第33卷 第13期 计 算 机 工 程 2007年7月
Vol.33 No.13 Computer Engineering July 2007
·多媒体技术及应用· 文章编号：1000—3428(2007)13—0216—02 文献标识码：A 中图分类号：TP37

基于潜在语义分析的视频检索

胡双演，李俊山，李建军
(第二炮兵工程学院研究生4队，西安 710025)

摘 要：潜在语义分析技术是建立在视频分析基础之上的，它通过某种映射关系来建立视频特征矩阵，实现了基于内容的视频检索。阐述了潜在语义分析技术，进行了视频颜色和纹理特征提取研究，实验结果表明，潜在语义分析对于视频内容检索效果较好。

关键词：基于内容的视频检索；潜在语义分析；奇异值分解

Video Retrieval Based on Latent Semantic Analysis

HU Shuangyan, LI Junshan, LI Jianjun
(Graduate Team 4, Second Artillery Engineering College, Xi'an 710025)

【Abstract】Latent semantic analysis is based on video analysis. It constructs the video feature matrix via a kind of mapping and realizes video based content retrieval. The paper discusses the main idea of latent semantic analysis and researches the method color texture extraction. The result proves latent semantic analysis performs preferable in content based video retrieval.

【Key words】content based video retrieval; latent semantic analysis; singular value decomposition

1990年，University of Chicago, Bell Communications Research和University of Western Ontario的Scott Deerwester, Thomas K.Landauer等5位学者共同提出了潜在语义分析(Latent Semantic Analysis，LSA)这一自然语言处理的方法。在自然语言理解、文本分析、信息过滤、情报检索等领域得到了广泛的应用。其基本思想是文本中的词与词之间存在某种联系，即存在某种潜在的语义结构，因此，采用统计的方法来寻找该语义结构，并且用语义结构来表示词和文本，这样的结果可以达到消除词与词间的相关性，并且能够化简文本向量。

近年来，随着潜在语义分析技术的逐渐成熟，人们已经开始将该方法引入到图像检索领域中，目前只有少量文献报道过相关方面的研究成果[1]。本文将潜在语义分析技术应用于视频检索中，以实现基于语义的视频检索。

1 潜在语义分析

LSA技术是建立在视频分析基础之上的，它需要对视频序列进行相应的分析，构造视频词典，通过某种映射关系来建立视频特征矩阵[2]，实现基于语义分析的视频检索。视频序列分解分为两种类型：(1)把像素、帧的区域等基本单元看作“字词”，把它们映射到一个或多个包含局部相似的视频序列的视频词典。文献[3]对视频词典的构造进行了全面的阐述。(2)“字词”凝聚成为一个实体，类似于上下文关系，比如视频帧、镜头、场景和语义结构等，构成索引和比较的主要实体。文章采用了第2种视频分解类型，将视频序列分解成帧的集合，表示为$F=\{f_i \mid i=1,2,\ldots\}$。

LSA利用截断的奇异值分解(Singular Value Decomposition，SVD)[4]生成的低维潜在语义空间来描述元素间的语义结构。它不同于向量空间模型(Vector Space Model，VSM)，潜在语义空间中对应于视频向量各个维度的不再是各个元素，而是存在于元素之间的潜在概念(Latent Concept)[5]。其本质就是用潜在概念的线性组合来描述视频序列和特征元素。

下面介绍潜在语义分析的实现过程：

Step1 构造一个$M\times N$的矩阵$\boldsymbol{A}$，其中M为视频词典中的“词汇”，N为视频序列中的上下文关系，$\boldsymbol{A}_{ij}$表示M中的“词汇”i在N_j中出现的次数。

Step2 利用矩阵的奇异值分解方法对矩阵$\boldsymbol{A}$进行变换。

设$A=USV^T$，其中$UU^T=VV^T=\mathrm{I}$，$S=diag(\sigma_1,\sigma_2,\ldots,\sigma_L)$，$\sigma_1\ge\sigma_2\ge\ldots\sigma_L\ge 0$，$L=\min(M,N)$。通过上述的奇异值分解，得到了矩阵$\boldsymbol{A}$的$L$个奇异值，其中的$k$个为非零值，由矩阵论的知识可知，矩阵$\boldsymbol{A}$可以用下式来近似表示：

$$\tilde{A}=U_kS_kV_k^T \tag{1}$$

其中，$\boldsymbol{U_k}$是矩阵$\boldsymbol{U}$的前k列构成的矩阵，S_k是由矩阵$\boldsymbol{A}$的k个大于零的奇异值构成的对角阵，$\boldsymbol{V_k}$是矩阵$\boldsymbol{V}$的前k行构成的矩阵。在给定k的条件下，利用SVD得到的近似矩阵$\tilde{A}$能够保证是最优的。

Step3 设待检索的视频序列为矩阵$\boldsymbol{Q}$，则

$$Q^T\tilde{A}=Q^TU_kS_kV_k^T=(Q^TU_k)(S_kV_k^T) \tag{2}$$

设$P_q=Q^TU_k$，P_j表示$S_kV_k^T$的第j个分量，因此，可以用P_q和P_j之间的夹角θ_j来衡量它们的相似度。$\cos(\theta_j)$值越大，相似度越高。

$$\cos(\theta_j)=m_c(p_j,q)=\frac{p_q\cdot p_j}{\|p_q\|\cdot\|p_j\|} \tag{3}$$

其中，$p_q\cdot p_j$为两个向量的内积

$$\|X\|=(X\cdot X)^{1/2}=(\sum_{i=1}^{n}x_i^2)^{1/2}$$

作者简介：胡双演(1978－)，男，博士研究生，主研方向：图像处理，视频检索；李俊山，博士生导师、教授；李建军，博士研究生

收稿日期：2006-07-10 **E-mail：**shuang0000shuang@163.com

2 视频特征提取

主要使用视频帧的颜色和纹理特征来构造视频词典。

2.1 颜色特征提取

传统的 RGB 颜色空间的可分辨色差是非线性的，而且和人的视觉特性相差较大，相反，HSV 颜色模型具有较好的线性伸缩性且比较接近人的视觉特性，其可感知的色差与颜色分量的对应样值的 Euclid 距离成比例，因此这种模型比 RGB 颜色模型更容易被接受。由于人的眼睛对颜色的分辨力存在很大局限性，因此必须对颜色空间进行适当的量化，从而减少特征计算时的计算量，提高检索速度和效率。将 h、s、v 3 个分量按人的颜色感知进行非等间隔的量化，根据对颜色模型的大量分析和计算，可把色调 h 分成 16 份，饱和度 s 和亮度 v 分别分成 4 份，并据色彩的不同范围进行量化，量化后的色调、饱和度和亮度值分别表示为 H、S、V。由于亮度分量 V 对彩色检索的贡献不是很大，因此将该分量忽略掉，用一维向量 H×S，即 64×1 维向量来表示视频的颜色特征。

2.2 纹理特征

纹理通常定义为图像的某种局部性质，或是对局部区域中像素之间关系的一种度量，可用来对图像中的空间信息进行一定程度的定量描述。纹理测量的方法有很多种，例如：自相关函数，功率谱，正交变换，灰度级共生矩阵，灰度级行程长，灰度级差分，滤波模板，相对极值密度，离散马尔可夫随机场模型，自回归模型，同时自回归模型，随机Mosaic模型等[6]，采取灰度级共生矩阵法进行视频帧纹理的计算，利用以下 6 个特征量表，沿着 0°、45°、90°和 135° 4 个角度方向来表示纹理特征。

极大概率：

$$f_1 = \max_{i,j} P(i,j,r,\theta) \tag{4}$$

能量：

$$f_2 = \sum_i \sum_j P(i,j,r,\theta)^2 \tag{5}$$

熵：

$$f_3 = \sum_i \sum_j P(i,j,r,\theta) \log P(i,j,r,\theta) \tag{6}$$

对比度：

$$f_4 = \sum_i \sum_j (i-j)^2 P(i,j,r,\theta) \tag{7}$$

相关：

$$f_5 = \sum_i \sum_j (i-u_x)(j-v_y) P(i-j) / \sigma_x \sigma_y \tag{8}$$

一致性：

$$f_6 = \sum_i \sum_j \frac{P(i,j,r,\theta)}{1+|i-j|} \tag{9}$$

其中，$P(i,j,r,\theta) = P(i,j,r,\theta)/R$ 为灰度共生矩阵中第 (i,j) 元素；$P(i,j,r,\theta)$ 为分别具有灰度为 i 和 j，两像素同时出现的频数；R 为可能提供的最多像素对数目，分别为 μ_x、μ_y、σ_x、σ_y 时 P_x 和 P_y 的均值和方差；P_x 和 P_y 分别定义为

$$P_x = \sum_i \frac{P(r,j,r,\theta)}{R} \tag{10}$$

$$P_y = \sum_i \frac{P(r,j,r,\theta)}{R} \tag{11}$$

3 实验结果

本文将颜色和纹理特征相结合，以此来作为视频词典，利用上述 LSA 技术在新闻、足球比赛、广告、MTV、游戏宣传片等 9 类视频库中进行了实验，实验结果表明，LSA 方法对于视频内容的检索非常有效。实验结果如图 1 所示，本文选取了 5s(25 帧/s)的新闻短片作为查询示例。检索结果按照 $\cos(\theta_j)$ 值从小到大排序。第 1 幅图像为待检索视频序列中的某帧图像，其他图像为检索出来的结果。

图 1 本文采用的 LSA 方法检索结果

4 总结

基于 LSA 的视频检索方法能显著地提高视频检索系统的性能，提高图像检索的质量，是一种具有良好前景的视频检索技术。但是，LSA 技术还不很完善，尚存在许多问题亟待进一步的研究和探讨。例如：视频词典的优化设计和完善，扩充视频特征元素等。通过今后的深入研究，LSA 技术将会在视频检索领域展现出它的特殊优势。

参考文献

1 Sdaroff S, Cascia M L, Sethi S. Unifying Textual and Visual Cues for Content-based Image Retrieval on WWW[J]. Computer Vision and Image Understanding, 1999, 75(1/2): 86-98.

2 Souvannavong F, Merialdo B, Huet B. Video Content Modeling with Latent Semantic Analysis[C]//Proc. of the 3rd International Workshop on Content-based Multimedia Indexing. 2003.

3 Pavel B. Survey of Clustering Data Mining Techniques[R]. San Jose, CA: Accrue Software Inc., 2002.

4 Liu Xin, Gong Yihong. Video Summarization and Retrieval Using Singular Value Decomposition[J]. Journal of ACM Multimedia Systems, 2003, 9(2): 157-168.

5 Deerwester S, Dumais S T, Landauer T K. Indexing by Latent Semantic Analysis[J]. Journal of the American Society for Information Science, 1990, 41(6): 391-407.

6 赵荣椿. 数字图像处理导论[M]. 西安：西北工业大学出版社, 2000.