

Emotional Image Classification Based on Color Scale Invariant Feature Transform Feature and Spatial Pyramid Model

Pengxiao Lv, Guanghua Gu*, Chengru Wang, Yangjun Li

School of Information Science and Engineering, Yanshan University, Qinhuangdao
Email: guguanghua@ysu.edu.cn

Received: Nov. 26th, 2013; revised: Dec. 9th, 2013; accepted: Dec. 14th, 2013

Copyright © 2014 Pengxiao Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for Hans and the owner of the intellectual property Pengxiao Lv et al. All Copyright © 2014 are guarded by law and by Hans as a guardian.

Abstract: The purpose of emotional image classification is that the computer can express the emotion reaction when observing the image, and classify the images into the different emotional categories automatically. In this paper, we proposed an emotional classification framework based on the spatial pyramid representation. First, we extracted the SIFT (Scale Invariant Feature Transform) feature of the colour, and performed the clustering method to form the codebook. Then, each image is described by using LLC (Local-constrained Linear Coding) scheme, and image representations were performed by the methodology of spatial pyramid. Finally, we performed the emotional categorization by the training classifier. As the colour information is significant to human visual perception we added an extra colour feature. In the experiments made on the IAPS (International Affective Picture System) and KDEF (Karolinska Directed Emotional Faces), an ideal classification result was obtained.

Keywords: Image Processing; Emotional Categorization; Colour-Scale Invariant Feature Transform; Spatial Pyramid; Local-Constrained Linear Coding

C-SIFT 特征结合空间金字塔描述的情感图像分类

吕鹏霄, 顾广华*, 王成儒, 李扬骏

燕山大学信息科学与工程学院, 秦皇岛
Email: guguanghua@ysu.edu.cn

收稿日期: 2013年11月26日; 修回日期: 2013年12月9日; 录用日期: 2013年12月14日

摘要: 情感图像分类的目的是希望计算机能够表述人类观察图像时所引起的情感反应, 并根据这种反应把图像分到不同的情感类别。本文提出了一种基于空间金字塔的情感图像分类方法, 首先对图像提取颜色尺度旋转不变特征, 并聚类形成视觉特征词典; 其次对图像进行空间金字塔分块, 使用局部约束线性编码方法表示各子块图像, 形成图像的空间金字塔描述; 最后通过训练分类器实现对情感图像的分类。该方法在 SIFT 的基础上加入了具有表征感情色彩的颜色特征, 提取了独特的图像情感特征。本文方法分别在国际情绪图片系统(IAPS)数据库和人脸情感数据库(KDEF)上进行了实验, 取得了较为理想的情感分类结果。

关键词: 图像处理; 情感分类; 颜色尺度旋转不变特征; 空间金字塔; 局部约束线性编码

*通讯作者。

1. 引言

图像能够唤起人类的情感，而目前的图像处理和应⤵大多忽略了情感因素。图像蕴含大量的情感信息，它们存在于图像语义分类层次^[1]的最高层情感层。人类在观察图像时，会产生不同的情感，比如，春色满园会让人觉得特别舒畅、教堂使人感觉到肃穆等。图像情感分类就是希望计算机能够描述人类观察图像时引起的情感反应，并根据这种反应把图像分到不同的情感类别。由于情感语义相对于认知层的语义来说具有更多的主观成分，这无疑是图像情感分类的一个巨大挑战，并且类间相似性也是分类的一大障碍。图像的情感分类隶属于图像语义理解，它能够有效的组织图像，并按一定的规则分类图像，方便人们快捷的浏览检索图像^[2]，虽然现在这项技术仍然处于初级阶段，但其具有深远的研究意义和广泛的应用前景。

情感图像分类的关键问题是提取有效的情感特征及生成正确表示特征分布的图像描述。除此之外，图像情感分析也是情感图像分类必不可缺的研究。图像情感分析是机器视觉、模式识别、图像理解、人工智能等研究领域中的重要研究方向，因此。国内外学者进行了广泛的研究，已经取得不少不少的成果。针对图像情感语义分析技术中的图像特征提取、视觉词典生成、特征映射和图像描述等问题，提出新颖的方法，有效的建立图像特征与情感之间的映射关系，提高情感图像分类识别率，使计算机识别图像的能力接近人的理解水平，这必将对图像分类与图像检索研究领域的发展起到促进的作用。

近年来随着计算机技术的深入发展，国内外许多研究机构和大学也开始研究如何利用计算机实现情感图像分类的研究。早期图像情感分类大多数都是基于机器学习理论的图像情感分类算法，就是采用直接采用低层图像特征，如颜色、纹理、形状等描述图像的内容，再结合监督学习方法利用分类器对图像的情感类别进行推理的一种情感图像分类方法。也就是我们通常所说的“黑匣子”的情感图像分类方法。例如 Yanulevskaya 等^[3]提取 Wiccest 特征和 Gabor 特征，采用有监督的支持向量机方法实现情感图像的分类。由于 Yanulevskaya 等^[3]提取图像的低层特征，与高层语义之间存在“语义鸿沟”，因此这种简单的利用机器

学习算法很难找到图像特征与情感属性之间的映射关系，所以这种基于“黑匣子”的情感分类算法所获取的结果总是不太理想

随后对情感图像分类的方法逐渐扩展到基于语义主题的情感图像分类算法。Li 等^[4]结合词包模型提出了一种情感图像分类方法，使用无监督的概率潜在语义分析(probabilistic Latent Semantic Analysis, pLSA)模型，将表示图像的高维向量映射到潜在语义空间的低维表示。pLSA 模型反映的是潜在主题和视觉单词本身的概念关系，该方法有效的构建了“语义鸿沟”的桥梁，在一定程度上提高了分类的正确率。刘研等^[5]提出了基于情感概率潜在语义分析(Affective-probabilistic Latent Semantic Analysis, Affective-pLSA)模型的图像情感分类算法。Affective-pLSA 分析模型扩展了 pLSA 模型试图挖掘出潜在语义主题和视觉单词共生频率的概念关系，而这种概念关系恰能有效地契合视觉认知理论^[5]。虽然文献^[4,5]考虑低层特征和高层语义之间的映射关系，但是都忽略了颜色信息。众所周知，颜色具有唤醒情感的力量，主要在于颜色的象征性往往使人们联想到某些特定的情感，颜色具有的感情色彩特征不能被忽略^[6]。另外，这些方法忽略了图像的空间结构信息。

考虑颜色及空间结构信息，本文基于 SIFT (Scale Invariant Feature Transform)引入具有表征感情色彩的颜色特征，分别提取 R、G、B 三个颜色通道的 SIFT 特征，串联在一起形成 384 维的 C-SIFT 特征；然后引入空间金字塔匹配模型，将图像分割成分辨率越来越精细的子区域，计算在多个空间分辨率下的视觉单词分布来形成图像的空间金字塔表示，有效的利用了图像的空间结构信息。

2. 情感图像分类

情感图像分类通常包含四个步骤：特征提取与描述、视觉词典生成、图像描述、分类。具体过程为：首先提取情感图像的稠密 C-SIFT 特征，对图像特征聚类生成视觉词典，将图像的特征映射到视觉词典中的视觉单词中，结合空间金字塔模型，分析特征分布形成图像描述；其次对训练图像进行学习，训练 SVM 分类器；最后根据学习得到的模型完成对测试图像的分类。过程如图 1 所示：

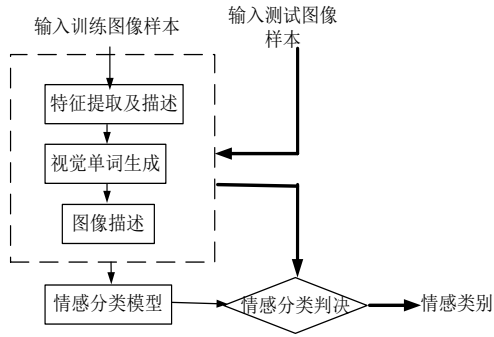


Figure 1. The structure of emotional classification system
图 1. 情感图像分类系统组成图

2.1. C-SIFT 特征提取与码本生成

2004 年 Lowe 提出尺度不变特征变换(Scale Invariant Feature Transform, SIFT)^[7], 该特征是图像的局部特征, 对尺度缩放、图像旋转、亮度变化具有不变性, 对视角变化、仿射变换、噪声也保持良好的鲁棒性。本文在 SIFT 的基础上增加颜色特征, 在提取图像特征时首先采用均匀网格分块来生成图像的局部区域, 图像块大小为 16×16 , 间隔为 8 像素。然后, 对每个图像块分别提取 R 、 G 、 B 三个通道的局部 SIFT 特征。在本文中设 x_r 、 x_g 、 x_b 分别表示为 R 、 G 、 B 三个通道的局部 SIFT 特征, 则图像的 C-SIFT 特征 $x = [x_r; x_g; x_b]$ 。由于 C-SIFT 特征为 384 维, 并且数据库图像信息量丰富, 特征数比较多, 为了提高处理速度及节省存储空间, 需要对特征进行降维^[8], 本文采用了主元素分析法(Principal Component Analysis, PCA)对特征降维, 在保留 95% 原信息的条件下, 把高维的特征映射到低维空间, 降维后的特征为 116 维。训练图像的特征个数为 n , 特征集为 $X = [x_1, x_2, \dots, x_n]$, 则协方差矩阵为:

$$E_s = \frac{1}{n} \sum_{k=1}^n x_k \cdot (x_k)^T \quad (1)$$

协方差矩阵奇异值分解:

$$E_s = U_s M_s U_s^T \quad (2)$$

式(2)中 M_s 是一个对角阵, 其包含 d 个奇异值(d 为 C-SIFT 特征维数), 并且按降序形式对奇异值进行排列 $m_1 \geq m_2 \geq \dots \geq m_d$, $U_s = \{u_1, u_2, \dots, u_d\}$ 中的特征值一一对应。

$$\min_{d_1} \sum_{k=1}^{d_1} m_k \geq T \sum_{k=1}^d m_k \quad (3)$$

低维空间特征维数 d_1 由式(3)决定, 其中 T 为控制变量代表保留信息量, 在本文中采用 $T = 0.95$ 表示在最小均方误差条件下对特征保留 95% 的信息量, 以避免丢失过多有用的信息而影响实验结果。取 U_s^T 的前 d_1 行特征向量作为转换矩阵 $P_s = \{u_1, u_2, \dots, u_{d_1}\}$, 则特征由高维空间向低维空间转换的过程如下:

$$y = P_s x \quad (4)$$

对每个情感类别的 C-SIFT 特征均采用传统的 k-means 方法聚类得到每个情感类别的类视觉词典 v_i , 则总的图像数据库的视觉词典为 $V = \{v_1, v_2, \dots, v_C\}$, C 为所有场景类别数。

2.2. 特征映射

特征映射指的是利用视觉单词对图像视觉特征进行量化编码, 生成视觉特征在视觉词典中的描述。令 $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$ 表示一幅图像的特征数据集, 其特征维数为 D , 特征个数为 N ; 令

$B = [b_1, b_2, \dots, b_M] \in R^{D \times M}$ 表示词典容量为 M 的视觉词典。

向量量化(vector quantization, VQ)方式是传统的特征量化编码方式, 采用最近邻方法, 对每一个特征向量计算其与 M 个视觉单词之间的欧式距离, 用与该特征向量距离最近的视觉单词替代。VQ 编码方式虽然简单, 但是编码结果太过粗糙, 容易丢失信息。为了克服缺点, 改善编码性能, Yang 等^[9]采用稀疏编码(Sparse Coding, SC)方法, 改善了 VQ 编码方式中有且只有一个非零元素的局限性; 随后 Yu 等^[10]改进了 SC 方式, 提出局部坐标编码(Local Coordinate Coding, LCC)方式, 阐述了稀疏性不能保证局部性, 局部性必然能够保证稀疏性的观点; Wang 等^[11]在 LCC 方式的基础上, 改进局部约束条件, 得到局部约束线性编码(Local-constrained Linear Coding, LCC)的方式。

$$\begin{aligned} \min_c \sum \|x_i - Bc_i\|^2 + \lambda \|d_i \Theta c_i\|^2 \\ s.t. 1^T = 1, \forall i \end{aligned} \quad (5)$$

式(5)为 LLC 编码方式优化的目标函数为, 其中 Θ 表示对应元素相乘运算, $d_i \in R^{M \times 1}$ 是局部适配器, c_i 是对应于特征包 $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$ 中每个特征的特征描述。 $\|d_i \Theta c_i\|^2$ 满足相似的特征描述符生成相似的编码。对于每个基向量 b_j , 根据其输入特征

x_i 的相似度, 得 $d_i = \exp\left(\frac{\text{dist}(x_i, B)}{\sigma}\right)$,

$$\text{dist}(x_i, B) = [\text{dist}(x_i, b_1), \text{dist}(x_i, b_2), \dots, \text{dist}(x_i, b_M)]$$

$\text{dist}(x_i, b_j)$ 表示特征 x_i 与视觉单词 b_j 之间的欧氏距离, σ 用于调整局部适配器权重的衰减速度。本文采用 LLC 编码方式, 并且在局部区域块形成描述时, 采用最大合并法。

2.3. 空间金字塔匹配

传统的词包模型(Bag of Word, BoW)把图像表示为无序的局部特征集合, 并且取得了很好的结果。但是 BoW 模型忽略了特征的空间分布信息, 严重限制了描述能力。为了克服其不足, Lazebnik 等在对图像的特征描述中加入了特征的空间信息, 提出了空间金字塔匹配模型^[12]。该模型将图像分割成分辨率越来越精细的子区域, 计算每个子区域内的局部特征直方图并组合得到图像的特征描述, 在图像描述中成功地融合了空间结构信息。

用空间金字塔方法对图像进行分块, 金字塔层次记为 $l = 0, 1, \dots, L$, L 是金字塔的最高层, 图 2 为图像三层金字塔网格结果, 其中第一幅图像代表原始图像, 中间图像表示将原始图像分割为 2×2 的网格, 第三幅表示将原始图像分割为 4×4 的网格。各层的子区域个数为 2^{2l} , 则 $L = 2$ 时分块区域总数为 $21(1 + 4 + 16)$ 。

本文采用空间金字塔分块方法, 统计在各层下所有网格单元内的每幅图像所有视觉单词的个数, 形成多级空间金字塔脂肪图描述 $\{h_i^l\}$, 其中 $l = 0, 1, \dots, L; i = 0, 1, \dots, 2^{2l}$, l 层的描述结果为 $h^l = [h_1^l; h_2^l; \dots, h_{2^{2l}}^l]$; 的将所得到的各层图像描述串联, 形成图像空间金字塔描述 $h = [\hat{h}^0, \hat{h}^1, \dots, \hat{h}^L]$, 对 h^l 加权 $\hat{h}^l = w_{h^l} \cdot h^l$, 其中加权函数为:

$$w_{h^l} = \begin{cases} \frac{1}{2^L}, & l = 0 \\ \frac{1}{2^{L-l+1}}, & l \neq 0 \end{cases} \quad (6)$$

3. 实验结果与分析

3.1. 实验设置

本文采用美国 NIMH 研究院情绪与注意研究中心

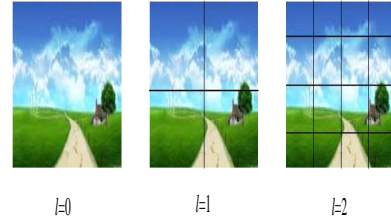


Figure 2. Display of three level spatial pyramid model
图 2. 三层金字塔网格示意图

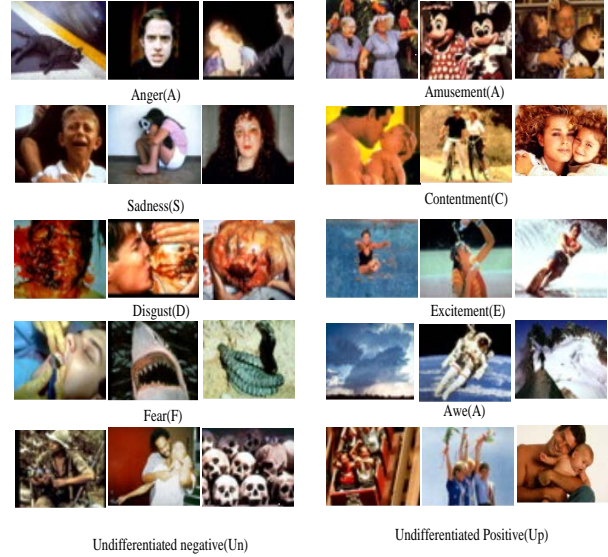


Figure 3. Example images of picture system of IAPS
图 3. IAPS 系统图像库示例图

提供的国际情绪图片系统^[13](International Affective Picture System, IAPS)作为实验数据库。不同图像集中的一些图像样本见图 3, IAPS 是提供情感实验调查研究的经过量化评定的情绪刺激图像系统, 包括美丽风光、鬼怪、可爱动物、娱乐旅游、亲情、受伤、恐怖动物、车祸等 1182 幅语义类别广泛的情感图像, 其图像大小均为 1024×768 像素。其中 396 幅图像被心理学家 Mikels 等人手工划分为 10 个情感类别: 包含 5 个消极情感(Anger, Disgust, Fear, Sadness, and Undifferentiated Negative)和 5 个积极情感类别(Amusement, Awe, Contentment, Excitement, and Undifferentiated Positive)。这 396 幅图像都有确定的感情标签, 目前已经成为图像情感类别领域中最常用的基准测试集。由于 anger 类别图像集中只有 8 幅图像, 样本过少, 在目前实验中暂时没有使用这一类图像集。其中左边是 5 个消极的情感类别(Anger, Disgust, Fear, Sadness, and Undifferentiated Negative), 右边是 5 个积极的情感类别(Amusement, Awe, Contentment, Excitement, and

Undifferentiated Positive)。在实验中,随机选取每个情感类别 70%的图像作为训练图像,剩下 30%的图像为测试图像。

另外,本文还采用 KDEF^[14](Karolinska Directed Emotional Faces,)数据库进行情感分类实验。该数据库包含 140 位业余演员 7 种不同的情感(Neutral, Happy, Angry, Afraid, Disgust, Sad, Surprise)的图像,并且每个情感分别从 5 个角度(-90°, -45°, 0°, 45°, 90°)拍摄,共 4900 幅人类面部表情。本文在实验时,分别把原数据库中同一情感的 140 个人分为一类,这样 7 种情感 5 个角度共为 35 类。实验时从每一种情感选取两个角度的图像构建新的数据集,共 14 类情感,这个数据库综合考虑了同一个人的不同表情以及同表情的不同角度。不同图像集中的一些图像样本见图 4,其中左边是左侧脸(角度为-90°)的 7 种情感,右边为正面脸(角度为 0°)的 7 种情感。图像大小均为 562 × 762 像素。在实验中,随机选取每个情感类别 50%的图像作为训练图像,剩下 50%的图像为测试图像。

在实验中,重复 10 次实验,最终的实验结果为 10 次的平均值。IAPS 和 KDEF 数据库为 RGB 彩色图像,以 32 像素为间隔,64 × 64 大小局部区域内对图像提取稠密 C-SIFT 特征,特征维数为 384 维,采用 PCA 降维方法,保留 95%特征信息,降维后的特征为 116 维。采用 k-means 对特征聚类生成视觉词典,LLC 的编码方式结合空间金字塔形成图像描述,Linear-SVM 实现情感图像的分类。本实验采用计算分类正确率的方法作为评价标准,若分类正确的图像数目为 n ,图像的总数目为 N ,则分类正确率 p 为:

$$p = \frac{n}{N} \times 100\% \quad (7)$$

3.2. 实验结果

本文对两个数据库均设置了三个实验,分别讨论码本容量、编码方式与最近邻个数、空间金字塔层数对分类正确率的影响。

实验一: C-SIFT 特征及码本容量对分类正确率的影响

图像的情感分类必须考虑到颜色能够唤醒情感的力量,颜色对情感表达的作用不容忽视,本文提出提取彩色图像的 C-SIFT 特征。基于词包模型^[15,16]的图像



Figure 4. Example images of picture system of IAPS

图 4. KDEF 系统图像库示例图

分类往往通过增大视觉词典的长度来达到提高分类精度的目的,虽然短码本计算简单,但判别力差,不能达到理想的分类效果;大码本具有良好的判别力,能较准确的区分不同类别,但是又会造成计算量过大和内存要求过高的问题,因此选取合适的视觉词典也是提高识别率的一个关键问题。实验一针对上述两个问题做了验证,分别提取带有情感色彩的 C-SIFT 特征及灰度图像的 SIFT 特征,并且对多个大小的类视觉词典进行了实验,每一类生成的视觉单词的数量分别为 25、50、100、150、200、250、300、350、400、450。

在实验一中分别提取图像的 C-SIFT 特征、SIFT 特征,均以 32 像素为间隔,64 × 64 大小局部区域内对图像提取稠密特征,C-SIFT 特征维数为 384 维,SIFT 特征维数为 128 维,IAPS 数据库的每幅图像的特征个数为 713 个,KDEF 数据库每幅图像的特征个数为 352 个。采用 PCA 降维方法,保留 95%特征信息,降维后的特征分别为 116 维、59 维。采用 k-means 对特征聚类生成视觉词典,LLC 的编码方式结合空间金字塔形成图像描述,LLC 编码方式取最近邻个数为 5,金字塔层数为 0,1 复合层,Linear-SVM 实现情感图像的分类。图 5 为分类正确率随类视觉词典大小变

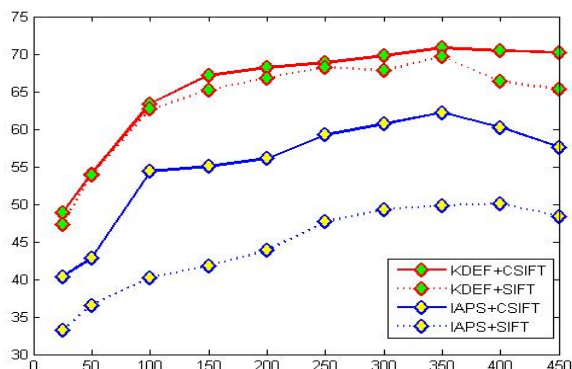


Figure 5. Performance comparisons with different category visual dictionaries

图 5. 分类性能随类视觉词典大小变化的趋势

化的趋势图，红色实线为 KDEF 数据库提取 C-SIFT 实验结果图，蓝色实线为 IAPS 数据库提取 C-SIFT 结果图，红色虚线为 KDEF 数据库提取 SIFT 实验结果图，蓝色虚线为 IAPS 数据库提取 SIFT 结果图。

由图 5 可以看出 C-SIFT 特征要比 SIFT 特征效果好，对于 IAPS 数据更为明显，由此可以看出颜色对于美丽风光、鬼怪、可爱动物、娱乐旅游、亲情、受伤、恐怖动物、车祸等图像情感表达的重要性。由于 KDEF 数据库是人脸数据库，情感差异主要由模特五官体现，色彩的效果不是很明显，但分类正确率也稍微提高了一点。同时可以看出类视觉词典的大小也影响着分类结果，通过两个数据库的分类结果曲线可以明显看出，当类视觉词典单词数量比较少的时候分类正确率比较低，随着单词数量的增大分类正确率也在逐渐提高，但是当类视觉词典单词数量为 350 时，达到了最大值，之后随着视觉单词数量的增大分类正确率逐渐降低。这是因为，类视觉词典比较小时，判别力比较差，不能达到理想的分类效果，随着类视觉单词的增多，词典判别力也在逐渐增强，但是类视觉词典比较大时，可能会造成信息冗余，同时会造成计算量过大和内存要求过高的问题，影响分类结果。

实验二：编码方式及最近邻个数对分类正确率的影响

在形成图像描述时有两个可变因素，一个是编码方式，一个是近似 LLC、SC 编码时用到的最近邻个数 knn 的选取，实验对这两个可变因素与分类正确率之间的关系进行了研究。实验二特征获取方法与实验一相同，降维后的 C-SIFT 特征为 116 维，采用 k-means 对特征聚类生成视觉词典，类视觉单词的数量为 300；

Table 1. Performance comparison with different coding schemes

表 1. 不同编码方式下分类正确率(%)

编码方式	VQ	SC	LLC
KDEF	55.36	65.95	70.61
IAPS	42.18	64.07	65.56

分别采用 LLC、SC、VQ 的编码方式结合空间金字塔形成图像描述，其中 LLC、SC 编码方式最近邻个数分别选取 2、5、8、11，金字塔层数为 0,1 复合层；采用 Linear-SVM 实现情感图像的分类。

表 1 为两个数据库的分类结果，可以看出在三种编码方式中，LLC 编码效果均最好，其次为 SC 编码，VQ 最差。原因在于 VQ 编码方式有且只有一个视觉单词来表示图像特征，这样比较容易丢失信息，所以分类结果比较不理想。LLC、SC 两种编码方式克服了这种不足，用多个视觉单词来表示图像特征，保留了大量信息，提高了分类正确率。而局部性必然保证稀疏性，但是稀疏性不能保证局部性，所以 LLC 编码方式的分类效果要好于 SC 编码方式。

图 6 列出了 knn 分别为 2, 5, 8, 11 时 LLC 和 SC 编码方式的分类正确率。其中菱形符的为 IAPS 数据库分类曲线图，圆圈为 KDEF 数据库分类曲线图，红色、青色线为 LLC 编码方式的正确率变化曲线，蓝色、绿色线为 SC 编码方式的正确率变化曲线。可以看出，随着 knn 数的增加分类正确率也在升高。这是因为随着 knn 个数的增大，向量量化过程中丢失的信息也在减少，因此分类正确率就会升高，对于 SC 编码方式当 knn 为 5 时，分类正确率达到了最高，对于 LLC 编码方式当 knn 为 8 时，分类正确率达到了最高，随后都又会随着 knn 的增大逐渐降低。这是由于当 knn 过大时一方面会造成信息冗余，另一方面特征并不能被正确的表示。例如，某个特征只有 6 个最近邻，但此时 knn 为 10，也许另外 4 个视觉单词就不能很好的表示特征，这样在形成图像描述的时候也会造成误差，影响正确率。

实验三：空间金字塔层数对分类正确率的影响

实验三验证了空间金字塔分层对分类正确率的影响。考虑到金字塔层数对分类正确率性能的影响，分别测试了不同层数时的分类正确率。随着金字塔层数的增加，对计算机内存的要求也越来越高，并且 KDEF 数据库样本多，为了保证实验时内存不溢出，减少了 KDEF 数据库的训练和测试样本，每类分别选

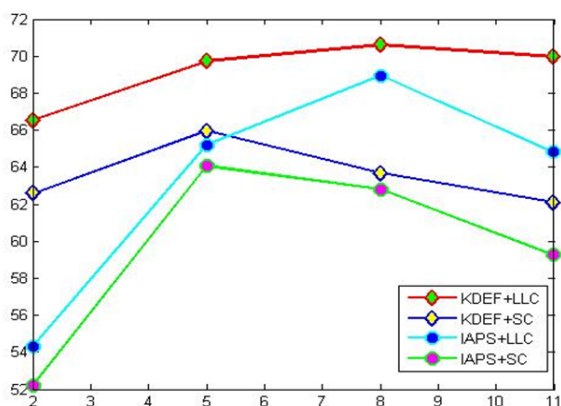


Figure 6. Performance comparison with different k nearest neighborhood

图 6. 不同近邻个数下的分类正确率(%)

取 30 幅训练图像和 30 幅测试图像，并且类视觉词典中的视觉单词为 100 个。需要指出的是，由于实验数据偏小，且类视觉词典也较小，实验正确率比较低，但是并不影响实验三突出体现金字塔分层对实验正确率影响这一目的。IAPS 数据库依然是随机选取每个情感类别 70% 的图像作为训练图像，剩下 30% 的图像为测试图像。

在实验三中特征获取方法与实验一相同，降维后的 C-SIFT 特征为 116 维。实验采用 k-means 对特征聚类生成视觉词典，两个数据库类视觉词典数量分别为 100、300；LLC 的编码方式结合空间金字塔形成图像描述，其中 LLC 编码方式最近邻个数为 8，金字塔层数分别为 0 层、1 层、2 层、0,1 复合层、0,1,2 复合层；Linear-SVM 实现情感图像的分类。0 层为原始图像，即没有采用空间金字塔分层，1 层为把图像分为 2×2 的网格，2 层为把图像分为 4×4 的网格，0,1 复合层为 5 ($1 + 2 \times 2$) 个网格，0,1,2 复合层为 21 ($1 + 2 \times 2 + 4 \times 4$) 个网格，分层后把各个区域的图像描述级联起来形成总的图像描述。表 2 为实验结果。

由表 2 可以看出，空间金字塔分层(0,1 复合层) 的实验结果由于加入了特征的空间信息，且保留了原始图像的全局信息，分类正确率较 0、1 层图像有明显提高。此外，1 层的也要比 0 层的正确率高，这是由于 1 层把图像分割为 2×2 个子区域，各子区域的图像描述级联起来为总的图像描述，增加了图像的空间信息，与实验算法理论相吻合。随着空间金字塔层数的增多，分类正确率逐渐升高，显然采用空间金字塔模型有助于提高识别正确率。在金字塔 0,1 复合层取

Table 2. Performance comparison with different levels of pyramid
表 2. 不同层数空间金字塔的分类正确

金字塔层数	0	1	0,1	2	0,1,2
KDEF	35.32	37.33	42.13	36.10	38.69
IAPS	63.70	64.07	64.30	62.19	60.24

分类识别率 70.61%

	0.37	0.00	0.00	0.00	0.04	0.00	0.01	0.00	0.06	0.00	0.19	0.00	0.33	0.00
AF	0.00	0.51	0.00	0.11	0.00	0.04	0.00	0.07	0.00	0.06	0.00	0.07	0.00	0.13
	0.03	0.00	0.74	0.00	0.06	0.00	0.00	0.00	0.06	0.00	0.11	0.00	0.00	0.00
AN	0.00	0.03	0.00	0.81	0.00	0.07	0.00	0.00	0.00	0.03	0.00	0.06	0.00	0.00
	0.07	0.00	0.09	0.00	0.53	0.00	0.11	0.00	0.03	0.00	0.17	0.00	0.00	0.00
DI	0.00	0.00	0.00	0.07	0.00	0.89	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.00
	0.04	0.00	0.01	0.00	0.09	0.00	0.77	0.00	0.04	0.00	0.01	0.00	0.01	0.01
HA	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.01	0.00	0.01	0.00	0.00	0.00
	0.00	0.00	0.04	0.00	0.00	0.00	0.01	0.00	0.73	0.00	0.17	0.00	0.04	0.00
NE	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.94	0.00	0.01	0.00	0.00	0.00
	0.07	0.00	0.11	0.00	0.11	0.00	0.03	0.00	0.17	0.01	0.47	0.00	0.01	0.00
SA	0.00	0.07	0.00	0.11	0.00	0.00	0.01	0.00	0.09	0.00	0.67	0.00	0.04	0.00
	0.13	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.81	0.00	0.00
SU	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.03	0.00	0.87	0.00
		AF	AN	DI	HA	NE	SA	SU						

Figure 7. The confusion matrix of emotional images categorization
图 7. 情感图像分类混淆矩阵

得最高分类结果，随后又随着层数的增多分类结果呈下降趋势。原因在于随着空间金字塔层数的增多，加入的空间信息越丰富，所以分类正确率越来越高。但是随着层数的增加，可能会产生过分割现象，即将图像中的目标过分细分，势必影响实验结果；同时区域块个数呈指数增加，区域块过多必然会造成计算量过大，也会使结果会有所下降。所以必须选择适当的金字塔分块层数才能达到较理想的分类效果。

经以上三个实验的结果与分析，本文的算法在类视觉词典大小为 400，LLC 编码最近邻个数为 8，金字塔 0,1 复合层时，取得最佳实验结果，IAPS 数据库平均分类识别率为 65.56%，KDEF 数据库平均分类识别率为 70.61%。KDEF 数据库的情感图像分类混淆矩阵如图 7 所示，该数据库由 7 类情感图像组成，并且每类情感图像分别包含正面人脸和侧面人脸情感图像，混淆矩阵奇数列为左侧面人脸分类正确率，偶数列为正面人脸分类正确率。

最后，本文算法与 Yanulevskaya^[3]算法、刘硕研^[5]算法进行比较。Yanulevskaya 和刘硕研采用 IAPS 图

Table 3. Performance comparison with different methods
表 3. 不同方法的情感分类算法比较

	分类模型	类别	分类正确率
Yanulevskaya ^[3] 算法	“词包模型” + SVM	8	54.00%
刘硕研 ^[5] 算法	Affective-pLSA + SVM	8	57.00%
本文算法	SPM + SVM	9	65.56%

像数据库中八类情感。本文采用 IAPS 图像数据库中九类情感。三种算法分类性能比较结果如表 3 所示,其中 Yanulevskaya 算法和刘硕研算法的实验结果源自文献^[5]。

4. 结论

本文提取了 C-SIFT 特征,利用 K-means 对特征聚类形成视觉词典,采用 LLC 编码方式结合空间金字塔匹配模型的图像表示方法,通过 SVM 分类器实现了情感图像的分类。本文提出的方法具有情感特征分辨力,原因在于既考虑到特征的情感信息,也结合了特征的空间结构信息。但是,由于数据库较小,在一定程度上限制了分类性能。扩充图像数据库和寻求更简单有效的算法,以提高情感图像分类识别率将是下一步研究的问题。

项目基金

国家自然科学基金青年科学基金项目(61303128)、河北省自然科学基金面上项目(F2013203220)、河北省高等学校科学研究青年基金项目(Q2012047)。

参考文献 (References)

- [1] 高隽, 谢昭 (2009) 图像理解理论与方法. 科学出版社, 北京, 1-30.
- [2] 徐思敏 (2012) 基于语义的图像检索关键技术研究. 硕士学位论文, 沈阳航空航天大学, 沈阳.

- [3] Yanulevskaya, V. and van Gemert, J.C. (2008) Emotional valence categorization using holistic image feature. *International Conference on Image Processing (ICIP)*, 101-104.
- [4] S. Li, Y.-J. Zhang and H.-C. Tan. (2010) Discovering latent semantic factors for picture categorization. *International Conference on Image Processing (ICIP)*, 1065-1068.
- [5] 刘硕研 (2011) 面向感知的图像场景及情感分类算法研究. 硕士论文, 北京交通大学, 北京.
- [6] 吴或 (2013) 基于颜色特征的网络不良视频检索技术研究. 成都检测台.
- [7] 付赛南 (2013) 基于特征降维的场景分类方法研究. 硕士论文, 上海交通大学, 上海.
- [8] Lowe, D. G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110.
- [9] Yang, J., Yu, K., Gong, Y. and Huang, T.S. (2009) Linear spatial pyramid matching using sparse coding for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 794-1801.
- [10] Yu, K. and Zhang, T. (2010) Improved local coordinate coding using local tangents. *Proceedings of the 27th International Conference on Machine Learning*, 215-1222.
- [11] Wang, J.J., Yang, J.C., Yu, K., et al. (2010) Locality-constrained linear coding for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 360-3367.
- [12] Lazebnik, S., Schmid, C. and Ponce, J. (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *IEEE Conference on Computer Vision and Pattern Recognition*, New York, 17-22 June 2006, 2169-2178.
- [13] Lang, P.J., Bradley, M.M. and Cuthbert, B.N. (1997) International affective picture system (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention.
- [14] Lundqvist, D., Flykt, A. and Öhman, A. (1998) The Karolinska directed emotional faces KDEF. Department of Clinical Neuroscience, Psychology Section, Karolinska Institute.
- [15] 李凤彩 (2012) 基于码本模型的场景图像分类研究. 硕士论文, 燕山大学, 秦皇岛.
- [16] 涂潇蕾 (2012) 上下文特征结合空间金字塔模型的场景分类算法研究. 硕士论文, 燕山大学, 秦皇岛.