

基于压缩感知理论的视频编解码方法

杨鑫诚 何 斌

(同济大学电子与信息工程学院 上海 201804)

摘 要 压缩感知理论是一种全新的信息处理方式,是近来国际上热门的研究方向,在信号处理等其他领域中具有很好的应用前景。结合压缩感知理论和传统的视频编解码技术,设计提出了一种基于压缩感知理论的视频编解码方案。用测试视频进行了仿真实验,结果表明该算法能够取得较好的效果。

关键词 压缩感知,视频编解码,稀疏性

中图分类号 TP391 **文献标识码** A

Video Codec Based on Compressed Sensing

YANG Xin-cheng HE Bin

(Electronic and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract Compressed sensing theory is a new information processing, recent popular research direction, and has a good prospect of application in signal processing and other fields. In this paper compressed sensing theory and traditional video codec technology, design a theory based on compressed sensing video codec program. The results show that the algorithm is able to achieve better results.

Keywords Compressed sensing, Video codec, Sparsity

在传统的数字信号处理领域,发送端采样率的设置要遵从奈奎斯特采样定律,以使接收端能够正确恢复数据。但是目前这一传统理论受到了极大的挑战。2004 年,Donoho 和 Candes 等人提出了压缩感知理论,压缩感知(CS Compressed sensing)是一种寻找欠定线性系统的稀疏解的技术,当信号具有稀疏性或可压缩性时,通过采集少量的信号投影值就可实现信号的准确或近似重构^[1,2]。其理论本身是“通过对信号的高度不完备线性测量的高精确的重建”,其采样数量远远小于 Nyquist 采样,从而降低了编码端的压缩负担,提高了编码效率。

利用压缩感知技术的优点和视频图像信号的稀疏性特点,本文将压缩感知技术融入现有视频解码框架,提出了一种基于压缩感知技术的视频解码方案,以便利用很少的测量数据重现视频图像源。

1 压缩感知理论

压缩感知理论对信号的采样、压缩编码发生在同一个步骤,利用信号的稀疏性,以远低于 Nyquist 采样率的速率对信号进行非自适应的测量编码^[1]。其编解码框图 1 所示。

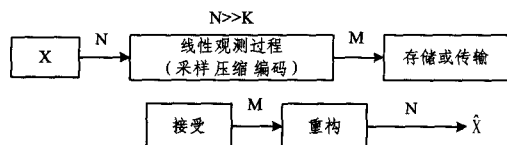


图 1 基于压缩感知的编解码框图

信号的稀疏性或可压缩性是压缩感知的重要前提和理论基础。压缩感知指出设长度为 $N \times 1$ 的信号 X ,其在时域的元素为 $x_n, n=1, 2, \dots, N$ 。假设 Ψ 域的一组标准正交基为 $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$,则信号 X 可以由 $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$ 线性表示为

$$X = \sum_{i=1}^N s_i \varphi_i \text{ 或 } X = \Psi s \quad (1)$$

s 为 $N \times 1$ 的列向量, Ψ 为 $N \times N$ 矩阵,并且满足 $s_i = \langle X, \varphi_i \rangle = \varphi_i^T X, \Psi = [\varphi_1, \varphi_2, \dots, \varphi_n]$ 。如果 X 只是 $K (K \ll N)$ 个基向量的线性组合,那么信号 X 就是 K -稀疏的(称 K 为信号 X 的稀疏度, Ψ 为 X 的稀疏基)。当式(1)中仅有少量的大系数,而大部分的系数都很小时,则称信号 X 是可压缩的。如果信号具有稀疏性或可压缩的,那么小系数的丢弃不会影响对原始信号的高概率重构。

图 2 压缩感知原理图中,线性观测过程可以用一个 $M \times N$ 的矩阵 Φ 表示,对信号 X 观测得到的 M 个观测值为 $y_j = \langle X, \varphi_j \rangle, j=1, 2, \dots, M$ 。其中 y_j 是向量 $Y (M \times 1)$ 中的元素, φ_j^T 是观测矩阵 $\Phi (M \times N)$ 的列向量。写成矩阵形式为:

$$Y = \Phi X \quad (2)$$

将式(1)代入式(2)中得:

$$Y = \Phi X = \Phi \Psi s = \Theta s \quad (3)$$

式中, $\Theta = \Phi \Psi$ 是 $M \times N$ 的矩阵。而由于式(3)中 s 是 K -稀疏的,即仅有 K 个非零系数,而 $K \ll M \ll N$,那么利用信号稀疏分解理论中已有的稀疏分解算法,可以通过求解式(3)的逆问题得到稀疏系数,再代回式(1)得到信号 X ^[1,2,3]。

2007 年 Candes 等研究者建立了著名的 Restricted Isometry Property (RIP) 理论, RIP 成为压缩感知奠基性理论:

矩阵 Φ 满足 s -阶 RIP 性质, 如果存在常数 $\delta_s \in [0, 1)$ 使得

$$(1 - \delta_s) \|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \delta_s) \|x\|_2 \quad (4)$$

对任意的 $x \in \Sigma_s$ 成立。式(4)等价 Gramian 矩阵 $\Phi^T \Phi$ 所有特征值位于区间 $[1 - \delta_s, 1 + \delta_s]$, 这里 $\#T \leq s$, 我们称 δ_s 为 RIP 常数^[6]。RIP 理论表明, 只要测量矩阵满足 RIP 特性, 那么 $K * \log(N/K)$ 的采样就能将 N 维信号的 K 个最大值稳定地重建出来。根据 RIP 理论, 可以通过求解 $L1$ 范数问题, 精确重建原始信号, 如下式

$$\hat{S} = \operatorname{argmin} s_0' \quad \text{s. t. } \Theta s' = y \quad (5)$$

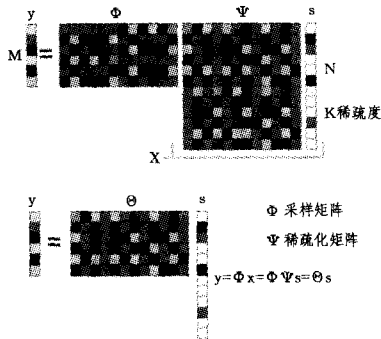


图2 基于压缩感知的原理图

2 视频编解码

传统的视频编解码器是基于宏块操作的, 下面以 H. 264 为例简单分析传统视频编解码的核心技术。H. 264 的编码器编码框架如图 3 所示, 从编码框架上可以看出, H. 264 视频标准采用混合的编码框架, 在预测的基础上编码压缩残差数据, 以实现高压缩性和适用性。输入的帧或场 F_n 以宏块为单位被编码器处理。首先, 按帧内或者帧间预测编码的方法进行处理。如果采用帧间预测编码, 其预测值 PRED 是由当前片中前面已编码的参考图像经运动补偿(MC)后得到的, 其中参考图像用 F_{n-1}' 表示。预测值 PRED 和当前块相减后, 产生一个残差块 D_n , 经块变换、量化后产生一组量化后的变换系数 X , 再经熵编码, 其与解码所需的一些头信息一起组成压缩后的码流, 经 NAL(网络自适应层)供传输和存储用^[7,8]。

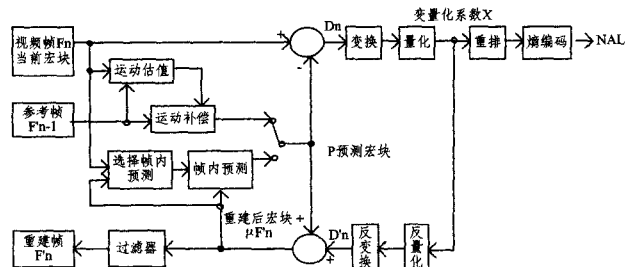


图3 H. 264 编码流程

H. 264 的解码器如图 4 所示。整个解码的过程是编码的逆过程。解码器将编码器的 NAL 输出的 H. 264 比特流通过熵解码得到量化后的一组变换系数 X , 再经反量化、反变换得到残差 D_n' 。利用从该比特流中解码出的头信息, 解码器就产生一个预测块 PRED, 它和编码器中的原始 PRED 是相同的。将该解码器产生的 PRED 与残差 D_n' 相加, 就得到了

uF_n' , 再经滤波, 最后就得到滤波后的解码输出图像 $F_n'^{[8]}$ 。

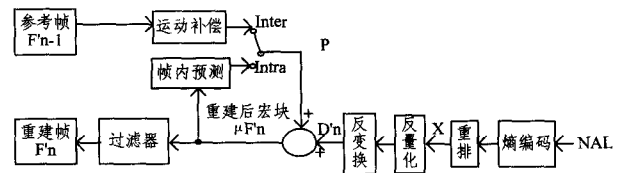


图4 H. 264 解码流程

从上述过程可知, 传统的编码过程复杂, 解码简单, 对编码端的要求更为严格。编码过程中图像变换后的系数大部分被抛弃, 造成数据和内存资源的浪费。

3 基于压缩感知的视频编解码方法

将压缩感知理论应用到视频编解码中, 关键是要准确稀疏地表示原始视图图像, 并构造合适的测量矩阵, 最后通过少量的测量值重构得到原始视图图像的信息。结合前两节中的压缩感知理论和传统的视频编解码核心技术, 本节在文献[12]研究的基础上, 设计一种基于压缩感知理论的视频编解码方式。

编码过程仍然沿用帧内帧间的编码思路, 采用运动估计与运动补偿的方案进行残差的获取, 对于 I 帧(帧内编码), 由于其本身的稀疏度有限, 且对于 P 帧(帧间编码)的参考作用, 将其采样率相对于 P 帧来说取高一些, 而对于 P 帧, x 信号稀疏度较大, 采用的帧间编码过程如图 5 所示。

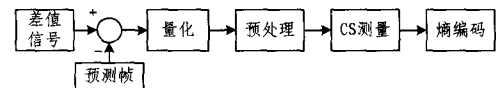


图5 基于压缩感知的视频编码框架

1. 获取帧间差值信号

对于视频帧间差值信号的获取借助于 H. 264 的预测编码算法, 以 16×16 像素的宏块为单位进行运动估值和运动补偿, 从而形成预测帧。

2. 预处理

预处理过程主要包括信号稀疏化、分块。

(1) 信号稀疏化: 依据信号稀疏化理论, 对信号进一步稀疏化。对输入的信号, 通过累加信号中非零的个数, 得出信号稀疏性度 K , 对小于稀疏化阈值 T 的帧差值统一置零(这样处理引起的失真比传统帧间信号编码中的有损压缩引起的失真要小很多, 同时这样处理有助于 CS 测量编码方法发挥作用)。本文测量过程中的稀疏基采用 Daubechies 9/7 小波基。

(2) 分块: 依据 BCS 理论, 通过分块的方式应用压缩感知测量编码进行视频帧间信号编码, 将一帧差图像划分成多个 $M \times M$ 的块, 逐个处理, 其中 M 取图像宽和高的一个公约数。

3. CS 测量

通过测量矩阵 Φ 与帧间信号的乘积获得观测值, 完成信号从高维度空间向低维度空间的映射。按照预处理环节划分的块, 逐块进行测量编码, 完成整帧帧间信号的压缩编码。观测矩阵的选取直接影响了压缩编码的重构质量, 通常有随机高斯观测矩阵、Fourier 观测矩阵和随机伯努利观测矩阵等。

随机高斯矩阵与大多数同定正交基构成的矩阵不相关,本文选用 32×32 的随机扰动分块 Hadamard 矩阵^[10]对信号的稀疏表示进行线性投影。

4. 熵编码

熵编码采用霍夫曼编码。霍夫曼编码无损压缩编码方式,可以从统计上对数据进行进一步压缩,同时不会因为失真问题而影响 CS 重构。

解码过程如图 6 所示。

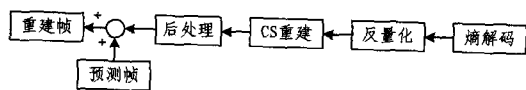


图 6 基于压缩感知的视频解码框图

1. 熵解码

进行霍夫曼解码,解出经过量化的 CS 测量编码观测值。

2. CS 重构

CS 重构是本方案解码过程中最为重要的一步。如本文第 2 节所述,CS 重构过程实际是求解一定条件下的最优化问题的过程,常用的算法有 MP, OMP 算法等。本文采用 GSPR 算法^[11]进行重构得到重构图像。

3. 解码帧重构

将解得的视频帧间信号与预测帧相加,得到重构帧图像。

相对于传统的视频编解码,基于压缩感知的视频编解码方案编码中的测量过程是将高维信号投影到低维空间的一个非自适应过程,与传统编码方式不同,基于压缩感知的编码方案得到的每个测量值包含了传统编码方式下所用样本的部分信息,可以有效地避免传统方式因丢弃高频分量而带来的图像信息丢失的现象;基于 BCS 理论的分块处理方式,虽然会带来重构后图像的清晰度有所损失,但是很大程度上降低了时间复杂度。对于解码过程,传统的解码过程是编码过程的逆过程,解码过程相对简单,编码过程相对复杂,而基于压缩感知理论的解码过程是一个求解欠定性方程组的最优解过程,相对于编码过程,解码过程较为复杂,这种处理方式将主要工作量放在解码过程,而解码过程基于强大的终端设备,这样减少了编码端的工作量,同时有利于在干扰较大的恶劣情况下进行编码测量,是现在视频编解码的主流发展方向。

4 实验仿真

仿真实验编码测量过程中的稀疏基采用 Daubechies 9/7 小波基,测量矩阵选用 Hadamard 随机扰动分块矩阵,重构算法采用 GSPR 算法,测试视频选择 352×288 标准试验视频“Flower.yuv”,实验在 Matlab 运行环境下实现。量化、反量化和熵编码、熵解码技术在实验中没有加以考虑。此外,由于压缩感知算法中采用的测量矩阵为随机矩阵,这会影响到图像重构效果,在同样的观测值 M 下,重构视频的平均 PSNR 值会有 0.3dB 左右的偏差,所以每次改变 M 值之后,都将程序运行 10 次取其平均值。

对于 I 帧的编码测量数分别取 $N=10000, 15000, 20000, 25000$, 并对某一固定的 N 值分别改变 P 帧的采样数, $N_1 =$

5000, 10000, 15000, 20000, 25000, 以“Flower.yuv”第 12 帧为例,绘制得到图 7,从图中可以看出,在同一视频序列中,关键的 I 帧编码的质量好坏对于整个视频的恢复具有重要的作用,而编码质量又与采样点数的多少基本对应。同时对于固定的 I 帧编码质量, P 帧的采样点数的多少对于视频帧间也具备对应的关系,但是其影响程度远远小于 I 帧的采样情况。原因是帧间编码的残差图像的稀疏性很强,在压缩感知理论下很少的测量值足以得到较好的重构图像。

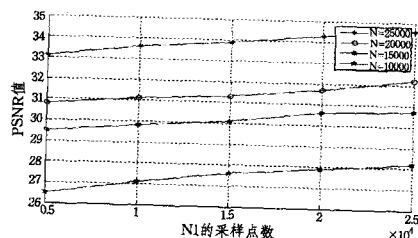


图 7 第 12 帧图像 PSNR 值比较

当 $N_1=15000$ 时,如图 8 所示,当 N 为 10000 时,PSNR 为 27.72dB,当 N 变成 15000 时,PSNR 为 30.14dB,当 N 变成 20000 时,PSNR 为 31.30dB,当 M 变成 25000 时,PSNR 为 33.84dB。

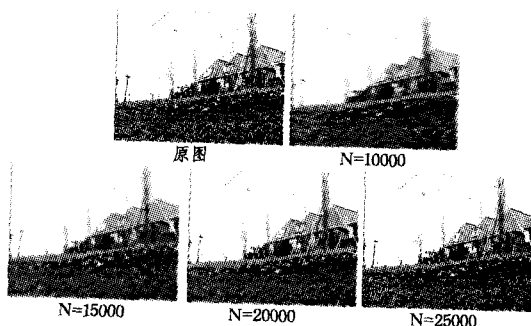


图 8 第 12 帧原图及其重构 ($N_1=15000$)

结束语 本文集合压缩感知理论和传统的视频编解码方案,提出了基于压缩感知的视频帧编解码方案,有针对性地进行了仿真试验。在以后的工作中,将重点研究如何根据信号的稀疏性,自适应地调整需要的观测矩阵的大小,研究压缩感知应用于视频编码中时与其它经典压缩方法的结合程度,相信压缩感知理论将为视频编码技术发展做出巨大的贡献。

参考文献

- [1] Donoho D. Compressed sensing [J]. IEEE Trans. Information Theory, 2006, 52(4): 1289-1306
- [2] Candes E, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information [J]. IEEE Trans. Information Theory, 2006, 52(4): 489-509
- [3] Tropp J, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Trans. Information Theory, 2007, 53(12): 4655-4666
- [4] Cotter S F, Rao B D. Sparse channel estimation via matching pursuit with application to equalization [J]. IEEE Trans. Comm, 2002, 50: 374-377

[5] Shen Y, Li S. Restricted p-isometry property and its application for nonconvex compressive sensing, *Adv Comput Math*, DOI 10.1007/s10444-011-9219-y

[6] Chartrand R, Staneva V. Restricted isometry properties and non-convex compressive sensing[J]. *Inverse Problems*, 2009, 24: 1-14

[7] ITU-T Rec. H. 264 and ISO/IEC 14496-10 AVC, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification[S]. Joint Video Team(JVT) of ISO/IEC MPEG& ITU-T VCEG 7th Meeting, Pattaya, Thailand, March, 2003;7-14

[8] Lim K-P, Sullivan G, Wiegand T. Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods[C]//Joint Video Team(JVT) of ISO/IEC MPEG and ITU-

T VCEG. Busan, Korea, April, 2005

[9] Chen Zhi-bo, Zhou Peng, He Yun, et al. Fast Motion Estimation for JVT[C]//Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q. 6) 7th Meeting, Pattaya II. Thailand, March, 2003;7-14

[10] Trnadt G L D. Fast compressive imaging using scrambled block Hadamard ensemble [C]//Proceedings of the European Signal Processing Conf. Lausanne Switzerland, 2008

[11] Figueiredom A T, Nowakrd R D, Wright SJ. Gradient projection for sparse reconstruction; application to compressed sensing and other inverse problems [J]. *IEEE Journal of Selected Topics in Signal Processing* 2007, 1(4):586-597

[12] Xie X, Yu L. A new video codec based on compressed sensing [Z]. 2011

(上接第 168 页)

机在 x 、 y 、 z 三个方向的加速度序列。对这些加速度进行绘图分析可以发现一个规律,即“相同的运动有相似的加速度曲线”。

从以上 5 幅图可以看出, z 轴加速度有明显的相似曲线,都是先维持在 10 左右的重力加速度,然后自由落体期间加速度为 0,一个与地面碰撞产生的巨大加速度,最后稳定在 10 左右的重力加速度;同时, x 和 y 轴方向的数据除了碰撞时,始终稳定在 0 左右。以上数据,证实了我们的算法是行之有效的。

在进行人体运动状态测试的验证中,我们选取了 3 种常见的运动模型。3 种常见的运动为:跌倒、散步、跑步。每组模型我们取 90~100 组值进行学习,随后进行验证性实验。每种动作重复 60 次并记录预期值。每次实验重复 10 次。实验结果如表 1 所列。

表 1 人体运动状态预测测试数据

预期\实际	跌倒	散步	跑步
跌倒	47	5	8
散步	6	39	15
跑步	10	8	42

其准确率如表 2 所列。

表 2 人体运动状态测试准确率

状态	跌倒	散步	跑步
正确率%	78.3	65	70

从表中可以看出,该方法能较为准确地预测出人体简单的运动状态,其中以跌倒的预测正确率最高。从表中还能发现,跑步和跌倒都属于激烈运动,在实时预测中会出现一定的误判,这也是我们整个算法有待改进的地方。

本段是对“跌倒,散步,跑步”3 种运动状态进行建模并通过实验验证其可靠性,而本方法仍适用于其他人体运动状态,只要通过本文介绍的方法进行数据采集并建模,就能得到不错的预测结果。

结束语 本文针对传统方法中对设备和庞大数据量的高

要求等问题,提出一种基于 SVM 的人体运动状态检测算法。其将传感器数据由“设备坐标系”映射到“标准坐标系”,然后将不同特征利用 SVM 分类建模。实验数据表明,该方法在移动终端传感器方面可以进行较好的运动状态检测,确实能够投入于应用之中,但就如何在“小数据集”的情况下进一步提高预测精度,有待进一步研究。

参 考 文 献

[1] 刘晓祥. 基于加速度传感器的动作捕捉系统的研究[D]. 长春: 吉林大学

[2] Van Gestel T, Suykens J A K, Baestaens D E, et al. Financial time series prediction using least squares support vector machines within the evidence framework[J]. 2001(04)

[3] Trotter M W B, Buxton B F, Holden S B. Support vector machines in combinatorial chemistry[J]. 2001(08)

[4] Wan Vincent, Campbell William M Support vector machines for speaker verification and identification[D]. 2000

[5] Aggarwal J K, Cai Q. Humanmotion analysis: A review [J]. *Computer Vision and Image Understanding*, 1999, 73 (3): 428-430

[6] 赵冬斌, 张文增, 孙振国. 基于高速摄像的人体上肢运动信息检测[J]. *生物医学工程杂志*, 2002(19): 76-79

[7] Steinhoff U, Schiele B. Dead reckoning from the pocket-An experimental study[C]// IEEE International Conference on Pervasive Computing and Communications (PerCom.), Ieee. doi: 10.1109/PERCOM.2010.162-170

[8] 王晓丹, 王机勤. 支持向量机训练和实现算法综述[J]. *生物医学工程杂志*, 2002(19): 76-79

[9] Chang Chih-chung, Lin Chih-jen. LIBSVM: A Library for Support Vector Machines[J]. *ACM Transactions on Intelligent Systems and Technology(TIST)*, 2001, 2(3): 27

[10] 张学工. 关于统计学习理论与支持向量机[J]. *自动化学报*, 2000, 26(1): 32-42