基于分块权值的语义图像检索

夏利民 朱 城 张海燕 彭东亮
(中南大学信息科学与工程学院 长沙 410075)

摘 要 图像低层视觉特征和高层语义间的"语义鸿沟"是图像检索的关键问题。为了进一步提高基于语义的图像检索系统工作效率,以分块权值和视觉词库为基础,结合图像低层特征和高层语义的相关性,提出了一种基于分块权值的语义图像模型,该模型用来反映图像的视觉特性,对图像的高层语义进行有效检测,从而提高语义图像的检索效率。实验结果表明,该方法提高了语义图像检索系统的查全率和查准率。

关键词 词袋,图像检索,子块,语义
中图法分类号 TP391 **文献标识码** A

Semantic Image Retrieval Based on Sub-block Weight

XIA Li-min ZHU Cheng ZHANG Hai-yan PENG Dong-liang
(School of Information Science and Engineering, Central South University, Changsha 410075, China)

Abstract The semantic gap between low-level visual feature and high-level semantic has become a primary problem. For improving the efficiency of semantic-based image retrieval system, this paper based on chunked weight and a visual vocabulary proposed a semantic image model which utilizes the correlation of low-level feature and high-level semantic. The model is used to interpret the image visual characteristic and detecte high-level semantic, which improves the efficiency of the semantic image retrieval. The experimental results show that this method improves the precision and recall of semantic image retrieval system.

Keywords Bags-of-word, Image retrieval, Sub-block, Semantic

1 引言

图像检索是一个古老的研究方向,从20世纪70年代末开始,基于文本的图像检索就已经产生。到20世纪90年代初,许多学者开始了基于内容的图像检索(CBIR)的研究[1]。随着研究的深入,为了克服语义鸿沟,人们提出了基于语义的图像检索方法[2]。基于语义的图像检索逐渐成为目前图像检索研究的热点。

目前,语义图像检索的方法已经有很多种,主要有基于支持向量机的面向语义图像检索、基于非负矩阵分解的隐含语义图像检索、基于模糊熵的空间语义图像检索模型研究、基于Bayes统计学习的语义图像检索等。本文首先建立一个视觉词库,它将包含图像库中能够遇到的大部分普通子块类型。然后,通过模式向量形成一个与给定图像相关联的视觉字典[3]。我们的研究目的是检索语义类似的图像。当我们去检索给定的一幅图像,图像的语义描述和返回的图像的语义概念相同时,就认为是相关的。如果图像的视觉效果类似,但是语义概念不同,就认为这些图像不相关。

2 视觉词库的建立

视觉词汇对图像特征进行量化分析可以处理多种问题。词库的建立需要先对图像库中的图像采用两种MPEG-7标准中的描述子来提取低层特征,也就是可扩展颜色描述子和同类纹理描述子[4,5]。在这一部分,我们将讨论视觉词汇的作用和建立视觉词汇的方法。

将图像库中的所有图像的低层特征进行提取,对这些低层特征相似的区域进行聚类。对于一个类采用同样语义概念来表达,而那些包含相同的高层语义描述词的图像也同样有类似的低层特征区域。也就是说,含有相似的低层特征的区域常常用相同的高层语义描述。例如,一个区域含有一个高层语义"森林",那么就会有和"森林"相对应的视觉低层特征,也就是说,这片区域的大部分面积颜色是"绿色"的。

为了构建一个结构化知识库,首先,对一个图像库中的图像区域进行适当量化;然后,图像库中的图像通过K-means聚类算法被分割成各种图像区域,我们将最常见的区域类型进行视觉词汇的描述;最后,将所有的词汇集中起来,创建一个视觉词汇数据库。这样,一幅图像将可以用一个图像区域集来表示。通过这种方式来建立低层特征和高层语义的联系。

图像库中分割后的所有图像区域将会构成一个特性向量集F,这个特征向量集和视觉词汇库中的词汇是相对应的。为了选择最常见的区域类型,对F采用K-means聚类算法,

到稿日期:2012-11-29 返修日期:2013-03-11 本文受国家自然科学基金(50808025),国家教育部博士点基金(20090162110057)资助。
夏利民(1963-),男,教授,博士生导师,主要研究方向为图像处理与模式识别;朱 城(1987-),男,硕士生,主要研究方向为图像处理;张海燕(1986-),女,硕士生,主要研究方向为模式识别;彭东亮(1989-),男,硕士生,主要研究方向为图像处理。

我们将得到 N_T 个聚类集群。我们选择距离每个集群质心位置最短的作为一个区域类型，最后得到的区域类型数量 N_T 就是聚类集群数量。我们定义视觉词汇 T 由这些区域类型的集合组成：

$$T=\{w_i\},i=1,2,\cdots,N_T,w_i\subset R \tag{1}$$

式中，w_i 表示第 i 个区域类型，N_T 表示区域类型的个数，R 表示所有区域的集合。

每个区域类型虽然没有包含概念语义信息，但是相比低层特征在描述上会有更加好的效果。例如，当我们直观地描述一个“粗糙纹理的绿色区域”的区域类型时，我们无法得到具体的概念如植被等。

3 基于分块权值的模式向量的构建

将一幅图像 I_i 的视觉内容表示成一个模式向量 m_i。这个向量将得到和这个给出图像所包含区域类型相关的视觉词汇[6]。为了构造这样一种模式向量，我们提出了一个可行的提取算法。首先，将一幅待检索图像 I_i 分割成 $N\times N$ 的子块，SBS_i(sub-block-sets)是这个图像的子块集，公式表达为：

$$SBS_i=\{sbs_{ij}\mid sbs_{ij}\subset I_i\} \tag{2}$$

sbs_{ij} $(j=1,2,\cdots,N\times N)$表示这个子块集中的一个子块。然后，构造一个模式向量，用这个图像全部的子块分别与所有的区域类型进行匹配。对于每一个子块，用“递归最短生成树”(RSST)算法[7]将其分割成 N_K 个小区域 r_{ij}，$r_{ij(k)}\in R_{ij}$，$k=1,2,\cdots,N_k$，为图像中 I_i 的第 j 个子块的分割区域集。采用欧氏距离对每一种区域类型的特征向量 $f(w_p)$ 与每个子块中的分割区域的特征向量 $f(r_{ij(k)})$ 进行比较，得到了所对应子块的模式向量[7]。

$$m_{ij}(p)=\min_{r_{ij(k)}\in R_i}\{d(f(w_p),f(r_{ij(k)}))\},\quad p=1,2,\cdots,N_T \tag{3}$$

$$m_{ij}=m\{m_{ij}(1),m_{ij}(2),\cdots,m_{ij}(N_T)\} \tag{4}$$

式中，N_K 表示子块区域的个数，m_{ij} 表示图像子块视觉内容在视觉字典中所对应的模式向量[8]。

对于子块中每个区域 $r_{ij(k)}\in R_{ij}$，$k=1,2,\cdots,N_k$，将其与每个区域类型用欧拉公式进行比较，得到相似程度为 $\Theta_{ij(k)}$：

$$\Theta_{ij(k)}=E\{f(w_p),f(r_{ij(k)})\} \tag{5}$$

其中，$p=1,2,\cdots,N_T$；$k=1,2,\cdots N_k$。

图 1 展现了一个图像例子，将这个图像分成 3×3 的子块，选取里面的第二个子块，将子块通过上述的聚类方法分割成 7 个区域，我们假设视觉词库由 10 个区域类型组成。这个图像的模式向量见式(6)。图左侧，展现了子块中的所有区域与第一个区域类型的距离。图右侧，展现了被标记为植被的区域和全部的区域类型之间的距离。根据上述模式向量的算法，对这些模式向量的分量进行计算，得到 $m(1)$为0.2。通过一系列的计算，得出图像子块的 7 个分割区域对于这 10 个区域类型的表格，见表 1。数值低的表示子块中的那些区域和区域类型的距离较近。

$$m=[m(1),m(2),\cdots,m(10)] \tag{6}$$

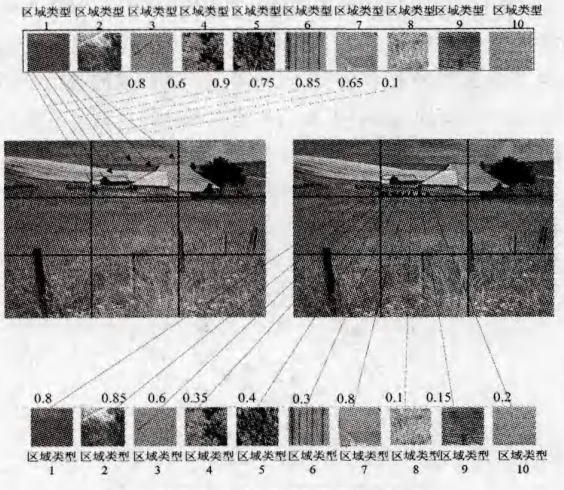

图 1 区域类型以及分块权值

表 1 区域类型分块权值

	类型 1	类型 2	类型 3	类型 4	类型 5	类型 6	类型 7	类型 8	类型 9	类型 10
子块 1	0.8	0.85	0.6	0.36	0.4	0.3	0.8	0.1	0.15	0.2
子块 2	0.6	0.2	0.5	0.35	0.4	0.6	0.7	0.45	0.1	0.22
子块 3	0.9	0.5	0.2	0.1	0.46	0.34	0.23	0.12	0.8	0.45
子块 4	0.75	0.5	0.2	0.15	0.42	0.11	0.45	0.65	0.12	0.4
子块 5	0.85	0.75	0.65	0.15	0.45	0.32	0.26	0.23	0.15	0.42
子块 6	0.65	0.84	0.45	0.45	0.15	0.15	0.42	0.15	0.48	0.42
子块 7	0.1	0.64	0.12	0.42	0.10	0.42	0.12	0.84	0.12	0.42
…	…	…	…	…	…	…	…	…	…	…
子块 Nk	…	…	…	…	…	…	…	…	…	…

通过上面的例子，我们了解到了模式向量在 m 图像处理中的主要作用。在得到了一份图像中每个子块的分割区域相似度后，我们进行权值的计算。

首先，设置一个相似度阈值，将其分成 3 个等级，分别为：$0\leqslant\Omega\leqslant 0.3$，$0.3<\Omega<0.7$，$0.7\leqslant\Omega\leqslant 1.0$；

接下来，得到每个子块区域的分割区域的大小，即像素点数量 $S_{(ij)k}$；

然后，通过式(7)得出每个子块中分割区域的权值大小：

$$w_{(ij)k}=[1-E\{f(w_p),f(r_{ij(k)})\}]\cdot S_{(ij)k} \tag{7}$$

最后，将每个区域类型的权值求和：

$$W_p=\sum_{p=1}^{N_T}w_{ip} \tag{8}$$

式中，$p=1,2,\cdots,N_T$。

通过这种计算得出每个区域类型在图像中的比重，映射到模式向量里，从视觉内容中得到它的视觉词汇。

4 高层语义的检测

在这一节中，我们介绍在图像数据库中高层概念的检索

方法。基于支持向量机(SVM)的检测器被用来区分每一个高层语义概念。首先,提取训练图像库的低层特征集,然后用SVM算法对这些特征集进行训练和机器学习,从而构造SVM多类分类器。然后,将该多类分类器示例的低层特征映射到高层语义特征。所谓的高层语义特征,即是一组向量,它代表的是图像属于各个类的概率,该高层语义特征将作为图像之间是否相似的度量方法。

为了能够更好地表示图像的内容,本文提取了3种性能较好的图像低层特征信息,如表2所列。

表2 3种图像低层特征信息的提取

低层视觉特征信息	维数
HSV非均匀量化直方图	9
RGB2阶颜色矩	72
RGHV共生矩阵	32

表中的特征信息特征方法可分别参见文献[9-11]。这样,对于任意一幅图像,可以得到它的这3种低层特征,将其进行综合并归一化后,得到一个113维数综合特征向量。

将在视觉词库中描述的待检索图像的模式向量输入到检测器中。检测器中输出的是给定图像包含具体概念的置信度,其输出值的大小在[0,1]范围内。置信度的数值靠近1时,这个对应的概念有很大的可能在图像中找到。需要注意的是,检测器的训练是基于每幅图像而并非每个子块区域。

在检索图像库高层语义特征集中查找示例图像高层语义特征 X_i 的近邻集 $N_i=\{X_{im},m=1,2,\cdots,L\}$,其中 L 是近邻集的大小,如图2所示。

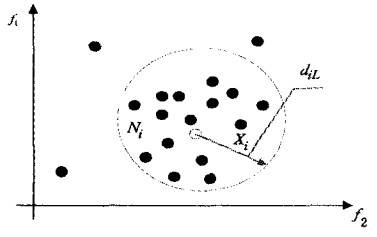

图2 近邻集的计算

图2中,$X_{i1},X_{i2},X_{i3},\cdots,X_{iL}$ 是距离 X_i 最近的 L 个检索图像库中的样本,设 $d_{i1},d_{i2},d_{i3},\cdots,d_{iL}$ 分别为空间中 X_i 到 $X_{i1},X_{i2},X_{i3},\cdots,X_{iL}$ 的二阶欧氏距离,满足 $d_{i1}\leqslant d_{i2}\leqslant d_{i3}\leqslant\cdots\leqslant d_{iL}$,则 $X_{i1},X_{i2},X_{i3},\cdots,X_{iL}$ 就是示例图像 X_i 的最终检索结果。

5 实验结果及分析

为了测试本文提出的方法的效率,我们选用了Oliva和Torralba创建的数据库。这个数据库共有2688个图像,共分为海岸、高速公路、森林等8个类。这些图像的大小为256×256个像素,这个数据库主要被用于场景识别的问题。对数据库中的所有图像进行全局标注。对于每个概念,我们计算它的平均查准率和平均查全率。我们首先要知道给定的图像是属于哪个语义范畴,检索出的图像只有属于同一范畴时才被认为是相关的。

查全率(P_{Recall})=[检出的相关图像数量($N_{related}$)/检索系统中相关图像总量($N_{Sumrelated}$)]×100%,如下式:

$$P_{Recall}=\frac{N_{related}}{N_{Sralated}} \tag{9}$$

查准率($P_{Precision}$)=[检出的相关图像数量($N_{related}$)/检出的图像总量($N_{Retrieval}$)]×100%,如下式:

$$P_{Precision}=\frac{N_{related}}{N_{Retrieval}} \tag{10}$$

从数据库中选取野外、高速公路、森林、海岸等8类语义图像类。选用800幅图像(每类平均100幅)作为训练集,对于每幅图像用分块权值的方法提取出特征向量,通过模式向量将其从低层特征映射到高层语义,然后通过学习构造相应的SVM分类器,分别对应于图像库中的8类语义。

为了证明有效性,我们将本文的方法和另外两种方法进行比较。这两种方法分别是全局视觉描述方法和局部描述方法。

表3 从8种不同的视觉词汇中计算出它们的查全率和查准率(大小由 Nt 定,K 种最接近的区域类)

	Nt=150						Nt=270					
	K=1		K=2		K=5		K=1		K=2		K=5	
	查全率	查准率	查全率	查准率	查全率	查准率	查全率	查准率	查全率	查准率	查全率	查准率
高建筑物	0.725	0.597	0.734	0.698	0.748	0.722	0.728	0.640	0.741	0.642	0.752	0.667
野外	0.654	0.650	0.662	0.620	0.671	0.652	0.661	0.635	0.654	0.629	0.670	0.672
高山	0.712	0.832	0.722	0.835	0.745	0.851	0.723	0.835	0.730	0.847	0.775	0.869
高速公路	0.628	0.578	0.630	0.582	0.633	0.707	0.645	0.671	0.650	0.682	0.656	0.699
森林	0.801	0.791	0.821	0.832	0.824	0.850	0.812	0.785	0.811	0.848	0.833	0.893
海岸	0.860	0.836	0.863	0.851	0.871	0.874	0.865	0.872	0.871	0.874	0.889	0.885
街道	0.688	0.680	0.720	0.716	0.754	0.746	0.770	0.774	0.808	0.806	0.865	0.859
内陆城市	0.711	0.708	0.750	0.718	0.762	0.741	0.760	0.755	0.755	0.742	0.826	0.815

表4 比较用本文方法(选用 $K=5,Nt=270$)与SIFT特征和Global特征之间检索结果的差异

	本文方法(K=5,Nt=270)	SIFT描述	Global描述
平均查全率	0.783	0.671	0.732
平均查准率	0.795	0.650	0.741

表3的数据是采用我们的算法的检索结果。我们通过观察,发现其对野外和高建筑物类的检索效果并不是非常的理想。然而,在海岸和森林的情况下,这个语义算法所得到的视觉块是可以很容易就区分出来的。随着子块 K 和区域类型 Nt 的增加,图像检索系统得到的查全率和查准率变大。

表4描述了本文方法与基于SIFT描述[11]和基于Global描述[12]两种算法的比较。通过对这3种方法所得到的平均查全率和平均查准率的结果进行比较,发现本文提出的算法对于语义图像的检索可以得到更加有效的结果。

结束语 本文提出一个基于分块权值的图像语义模型,通过这个模型构造一个基于区域子块的视觉词汇的语义图像

检索系统。本方法主要通过图像区域中的视觉属性获得词袋，通过子块集合聚类得到一个最接近图像区域的区域类型，不同于其它采用整个视觉特征来获取词袋的方法。这样更能够简单有效、准确地表示图像特征，这样使得图像特征更加靠近视觉词汇的概念。本文提出的方法在语义图像检索的实验中取得了良好的效果。实验证明，我们提出的方法优于普通的基于局部或全局描述子构造的词袋模型。

参 考 文 献

[1] Hiremath P S, Pujari J. Content based image retrieval using color, texture and shape features[C]// Advanced Computing and Communication. Gulbarga: Gulbarga University. 2007

[2] 周明全，耿国华，韦娜. 基于内容图像检索技术[M]. 北京：清华大学出版社，2007

[3] Duygulu P, Barnard K, De Freitas J F G, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary[J]. Lecture Notes in Computer science, 2002(2353): 97-112

[4] Spyrou E, Tolias G, Mylonas P, et al. Concept detection and keyframe extraction using a visual thesaurus[J]. Multimedia Tools and Applications, 2009, 41(3): 337-373

[5] Manjunath B, Ohm J, Vasudevan V, et al. Color and texture descriptors[J]. IEEE Trans Circuits Syst Video Technol, 11(6): 703-715

[6] Spyrou E, Tolias G, Mylonas P, et al. Concept detection and keyframe extraction using a visual thesaurus[J]. Multimedia Tools and Applicationgs, 2009(41): 337-373

[7] Avrithis Y, Doulamis A, Kollias S. A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases [J]. Computer Vision and Image Understanding, 1999(75): 3-24

[8] Mylonas P, Spyrou E, Avrithis Y, et al. Using Visual Context and Region Semantics for High-Level Concept Detection[J]. IEEE Transaxtions on Multimedia, 2009, 11(2): 229

[9] 曹利华，柳伟，李国辉. 基于多种主色调的图像检索算法研究与实现[J]. 计算机研究与发展，1999(36)：96-100

[10] 张瑜慧. 基于 SVM 的语义图像检索技术的研究与实现[D]. 扬州：扬州大学，2007

[11] Chang E, Kingshy G, Sychay, et al. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003(13): 26-38

[12] Lowe D G. Object recognition from local scale-invariant features [J]. Computer Vision, 1999(2): 1150-1157

[13] Wang Jun-qiu. Vision-based Global Localization Using a Visual Vocabulary[C]// Robotics and Automation. Beijing: Peking University, 2005

(上接第 256 页)

从实验结果图我们可以看出，随着样本采样率的增大，采样质量逐渐提高，运行时间也随之增大。通过两种算法实验结果的对比，我们能够明显看出本文所提算法 SWMDS 的采样质量远远高于传统的 Resample 算法，并且当采样率为 5%～15%时，SWMDS 算法的采样质量几乎是 Resample 算法的 5 倍。在运行时间上，由于本文算法 SWMDS 采样前加入了聚类处理，因此相比于传统 Resample 算法要花费更多的时间，但依然保持与采样率之间的线性关系。

结束语 本文针对传统均匀采样在轨迹数据流摘要构造过程中易丢失关键信息的问题，提出一种基于概率密度聚类的数据流偏倚采样算法。该算法在滑动窗口模型下，结合了偏倚采样算法思想，首先基于数据存在密度进行聚类分析，将滑动窗口划分为强簇、弱簇和过度簇，然后针对不同的簇给予不同的采样率，依据各自的采样率对窗口内数据进行偏倚采样，进而构造出更为完善的数据流摘要。算法充分考虑了轨迹数据流自身的分布特性，能够在较低的采样率下获得较高的采样质量。

参 考 文 献

[1] Kun-Ta C, Hung-Leng C, Ming- Syan C. Feature-preserved sampling over streaming data[J]. ACM trans. Knowl. Discov. Data, 2009, 2(4): 1-45

[2] 张春阳，周继恩，钱权，等. 抽样在数据挖掘中的应用研究[J]. 计算机科学，2004，31(2)：126-128

[3] Dimitris S, Antonios D, Timos S. Hierachically compressed wavelet synopses[J]. The VLDB Journal, 2009, 18(1): 203-231

[4] 余波，朱东华，刘嵩，等. 密度偏差抽样技术在聚类算法中的应用研究[J]. 计算机科学，2009，36(2)：207-209

[5] 戴东波，赵杠，孙圣力. 基于概率数据流的有效聚类算法[J]. 软件学报，2009，20(5)：1313-1328

[6] 常建龙，曹锋，周傲英. 基于滑动窗口的进化数据流聚类[J]. 软件学报，2007，18(4)：905-918

[7] 程转流，胡为成. 滑动窗口模型下的概率数据流聚类[J]. 计算机工程与应用，2011，47(4)：141-145

[8] B Ying-yi, C Lei, Wai-Chee F A, et al. Efficient anomaly monitoring over moving object trajectory streams[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, ACM, 2009