

基于多线性子空间 KL 变换的可伸缩视频压缩方法

刘健^① 吴飞^① 姚蕾^② 庄越挺^①

^①(浙江大学计算机科学与技术学院 杭州 310027)

^②(宾夕法尼亚州立大学计算机科学与工程系 University Park PA 16802)

摘要: 该文提出一种基于多线性子空间 KL(Karhunen-Loeve)变换的可伸缩视频压缩算法。该算法将纹理不同的图像子块投影到多个线性子空间中去,然后分别进行 KL 变换编码,使得每个子空间内变换系数更接近高斯分布,从而最大限度地变换编码阶段提高压缩效率。同时,该算法结合多尺度小波分解,实现了质量可伸缩的视频编码,提高了视频压缩率。通过与 DCT、小波变换以及多向 DCT 算法进行比较,证明该算法可获得更好的率失真性能。

关键词: 可伸缩视频编码; KL 变换; 多线性子空间

中图分类号: TN919.81

文献标识码: A

文章编号: 1009-5896(2010)05-1188-05

DOI: 10.3724/SP.J.1146.2009.00463

A Scalable Video Coding Scheme Using KL Transform in Multiple Linear Subspaces

Liu Jian^① Wu Fei^① Yao Lei^② Zhuang Yue-ting^①

^①(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

^②(Computer Science and Engineering Department, Pennsylvania State University, University Park, PA 16802, USA)

Abstract: This paper proposes a novel transform coding method based on the Karhunen-Loeve (KL) transform in multiple linear spaces. This method uses multiple linear subspaces to approximate image signals, and use a KL transform for each linear subspace, so that the transform coefficients in each subspace is close to Gaussian, which effectively de-correlates the coefficients. An SNR-scalable video coding scheme is developed by combining the method with wavelet decomposition. Experiments prove that this scheme accomplishes better coding efficiency than the schemes employing DCT, DWT, and Directional DCT methods.

Key words: Scalable video coding; KL transform; Multiple subspaces

1 引言

在视频数据压缩中,变换编码是各编码标准普遍采用的主要技术。由于自然视频图像在相邻像素间存在着很大的空间相关性,通过变换,可以减小或去除像素间的相关性,把有记忆信源转换为无记忆信源,从而使得标量量化结合熵编码可以获得良好的压缩效率。常用的变换编码技术有 Hadamard 变换、KL 变换、DCT 变换和离散小波变换(DWT)等。在这些变换方法中,DCT 变换由于具有良好的去相关性和能量集中效果,并且存在快速的算法,因此被广泛应用于 JPEG, MPEG1/2/4, H.263/H.264 等国际标准中。

然而 DCT 在理论上存在一定局限性。首先,传统 DCT 变换在水平和垂直两个方向上分别进行一维变换,这对于水平和垂直纹理具有较好的变换

效果,而对其他方向纹理的保持度不佳,易产生 Gibbs 伪影。文献[1]和文献[2]分别研究了多向 DCT 算法,其在变换中采用多种方向模式,文献[3]则在小波变换中引入方向性,这些方法都在一定程度上提高了压缩效率。

另一方面,根据编码理论,理想的变换应使得变换系数之间达到统计独立,这时变换系数的一阶熵就表示了该信源无失真编码码率的下限,采用标量量化和熵编码即可达到良好压缩效果。对于平稳信号(统计特性不随时间变化的信号),KL 变换是最优的变换,它可以使变换系数达到不相关^[4]。假使变换系数满足高斯分布,统计意义上的不相关和独立是等价的。然而,实际图像的变换系数常常不满足高斯分布,而是更接近于混合高斯分布,需要用多个高斯分布来拟合。因此如果用混合线性模型来对视频信号进行压缩,将可以更加接近理论上的最优效果。

近年来出现的广义主成分分析(Generalized Principle Component Analysis, GPCA)方法^[5-7],

2009-04-02 收到, 2009-11-10 改回

国家自然科学基金(60533090, 60603096), 长江学者和创新团队发展计划资助课题

通信作者: 刘健 ken3300@zju.edu.cn

可以把位于多个线性子空间中的数据分割出来,并分别估计其所在线性子空间的维数和基向量。通过GPCA可将视频图像信号映射到多个线性子空间中,使得在每一个线性子空间中的变换系数都接近于高斯分布,于是应用这一方法进行图像和视频编码可获得更接近于理论上最优的压缩效率。文献[8]和文献[9]利用该思路在图像编码上获得了很好的结果。

本文把GPCA方法和KL变换应用于视频压缩,提出了基于多线性子空间KL变换的可伸缩视频编码方法。该方法的第1步仍然是进行I帧和P/B帧的模式选择以及运动预测,而后对I帧和P/B帧的预测残差进行变换编码。具体而言,该方法首先用离散小波变换得到原始帧的一个多尺度分解,其中低频信号作为基本层,高频信号作为增强层。然后用GPCA方法对各尺度上的图像信号进行子空间分割,最后对各子空间内的数据分别进行KL变换编码。解码时如果只截取低频小波信号,则得到基本层的解码;如果加上高频小波信号,则得到较完整的增强层解码。在实验中将本文方法与H.264的可伸缩视频编码^[10]进行了比较,初步表明这种方法的压缩效率优于基于DCT的编码方法。

本文的第2节介绍用于多线性子空间分割的GPCA算法,第3节阐述如何将多线性子空间KL变换应用于可伸缩视频编码,第4节给出实验结果,第5节为结论。

2 广义主成分分析方法(GPCA)概述

GPCA是一种把样本点集合分割并投影到多个线性子空间中去的代数几何方法。假设有子空间 $S_i \subset \mathbb{R}^D$,其维数是 d_i , $0 < d_i < D$, S_i 可以用 $D - d_i$ 个一次多项式来表示。

$$S_i = \{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{C}_i^T \mathbf{x} = 0 \Leftrightarrow \bigwedge_{j=1}^{D-d_i} (\mathbf{c}_{ij}^T \mathbf{x} = 0)\} \quad (1)$$

其中 $\mathbf{C}_i \doteq [c_{i1}, \dots, c_{i(D-d_i)}] \in \mathbb{R}^{D \times (D-d_i)}$ 是 S_i 的补空间 S_i^\perp 的基。类似地, n 个子空间 $\{S_i \subset \mathbb{R}^D\}_{i=1}^n$ 也可以用一系列齐次多项式来表示。

$$\begin{aligned} \bigvee_{i=1}^n (\mathbf{x} \in S_i) &\Leftrightarrow \bigvee_{i=1}^n \bigwedge_{j=1}^{D-d_i} (\mathbf{c}_{ij}^T \mathbf{x} = 0) \\ &\Leftrightarrow \bigwedge_{\sigma} \bigvee_{i=1}^n (\mathbf{c}_{i\sigma(i)}^T \mathbf{x} = 0) \end{aligned} \quad (2)$$

式(2)右边是应用德摩根律交换与、或运算得到的, σ 表示从基 \mathbf{C}_i 中选出的一个特定的法向量 $\mathbf{c}_{i\sigma(i)}$ 。可以看出式(2)右边展开后共有 $\prod_{i=1}^n (D - d_i)$ 个形如 $\bigvee_{i=1}^n (\mathbf{c}_{i\sigma(i)}^T \mathbf{x} = 0)$ 等式,而

$$\bigvee_{i=1}^n (\mathbf{c}_{i\sigma(i)}^T \mathbf{x} = 0) \Leftrightarrow \prod_{i=1}^n \mathbf{c}_{i\sigma(i)}^T \mathbf{x} \doteq p_{n\sigma}(\mathbf{x}) = 0 \quad (3)$$

式(3)是 n 个一次多项式的乘积,它展开后是一个含有 D 个变量的 n 次齐次多项式,其中共有 $M_n =$

$\binom{n+D-1}{n}$ 个 n 次单项式。把多项式中的 x 看成已知变量,其系数看成未知变量,则多项式(3)可以写成系数 β_n 的线性组合:

$$p_n(\mathbf{x}) = \beta_n^T \mathbf{v}_n(\mathbf{x}) = \sum_{n_1 + \dots + n_D = n} \beta_{n_1, \dots, n_D} x_1^{n_1} \cdots x_D^{n_D} \quad (4)$$

其中 $\mathbf{v}_n(\mathbf{x}) \in \mathbb{R}^{M_n}$ 是 n 次的维罗内斯嵌入(Veronese embedding)^[11]。根据公式(4),给定足够数量的点 $\{\mathbf{x}_i\}_{i=1}^N$,可以解出所有的多项式 $\{p_n(\mathbf{x})\}$,并可以通过多项式 $p_n(\mathbf{x})$ 在子空间内一点 \mathbf{x}_i 上的导数 $\{Dp_n(\mathbf{x})\}$,解出各子空间的法向量 $\{\mathbf{c}_i\}$ 。

关于GPCA的详细算法和证明请参见参考文献[5]。

3 基于多线性子空间KL变换的多尺度视频编码

根据编码理论,在变换系数符合或接近高斯分布的情况下,KL变换可使变换系数达到统计独立,因而在去相关性上是最优变换。但是获取的实际图像信号中,图像子块变换系数常常与高斯分布不一致,而是更接近于多个高斯分布的混合体。因此,如能把图像子块映射到多个线性子空间中去,再分别进行KL变换编码,会取得更好的压缩效果。本节首先介绍如何对图像信号进行多线性子空间KL变换编码,然后在此基础上结合小波分解对视频中I帧和P/B帧的预测残差进行多尺度编码,实现质量可伸缩的视频压缩。

3.1 单一尺度图像变换编码

假设一帧图像大小为 $W \times H$,每个像素有Luma, Cb和Cr 3个颜色分量。将每个视频帧划分成 $l \times m$ 大小的子块,则每个子块中各像素值可按一定顺序组成 D 维向量 $\mathbf{v} \in \mathbb{R}^D$ ($D = 3 \cdot l \cdot m$),向量个数总共是 $N = (W \times H) / (l \times m)$ 。由上文讨论可知,由于图像包含纹理特征迥异的不同区域,在一个统一的线性空间中对这些向量进行稀疏表示是有局限性的,然而对于每一个区域应用一个线性模型是合理的。假使我们已经把 N 个向量划分为 n 个不同的组: $G = \bigcup_{i=1}^n g_i$,每一个组 g_i 都构成一个线性子空间 S_i ,对每一个子空间 S_i 找到基向量组 $B_i = \{b_{ij}\}_{j=1}^{d_i}$,使得 g_i 中的向量 \mathbf{v} 都可以表示成 $\mathbf{v} = \sum_{j=1}^{d_i} \alpha_j b_{ij}$,其中 d_i 是线性子空间 S_i 的维数, α_j 是变换系数。由于变换系数 α_j 间的相关性很小,能量集中在前几个系数上,多数系数接近于零,因此通过舍弃部分系数或者进行量化和熵编码即可实现较好的数据压缩。如果子空间划分得当,则在每个子空间中变换系数 α_j 都更接近高斯分布,那么系数的

统计独立性更好, 熵编码的效率更高。

3.2 小波域内的多尺度变换编码

3.1 节中所述的多线性子空间 KL 变换编码方法直接对图像子块进行处理, 这种做法有一定局限性。首先, 对于普通图像, 该方法并没有充分消除相邻图像块之间的相关性。在视频编码中, 通常先对图像帧进行帧内或帧间预测, 而后对预测残差进行变换编码。运动预测的目的就是消除图像子块之间的时间相关性或空间相关性。其次, 它没有区分图像信号的高频和低频分量。图像子块中既有低频分量(较平滑区域), 也有高频分量(梯度较大的纹理区域), 低频分量只需用较低分辨率表示, 只有高频分量需要用到较高分辨率。因此, 本文进一步提出利用小波的多分辨率分析功能, 对低频和低频分量进行分层处理, 以进一步提高压缩效率。

如图 1 所示, 一帧图像经过 3 层小波分解, 形成了一个低频分量(左上角的系数块)和 3 层各 3 个子带的高频分量。由于位于 3 个子带相同位置上的系数对应着原图像中同一空间位置上的像素, 因此把同一层中各子带相同位置上的系数组成一个向量。低频分量单独进行 KL 变换, 与高频分量分开处理。高频分量中很多向量的能量接近于零, 因此设定能量阈值 ε , 当 $\|v\| > \varepsilon$ 时, 向量 v 参与运算, 否则向量 v 用 0 表示。这样可以大大减少参与运算的向量个数, 提高运算效率。同时不同小波层中的向量分别进行 GPCA 分组和 KL 变换, 得到的系数按层序排列, 可以实现质量可伸缩性。编码流程总结如下:

算法 1 多线性子空间 KL 变换编码

输入 视频序列中的 I 帧和 P、B 帧预测残差

Begin

(1)对输入帧进行 L 层小波分解;

(2)对步骤 1 中得到的每一层小波系数分别构建向量, 形成 L 个向量集;

(3)用 GPCA 方法对步骤 2 中得到的向量集分别进行分组;

(4)对步骤 3 中得到的向量组分别进行 KL 变换, 得到变换矩阵和变换系数;

(5)对步骤 4 中得到的系数进行量化和熵编码, 并按顺序输出形成码流;

End

3.3 复杂度分析

GPCA 方法的时间复杂度较高。据第 2 节中所述算法, 要解出各子空间的法向量, 须求得多项式 $\{p_n(x)\}$ 的系数。每个多项式 $p_n(x)$ 含有 M_n 项, 由于样本点并不是精确地位于各子空间内, 我们需要足够多的 N 个样本点来估计 $p_n(x)$ 的系数, 因此须求解一个 $M_n \times N$ 的方程组。在有误差的情况下, 用 SVD 分解的方法求解, 复杂度是 $O(M_n^2 N)$ 。 N 的大小可以与子空间个数相当, $M_n = \binom{n+D-1}{n}$ 则是一个

n^D 数量级的量。当输入向量维数 D 过大时, 算法复杂度是不可接受的。因此, GPCA 只适用于子空间维数很小的情况。这时, 需先对输入向量用 PCA 方法进行初步降维, 以减小 D 。在视频帧编码的实验中, 多数情况下子空间个数 $2 \leq n \leq 6$, $D \leq 6$, $M_n \leq 462$ 。而 DCT 算法的复杂度则较低, 是 $O(b^3)$, 其中 b 是变换块的大小, 且 DCT 还有优化的快速算法。多向 DCT(D-DCT)由于需要对多种模式进行选择, 因此时间复杂度略高。

尽管 GPCA 复杂度较高, 但是通过预先降维、限制子空间个数等优化方法, 可使其运算时间减少到可以接受的范围内。随着处理器运算能力的不断提高, 应用这一算法进行图像视频压缩将逐渐变得可行。表 1 对比了在 MATLAB 7.5 上用 DCT 和 GPCA 方法分别压缩一幅视频帧所消耗的时间, 实验程序在 intel P4 3.0GHz 的 CPU 上运行。

表 1 DCT, D-DCT 和 GPCA 运行时间对比(s)

	DCT	D-DCT	GPCA
Akiyo(CIF)	0.4965	2.7825	6.4526
Foreman(QCIF)	0.1310	1.1352	3.6238

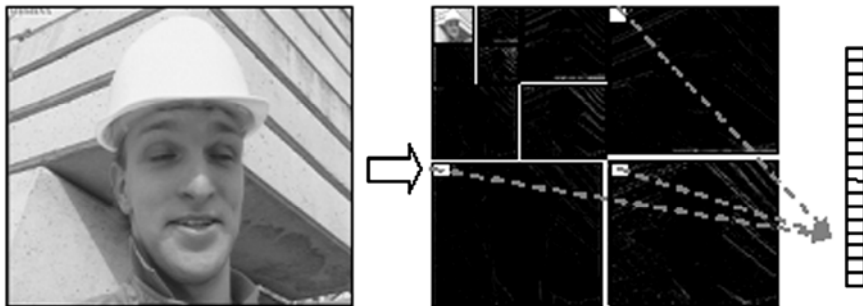


图 1 用小波分解进行多尺度编码

4 实验结果

本文采用一系列实验来验证多线性子空间 KL 变换(以下简称为 M-KL 变换)在视频编码中的有效性。首先在静态图像上测试 M-KL 变换编码。本文选取了“Akiyo”,“Mobile”, 2 个视频序列的首帧, 分别用 DCT 变换、多向 DCT 变换^[2](以下简称 D-DCT)、离散小波变换(以下简称 DWT)和多线性子空间 KL 变换进行压缩, 并比较在保留相同比例变换系数时压缩图像的 PSNR 值, 图 2 给出了不同变换方法下 PSNR 值比较, 其中离散小波变换编码采用了 JPEG 2000 的算法。

图 3 是小波分解后所得的高频图像块在各子空间内的分布情况, 右边图中不同灰度的图像子块代表了不同的子空间, 可看出 GPCA 具备一定的纹理分割能力。

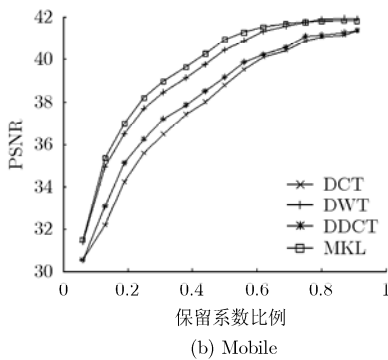
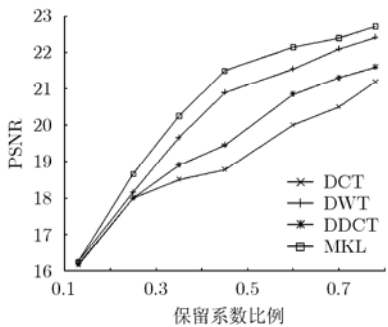


图 2 静态图像 M-KL 变换编码与 DCT, D-DCT 和 DWT 编码的结果比较

图 4 是多线性子空间 KL 变换与 DCT 变换编码的主观效果比较, 左图是 DCT 变换的结果, 右图是 M-KL 变换的结果, 分别保留相同比例的系数。可以看出 DCT 变换容易在文字和其他纹理周围形成明显的 Gibbs 伪影, 而多线性子空间 KL 变换压缩得到的图像使得伪影大大减弱。

最后对 P/B 帧的预测残差进行 M-KL 变换编码。实验中采用了 H.264/AVC 的可伸缩视频编码参考模型 JSVM 9.0^[12], 分别用多向 DCT 和 M-KL 变换替代原有的整数 DCT 变换, 对 Akiyo 和 Mobile 视频序列的前 32 帧预测残差进行压缩, 所得结果如图 5 所示。

5 结束语

本文针对传统 DCT 变换和 KL 变换在图像视频

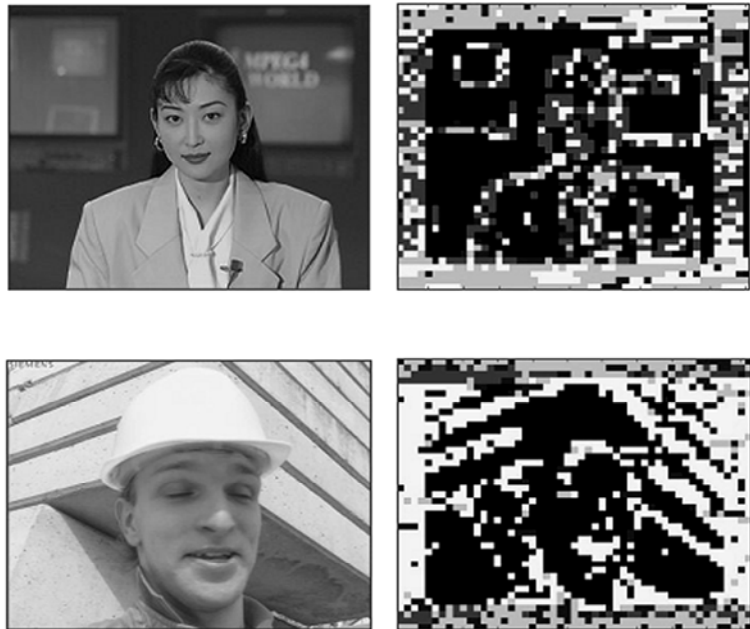


图 3 静态图像帧经过小波分解后所得的高频信号在多线性子空间中的分布情况



图 4 视频例子 Mobile 中的 I 帧分别采用 DCT 变换和 M-KL 变换的主观效果比较

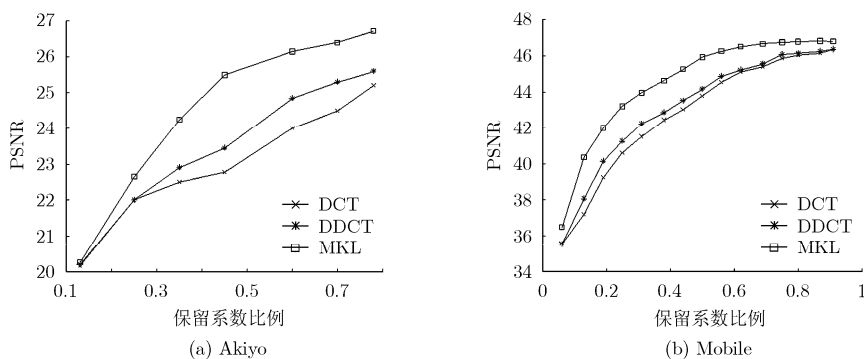


图 5 视频预测残差 M-KL 变换编码与 DCT 和 DDCT 编码的结果比较

编码中的不足,提出了多线性子空间 KL 变换方法。在假设自然图像信号具有多线性子空间分布特性前提下,该方法在理论上可更接近最优的变换效果。结合多尺度小波分解,本文进一步构建了一种质量可伸缩的视频编码方法,并通过实验证明该方法的有效性。尽管该方法在计算上比 DCT 和小波变换方法更为复杂,但是随着硬件计算能力的不断提高,相信该方法会有更好的应用前景。在将来的工作中,GPCA 的多线性子空间分割功能可以进一步用来分割出具有语义的图像区域,从而有助于基于区域的编码等应用。

参 考 文 献

- [1] Zeng Bing and Fu Jing-jing, *et al.* Directional discrete cosine transforms — A new framework for image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, 18(3): 305–313.
 - [2] Xu Hao, Xu Ji-zheng, and Wu Feng, *et al.* Lifting-based directional DCT-like transform for image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, 17(10): 1325–1335.
 - [3] Ding Wen-peng, Wu Feng, and Wu Xiao-lin, *et al.* Adaptive directional lifting-based wavelet transform for image coding. *IEEE Transactions on Image Processing*, 2007, 16(2): 416–427.
 - [4] 张旭东, 卢国栋, 冯健等. 《图像编码基础和小波压缩技术——原理、算法和标准》. 第一版, 北京: 清华大学出版社, 2004: 56–75.
 - [5] Vidal R, Ma Yi, and Sastry, *et al.* Generalized Principal Component Analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(12):1945–1959.
 - [6] Liou Cheng-yuan and Musicus B R, *et al.* Cross entropy approximation of structured gaussian covariance matrices. *IEEE Transactions on Signal Processing*, 2008, 56(7): 3362–3367.
 - [7] Vidal R and Hartley R, *et al.* Three-view multibody structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 214–227.
 - [8] Huang Kun, Yang A Y, and Ma Yi, *et al.* Sparse representation of image with hybrid linear models, ICIP, Singapore, 2004: 1281–1284.
 - [9] Hong Wei, Wright J, and Huang Kun, *et al.* A multi-scale hybrid linear model for lossy image representation. *IEEE Transactions on Image Processing*, 2006, 15(12): 3655–3671.
 - [10] Schwarz H and Wien M, *et al.* The scalable video coding extension of the H.264/AVC standard {Standards in a nutshell}. *IEEE Signal Processing Magazine*, 2008, 25(2): 135–141.
 - [11] Harris J. Algebraic Geometry: A First Course. New York, Springer-Verlag, 1992, 24.
 - [12] ISO/IEC, JTC1/SC29/WG11, Joint scalable video model JSVM-9. 79th MPEG Meeting, Marrakech, Morocco, JVTU202, Jan. 2007.
- 刘 健: 男, 1981 年生, 博士, 研究领域为视频数据编码、视频内容分析。
- 吴 飞: 男, 1973 年生, 副教授, 研究领域为多媒体分析与检索、统计学习理论。
- 姚 蕾: 女, 1983 年生, 博士, 研究领域为视频数据编码、图像内容分析和检索。
- 庄越挺: 男, 1965 年生, 教授, 博士生导师, 研究领域为多媒体分析与检索、计算机动画、数字图书馆。