

基于情感语义的图像注释与检索

武 频¹, 陶 聪¹, 朱永华¹, 颜宏杰¹, 高宏浩²

(1. 上海大学 计算机工程与技术, 上海 200444;

2. 上海大学 计算中心, 上海 200444)

摘 要: 图像情感语义的注释与检索起步不是很久, 涉及了很多学科的综合知识, 需要对心理学、计算机科学、生理学等各个学科的知识 and 前沿成果都有比较深入的了解, 这个领域的研究充满了挑战和难度, 同时其后续研究也存在着很大的可能性。情感语义是图像语义的最高层次, 在图像情感语义注释和检索中起着很重要的作用。文中具体研究了底层特征提取中现有的一些常用方法, 构建出图像的底层特征数据库。应用因子分析法对实验收集的用户情感数据库进行分析, 构建出情感空间作为图像情感语义注释的基础。首次将 LSSVM 应用于图像情感语义注释上, 实现了图像底层特征到高层情感语义的映射。然后通过相似度计算, 在情感空间中完成图像的情感检索。实验结果取得了不错的用户满意度。

关键词: “维量”思想; 图像检索; 情感语义注释; 因子分析; LSSVM 神经网络

中图分类号: TP39

文献标识码: A

文章编号: 1673-629X(2015)10-0013-06

doi: 10.3969/j.issn.1673-629X.2015.10.003

Image Annotation and Retrieval Based on Emotion Semantic

WU Pin¹, TAO Cong¹, ZHU Yong-hua¹, YAN Hong-jie¹, GAO Hong-hao²

(1. School of Computer Engineering and Science, Shanghai University,

Shanghai 200444, China;

2. Computer Center, Shanghai University, Shanghai 200444, China)

Abstract: Image annotation and retrieval involves comprehensive knowledge of many disciplines, needs to be clear in the heart about psychology, physiology, computer science and other subjects of knowledge and cutting-edge results. The research in this field is full of challenge and difficulty, but at the same time its follow-up studies also exists a lot of possibilities. Emotion semantic is the highest level of image semantics, it is extremely important in image emotion semantic annotation and retrieval. In this paper, study some existing commonly used methods of underlying feature extraction, build up the image characteristics of the underlying database. Apply factor analysis method to analyze the collection of user emotional database, and build emotional space as the basis of image emotional semantic annotation. LSSVM is applied to image semantic annotation for the first time, realizing the image characteristics of the underlying semantic mapping to the top. Then, through the calculation of similarity in the emotional space, complete the image retrieval. The experimental results have achieved good user satisfaction.

Key words: “Dimensional” thinking; image retrieval; emotional semantic annotation; factor analysis; Least Squares Support Vector Machine (LSSVM) neural network

0 引 言

人之所以能成为主宰世界的精英, 是因为人具有情感这一标志人类智能的能力。看, 世界万物均是图像景象。“眼见为实, 耳听为虚”、“一览无余”, 都体现出图像本身蕴含着丰富的情感, 体现出图像在信息传递中是存在特别之处的。长久以来, 人类都一直在努

力研究, 尝试设计出一台机器人, 它能拥有人类的思维, 能代替人类进行各种各样的工作。如今, 计算机技术越来越发达, 逐渐实现由计算机认识世界。由于主观性是情感的主要特点, 以对人类的认知机理研究为基础来研究图像情感语义分析技术是很有必要的^[1]。

收稿日期: 2014-12-12

修回日期: 2015-03-18

网络出版时间: 2015-08-26

基金项目: 上海市科学技术委员会资助项目 (14590500500); 上海市自然科学基金 (15ZR1415200); 上海高校青年教师培养资助计划 (ZZSD13008)

作者简介: 武 频 (1975-), 女, 博士后, 副教授, 研究方向为 CFD 数值计算、高性能计算、图像处理、深度学习等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150826.1604.088.html>

因此,在图像检索和图像理解领域,对图像情感语义分析技术里如何建立情感模型、如何提取有用的图像特征、如何选择映射算法等问题的研究,对如何通过建立图像底层特征和高层情感语义之间的关系,提升情感识别的准确率,从而让计算机更加智能,是具有一定的理论意义和学术价值的。

文中建立的情感模型以心理学中“情绪维量”为基础,通过对情感形容词的语义量化,并利用因子分析法分析收集到的用户数据来建立情感空间。学习了解了如何在情感空间中通过相似性度量的方式求得所需图像。文中的图像检索系统就是在这个情感模型基础上设计出来的。文中抽取了图像的颜色特征、纹理特征和形状特征作为图像的底层特征,通过 LSSVM 神经网络(Least Squares Support Vector Machine Neural Network)^[2]实现图像由底层特征空间到情感空间的映射,图像的检索通过在情感空间中进行相似性计算完成。

1 图像检索的研究历史

20 世纪 70 年代末,诞生了基于人工标注的文本检索方式^[3]。这种方法首先对图像数据进行人工标注,然后基于所标注的关键字采用文本检索技术实现图像的检索。但当图像数目非常庞大时,首先,人工标注过程中的工作量非常大;其次,很难通过区区几个关键字就表达清楚一幅图像的内容;再者,由于图像是分布在网络环境中,不同地区的人所使用的语言文字也会有差别,所以只用一种语言对图像进行标注也不切实际;最后,就是同一张图像,由于图像内容的丰富性,不同的人也会有不同的感受,所以标注的结果也可能不是很精确,这样就会影响检索的结果。

20 世纪 90 年代,随着大规模数字图像库的出现,为了克服文本检索技术的不足^[4],学者们提出了一种基于内容的图像检索(Content Based Image Retrieval, CBIR)^[5],这种方法通过使用颜色、形状、纹理等视觉特征^[6]来代表图像实现图像检索。经过十几年的努力,在一些特殊领域,学者们研究设计的很多原型系统取得了很大的成功。跟文本检索技术相比,基于内容的图像检索技术有了很大的进步^[7]。但是在通用领域,由于图像中包含各种各样复杂的对象,图像内容丰富,这种技术取得的效果也不是很理想。因为基于内容的图像检索是依据图像底层的视觉信息来查找的,人们却通常是利用图像中包含的语义信息来查找他们所需要的图像。然而人们所理解的图像语义和图像底层特征之间是存在“语义鸿沟”的^[8-10]。因为这种语义鸿沟的存在,基于内容的图像检索技术在复杂的实际应用中并没有取得太好的成果。

随着研究的深入,学者们提出了基于语义的图像检索方法来克服语义鸿沟。1996 年,Eakins 指出情感语义是图像语义的最高层次^[11],在基于图像语义的图像检索领域中属于一个比较新的研究方向。王惠锋等提出了“面向对象图像模型”^[12],这成为了图像情感检索研究的出发点。文献[13]中,王上飞等从心理学“维量”思想出发,基于径向基函数神经网络进行图像情感注释,在情感空间内实现图像的感性检索。文献[14]中,作者利用支持向量机算法在解决小样本、非线性及高维模式识别问题中具有的显著优势,实现图像的自动情感注释,设计了基于情感语义的图像检索系统。

除了在图像检索中的应用,这个领域的研究成果还能用于广告、产品设计、电子商务等方面,不管是应用前景还是经济价值都非常大。

2 情感模型的建立

维量分析法是心理学家在研究表情时提出的,他们将人类的情感分为六大类:高兴、惊奇、恐惧、悲伤、厌恶和愤怒。以这六种基本情感为基础,通过维量分析的方法可以表达人类的所有情感。应用到情感图像检索领域,研究者们从维量空间思想中得到提示,可以采用类似 RGB 颜色空间的方式来构建用户的情感语义空间,通过这种方法,每一类情感在这个情感语义空间中都对一个确定的向量。

要建立情感模型,必须对形容词进行量化。文中主要分了三个步骤来构建情感模型:首先是针对特定的图像库搜集一些能有效表达情感的形容词,文中是针对自然风景图像库进行图像检索,所以采用了文献[13]中所使用的 18 对形容词对^[13],如图 1 所示。第二步就是进行数据采集。选取一定数量的用户针对上面选择的 18 对形容词对,对样本图像进行评价。把评价分为五个等级:形容词对“协调的—不协调的”,用户的评价可以分为非常协调、有点协调、中性、有点不协调、非常不协调,它们分别对应 0.0,0.25,0.5,0.75,

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

图 1 形容词对

1.0。通过对这些数据的收集,建立用户的情感语义数据库。最后就是通过因子分析法对采集到的数据进行分析,建立情感空间。

3 系统模型构建

底层特征提取、建立情感空间、图像情感语义注释和图像检索是基于情感语义图像检索的四大组成部分。

3.1 图像特征提取

在现实世界中,情感活动离不开周围的环境,而现实中的环境是由一幅幅的静态图像或动态画面组成的。图像最基本的低层可视化内容是颜色、纹理和形状特征,人们在观察图像时,会受这些视觉特征的影响而产生不同的情感体会。在底层特征提取的时候应该注重对那些容易使人产生情感波动的特征进行提取,对于图像语义识别是非常重要的。

(1)颜色特征。

①颜色直方图。

全局颜色直方图^[15]描述不同色彩在整幅图像中的比例,具有平移、尺度、旋转不变性,简单方便,自从被提出来之后就一直是流行的图像检索技术之一。一般图像的全局直方图实际上是一个 1-D 离散函数,定义如式(1):

$$H(k) = \frac{n_k}{N}, k = 0, 1, \dots, L - 1 \tag{1}$$

取颜色直方图中频度最大像素和频度平均值,共 6 维。

②颜色矩。

颜色矩由 Stricker 和 Orengo^[16]提出,是一种非常有效却很简单的颜色特征。此方法基于用矩可以表示出图像中所有颜色分布这一数学基础。而且,图像的颜色分布只需要用颜色的一阶矩、二阶矩以及三阶矩这三个低阶矩就足以表示出来,因为颜色分布信息大部分集中于低阶矩中。这种方法相比于颜色直方图的另外一个优势:不需要对特征进行向量化。因此,图像的颜色矩一共只需要 9 个分量(3 个颜色分量,每个分量上 3 个低阶矩)。

(2)分布特征。

首先根据式(2)计算出每个像素点的亮度,接下来要将彩色图像转化成二值图像。

具体方法为:先计算出整幅图像亮度的平均值,然后把每个像素点的亮度值与这个均值比较,如果大于均值,则记作 1;其他情况记作 0。完成这一步后,再把图像分解成大小相等的小块,统计每一个小块中像素点值是 1 的数目,对其进行归一化,结果即为图像的灰度分布特征。

$$H = 0.299R \times 0.578G \times 0.114B \tag{2}$$

文中取了 64 维。

(3)形状特征。

不一样的形状也会给人带来不一样的感受,如方形会给人以“规矩、整齐”的感受。在形状特征这块,采用形状特征区域描述方法中的七个不变矩作为图像的形状特征。文中取了 7 维。

(4)纹理特征。

利用全局灰度直方图的矩来描述纹理是最简单的统计方法。这里围绕灰度直方图,提取出一组参数平滑度、一致性、熵来描述纹理的特性。文中取 3 维。

综上特征提取,每幅图像共有 89 维个特征。分别为颜色直方图在中频度最大和频度平均值(6),颜色矩(9),分布特征(64),形状特征(7),纹理特征(3)。

3.2 基于 LSSVM 的情感语义注释

通过因子分析,已经建立了 L 维的情感空间,在这个情感空间中,每一幅样本图像和每一组形容词对都对应于空间其中的一个向量。先用样本图像在 LSSVM 神经网络中进行情感语义注释,然后在神经网络完成学习之后再对剩余图像库中的图像自动进行情感语义注释,这个过程就是图像的情感注释。

从心理实验中可知,在建立情感语义空间的过程中,如果样本过多的话,容易引起用户疲劳,而使评估结果不准确。为了避免引起用户疲劳,待评估的样本图像不宜太多。所以可以把情感语义注释问题定义为小样本学习问题。而在小样本学习方面,支持向量机方法取得了较好的实验结果。而改进的 SVM—最小二乘 SVM,与 SVM 相比主要是采用不同的优化目标函数,用等式约束代替不等式约束,将求解的优化问题转成线性方程,大大减少算法的复杂性。鉴于此,文中采用最小二乘支持向量机算法实现图像的情感语义注释。

将提取来经过预处理的训练样本图像的底层特征作为 LSSVM 的输入,把训练样本图像某一维的因子得分作为 LSSVM 的输出(就是训练样本图像在情感空间中所对应向量的坐标)。经过对训练样本的学习,LSSVM 慢慢记忆了训练样本底层特征与其高层情感语义之间的联系。然后将未注释样本图像的底层特征作为输入输进去,LSSVM 就能够自动计算出该输入所对应的输出,即输入样本图像在情感语义空间所对应的坐标,这样就实现了图像的情感语义注释。

在对所有样本库中图像都完成情感语义注释后,就将所有样本图像都放到了情感空间中,每一幅图像都对应情感空间中一个向量。而经过因子分析,每一组形容词对也对应情感空间中的一个向量。这些向量为情感图像检索打下了基础。

3.3 图像的检索

随便选一个库中的形容词输入,比如“荒凉的”。在输入形容词后,系统会在情感空间计算该形容词和每幅图像的相似度,找出相似度最大的前 12 幅图像作为输出显示给用户。下面介绍一下相似度的定义。

假设 $f_m = (f_{m1}, f_{m2}, \cdots, f_{mL})$ 为图像 m 在情感语义空间的坐标, $a_n = (a_{n1}, a_{n2}, \cdots, a_{nL})$ 为形容词 n 在情感空间的坐标,则图像 m 与形容词 n 的相似度 d_{mn} 为:

$$d_{mn} = a_n \cdot f_m / \|a_n\| \|f_m\| \tag{3}$$

系统按照在情感语义空间中计算得出的相似度从高到低排序,找出最相似的 12 幅图像作为检索结果返回给用户。

4 实验分析

4.1 情感空间实验

实验邀请了 15 位在校研究生作为测试用户,记录了他们的性别、年龄等基本信息以方便以后的整理归类。首先让参加测试的用户先对所有样本图像进行一次快速的浏览,在对整体了解后,再开始对测试样本图像的评价实验。先对每一幅图像用图 1 中的 18 组形容词对进行评价,收集用户评价数据,随后对数据进行因子分析,得到载荷矩阵,其累计贡献率达到了 86.54%,如图 2 所示。

0.911 385 4	-0.224 494 2	-0.111 034 0	0.013 030 5
0.881 372 2	-0.248 817 6	-0.114 750 9	0.036 167 6
0.843 912 1	0.200 274 1	-0.143 685 5	-0.030 866 8
0.840 845 9	-0.346 278 3	-0.022 671 2	-0.034 506 3
0.835 574 2	-0.214 820 8	-0.149 723 1	0.105 914 3
0.718 936 2	0.313 828 3	-0.183 338 7	-0.208 035 9
0.711 133 1	0.192 458 9	-0.198 090 9	-0.002 251 4
0.702 033 9	-0.324 449 6	0.147 363 2	-0.129 817 9
0.701 025 6	-0.146 008 8	0.205 933 7	-0.057 262 6
0.689 482 5	0.418 227 1	0.088 595 4	0.007 164 9
0.635 925 1	0.014 762 9	0.306 928 7	0.131 946 8
0.574 790 5	-0.571 463 1	-0.192 274 6	-0.125 821 6
0.558 593 5	0.470 037 4	-0.096 561 1	-0.032 279 2
0.538 385 6	-0.019 429 4	0.307 432 0	0.193 761 6
0.342 954 4	0.763 847 8	0.162 339 0	-0.021 175 9
0.219 625 9	0.637 879 4	-0.145 696 1	-0.350 888 7
0.332 763 1	0.086 455 7	0.809 366 3	0.162 620 8
0.160 086 9	0.256 061 6	-0.360 524 6	0.828 633 8

图 2 因子载荷矩阵

得到因子载荷矩阵后,就顺利地将本来 18 维的形容词空间降成 4 维,每个形容词都对应情感空间中的一个向量,载荷值就是向量每维上的坐标。接下来利

用主成分分析法求得因子得分矩阵,这样就把每幅图像也成功降成 4 维,每幅图像也对应情感空间中一个向量,得分矩阵每列的值就是向量每维上的坐标。这样就把图像底层特征和高层情感语义统一到了一个情感空间中,每个形容词和每幅图像都是这个空间中的一个向量。

4.2 基于 LSSVM 的情感语义注释实验

用到的实验样本图像库是从整个图片库中精心挑选出来的 150 幅图像,然后训练样本图像和测试样本图像从中各随机分配 75 幅。根据上面的方法,先利用 LSSVM 神经网络对训练样本进行训练学习,完成情感注释。对于剩下的测试样本图像则利用训练好的 LSSVM 神经网络自动进行注释。

为了验证 LSSVM 神经网络情感注释的有效性,再次请了 15 位研究生对 75 幅测试样本图像根据 18 组形容词对进行人为的情感语义标注,这些标记结果作为原始精确结果。而经 LSSVM 自动标注的结果作为实验结果,将实验结果和原始精确结果进行比较。

观察某一个研究生对全部测试样本图像的结果,将其与通过 LSSVM 神经网络注释得到的结果进行比较。设两者结果一样的图像数目为 m ,则这个研究生的图像吻合率为: $P = \frac{m}{75} \times 100\%$ 。图 3 展示了 15 位研究生的图像吻合情况。

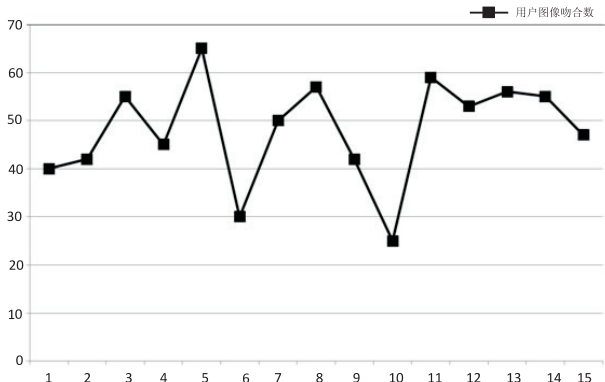


图 3 图像吻合率

图中,横轴代表研究生的编号(1 到 15),纵轴代表与每个研究生对应的图像吻合率。从图中可以发现,图像吻合率平均在 50% 左右。

对于某一个形容词,假设在人工注释和 LSSVM 神经网络注释中都是使用这个形容词完成注释的图像数目是 n ,则该形容词的吻合比率为: $P = \frac{n}{75} \times 100\%$ 。

图 4 是 6 个情感形容词“充满生机的”“荒凉的”“冷色调的”“暖色调的”“视野宽广的”“狭窄的”注释的吻合比率。

图中,横轴代表 6 个情感形容词,纵轴代表形容词的吻合率。可以发现,形容词的吻合比率达到 50%

以上。

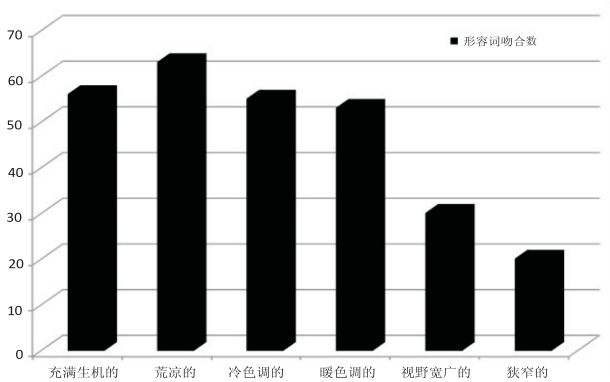


图 4 形容词吻合率

4.3 基于情感语义的图像检索实验

在完成建立情感空间和情感注释后,顺利地将每一幅图像和每一对情感形容词都对应到情感空间中的一个向量,在用户进行图像检索时,系统通过相似度计算,输出相似度最接近的 12 幅图像。图 5、6 列出了对形容词“荒凉的”“充满生机的”的检索结果,实验中用户对结果比较满意。

5 结束语

文中在对国内外相关领域最新成果学习研究的基

础上,深入探讨了图像情感语义注释与检索方面的一些关键技术和常用算法。底层特征是整个图像情感语义注释与检索过程中的基础,文中从颜色、纹理、形状等方面用当代比较有效的一些方法对底层特征进行了提取,随后详细分析了每种特征会对人类情感产生什么样的影响。

在此基础上,第一步,文中选择了比较权威的风景图像数据库,筛选出比较典型的用来描述图像情感的 18 组形容词对,通过邀请测试用户对样本进行评价,收集评价数据,对数据应用因子分析法构建出了情感空间。第二步,从之前提取出的底层视觉特征中选出一些会引起情感波动的特征作为图像的底层特征。在完成这两步以后,利用最小二乘支持向量机算法来构建出底层特征和高层情感语义间的映射关系,通过对样本图像的学习训练,自动完成对图像数据库剩余图像的注释。实验结果表明,利用最小二乘支持向量机算法实现图像注释得到的结果是可以接受的。

在完成对图像数据库中所有图像的注释后,就成功地把图像和形容词对统一到了同一个情感空间中,每张图像和每个形容词都对应这个空间中的一个向量,通过对相似度的计算,在情感空间中完成图像的情感检索。实验结果取得了不错的用户满意度。

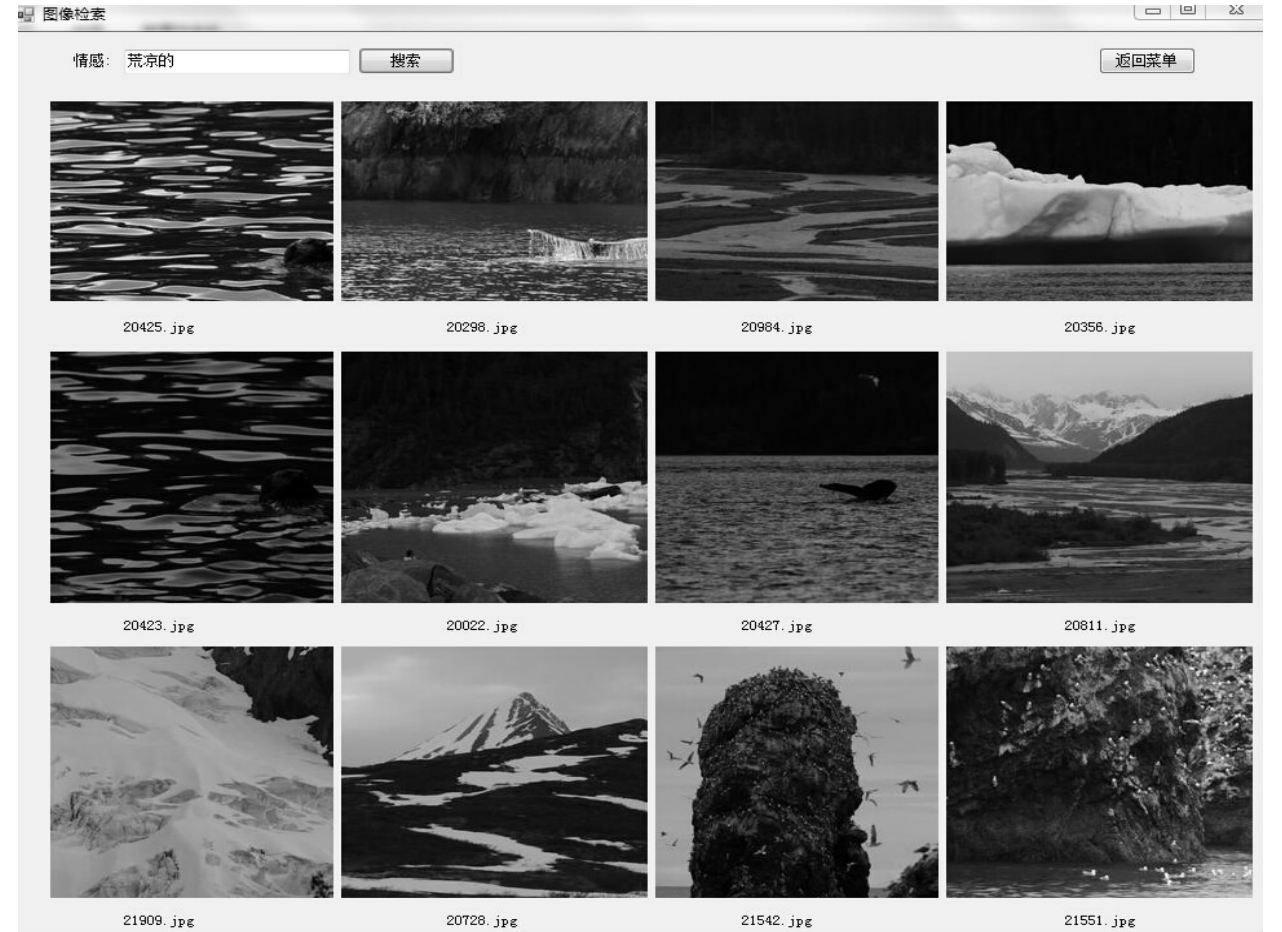


图 5 荒凉的图像

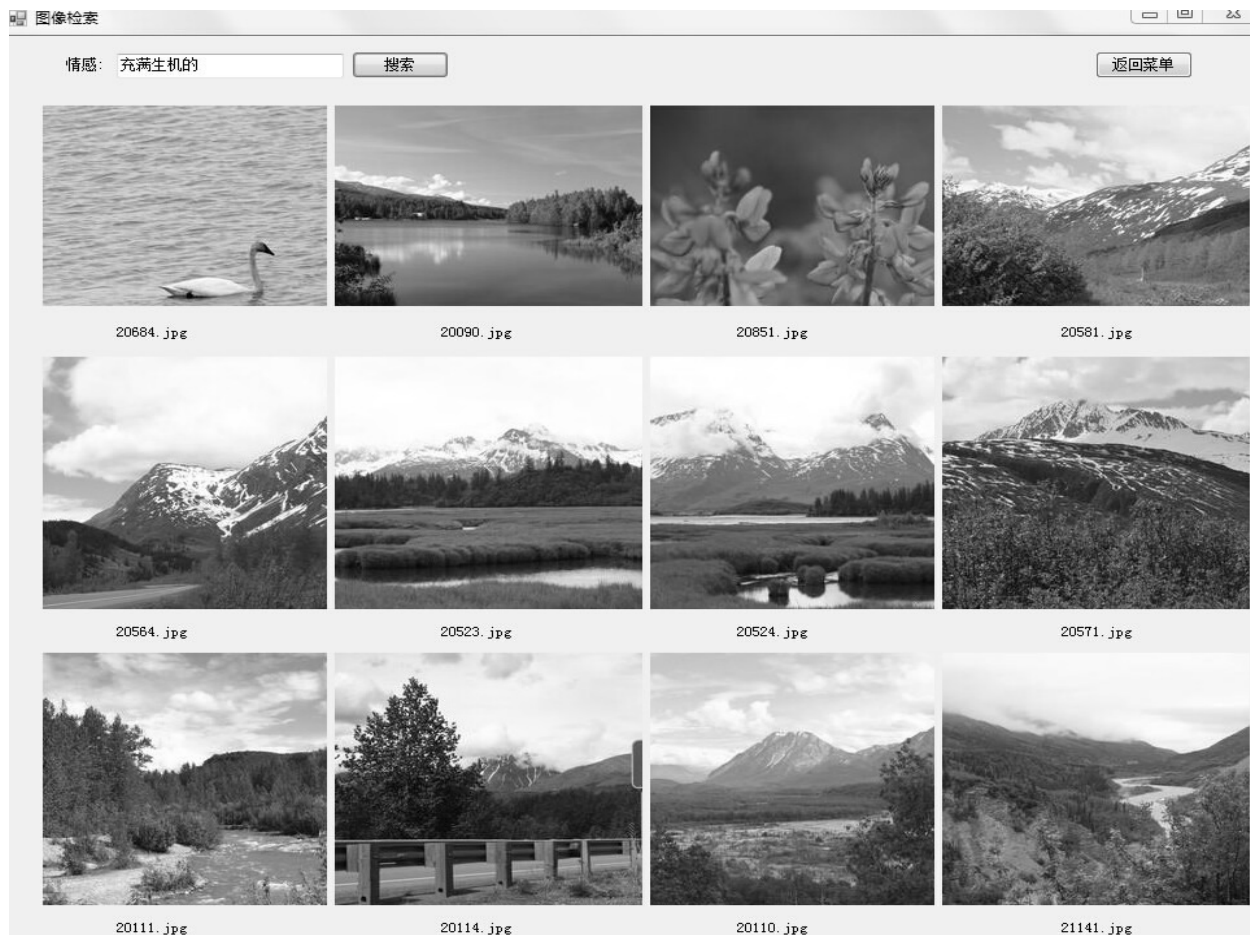


图 6 充满生机的图像

参考文献:

- [1] 向友君,谢胜利. 图像检索技术综述[J]. 重庆邮电学院学报:自然科学版,2006,18(3):348-354.
- [2] 郭新辰. 最小二乘支持向量机算法及应用研究[D]. 长春:吉林大学,2008.
- [3] Chen H L, Rasmussen E M. Intellectual access to images[J]. Library Trends, 1999, 48(2):291-302.
- [4] 李向阳,庄越挺,潘云鹤. 基于内容的图像检索技术与系统[J]. 计算机研究与发展,2001,38(3):344-354.
- [5] 王伟凝,余英林. 图像的情感语义研究进展[J]. 电路与系统学报,2003,8(5):101-109.
- [6] Neumann D, Gegenfurtner K R. Image retrieval and perceptual similarity [J]. ACM Transactions on Applied Perception, 2006, 3(1):31-47.
- [7] Ritendra D, Dhiraj J, Jia L, et al. Image retrieval: ideas, influences, and trends of the new age [J]. ACM Computing Surveys, 2008, 40(2):1-60.
- [8] Arnold W M, Marcel W, Simone S, et al. Content-based image retrieval at the end of the early years [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(12):1349-1380.
- [9] Peter E, Christine S. Towards a comprehensive survey of the semantic gap in visual image retrieval [C]//Proc of LNCS. [s. l.]:[s. n.], 2003.
- [10] Datta R, Ge W, Li J, et al. Toward bridging the annotation-retrieval gap in image search [J]. IEEE Multimedia, 2007, 14(3):24-35.
- [11] Eakins J P. Automatic image content retrieval—are we getting anywhere [C]//Proceedings of the third international conference on electronic library and visual information research. Milton Keynes: De Montfort University, 1996:123-135.
- [12] 王惠锋,孙正兴,王 箭. 语义图像检索研究进展[J]. 计算机研究与发展, 2002, 39(5):513-523.
- [13] 王上飞,陈恩红,李金龙,等. 基于感性的图像评估与检索[J]. 模式识别与人工智能, 2001, 14(3):297-301.
- [14] 吴星和. 图像情感语义注释与检索研究[D]. 杭州:浙江大学, 2013.
- [15] Swain M J, Ballard D H. Color indexing [J]. International Journal of Computer Vision, 1991, 7(1):11-32.
- [16] Stricker M, Orengo M. Similarity of color images [C]//Proc of SPIE. [s. l.]:[s. n.], 1995.

作者：[武频](#)，[陶聪](#)，[朱永华](#)，[颜宏杰](#)，[高宏浩](#)，[WU Pin](#)，[TAO Cong](#)，[ZHU Yong-hua](#)，[YAN Hong-jie](#)，[GAO Hong-hao](#)

作者单位：[武频, 陶聪, 朱永华, 颜宏杰, WU Pin, TAO Cong, ZHU Yong-hua, YAN Hong-jie\(上海大学 计算机工程与技术, 上海, 200444\)](#)，[高宏浩, GAO Hong-hao\(上海大学 计算中心, 上海, 200444\)](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2015(10)

引用本文格式：[武频](#).[陶聪](#).[朱永华](#).[颜宏杰](#).[高宏浩](#).[WU Pin](#).[TAO Cong](#).[ZHU Yong-hua](#).[YAN Hong-jie](#).[GAO Hong-hao](#)

[基于情感语义的图像注释与检索](#)[期刊论文]-[计算机技术与发展](#) 2015(10)