

基于稀疏贝叶斯模型的视频恢复方法研究*

朱 立

(广东技术师范学院工业实训中心, 广东 广州 510665)

摘 要: 结合压缩感知的稀疏表示理论和贝叶斯模型, 提出一种在数据库中进行快速有效的视频信息查询和视频恢复的方法。根据最小 l_1 范数重构数据信息的要求, 方法利用贝叶斯模型对参数进行估计, 并通过稀疏特性对视频分类和特征提取, 达到恢复视频信息的目的。实验结果表明, 比较主成份分析和随机投影算法, 该文的实现方法具有更好的恢复性能。

关键词: 稀疏; 贝叶斯模型; 压缩感知; 视频恢复

中图分类号: TN911 **文献标志码:** A **文章编号:** 0529-6579 (2012) 05-0036-04

Video Retrieval Method Using Sparse Bayesian Modeling

ZHU Li

(Industrial Training Center, Guangdong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: Combining Compressed Sensing sparse theory with Bayesian modeling, a faster and more effective method of video information retrieval is proposed in database. Based on minimizing l_1 principle for classification and feature extraction, the method implements video retrieval requirement by utilizing Bayesian modeling to automatically estimate the regularization parameters. Experimental results demonstrate that the proposed method performs better retrieval performances than PCA and RP algorithm.

Key words: sparse; Bayesian modeling; compressed sensing; video retrieval

网络技术的迅速发展和大量视频数据的产生和传播, 使视频信息在数据库中快速、准确的有效查找与恢复成为信息领域很重要的一种技术需求。例如, 从优酷网站搜索感兴趣的节目内容, 媒体公司要限定版权保护及其视频源位置, 而加密系统也要检测其合法性请求。目前, 一些算法通过索引或 Hash 来改进搜索效率。例如, 文献 [1-4] 使用了索引树结构进行视频查找和恢复, 文献 [5] 利用几何 Hashing 建立数据库的索引对视频源进行处理。在文献 [6] 中, 结合 PCA (主成份分析) 和空间划分技术, Gao. L 推出了一种 kd-树索引机制的视频源追踪恢复方法。文献 [7] 使用 RP (随机投影) 技术把视频图像投影到不同的搜索空

间, 并在每个空间上分别使用 kd-树进行视频恢复。

最近, 稀疏性和压缩感知理论在信号分类及其信号恢复的判别性能领域得到了广泛研究, 并引起了学术界的巨大兴趣^[8-10]。借鉴其最新的研究成果, 本文主要研究视频信息恢复中的稀疏表示方法和判别属性。我们的目的是从视频数据库样本中寻找所需要的视频节目的最稀疏表示。首先对开始视频帧建立统一的数据库, 然后利用压缩感知中的稀疏 (Sparse) 理论和贝叶斯 (Bayesian) 参数估计算法在数据库中研究视频的稀疏表示, 最后对估计的视频稀疏系数进行分类并通过稀疏矩阵解决视频的恢复问题。

* 收稿日期: 2012-04-20

基金项目: 国家自然科学基金资助项目 (90604008); 广东省科技计划资助项目 (2012B010100035); 广东省高等学校人才引进项目

作者简介: 朱立 (1969 年生), 女, 高级工程师; E-mail: zhuli@gdin.edu.cn

1 视频信息的稀疏表示

通过级联不同的视频内容, 视频数据库的信息可表示为如下的矩阵形式

$$\mathbf{A} = [a_{1,1}, a_{1,2}, \dots, a_{1,N_1}, \dots, a_{K,1}, \dots, a_{K,N_K}] \quad (1)$$

这里, $a_{i,j}, i = 1, \dots, K, j = 1, \dots, N_i$, 代表了第 i 个视频的第 j 帧。每个 $a_{i,j}$ 假设是大小为 $M = VH$ 的列向量, V 和 H 代表每个帧的垂直、水平的维数。简便起见, (1) 式可写为

$$\mathbf{A} = [a_1, a_2, \dots, a_N] \quad (2)$$

其中, $N = \sum_{i=1}^K N_i$ 。设 y 表示数据库中视频夹 (Video-Clip) 信息, 则其矢量形式为

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_S \end{bmatrix} \quad (3)$$

y_i 表示了第 i 帧, 而 S 为视频夹的长度。通过移位 (1) 中的列, 我们可以建立如下的矩阵

$$\bar{\mathbf{A}} = \begin{pmatrix} a_1 & a_2 & \cdots & a_{N-(S-1)} \\ a_2 & a_3 & \cdots & a_{N-S} \\ \vdots & \vdots & \ddots & \vdots \\ a_S & a_{S+1} & \cdots & a_N \end{pmatrix} \quad (4)$$

那么, \mathbf{y} 可以表示为一种稀疏形式

$$\mathbf{y} = \bar{\mathbf{A}}\mathbf{x}_0 \quad (5)$$

$\mathbf{x}_0 = (0, \dots, 0, 1, 0, \dots, 0)^T$ 是一种稀疏矢量, 其中对应的某个视频夹 y 在数据库的位置为 1, 其余所有系数都等于 0。对于需要搜索的视频夹 y , 寻找对应的 \mathbf{x}_0 的方法可以通过压缩感知理论 CS (Compressed Sensing) 求解。在噪声环境下, CS 求解系统可以模型化为

$$\mathbf{y} = \bar{\mathbf{A}}\mathbf{x} + \mathbf{n} \quad (6)$$

\mathbf{n} 表示 $(SM \times 1)$ 维零均值方差为 β^{-1} 的高斯噪声矢量。使用 CS 理论中的最小 l_1 范数解决 (6) 式的形式为

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \{ \|\mathbf{y} - \bar{\mathbf{A}}\mathbf{x}\|_2^2 + \alpha \|\mathbf{x}\|_1 \} \quad (7)$$

其中, $\|\cdot\|_1$ 表示 l_1 范数。

2 基于 S-Bayesian 视频恢复算法

2.1 S-Bayesian 视频参数的估计

求解 (7) 式中 $\hat{\mathbf{x}}$ 的优化方法有多种^[8,11], 在此我们利用 CS 理论的稀疏方法和 Bayesian 模型,

通过自适应调整估计参数, 提出一种稀疏 - 贝叶斯 S-Bayesian (Sparse-Bayesian) 方法解决此问题。在 S-Bayesian 方法中, 所有未知数都作为一定概率分布的随机变量, 其联合概率分布形式为

$$P(\mathbf{x}, \alpha, \beta, \mathbf{y}) = P(\mathbf{y} | \mathbf{x}, \beta) P(\mathbf{x} | \alpha) P(\alpha) P(\beta) \quad (8)$$

这里, α 为视频类型属性, β 为 0 均值、方差为 β^{-1} 的某类视频信息。根据噪声特性, (8) 式右边第一项可以写作为

$$P(\mathbf{y} | \mathbf{x}, \beta) = N(\mathbf{y} | \bar{\mathbf{A}}\mathbf{x}, \beta^{-1}) \quad (9)$$

由文献 [12] 可知, 最小 l_1 范数等价于分级优先拉普拉斯算法对系数 x 的处理, 也就是

$$P(x | \alpha) = \prod_{i=1}^N N(x_i | 0, \alpha_i) \quad (10)$$

$$P(\alpha_i | \lambda) = \frac{\lambda}{2} \exp(-\frac{\lambda \alpha_i}{2}), \quad \alpha_i \geq 0, \quad \lambda \geq 0 \quad (11)$$

这里, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$, λ 为调节因子。根据此条件, 信号分布 $P(\mathbf{x} | \mathbf{y}, \lambda, \beta)$ 可以被估计为多变量高斯分布 $N(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$, 具有参数:

$$\boldsymbol{\Sigma} = [\beta \bar{\mathbf{A}}^T \bar{\mathbf{A}} + \boldsymbol{\Lambda}]^{-1}, \quad (12)$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma} \beta \bar{\mathbf{A}}^T \mathbf{y} \quad (13)$$

$\boldsymbol{\Lambda} = \operatorname{diag}(1/\alpha_i)$, 参数 α 通过似然函数可以估计

$$\boldsymbol{\Gamma} = -\frac{1}{2} \log |C| - \frac{1}{2} \mathbf{y}^T C^{-1} \mathbf{y} + N \log \frac{\lambda}{2} - \frac{\lambda}{2} \sum_i \alpha_i \quad (14)$$

其中, $\boldsymbol{\Gamma}$ 表示似然函数, $C = (\beta^{-1} \mathbf{I} + \bar{\mathbf{A}} \boldsymbol{\Lambda}^{-1} \bar{\mathbf{A}}^T)$ 。通过 (14) 式, 可得到参数的更新过程:

$$\alpha_i = -\frac{1}{2\lambda} + \sqrt{\frac{1}{4\lambda^2} + \frac{\langle x_i^2 \rangle}{\lambda}} \quad (15)$$

$$\lambda = \frac{N-1}{\sum_i \alpha_i / 2} \quad (16)$$

这里, $\langle x_i^2 \rangle = x_i^2 + \sum_{ii}$ 。在算法每次迭代时, 首先给出估计参数 α, λ , 然后通过 (12) 和 (13) 式计算 \mathbf{x} 的分布, 最后由 (15) 和 (16) 式 α_i, λ 得到视频信息稀疏表示的性能。

2.2 视频信息分类

为了判断视频信息是否在数据库, 我们通过视频第一帧的位置对其进行分类定位。如果视频夹在数据库中, 那么只有视频第一帧位置的稀疏表示为非 0 值且等于 1^[13]。设 $\hat{\mathbf{x}}$ 是矢量 $\boldsymbol{\mu}$ 的 Bayesian 算法收敛值, $m = \max_i \hat{x}_i$ 。那么, 我们定义视频矢量信息的分类形式为:

$$x_i^{\text{comp}} = \begin{cases} 1 & \text{如果 } \hat{x}_i = m \\ 0 & \text{其它} \end{cases}, i = 1, 2, \dots, N \quad (17)$$

定义一个阈值 δ , 当且仅当 X^{comp} 只有一个非 0 值, 并满足 (18) 式时, 我们认为视频信息在数据库中

$$\|\hat{X} - X^{\text{comp}}\|_1 \leq \delta \quad (18)$$

2.3 视频特征提取

为了实现不同视频帧之间的有效搜索, 需要对视频信息的特征进行提取。假设使用 PCA (主成份分析) 生成的矩阵 $D \in \mathbf{R}^{T \times M}$ ($T \ll M$) 来表示图像空间到特征空间的投影, 并有

$$D_{T \times M} A_{M \times N} = A_{T \times N}^{\sim} \quad (19)$$

那么, 根据 $A_{T \times N}^{\sim}$ 大小, 运用 S-Bayesian 参数估计和视频信息分类方法, 我们可以快速查找并恢复需要的视频信息。其中, 维数 T 的大小决定了 PCA 变换后能量损失的情况。

3 仿真结果

实验把 12 个视频 (每个长约 30 min) 划分为 2 个数据库, 用刻度因子 32 进行视频帧下采样产生 $11 * 8$ 个视频图标。然后, 使用 CS 理论的稀疏变换 ($T=4$) 对视频帧投影得到稀疏矩阵 $A_{T \times N}^{\sim}$ 。在两个数据库中随机选择 250 个视频夹进行搜索, 其中视频夹的长度为 $S=3, 5, 7$ 帧。实验在 Intel-2 GHz CPU 和 2 GB-RAM 笔记本电脑中进行。S-Bayesian 算法搜索视频节目的稀疏表示的平均时间如表 1。

表 1 搜索一个视频节目的 CPU 时间
Table 1 CPU time used to find a video query

帧序	在数据库的帧/s	不在数据库的帧/s
3	1.45	2.90
5	2.62	4.44
7	5.72	8.07

实际上, 由于编码、通信和图像格式转换, 视频夹经常受到噪声攻击。为此, 我们加入不同级的高斯噪声到视频图像, 使图像 PSNR = 20 dB, 25 dB, 30 dB 和 35 dB。图 1 说明了一个数据库中的原始视频帧图像及其对应的 20 dB 噪声帧图像。

为了和 PCA、RP (随机投影) 等相关的算法进行比较, 我们采用精度 - 恢复率曲线 (Precision-recall ratio curves) 来分析 S-Bayesian 算法实现的性能。对于给定的阈值 δ , 设 a 表示恢复了相关视频



(a)



(b)

图 1 噪声测试视频帧

Fig. 1 Video image at noisy test cases
(a) 数据库中的原始帧, (b) 噪声帧 (20 dB)

夹的数目, b 代表相关视频夹没有被恢复的数目, c 为恢复的非相关视频夹数目。则, $\text{Precision} = a / (a + c)$, 而 $\text{recall ratio} = a / (a + b)$ 。再改变阈值 δ , 计算相应方法的 Precision 和 recall 值, 就可以描绘一定算法实现视频夹的精度 - 恢复率曲线。

图 2 (a) 和 (b) 中, 分别设置 $T=4, S=3$, PSNR = 20 dB 和 $T=4, S=7$, PSNR = 20 dB, 比较了不同算法确定视频夹的精度 - 恢复率曲线。由图 2 (a) 可见, PCA 性能优于 RP 实现算法, 而 S-Bayesian 算法又优于 PCA 实现算法。相同的条件下, 由图 2 (b) 可见, RP 性能优于 PCA 实现算法, 而 S-Bayesian 算法又优于 RP 实现算法。在此情况下, S-Bayesian 算法的 Precision = 1, 而 Recall ratio ≤ 0.98 。

4 结论

借鉴 CS 稀疏表示理论和 Bayesian 参数估计模型, 本文提出了一种快速有效的视频信息搜索和恢复的稀疏贝叶斯 (S-Bayesian) 算法。算法利用 Bayesian 参数估计得到视频信息的稀疏表示特性, 并通过稀疏性对视频进行分类, 最后由稀疏矩阵把视频图像空间投影到特征空间确定解决了视频信息恢复问题。实验结果表明, 本文提出的方法比 PCA 和 RP 具有更好的恢复性能和更强的鲁棒性。

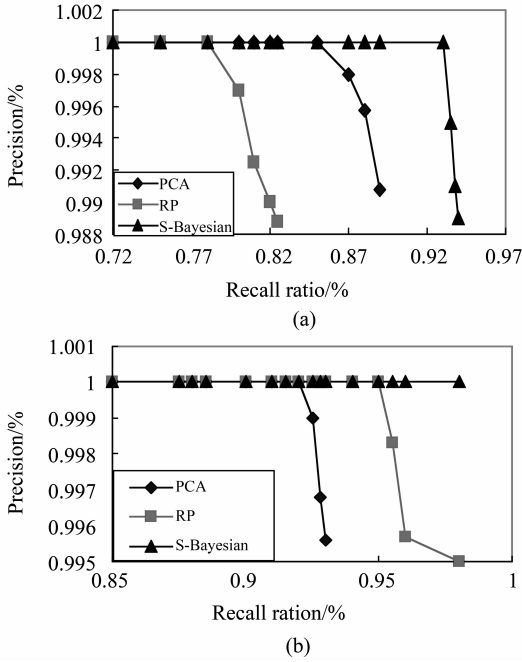


图 2 PCA, RP 和 S-Bayesian 算法实现视频夹的精度 - 恢复率曲线

Fig. 2 Precision-recall ratio curves for PCA, RP, and S-Bayesian algorithm

(a) $T=4, S=3, PSNR=20\text{ dB}$,

(b) $T=4, S=7, PSNR=20\text{ dB}$

参考文献:

[1] HSU C T, TENG S J. Motion trajectory based video indexing and retrieval [C] // Proc International Conference on Image Processing, 2002, 1(1): 605 - 608.

[2] MEZARIS V, KOMPATAIARIS I, BOULGOURIS N V, et al. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval [J]. IEEE Trans. on Circuits and Systems for Video Technology, 2004, 14 (5) : 606 - 621.

[3] YUAN J, DUAN L Y, QI Tian, et al. Fast and robust short video clip search using an index structure [C] // Proc of 6th ACM SIGMM International Workshop on Multimedia Info Retrieval (MIR), 2005: 61 - 68.

[4] SNOEK C G M, WORRING M. Multimedia event-based video indexing using time intervals [J]. IEEE Trans on Multimedia, 2005, 7(4): 638 - 647.

[5] FAN J, ELMAGARMID A K, ZHU X, et al. Classview: Hierarchical video shot classification, indexing, and accessing [J]. IEEE Trans on Multimedia, 2004, 6(1): 70 - 86.

[6] GAO L, LI Z, KARSAGGELOS A K. Fast video shot retrieval with luminance field trace indexing and geometry matching [C] // Proc IEEE Int Conf on Image Processing (ICIP), 2006: 1497 - 1500.

[7] GAO L, LI Z, KARSAGGELOS A K. A video retrieval algorithm using random projections [C] // Proc IEEE Int Conf on Image Processing (ICIP), 2009: 797 - 800.

[8] STARCK J L, MURTAGH F, FADILI J. Sparse image and signal processing: wavelets, curvelets, morphological diversity [M]. Cambridge: Cambridge University Press, 2010.

[9] ELAD M, FIGUEIREDO M A T, YI M. On the role of sparse and redundant representations in image processing [J]. Proceedings of IEEE, 2010, 6(3): 972 - 982.

[10] WRIGHT J, YI M, MAIRAL J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of IEEE, 2010, 6(1): 1031 - 1044.

[11] BABACAN S D, MOLINA R, KARSAGGELOS A K. Parameter estimation in TV image restoration using variational distribution approximation [J]. IEEE Transactions on Image Processing, 2008, 17(5): 326 - 339.

[12] BABACAN S D, MOLINA R, KARSAGGELOS A K. Bayesian compressive sensing using laplace priors [J]. IEEE Trans on Image Processing, 2010, 19(1): 53 - 63.

[13] 赵慧民, 郭一缜, 丁晓艳, 等. 用于视频多播传输的压缩传感实现方法研究 [J]. 中山大学学报: 自然科学版, 2012, 51(1): 45 - 49.