472	0.106824	-0.0462042
2	0.932427	-0.25295	0.0973704	-0.0412107
3	0.907888	-0.287328	0.00506772	-0.034545
4	0.884263	0.157373	0.0106695	0.299862
5	0.96292	-0.0457377	-0.0359891	0.144142
6	0.140618	0.716816	0.583489	-0.125139
7	0.870879	0.310206	-0.0971494	0.313076
8	0.706511	0.327731	-0.0832736	-0.511208
9	0.910451	0.0499248	-0.142059	0.253018
10	0.798501	0.154013	-0.337844	-0.130346
11	0.697604	0.14299	-0.243628	-0.565553
12	0.758127	-0.31223	-0.37056	0.0936458
13	0.854406	-0.295363	0.169029	-0.102517
14	0.924435	0.226553	-0.0280306	0.0672666
15	0.421595	-0.395991	0.738923	0.0189711
16	0.748964	0.370119	0.215694	0.124278
17	0.330314	-0.640285	0.311852	-0.190736
18	0.285729	0.826028	0.178735	-0.00744723

求得因子载荷矩阵后,也就将原来的18维的形容词空间降为4维,每个形容词可以看成该空间的一个矢量,各维坐标即为载荷值。由于是用主成分法求的载荷,因此可以同时得到因子得分矩阵,即每幅图片也同时降至4维空间,这样情感空间就成为一个统一的空间,图片和形容词都可以看成该空间中的矢量。

6.2 相似度实验

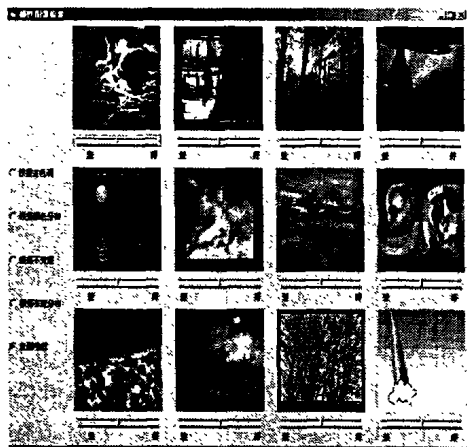
相似性的度量如5.1中(7)式所定义。实验方法如下:将用户按性别分为两类,男性用户和女性用户,分别求这两类用户的因子载荷和得分矩阵,然后按照式(7)计算出图片与每一个形容词相似度,将相似度和原始的评估数据分别排序,对于每一个形容词对分别找出排序后的前12幅图片与原始评估的前12副图片以及后12幅图片与原始评估的后12副图片的吻合个数,实验结果如表3。由表可知,平均吻合率可以达到73.73%。这表明 dm_n 可以较好地反映情感空间中向量的相似度。

6.3 情感检索实验

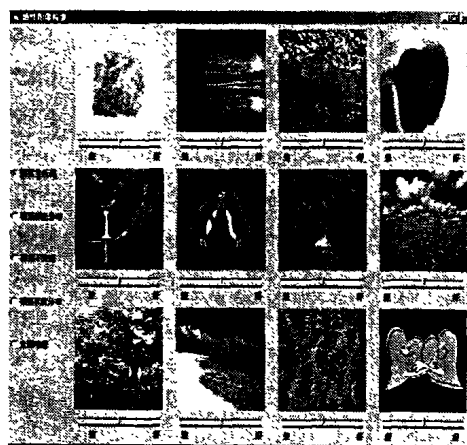
情感空间建立之后,通过情感注释将每幅图像映射成情感空间的一个向量,当用户提出检索要求时,根据相似度,从图像库中搜索出最接近的12幅图像,下面列出了女性用户的一些检索结果,分别是“阴暗的”,“浪漫的”,实验中用户对此检索结果较满意。

表3 相似度实验结果

形容词对序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
男性用户	10	10	9	9	7	8	10	10	8	9	10	9	8	8	9	9	8	9
吻合率:74.77%																		
	8	10	9	9	9	10	10	10	8	7	9	11	9	9	10	9	8	8
女性用户	10	10	7	9	10	9	9	10	9	8	11	8	9	7	11	6	10	9
吻合率:72.68%																		
	9	6	8	7	7	8	8	10	9	10	9	7	9	10	8	9	9	9



“阴暗的”



“浪漫的”

图7 公共情感图像获取结果

为进一步检验本文方法的有效性,我们请了二十位20岁左右的学生使用本系统,分别进行公共情感图像获取和个性化情感图像获取实验,在个性化情感图像获取中根据 VIGA 初始群体的不同产生方式,又分为两种,初始群体随机产生方式的个性化情感图像获取和基于公共情感检索结果的个性化情感图像获取。实验结果如表4所示。

表4 情感图像获取实验

	公共情感 图像获取	个性化情感图像 获取(基于公 共获取结果的)	个性化情感图像 获取(随机的)
平均交互次数	0	4.54	5.98
准确率	40%	85%	75%
成功率	78%	95%	93%

表4中准确率和成功率的定义分别如下:

$$\text{准确率} = \frac{\text{用户满意的图像个数}}{\text{所有检索到的图像个数}} \quad (11)$$

$$\text{成功率} = \frac{\text{能检索到满意图像的用户人数}}{\text{所有用户人数}} \quad (12)$$

由表4可以看出,相比于公共情感图像获取,个性化图像获取的准确率和成功率都有所提高,特别是准确率,这表明用户不仅可以检索到满意的图像,并且获得更多满意的图像。在两种个性化情感图像获取中,基于公共情感检索结果的图像获取的交互次数较少,而准确率和成功率较高。可见公共情感图像获取可以迅速捕捉用户的情感信息,减轻系统在线学习的压力;而个性化的情感图像获取可以提高检索的准确率,使检索的结果更加符合用户的个性。同时平均交互次数为4.54,也是用户可以接受的。

结束语 如何处理用户的情感是多媒体数据库中的一个重要问题。目前的大部分情感信息管理系统没有同时考虑情感的两个方面:公共情感和个性化情感。本文在基于图像内容的基础上,提出了一种包括公共情感和个性化情感的情感图像获取模型,检索符合用户情感需求的图像。首先,借鉴心理学中的“维量”思想,建立公共情感空间;接着,抽取较容易引起情感变化的特征作为图像的视觉特征,建立图像的特征空间,并采用支持向量机的方法根据用户的公共情感自动注释图像,这样即可在情感空间进行公共情感检索,快速捕捉用户情感。此外,采用可视化交互式遗传算法实现动态、交互、自适应的个性化的情感检索。该模型应用于风景和服装图像的情感检索,取得了较好的实验结果。本文所提出的模型也可用于服装设计、网页设计、电子贺卡设计等其它领域。

参考文献

- 1 Nagamachi M. Emotion engineering. Kaibundo publishing, Tokyo, Japan, 1997
- 2 Picard R W. Synthetic emotion. IEEE Computer Graphics and Applications, 2000, 20(1): 52~53
- 3 马谋超. 心理学中的模糊集分析. 贵州科技出版社, 1993
- 4 Norihiko M, Yasuo N, Hideo J, et al. Talking of the Dreams of "Engineering of Impression", Impression'98
- 5 李璞琨. 心理学与艺术. 首都师范大学出版社, 1996
- 6 汪田明. 浅析色彩的情感语言. 株洲工学院学报, 2000, 15(2): 18~19
- 7 曹莉华, 柳伟, 李国辉. 基于多种主色调的图像获取算法研究与实现. 计算机研究与发展, 1999, 36(1): 96~100
- 8 Hsia T C. A Note on Invariant Moments in Image Processing. IEEE Trans. On SMC, 1981, 11(12): 831~834
- 9 [美]斯托曼 K T 著, 张燕云译. 情绪心理学. 辽宁人民出版社, 1987
- 10 Vapnik V N. Statistical Learning Theory. Wiley, New York, 1998
- 11 王上飞, 王煦法. 基于支持向量机的图像情感语义注释和检索算法的研究. 模式识别与人工智能, (已录用)
- 12 Hideyuki T. Visualized IEC: Interactive Evolutionary Computation with Multidimensional Data Visualization[A]. In: IEEE Intl. Conf. on Industrial Electronics, Control and Instrumentation[C]. Nagoya, Japan, 2000, 4: 22~28