

基于语义测度的图像相似性计算研究

陈久军 肖刚 高飞 张元鸣

(浙江工业大学信息学院, 杭州 310014)

摘要 针对图像检索中的低层视觉特征相似性度量问题,提出一种基于语义测度的图像相似性计算方法。该方法在图像区域分割的基础上,通过构建图像区域子块与语义元数据之间的统计映射关系,实现图像内容的统计语义描述,建立图像之间、图像与语义类别、语义类别之间的分层语义相似测度。通过对自然图像库的实验结果表明,该方法在相似图像检索中具有更好的性能。

关键词 语义测度 相似性 图像语义 图像检索

中图法分类号: TP391 **文献标识码:** A **文章编号:** 1006-8961(2007)10-1877-04

Research on Image Similarity Calculation Based on Semantic Measure

CHEN Jiu-jun, XIAO Gang, GAO Fei, ZHANG Yuan-ming

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014)

Abstract In order to solve the problem that the low-level visual features can not express the high-level semantic contents accurately in image retrieval, the paper presents a new image similarity calculating approach using the proposed semantic measure. Based on the segmented image regions, the statistical mapping relationships between image blobs and semantic metadata are constructed and the layered semantic measurements between images are then defined. Experiment results on natural images show good retrieval quality based on the semantic similarity measure method.

Keywords semantic measure, image similarity, image semantic, image retrieval

1 引言

图像间的相似性度量是图像检索中的关键技术之一,一般通过衡量其特征向量之间的“距离”来分析其相似性^[1],常采用的特征向量包括颜色^[2]、形状^[3]、纹理^[4]、区域^[5]等。这种应用低层特征相似性度量的方式,仅仅体现了图像在视觉上的相似性,而不能直接反映图像的语义描述。因此研究语义层面的图像相似测度方法,将有利于提高图像检索结果与人类理解的吻合程度,提高检索性能。

本文研究一种基于语义测度的图像相似性计算方法。该方法在图像区域分割的基础上,通过构建图像子块与语义数据之间的统计映射关系,实现图

像内容统计语义描述,建立图像之间、图像与语义类别、图像语义类别之间的分层语义相似测度。

2 图像语义相似测度分析

2.1 图像语义类别思想

目前图像语义类别的研究已经取得了一定的成果: Szummer 将图像分为室内图像与户外图像两大类^[6]。Gorkani 将户外图像定义为城市图像与风景图像的集合^[7]。Vailaya 则在此基础上重点研究了日出、山脉与森林的图像类别^[8]。Li 研究了 13 类复杂场景的风景图像^[9]。

本文针对风景图像,预定义了 9 个元语义数据: 水 (Water)、草 (Grass)、云彩 (Cloud)、植被

基金项目: 国家重点基础研究发展计划 973 项目 (2003CB317000); 浙江省科技厅项目 (2005C31010)

收稿日期: 2007-07-05; 改回日期: 2007-07-11

第一作者简介: 陈久军 (1977 ~), 男, 2006 年获浙江大学博士学位, 讲师。主要研究方向为图像分析、机器视觉。E-mail: rackycj@zjut.edu.cn

(Foliage)、树木(Tree)、田野(Field)、岩石(Rock)、花朵(Flower)、沙土(Sand)来描述 6 个高层语义类别,包括海岸(Coast)、河流(River)、山脉(Mountain)、森林(Forest)、草原(Prairie)、天空(Sky)。表 1 为风景图像数据库中 500 幅图像的所有标注的语义对象,共计 9 大类 1 536 个元语义实例。

表 1 500 幅自然图像中语义对象标注统计

Tab. 1 Statistical result of semantic object labeled in 500 natural images

语义对象	数量
Water	217
Grass	124
Tree	167
Foliage	137
Field	133
Rock	245
Flower	125
Sand	121
Cloud	267
总计	1 536

2.2 语义相似测度定义

目前在构建语义相似度量时,一般直接计算低层特征与高层语义的加权平方和,如式(1)所示:

$$Dist = (w_1 D_{\text{Semantic}}^2 + w_2 D_{\text{Feature}}^2)^{1/2} \quad (1)$$

式中, w_1, w_2 分别表示语义特征距离和低层特征距离的权重,通过调节 w_1, w_2 的值来确定两种特征的重要程度。但由于低层特征与语义特征在图像内容的表现能力上的差异性,两者往往不在同一个数量级上。如果将两个权值设定在同一个数量级上,则低层特征带来的误差可能会抵消语义相似性的正确性。

本文采用以语义特征为主,低层特征为辅的方式,将图像特征分为核心特征集合与辅助特征集合两个类别。其中核心特征是指那些最能表现图像内容的语义特征,辅助特征则包括了图像低层特征以及部分非核心的语义特征。这样处理的优点为:(1)低层特征的不足不会影响到高层语义的相似性;(2)在语义网络不完善时低层特征又可以起补

充作用。

定义 3 个层次的语义相似测度方案:图像之间、图像与语义类、语义类别之间相似性测度。

图像由元语义集合描述。定义元语义集合 $S = \{S_1, \dots, S_k, \dots, S_p\}$, 其中, $S_k \subset S$ 表示元语义的子集。 M 表示元语义 $s(s \in S_k)$ 与特征空间 F_j 的映射函数,如式(2)所示。

$$F_j = \{f = M(s) \in R^n; s \in S_k; k = 1: p\} \quad (2)$$

定义图像类别 c_i , 给出类别的特征描述:

$$f(c_i) = \langle F_1^R(c_i) \cdots F_{M_i}^R(c_i), F_1^O(c_i) \cdots F_{N_i}^O(c_i) \rangle \quad (3)$$

式中, $\{F_j^R(c_i) | j = 1, \dots, M_i\}$ 描述图像类别 c_i 的核心特征; $F_j^O(c_i) | j = 1, \dots, N_i$ 表示 c_i 的辅助特征。

定义图像 x 与语义类别 c_i 相关的特征:

$$f(x | c_i) = \langle F_1^R(x | c_i) \cdots F_{M_i}^R(x | c_i), F_1^O(x | c_i) \cdots F_{N_i}^O(x | c_i) \rangle \quad (4)$$

建立图像与语义类别 c_i 语义测度函数:

$$S(x | c_i) = S(f(x | c_i), f(c_i)) \\ = \frac{1}{N_i} \prod_{j=1}^{M_i} \tau(F_j^R(x | c_i), F_j^R(c_i)) \cdot \prod_{j=1}^{N_i} \tau(F_j^O(x | c_i), F_j^O(c_i)) \quad (5)$$

其中, $\tau(a, B) = \begin{cases} 1, & \exists i: a = b_i \\ 0, & \forall i: a \neq b_i \end{cases}, B = \{b_i | i = 1, \dots, I$ 表示将所有特征取值 ($B = \{b_i\}$) 与类型特征进行比较的二元表示。在语义类别 c_i 中,度量两个图像 x 与 y 相似程度:

$$S(x, y | c_i) = \frac{1}{N_i} \prod_{j=1}^{M_i} \tau(F_j^R(x | c_i), F_j^R(y | c_i)) \cdot \prod_{j=1}^{N_i} \tau(F_j^O(x | c_i), F_j^O(y | c_i)) \quad (6)$$

根据式(6)获取图像 x, y 之间的相似程度:

$$S(x, y) = \max_i (S(x, y | c_i)) \quad (7)$$

式(6)描述了两个图像在给定的语义类别下所具有的相似程度;式(7)描述了两个图像之间的相似程度。

2.3 图像语义类别统计描述

基于表 1 所给定的自然图像数据库,定义图像的元语义统计向量,如式(3)所示。

将对应于图像 I 的核心特征 $F_j^R(I)$ 定义为二元组 $\langle \text{MetaSemantic}_i, OR_i \rangle$, 其中 MetaSemantic_i 表示图像中所包含的元语义, OR_i 表示对应元语义在图像中的出现统计概率。根据人类对图像内容的认知方

式,图像具体对象所占的区域面积越多,那么该对象的关注程度也就越高。因此, OR_i 定义为元语义所对应对象的区域面积与所有语义对象区域面积的比值,如式(8)所示。

$$OR_i = \frac{Area(Object_i(MetaSemantic_i))}{\sum_j Area(Object_j)} \quad (8)$$

若 OR_i 值很小,则所对应元语义对图像整体的语义特征描述影响较小,过滤此类元语义。

假设元语义出现概率的最低阈值为 $\theta = 0.20$, 将 $OR_i \geq \theta$ 的元语义作为图像类别的核心特征,即

$$F_j^R(c_i) = \{ \langle MetaSemantic_k, OR_k \rangle : OR_k \geq 0.20 \}$$

定义辅助特征集合,如式(9)所示。

$$F_j^O(c_i) = \{ F_1^O(\langle MetaSemantic_k, OR_k \rangle : OR_k < 0.20), F_2^O(ColorHist), F_2^O(EdgeHist) \} \quad (9)$$

给定图像类别 c 和图像 I 的核心特征集合,用式(10)与式(11)表示:

$$F^R(c) = \{ F_i^R : \langle MetaSemantic_i, OR_i \rangle i = 1, \dots, n \} \quad (10)$$

$$F^R(I) = \{ F_j^R : \langle MetaSemantic_j, OR_j \rangle j = 1, \dots, n \} \quad (11)$$

根据式(5), $\tau(a, B)$ 判断图像 I 包含相关元语义的情况,如果图像中不包含图像类别 c 所定义的核心语义特征,返回 $S(I|c) = 0$,表示图像不属于该图像类别;如果通过比较, I 包含相关图像类别的核心特征,计算 OR_i 与 OR_j 之间的标准差,衡量两者的相似程度。

$$Dis(I, c) = \sqrt{\frac{1}{n - 1} \sum_{k=\min(i, j)}^{\max(i, j)} (OR_{j(k)} - OR_{i(k)})^2} \quad (12)$$

同理,式(6)、式(7)计算图像间的相似程度。

3 自然图像语义相似测度实验

3.1 自然风景图像语义统计

分析表 1 自然图像数据库中的相关语义信息,获取风景图像语义类别的统计描述,如表 2 所示。表中包含 6 列 9 行,每一列对应一个风景图像语义类别,每一行对应一个元语义。用 9 个元语义的概率值,描述 6 个风景图像类别。

表 2 自然风景图像语义统计描述
Tab. 2 Statistical semantic representation of natural images

	Coast	Forest	Prairie	Mountain	River	Sky
Water	38% 0.16 ~ 0.40	0	1% 0 ~ 0.02	1% 0 ~ 0.02	26% 0.11 ~ 0.40	1% 0 ~ 0.04
Grass	4% 0.0 ~ 0.16	5% 0 ~ 0.17	18% 0.03 ~ 0.40	2% 0 ~ 0.18	3% 0 ~ 0.60	2% 0 ~ 0.05
Tree	0	20% 0.02 ~ 0.40	1% 0 ~ 0.01	1% 0 ~ 0.01	0	0
Foliage	7% 0 ~ 0.27	63% 0.40 ~ 0.94	10% 0 ~ 0.22	22% 0 ~ 0.31	22% 0.03 ~ 0.50	1% 0 ~ 0.03
Fields	2% 0 ~ 0.09	3% 0 ~ 0.07	20% 0.19 ~ 0.50	7% 0 ~ 0.19	1% 0 ~ 0.04	1% 0 ~ 0.04
Rock	18% 0.02 ~ 0.40	0	3% 0 ~ 0.11	42% 0.2 ~ 0.70	18% 0.02 ~ 0.40	0
Flower	0	0	18% 0.1 ~ 0.43	5% 0 ~ 0.10	0	0
Sand	10% 0 ~ 0.20	0	5% 0.03 ~ 0.30	0	0	0
Cloud&Sky	21% 0.17 ~ 0.40	5% 0 ~ 0.09	23% 0.05 ~ 0.40	25% 0.16 ~ 0.40	18% 0.01 ~ 0.20	95% 0.78 ~ 1

3.2 自然风景图像语义类别之间的相似性

考察语义类别之间的相似性,如表 3 所示。从表

中数据可知,在“Coast”类型的图像,有 88% 的“Coast”图像可以准确的归类,而有 12% 的图像可能

被误认为是“River”类型的图像。这主要是由于“Coast”与“River”类型的图像具有相似的元语义组成。

表 3 图像语义类别间相似性

Tab.3 Similarity measure of image categories

	Coast	Forest	Prairie	Mtn	River	Sky
Coast	88.0	0	0	0	12.0	0
Forest	0	92.5	1.9	0	5.6	0
Prairie	1	0	96.5	2.5	0	0
Mountain	5.7	0	6.5	80.6	7.2	0
River	8	0	0	0	92	0
Sky	0	0	2.5	2.5	0	95

3.3 基于语义相似测度的图像检索

基于表 2 的数据,定义风景图像 6 个语义类别的核心语义特征描述:

$$F^R(\text{Coast}) = \{F_1^R(\langle \text{Water}, 0.38 \rangle), F_2^R(\langle \text{Sky}, 0.21 \rangle)\}$$

$$F^R(\text{Forest}) = \{F_1^R(\langle \text{Foliage}, 0.63 \rangle), F_2^R(\langle \text{Tree}, 0.20 \rangle)\}$$

$$F^R(\text{Prairie}) = \{F_1^R(\langle \text{Sky}, 0.23 \rangle), F_2^R(\langle \text{Field}, 0.20 \rangle)\}$$

$$F^R(\text{Mountain}) = \{F_1^R(\langle \text{Rock}, 0.42 \rangle), F_2^R(\langle \text{Sky}, 0.25 \rangle), F_3^R(\langle \text{Foliage}, 0.22 \rangle)\}$$

$$F^R(\text{River}) = \{F_1^R(\langle \text{Water}, 0.26 \rangle), F_2^R(\langle \text{Foliage}, 0.22 \rangle)\}$$

$$F^R(\text{Sky}) = \{F_1^R(\langle \text{Sky}, 0.95 \rangle)\}$$

假设有 4 个图像 a, b, c, d , 其统计量如下所示:

$$F^R(a) = \{F_1^R(\langle \text{Water}, 0.33 \rangle), F_2^R(\langle \text{Sky}, 0.49 \rangle)\}$$

$$F^R(b) = \{F_1^R(\langle \text{Water}, 0.77 \rangle)\}$$

$$F^R(c) = \{F_1^R(\langle \text{Sky}, 0.68 \rangle), F_2^R(\langle \text{Fields}, 0.32 \rangle)\}$$

$$F^R(d) = \{F_1^R(\langle \text{Sky}, 0.72 \rangle)\}$$

根据前面的定义, (1) 通过核心语义特征筛选, 图像 a 与图像类别“Coast”类别相关, 相似距离为 $Dis(a, \text{Coast}) = 0.28$; (2) 图像 b 与图像类别“Coast”类别相关 $Dis(b, \text{Coast}) = 0.39$; (3) 图像 c 与图像类别“Prairie”类别相关 $Dis(c, \text{Prairie}) = 0.197$; (3) 图像 d 分别与图像类别“Coast”、“Prairie”、“Mountain”、“Sky”类别相关, 分别计算相似距离 $Dis(d, \text{Coast}) = 0.51$, $Dis(d, \text{Prairie}) = 0.49$, $Dis(d, \text{Mountain}) = 0.47$, $Dis(d, \text{Sky}) = 0.23$; 根据相似距离判断, 图像 d 应该属于“Sky”图像类别。通过上述分析, 相对于低层相似度量方法, 应用语义相似测度准则, 所获取的图像具有更好的语义相关性。

4 结 论

本文研究了图像语义相似测度问题, 深入探讨了风景图像的语义类别的统计描述。通过构建图像区域子块与元语义、语义类别之间的统计映射关系, 实现图像内容的统计语义描述, 建立图像之间、图像与语义类别、图像语义类别之间的分层语义相似测度。通过实例分析, 验证了语义相似测度方法的有效性。

参考文献 (References)

- 1 Maxwell B A, Buddemeier R W. Coastal typology development with heterogeneous data sets[J]. *Regional Environmental Change*, 2002, 3(1-3): 77~87.
- 2 Swain M, Ballard D. Color indexing[J]. *International Journal of Computer Vision*, 1991, 7(1): 11~32.
- 3 Sun Jun-ding, Cui Jiang-tao, Wu Xiao-sheng, et al. Color image retrieval based on color and shape features[J]. *Journal of Image and Graphics*, 2004, 9(7): 820~827. [孙君顶, 崔江涛, 毋小省等. 基于颜色和形状特征的彩色图像检索方法[J]. *中国图象图形学报*, 2004, 9(7): 820~827.]
- 4 Wan Hua-lin, Chowdhury M U. Texture feature and its application in CBIR[J]. *Journal of Computer-aided Design & Computer Graphics*, 2003, 15(2): 195~199. [万华林, Chowdhury M U. 图像纹理特征及其在 CBIR 中的应用[J]. *计算机辅助设计与图形学学报*, 2003, 15(2): 195~199.]
- 5 Zhao Yu-feng, Zhao Yao, Zhu Zhen-fen. New region-based image similarity calculation[J]. *Journal of the China Railway Society*, 2007, 29(1): 55~60. [赵玉凤, 赵耀, 朱振峰. 一种新的基于区域的图像相似性计算[J]. *铁道学报*, 2007, 29(1): 57~62.]
- 6 Szummer M, Picard R. Indoor-outdoor image classification[A]. In: *Proceedings of IEEE International Workshop on Content-based Access of Image and Video Databases*[C], Bombay, India, 1998: 42~51.
- 7 Gorkani M, Picard R. Texture orientation for sorting photos at a glance[A]. In: *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR1994)*[C], Jerusalem, Israel, 1994: 459~464.
- 8 Vailaya A, Figueiredo M, Jain A K, et al. Image classification for content-based indexing[J]. *IEEE Transactions on Image Processing*, 2001, 10(1): 117~130.
- 9 Li F F, Pietro P. A bayesian hierarchical model for learning natural scene categories[A]. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*[C], 2005, 2(2): 524~531.