

视频压缩感知中基于结构相似的帧间组 稀疏表示重构算法研究

和志杰,杨春玲,汤瑞东

(华南理工大学电子与信息学院,广东广州 510640)

摘 要: 基于视频帧内图像的非局部相似性和帧间信号的相关性,本文提出了一种基于结构相似的帧间组稀疏表示重构算法(SSIM-InterF-GSR),有效地提高了视频压缩感知的重构性能.在SSIM-InterF-GSR算法中,提出以结构相似度(SSIM)作为相似块匹配准则,在当前帧和参考帧内搜索匹配块生成相似块组,以相似块组的稀疏性作为正则项重构当前帧.同时,还提出了阶梯递减匹配块个数调整方案用于SSIM-InterF-GSR重构算法的迭代过程.仿真结果表明,相比于目前最好的视频压缩感知重构算法(Up-Se-AWEN-HHP),本文算法获得了更好的重构质量,最多可提升4~5dB.

关键词: 非局部相似性; 视频压缩感知; 组稀疏表示; 相似块组

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2018)03-0544-10

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.03.005

Research on Structural Similarity Based Inter-Frame Group Sparse Representation for Compressed Video Sensing

HE Zhi-jie, YANG Chun-ling, TANG Rui-dong

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China)

Abstract: Based on the nonlocal similarity and the correlation among inter-frames in video sequences, this paper proposes an algorithm of structural similarity based inter-frame group sparse representation (SSIM-InterF-GSR), which effectively improves the reconstruction performance for compressed video sensing. In SSIM-InterF-GSR, the structural similarity (SSIM) is utilized as block matching criterion to generate the group of similar blocks from the current frame and reference frames. And then, the sparsity of the groups is used as the regularization term to reconstruct the current frame. Meanwhile, the step-decreasing scheme for number of matching blocks is proposed during the iteration process of SSIM-InterF-GSR. Simulation results show that, compared to the state-of-the-art compressed video sensing reconstruction algorithm (Up-Se-AWEN-HHP), the SSIM-InterF-GSR algorithm obtains a better reconstruction quality. The most gap is up to 4~5dB.

Key words: nonlocal similarity; compressed video sensing; group-based sparse representation; the group of similar blocks

1 引言

压缩感知^[1,2] (Compressed Sensing, CS)是一种新颖的信号采集理论,其核心思想是:在信号是稀疏的或者可压缩的条件下,我们能够以低于奈奎斯特的采样率进行采样,在重构端依然可以精确地重建原始信号. CS能同时完成对信号的采样和压缩,这一过程特别适合于采集端资源受限的数据采集系统. 视频压缩感知 (Compressed Video Sensing, CVS)是基于CS理论的视频

采集与处理方法,它为视频采集与处理理论的发展提供了一个新方向.

压缩感知的研究包括观测矩阵的构造,信号稀疏表示和重构算法. 重构算法作为压缩感知理论的关键技术之一,自从CS理论提出以来,它一直是国内外的研究热点,特别是近几年对高维的图像信号压缩感知重构算法的研究有了长足的进步. 图像压缩感知重构算法主要有以下几类:凸优化算法,贪婪迭代算法,和迭代阈值算法. 这些算法一般是利用图像的先验知识构

造正则项,求解优化问题.属于凸优化算法的基追踪(Basis Pursuit, BP)和基追踪去噪(Basis Pursuit DeNoising, BPDN)模型可以分别通过线性规划和二阶锥规划求解^[3],但是这两种解法计算复杂度高,特别是在处理高维图像和视频时更是如此.为了减小计算复杂度,梯度下降法被用于求解 BPDN 问题,如可分离逼近稀疏重构(Sparse Reconstruction by Separable Approximation, SpaRSA)^[4],梯度投影稀疏重构(Gradient Projection for Sparse Reconstruction, GPSR)^[5]等.正交匹配追踪(Orthogonal Matching Pursuits, OMP)^[6],压缩采样匹配追踪(Compressive Sampling Matching Pursuit, CoSaMP)^[7]等贪婪迭代算法也被用于 CS 重构,相比于凸优化算法,贪婪迭代算法计算复杂度显著降低,但是图像重构质量会有一定的损失.贪婪迭代算法的一种替代算法是迭代阈值算法,它通过连续投影和阈值处理来实现重构,其优点在于不仅能够降低算法复杂度,而且还可以结合额外的优化准则(如 Wiener 滤波)来提高图像重构质量,迭代硬阈值(Iterative Hard-Thresholding, IHT)^[8]是一种经典的迭代阈值重构算法,在此基础上提出的基于块 CS 平滑投影 Landweber(Block-based CS sampling and Smoothed Projected Landweber, BCS-SPL)^[9]算法具有高质量的重构效果且计算复杂度低,得到了广泛的关注和应用^[10-12].最近,基于图像非局部相似性的压缩感知重构算法被提出(如文献[13,14]).文献[13]利用图像的非局部相似性对图像块进行聚类,把相似块组在三维空间(如 3D-DCT, 3D-DWT)的稀疏性作为正则项以实现图像重构.文献[14]提出了一种基于组稀疏表示的图像重构算法(Group Based Sparse Representation for Image Restoration, GSR),即把非局部相似块组作为稀疏表示的基本单元,通过高效的自适应稀疏字典学习,实现了图像的重构. GSR 算法的图像重构性能优于目前主流的图像压缩感知重构算法 SPL,是当前性能最好的图像压缩感知重构算法.

视频压缩感知重构算法是在图像压缩感知重构算法基础上的研究.文献[15]将二维图像 CS 直接推广到视频 CS,对视频的每一帧独立观测,然后利用二维稀疏变换(如 2D-DCT, 2D-DDWT)重构.但是,这个简单的推广没有利用到视频帧间的时间相关性.为了增强信号的稀疏性,文献[16]利用视频时间上的冗余信息,用三维稀疏变换(如 3D-DWT)重构,得到了比二维稀疏变换更好的重构效果.但是,三维稀疏变换重构计算复杂度高,内存需求量大且重构性能不高.文献[10]将传统视频压缩算法中的运动估计和运动补偿技术应用到视频压缩感知中,利用补偿残差信号比原始信号更稀疏的原理,提出了一种基于块的视频压缩感知残差(Motion Compensated BCS-SPL, MC-BCS-SPL)重构算法.相比于

三维稀疏变换重构算法, MC-BCS-SPL 在保证重构质量的同时减少了计算复杂度.为了进一步提高预测图像质量,以增强残差信号的稀疏度,文献[17]提出了多假设预测(Multi-Hypothesis predictions BCS-SPL, MH-BCS-SPL)重构算法,每个当前块的多假设预测集是来自一个或者多个已重构的参考帧,通过 Tikhonov 正则化得到预测块,然后残差重构得到当前帧, MH 重构算法得到了更高的重构效率.文献[11]及本实验室研究^[12]在 MH 的基础上,各提出了一种假设集优化方案,可以提高多假设预测精度增强重构质量同时减少预测算法复杂度.以上视频压缩感知重构算法都是基于 SPL 实现的,在 GSR^[14]的基础上, X. Gao 等利用视频序列的非局部相似性,提出了一种基于时空分层帧的视频压缩感知编码(Hierarchical frame based spatial-temporal recovery for video compressive sensing coding, Hi-STG-SR)^[18]方案,获得了较好的重构效果.但是,这一方案存在以下缺点:(1)分层帧观测结构复杂;(2)欧氏距离作为相似块匹配准则,不能很好的反映块之间的结构相似性;(3)由于质量越差的带噪图像结构细节信息越少,固定匹配块个数的相似块组构造方法,不能很好的适应重构算法.

基于以上分析,本文提出了一种基于结构相似的帧间组稀疏表示视频压缩感知重构算法(an algorithm of Structural Similarity based Inter-Frame Group Sparse Representation, SSIM-InterF-GSR).在重构非关键帧时,同时利用帧内图像的非局部相似性和帧间信号的相关性构造相似块组,再利用相似块组的稀疏性构造正则项重构当前帧.为了寻找更好的匹配块,本文提出使用更加符合人眼视觉系统(Human Visual System, HVS)的结构相似度(Structural SIMilarity, SSIM)作为相似块匹配准则.另外研究中我们发现随着重构算法的迭代,重构视频图像质量不断提高,细节信息增加,帧内图像块之间的相似性变小.基于这一理论,本文提出了阶梯递减匹配块个数调整方案用于本文所提的 SSIM-InterF-GSR 重构算法,这一方案比固定匹配块个数方案得到了更高精度的重构效果.

2 GSR 重构算法及 SSIM

2.1 图像压缩感知中组稀疏表示重构算法

信号压缩感知的求解可以表示为式(1)的一般形式:

$$\operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \Gamma(\mathbf{x}) \quad (1)$$

其中, \mathbf{y} 为观测值, \mathbf{x} 是待重构信号, $\Phi \in \mathbf{R}^{M \times N}$ 是观测矩阵. $\frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2$ 是 2 范数约束的数据保真项, $\Gamma(\mathbf{x})$ 是根据图像先验信息所加的正则项, λ 是正则化权重因子.

文献[14]利用图像的非局部相似性构造相似块组,然后利用相似块组的稀疏性作为正则项来求解优化问题式(1),式(1)可以转化为:

$$\hat{\alpha}_c = \operatorname{argmin}_{\alpha_c} \frac{1}{2} \|\Phi D_c \circ \alpha_c - y\|_2^2 + \lambda \|\alpha_c\|_0 \quad (2)$$

其中, D_c 是对每一个相似块组自适应学习稀疏字典的串联, α_c 是对应的稀疏系数的串联. 上式可以通过分裂布雷格曼迭代 (Split Bregman Iteration, SBI)^[19] 求解, 重构图像即为 $\hat{x} = D_c \circ \hat{\alpha}_c$.

2.2 结构相似性 (SSIM)

图像中相邻像素灰度值是相关的,而这种相关性决定于图像内的结构信息. 相比图像像素灰度,人眼视觉对场景中的结构信息更加敏感. 文献[20]提出了一种结构相似度 (SSIM) 图像质量评价准则,比 PSNR、SAD 等经典图像质量测度更加符合人眼视觉系统 (HVS). 两个图像块之间的 SSIM 定义为:

$$\begin{aligned} \text{SSIM}(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x, y) &= \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{aligned} \quad (3)$$

其中, x, y 分别表示原始图像块和失真的带噪图像块, $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ 分别表示 x 和 y 的均值、方差和协方差. C_1, C_2 和 C_3 是为了避免分母为零而添加的小常数.

失真图像的 SSIM 计算过程如下: 首先对图像分为不重叠的大小相同的子块, 然后对每一个子块分别利用式(3)计算子块的 SSIM, 最后对所有的子块的 SSIM 求平均值, 即为图像的 SSIM. SSIM 值越高, 表示原图像和失真图像的相似度越高, 失真图像的质量越好.

3 基于结构相似的帧间组稀疏表示视频压缩感知重构

本小节将详细描述本文提出的基于结构相似的帧间组稀疏表示重构算法. 首先提出了基于结构相似的帧间组稀疏表示的 CVS 重构框架, 然后描述了本文提出的帧间组稀疏重构算法, 最后给出了基于 SSIM 的非关键帧相似块组构造方案.

3.1 SSIM-InterF-GSR 算法描述

3.1.1 算法整体框架

在编码端, 视频序列被分为图像组 (Group Of Picture, GOP), 每个 GOP 含有一个关键帧和多个非关键帧. 所有帧均采用非重叠块 CS (Block-based CS, BCS)^[8] 的方法进行观测. 关键帧高采样率观测, 非关键帧低采样率观测.

在解码端, 关键帧首先使用 BCS-SPL^[9] 初始重构, 然后使用帧内 GSR^[14] 重构; 非关键帧首先以前一个重

构帧作为参考帧, 使用单假设 (Single Hypothesis, SH)^[21] 重构, 然后利用当前帧和一个或者多个已重构的参考帧构造相似块组, 通过帧间 GSR (SSIM-InterF-GSR) 重构. 算法实现的整体框架如图 1 所示.

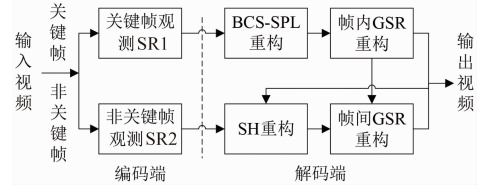


图1 基于帧间组稀疏表示的CVS

3.1.2 SSIM-InterF-GSR 重构算法

对当前帧做初始重构时, 我们通常把每个非重叠图像块作为独立的 CS 子问题进行求解, 但是这种方法不可避免地会丢失图像的全局特性. 为了消除重构图像的块效应, 在 GSR 重构算法中采用重叠分块的方法. 以一个参考帧为例, 非关键帧相似块组的构造方法如图 2 所示. 第 k 个图像块的相似块组由两部分构成: 来自当前帧的匹配块和来自参考帧的匹配块, 分别称之为帧内相似块组 x_{ck1} 和帧间相似块组 x_{ck2} . 帧内相似块组和帧间相似块组组合, 一起构成非关键帧的相似块组 x_{ck} .

非关键帧的相似块组 x_{ck} 可以用自适应学习字典 D_{ck} 的很少原子精确的表示, 即:

$$x_{ck} = D_{ck} \alpha_{ck} \quad (4)$$

其中, α_{ck} 是相似块组 x_{ck} 基于稀疏字典 D_{ck} 的稀疏表示系数. 自适应稀疏字典 D_{ck} 的构造参见文献[14].

利用所有相似块组的稀疏系数 $\alpha_c = \{\alpha_{ck}\}$ 恢复图像块的表达式为:

$$x_c = D_c \circ \alpha_c \quad (5)$$

其中, D_c 表示所有 D_{ck} 的串联, α_c 表示所有 α_{ck} 的串联, x_c 表示所有相似块组 x_{ck} 的串联.

因此, 对非关键帧重构, 即为求解式(6)的优化问题:

$$\begin{aligned} \operatorname{argmin}_{\alpha_c} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \|\alpha_c\|_0 \\ \text{s. t. } x = R(x_c) \end{aligned} \quad (6)$$

其中, R 表示从 x_c 中提取帧内相似块组的操作. 为了保持各变量的物理意义明确, 求解式(6)最优化问题时, 引入向量 u , 则式(6)可以重写为:

$$\begin{aligned} \operatorname{argmin}_{\alpha_c} \frac{1}{2} \|\Phi u - y\|_2^2 + \lambda \|\alpha_c\|_0 \\ \text{s. t. } u = R(D_c \circ \alpha_c) \end{aligned} \quad (7)$$

为了求解等式约束的优化问题, 式(7)可以写为增广拉格朗日的形式:

$$\begin{aligned} L(u, \alpha_c) &= \frac{1}{2} \|\Phi u - y\|_2^2 + \lambda \|\alpha_c\|_0 \\ &+ \frac{\mu}{2} \|u - R(D_c \circ \alpha_c) - b\|_2^2 \end{aligned} \quad (8)$$

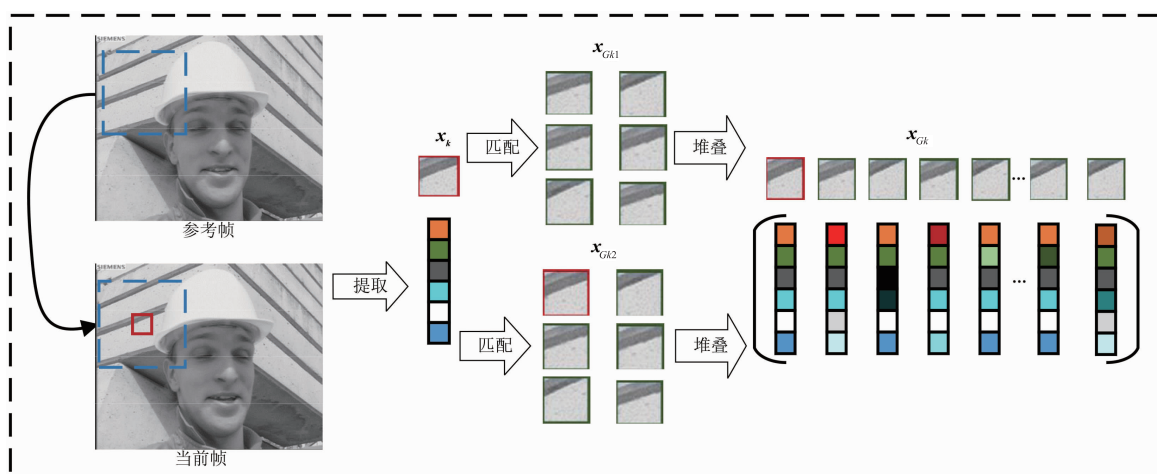


图2 非关键帧相似块组构造示意图

使用 SBI 求解式 (8), 可以分解为如下迭代子问题:

$$\begin{cases} \mathbf{u}^{(t+1)} = \operatorname{argmin}_{\mathbf{u}} \frac{1}{2} \|\Phi \mathbf{u} - \mathbf{y}\|_2^2 \\ \quad + \frac{\mu}{2} \|\mathbf{u} - R(\mathbf{D}_c \circ \boldsymbol{\alpha}_c^{(t)}) - \mathbf{b}^{(t)}\|_2^2 \\ \boldsymbol{\alpha}_c^{(t+1)} = \operatorname{argmin}_{\boldsymbol{\alpha}_c} \lambda \|\boldsymbol{\alpha}_c\|_0 \\ \quad + \frac{\mu}{2} \|\mathbf{u}^{(t+1)} - R(\mathbf{D}_c^{(t+1)} \circ \boldsymbol{\alpha}_c) - \mathbf{b}^{(t)}\|_2^2 \\ \mathbf{b}^{(t+1)} = \mathbf{b}^{(t)} - (\mathbf{u}^{(t+1)} - R(\mathbf{D}_c \circ \boldsymbol{\alpha}_c^{(t+1)})) \end{cases} \quad (9)$$

本文提出的基于结构相似的帧间组稀疏表示视频压缩感知重构算法的具体实现流程如图 3 所示. 在流程图中, 组稀疏表示和更新当前帧两步是对式 (9) 的求解, 实现过程与文献 [14] 中相似, 即对 \mathbf{u} 子问题通过梯度下降法求解, 对 $\boldsymbol{\alpha}_c$ 子问题通过硬阈值求解, 本文不再赘述. 基于 SSIM 生成相似块组和阶梯递减调整匹配块个数这两个技术是本文所提 SSIM-InterF-GSR 算法的关键, 它们能够显著的提升算法重构质量. 下一小节将详细介绍以上两个创新点.

3.2 基于 SSIM 的非关键帧相似块组构造方案

对比帧内 GSR, SSIM-InterF-GSR 的相似块组由帧内相似块组和帧间相似块组两部分组成. 基于视频序列帧间相关性, 在参考帧中选择帧间相似块组的作用是, 利用已重构的高质量参考帧, 更新当前帧内匹配的相似块, 以提高当前帧的重构质量. 然而, 当帧内相似块组和当前块的匹配程度低于帧间相似块组时, 帧内相似块组就会被错误更新. 因此, 非关键帧相似块组的构造是影响帧间 GSR 算法重构质量的关键. 基于以上分析, 本文提出了基于 SSIM 的相似块匹配准则和阶梯递减匹配块个数调整方案, 下面分别详细描述.

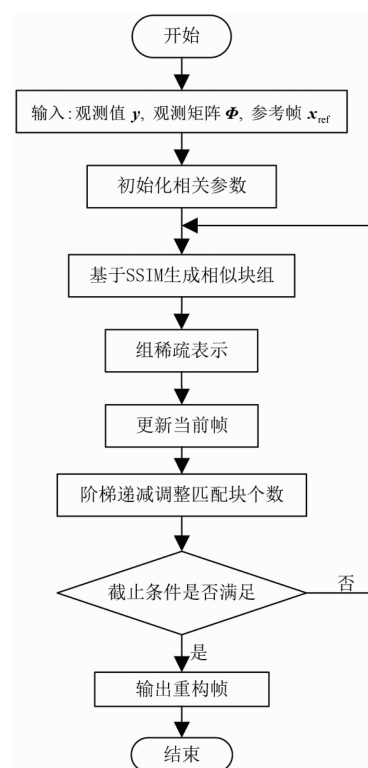


图3 SSIM-InterF-GSR算法实现流程图

3.2.1 基于结构相似(SSIM)的匹配块选择

相似块匹配准则直接决定着选择的匹配块与当前图像块的相似程度. 目前相关文献^[13,14,22]中都采用最直接的欧氏距离匹配准则, 其优点在于计算复杂度低, 物理意义清晰. 它特别适合于变化比较平坦的图像块之间的匹配, 但是对边缘或者纹理信息比较多的图像块, 仅从灰度值来考察图像块之间的相似程度就不是一个很好的匹配准则了. 为了综合考虑图像块之间灰度值, 对比度以及结构特征, 本文提出了利用结构相似度

(SSIM)^[20] 作为相似图像块匹配准则。

为了减少计算复杂度,我们先利用欧式距离作为匹配准则预选一定数量相似块,然后再用 SSIM 选定最终的相似块.以当前帧的一个图像块 \mathbf{x}_k (图 2 中红色标记的方块) 为例,它的候选相似块集 C_k 由当前帧和参考帧窗内的图像块构成.假设最终选定的相似块个数为 c ,首先利用欧式距离作为相似块匹配准则从候选集 C_k 中选出 $3 \times c$ 个与 \mathbf{x}_k 最匹配的图像块,作为最佳候选相似块集 O_{c_i} ; 然后以 SSIM 作为相似块匹配准则,从 O_{c_i} 中选出最匹配的 c 个图像块构成图像块 \mathbf{x}_k 最终的相似块组 \mathbf{x}_{ck} .

3.2.2 帧内匹配块个数阶梯递减调整方案

在非关键帧重构过程中,随着算法的迭代,视频信号的重构质量不断提升,图像细节信息越来越多,帧内与当前块相似的图像块个数会越来越少.如果固定匹配块个数,那么排在后面的匹配块与当前块的相似程度很低.把这些块作为有用信息更新当前帧,那么一定会降低视频信号的重构质量.

为了解决上面的问题,本文提出了一种阶梯递减匹配块个数调整方案,即在算法迭代开始时,由于当前帧图像含有噪声多,细节信息不太明显,相似块匹配时不会很精确,可选择较多的匹配块;随着迭代次数的增加,当前帧图像质量提升,图像内的细节信息越来越多,与当前图像块匹配的相似块个数会越来越少,匹配块个数减少.该调整方案的实现过程如图 4 所示,其中 \mathbf{x}^0 为当前帧的单假设初始重构, c_0 为当前帧初始匹配块个数, Step 为阶梯递减步长, IterNum 是帧间 GSR 重构最大迭代次数, T 为匹配块个数调整次数, τ_1 是控制匹配块个数调整的阈值, τ_2 是控制重构算法迭代截止的阈值 ($\tau_1 > \tau_2$).

在图 4 中,帧间 GSR 重构对应着问题式(9)的求解过程,由此可以得到第 t 次迭代的重构帧 $\mathbf{x}^{(t)}$. 在匹配块个数调整的过程中,如果前后两次重构帧的 SSIM 差的绝对值小于一个固定的阈值 τ_1 , 则说明在该匹配块个数条件下,重构质量的提升已经到达了上限,应该调整匹配块个数.但是,匹配块个数递减调整也不是无限制的,必须要保证大于零(当前处理图像块 \mathbf{x}_k 一定是自身的一个相似块),本文通过限制调整次数 ($\text{count} < T$) 来保证这一点.

表 1 不同帧间匹配块个数调整方案性能 (PSNR/dB) 对比

	按式(10)调整 帧间匹配块个 数其中($k=2$)	固定帧间匹配块个数					
		10	20	30	40	50	60
soccer	32.97	32.12	32.30	32.83	32.76	32.76	32.68
football	27.29	26.42	27.17	27.37	27.42	27.25	27.12
foreman	39.20	39.07	38.93	38.89	38.76	38.61	38.45

3.2.3 帧间匹配块个数

来自参考帧的匹配块直接关系到最终的重构质

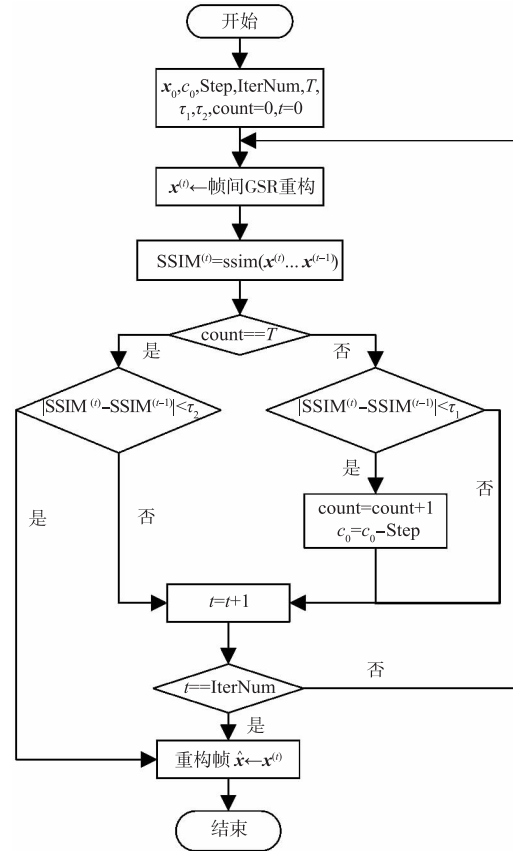


图 4 匹配块个数调整方案的实现框图

量,个数选择少时,不足以更新帧内相似块组;相反,选择过多时,参考帧中的匹配块必然存在一部分与当前图像块的相似度很小,以这些匹配块更新帧内相似块组,势必会越来越差.另外,随着算法迭代的进行,当前帧重构质量提升,细节信息增多,帧间与当前块相似的图像块个数减少,即帧间匹配块个数也应随算法迭代进行而调整.综上所述,本文利用 3.2.2 节已经确定的帧内匹配块个数,通过比例关系来确定帧间匹配块个数.

$$c_{\text{ref}} = k c_0 \quad (10)$$

其中, k 为比例系数, c_0 为当前帧内匹配块个数, c_{ref} 为来自某一参考帧的帧间匹配块个数.比例系数 k 的取值将在仿真部分给出.

为了验证式(10)的有效性,以 soccer、football 和 foreman 序列的第二帧为例,本文对比了固定帧间匹配块个数(取值分别为 10、20、30、40、50、60 块)和按式(11)调整帧间匹配块个数的重构性能 (PSNR/dB), 如表 1 所示.从仿真结果可见,按式(10)确定帧间匹配块个数具有更好的适应性.

4 仿真结果及分析

在仿真中,我们用到的视频序列有 4CIF 格式的

crew, CIF 格式的 bus、foreman、football、coastguard、hall、mother-daughter, 以及 QCIF 格式的 soccer、foreman 和 football. 在信号采集端, 使用鲁棒性较强的随机高斯观测矩阵采集原始信号. 在基于组稀疏表示重构中, 图像分块均设置为大小是 8×8 , 滑动步长为 4 个像素的重叠矩形块. 关键帧重构时, 相似块组个数设置为固定的 40 块^[14]. 非关键帧重构时, 当前帧的参考帧选择如图 5 所示(以 GOP = 8 为例). 每个当前帧的参考帧由一个或者两个已重构帧构成, 把时间上最邻近的参考帧称为第 1 参考帧, 把时间上次邻近的参考帧称为第 2 参考帧(以第 2 帧为例, 第 1 帧为它的第 1 参考帧; 以第 3 帧为例, 第 2 帧为它的第 1 参考帧, 第 1 帧为它的第 2 参考帧). 非关键帧相似块组的匹配块总数 $c = (1 + k_1 + k_2)c_0$, 其中 k_1 为从第 1 参考帧中选取匹配块个数的比例系数, k_2 为从第 2 参考帧中选取匹配块个数的比例系数, c_0 为从当前非关键帧中选取的帧炮相似块个数. 考虑到时间上越邻近的视频帧相关性越大, 通常 $k_1 > k_2$. 仿真中用到的比例系数 k_1, k_2 取经验值, 取值如表 2 所示(以 GOP = 8 为例).

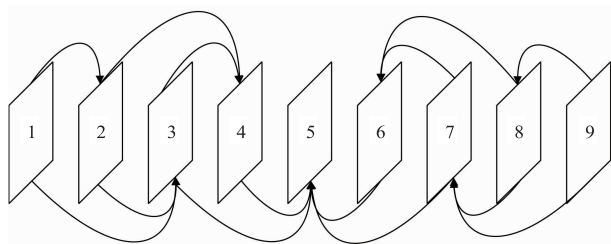


图5 非关键帧参考帧示意图

表 2 确定帧间相似块组匹配块个数的比例系数

视频序列号	2,8	3,4,6,7	5
k_1	2.0	1.8	1.3
k_2	0.0	1.0	1.0

4.1 无量化仿真结果及分析

本小节, 本文算法和两个 CVS 重构算法做了对比, 一个是基于 GSR 的 CVS 重构算法 Hi-STGSR^[18], 另一个是目前性能最好的 CVS 重构算法 Up-Se-AWEN-HHP^[12]. 对比中相关算法的实验结果均来自于对应文献.

4.1.1 与 Hi-STGSR 的仿真结果对比分析

为了与文献[18]中 Hi-STGSR 方案仿真结果做对比, 这部分实验条件与文献[18]完全一致. 视频序列是 CIF 格式的 bus、foreman 和 football 以及 4CIF 格式的 crew. 一个 GOP 包含 8 帧, 共取 12 个 GOP, 即 96 帧. 观测时分块大小是 32×32 . 对每个序列, 平均采样率分别设置为 0.3, 0.4, 0.5, 即在本文算

法中, 分别设置关键采样率为 0.5, 0.6, 0.7, 对应的非关键帧采样率为 0.2714, 0.3714, 0.4714. 仿真结果如表 3 所示.

从表 3 的仿真结果可见, 对比 Hi-STGSR, 本文算法在重构质量 PSNR 上均有提升. 在平均采样率是 0.3, 0.4, 0.5 的条件下, SSIM-InterF-GSR 较 Hi-STGSR 分别有 2.89dB, 2.56dB, 2.39dB 的提升. 对 football 序列, 在平均采样率 0.5 时, 本文算法比 Hi-STGSR 甚至高出 4.79dB. 这主要是因为: (1) 在 Hi-STGSR 中采用欧氏距离作为相似块匹配准则, 而欧氏距离只是像素灰度值之间误差的一种统计方法, 并没有考虑到像素之间的空间相关性以及块之间的结构相似性, 相反 SSIM 能够很好的反映图像块的结构信息; (2) 在 Hi-STGSR 非关键帧重构时, 采用固定相似块个数的相似块组构造方案, 这种方案忽略了重构视频帧之间图像质量的差异, 以及当前帧随迭代进行图像质量提升的这一事实. 而本文提出的阶梯递减相似块个数调整方案正是针对这些问题的一种解决方案.

4.1.2 与 Up-Se-AWEN-HHP 的仿真结果对比分析

SSIM-InterF-GSR 重构算法与文献[11]中 Up-Se-AWEN-HHP 仿真结果的对比如表 4 所示. 实验条件与文献[11]完全相同, 即 GOP = 8, 关键帧采样率是 0.7, 非关键帧采样率从 0.1 到 0.5 变化.

从表 4 仿真结果可见, SSIM-InterF-GSR 在大多数情况下均优于 Up-Se-AWEN-HHP. 对 foreman, hall 和 coastguard 序列, 在不同非关键帧采样率条件下, SSIM-InterF-GSR 重构质量 PSNR 均高于 Up-Se-AWEN-HHP. 而且非关键帧采样率越高, 本文的解码方案对视频序列的重构质量 PSNR 提升越明显. 如对 hall 序列, 在非关键帧采样率从 0.1 到 0.5 变化时, SSIM-InterF-GSR 相比 Up-Se-AWEN-HHP 分别提升 3.34dB, 4.25dB, 4.44dB, 4.72dB, 5.19dB. 这是因为: (1) 本文重构算法利用相似块组在自适应稀疏字典下的稀疏性, 而 Up-Se-AWEN-HHP 重构算法利用预测残差信号在变换域(2D-DCT, DDWT)的稀疏性. 和固定稀疏变换相比, 基于相似块组自适应学习得到的稀疏字典能够更稀疏的表示当前信号. 而且采样率越高, 非关键帧初始重构细节信息越多, 在 SSIM-InterF-GSR 中, 非关键帧相似块组的各匹配块之间相似程度越高, 其稀疏性越好, 因此重构质量提升越明显; (2) Up-Se-AWEN-HHP 是一种观测域的多假设重构算法, 在重构端以每一个图像块作为 CS 求解子问题, 会产生块效应影响重构质量. 在 SSIM-InterF-GSR 重构算法中, 相似块组的构造是基于像素域的重叠分块, 它很好地消除由于不重叠分块导致的重构块效应.

表 3 仿真视频序列前 96 帧的平均 PSNR

平均采样率	算法	bus_cif	foreman_cif	football_cif	crew_4cif	平均
0.3	Hi-STGSR ^[18]	30.32	37.54	33.94	39.76	35.39
	SSIM-InterF-GSR	34.01	39.78	38.73	40.59	38.28
0.4	Hi-STGSR ^[18]	33.99	39.68	36.55	41.10	37.83
	SSIM-InterF-GSR	36.73	41.48	41.19	42.14	40.39
0.5	Hi-STGSR ^[18]	37.24	41.46	38.83	42.40	39.98
	SSIM-InterF-GSR	39.08	43.17	43.42	43.81	42.37

表 4 CIF 格式视频序列关键帧采样率 0.7, 前 88 帧的平均 PSNR

序列	算法	非关键帧采样率				
		0.1	0.2	0.3	0.4	0.5
foreman_cif	Up-Se-AWEN-HHP ^[11]	36.17	38.15	39.65	40.86	42.08
	SSIM-InterF-GSR	37.35	39.53	40.95	42.24	43.53
hall_cif	Up-Se-AWEN-HHP ^[11]	35.15	36.50	37.51	38.37	39.11
	SSIM-InterF-GSR	38.49	40.75	41.95	43.09	44.30
coastguard_cif	Up-Se-AWEN-HHP ^[11]	30.86	32.94	34.67	36.29	37.99
	SSIM-InterF-GSR	31.09	33.83	35.67	37.39	39.07
mother-daughter_cif	Up-Se-AWEN-HHP ^[11]	42.29	43.80	45.04	46.15	47.25
	SSIM-InterF-GSR	41.56	43.77	45.20	46.56	47.88

4.2 帧间残差 DPCM 量化仿真结果及分析

在以上仿真中,我们都假定重构端有精准的观测值.然而在实际应用中,观测值量化后才能存储或传输.为了充分利用视频序列的帧间相关性,本文利用文献[12]的观测值帧间残差量化方法对观测值进行量化.

在以下视频压缩感知重构算法中,设置 GOP = 8, 测试 12 个 GOP, 共 96 帧, 仿真序列为 15Hz 的 QCIF 格式的 foreman、football 和 soccer. 观测时采用非重叠分块, 分块大小是 16 × 16. 在 SSIM-InterF-GSR 仿真中发现, 关键帧采样率 k_{sr} 和量化深度 k_B , 及非关键帧采样率 nk_{sr} 和量化深度 k_b 的取值对率失真性能有较大的影响, 本文通过遍历多个采样率和量化深度得到最优的率失真性能曲线. 采样率和量化深度 (k_{sr} , nk_{sr} , k_B , k_b) 的取值如表 5 所示, 共 30 个点. 仿真结果如图 6 所示, 分别给出了 PSNR 和 SSIM 作为评判标准的性能曲线. 图 6 还给出了最新文献中基于多参考帧的最优多假设预测算法 (Multireference-based Optimal Multihypothesis Prediction, MRMH)^[12], 和 MH-BCS-SPL^[17] 的视频压缩感知重构算法率失真性能曲线. 其中 MRMH 和 MH-BCS-SPL 算法与 SSIM-InterF-GSR 在相同的观测条件下进行仿真, 即遍历表 5 的所有点, 筛选最优点得到最终率失真曲线. 另外, 为了与传统视频压缩性能对比, 图 6 给出了 H. 264^[23] 帧内编码率失真性能, 其中量化步长 QP 分别取 16, 20, 24, 28, 32, 36.

从图 6 的仿真结果可以得出三个结论: (1) 相比于

MRMH 和 MH-BCS-SPL, 基于 PSNR 或 SSIM 质量评判准则分析, 本文算法 SSIM-InterF-GSR 具有最好的重构性能. PSNR 最多达到 2dB 的性能提升, SSIM 最多可达 0.0381 的性能提升. 这是因为 MRMH 和 MH-BCS-SPL, 与 Up-Se-AWEN-HHP 一样, 均是观测域的多假设重构算法. 固定非重叠块的预测限制了其重构性能; (2) 相比于基于 MH 的重构算法, 本文算法对运动比较快的序列提升更加明显. 这是因为对运动比较快的序列, 视频帧之间的时间相关性很小, 在参考帧中很难找到好的匹配块; 而对于 SSIM-InterF-GSR 算法, 相似块组由帧内相似块组和帧间相似块组两部分构成, 对于运动比较快的序列, 由于帧内相似块组的存在依然可以保证重构质量; (3) 相比于传统的 H. 264, 本文的 CVS 重构方案在率失真性能上仍然有很大的差距. 对运动比较快的 football 序列在比特率是 547.18kbps 时, 基于 PSNR 评判标准, H. 264 较本文算法高出 6.79dB; 基于 SSIM 评判标准, H. 264 较本文算法高出 0.1017.

在比特率是 643.90kbps 时, 图 7 给出了 football 序列第 53 帧原始图像 (图 7(a) 原始帧) 视觉效果和在不同解码方案 (图 7(b) H. 264^[23], 图 7(c) SSIM-InterF-GSR, 图 7(d) MRMH^[12], 图 7(e) MH-BCS-SPL^[17]) 下的重构视觉效果. 从图 7 可以观察到, 传统 H. 264 编解码方案具有最好的重构视觉效果, 而本文所提的 SSIM-InterF-GSR 较基于 MH 的视频压缩感知重构算法 MRMH 和 MH-BCS-SPL 具有更好的重构效果.

表 5 采样率和量化深度取值

k_sr	0.50	0.50	0.55	0.60	0.60	0.70	k_B	8	8	8	8	8
nk_sr	0.10	0.15	0.20	0.25	0.30	0.40	nk_b	8	7	6	5	4

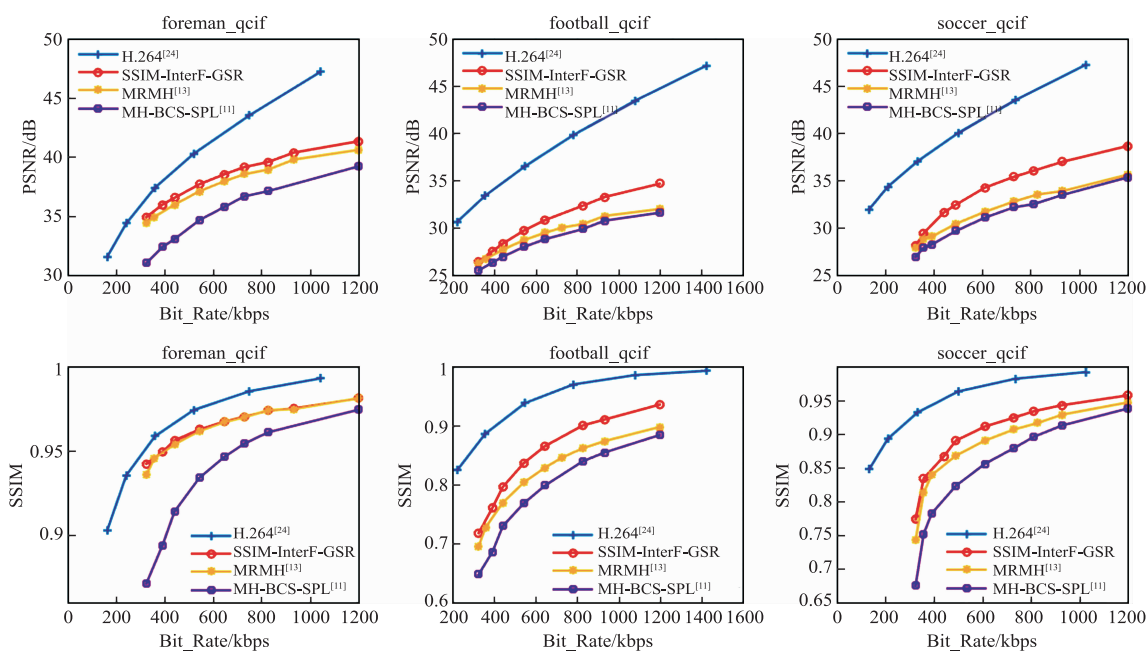


图6 Foreman, Football和Soccer率失真性能仿真结果

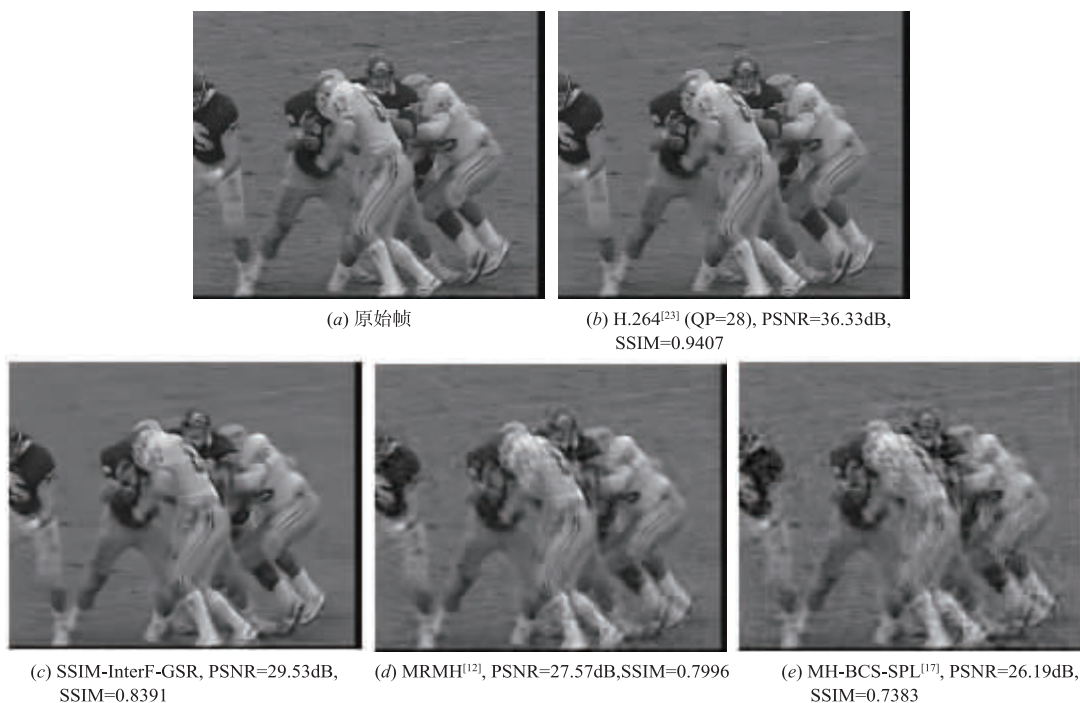


图7 Football序列第53帧不同解码方案的输出(比特率: 643.90kbps)

5 结论

在本文中,我们提出了一种基于结构相似的帧间

组稀疏表示视频压缩感知重构算法(SSIM-InterF-GSR).首先,描述了算法的实现流程,即在当前帧和一个或者多个已重构参考帧中匹配相似块生成相似块

组,通过相似块组构造自适应稀疏字典,以相似块的稀疏性作为先验知识构造正则项,通过 SBI 求解优化问题实现视频的高精度重构. 其次,为了寻找更好的匹配块,本文提出使用更加符合人眼视觉系统的结构相似度(SSIM)作为相似块匹配准则. 最后,在算法仿真中发现,相似块组的匹配块个数对当前帧重构质量有很大的影响,我们还提出了阶梯递减匹配块个数调整方案用于确定相似组的匹配块个数,该方案能够很好的适应随着算法迭代,重构视频图像质量提高,细节信息增加,帧内图像块之间的相似度变小这一理论. 仿真结果表明,相比于现有的视频压缩感知重构算法,本文算法获得了更好的视频重构质量. 但是,相比于传统的视频编解码方案仍然有较大的差距. 此外,从仿真结果还可以看出,在 CVS 中采样率和量化深度之间存在一个折中的关系. 如何构造观测矩阵以及权衡采样率和量化深度之间的关系将是下一步的研究重点.

参考文献

- [1] DONOHO D L. Compressed sensing [J]. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289 – 1306.
- [2] 石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展 [J]. *电子学报*, 2009, 37(5): 1070 – 1081.
SHI Guang-ming, LIU Dan-hua, GAO Da-hua, et al. Advances in theory and application of compressed sensing [J]. *Acta Electronica Sinica*, 2009, 37(5): 1070 – 1081. (in Chinese)
- [3] Candes E, Romberg J. l1-magic: Recovery of Sparse Signals via Convex Programming [OL]. URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf, 2005-4-14.
- [4] WRIGHT S J, NOWAK R D, FIGUEIREDO M A T. Sparse reconstruction by separable approximation [J]. *IEEE Transactions on Signal Processing*, 2009, 57(7): 2479 – 2493.
- [5] FIGUEIREDO M A T, NOWAK R D, WRIGHT S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems [J]. *IEEE Journal on Selected Topics in Signal Processing*, 2007, 1(4): 586 – 597.
- [6] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. *IEEE Transactions on Information Theory*, 2007, 53(12): 4655 – 4666.
- [7] ANDRIOLE S J S J. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples [J]. *Communications of the ACM*, 2010, 53(12): 67 – 79.
- [8] LU GAN. Block compressed sensing of natural images [A]. 2007 15th International Conference on Digital Signal Processing [C]. Cardiff, UK: IEEE, 2007. 403 – 406.
- [9] MUN S, FOWLER J E. Block compressed sensing of images using directional transforms [A]. *I Data Compression Conference (DCC)* [C]. Snowbird, USA: IEEE, 2009. 3021 – 3024.
- [10] SUNGKWANG M, FOWLER J E. Residual reconstruction for block-based compressed sensing of video [A]. *Data Compression Conference (DCC)* [C]. Snowbird, USA: IEEE, 2011. 183 – 192.
- [11] Kuo Y, Wu K, Chen J. A scheme for distributed compressed video sensing based on hypothesis set optimization techniques [J]. *Multidimensional Systems and Signal Processing*, 2017, 28(1): 129 – 148.
- [12] 杨春玲, 欧伟枫. CVS 中基于多参考帧的最优多假设预测算法 [J]. *华南理工大学学报: 自然科学版*, 2016, 44(1): 1 – 8.
YANG Chun-Ling, OU Wei-Feng. Research on multireference-based optimal multihypothesis prediction in compressed video sensing [J]. *Journal of South China University of Technology (Natural Science Edition)*, 2016, 44(1): 1 – 8. (in Chinese)
- [13] 沈燕飞, 李锦涛, 朱珍民, 等. 基于非局部相似模型的压缩感知图像恢复算法 [J]. *自动化学报*, 2015, 41(2): 261 – 272.
SHEN Yan-Fei, LI Jin-Tao, ZHU Zhen-Min, et al. Image reconstruction algorithm of compressed sensing based on nonlocal similarity model [J]. *Acta Automatica Sinica*, 2015, 41(2): 261 – 272. (in Chinese)
- [14] ZHANG J, ZHAO D, GAO W. Group-based sparse representation for image restoration [J]. *IEEE Transactions on Image Processing*, 2014, 23(8): 3336 – 3351.
- [15] Do T T, Chen Y, Nguyen D T, et al. Distributed compressed video sensing [A]. *Image Processing (ICIP)* [C]. Cairo, Egypt: IEEE, 2009. 1393 – 1396.
- [16] Cossalter M, Valenzise G, Tagliasacchi M, et al. Joint compressive video coding and analysis [J]. *IEEE Transactions on Multimedia*, 2010, 12(3): 168 – 183.
- [17] Chen C, Tramel E W, Fowler J E. Compressed-sensing recovery of images and video using multihypothesis predictions [A]. *Signals, Systems and Computers (ASILOMAR)* [C]. Pacific Grove, USA: IEEE, 2011. 1193 – 1198.
- [18] Gao X, Jiang F, Liu S, et al. Hierarchical frame based spatial-temporal recovery for video compressive sensing coding [J]. *Neurocomputing*, 2016, 174: 404 – 412.
- [19] GOLDSTEIN T, OSHER S. The split bregman method for L1-regularized problems [J]. *SIAM Journal on Imaging Sciences*, 2009, 2(2): 323 – 343.
- [20] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity

- [J]. IEEE Transactions on Image Processing, 2004, 13 (4):600-612.
- [21] TRAMEL E W, FOWLER J E. Video compressed sensing with multihypothesis [A]. Data Compression Conference (DCC) [C]. Snowbird, USA: IEEE, 2011. 193-202.
- [22] LI S, QI H. A douglas-rachford splitting approach to compressed sensing image recovery using low-rank regularization [J]. IEEE Transactions on Image Processing, 2015, 24 (11):4240-4249.
- [23] SÜHRING K. H. 264/AVC Software JM 19.0 [OL]. <http://iphome.hhi.de/suehring/tml/>, 2015.

作者简介



和志杰 男,1990 年生于河南漯河,华南理工大学电子与信息学院在读研究生. 研究方向:视频压缩感知.



杨春玲(通信作者) 女,1970 年生于河南新乡,华南理工大学电子与信息学院博士生导师. 研究方向:图像/视频压缩研究.

E-mail: eeelyang@scut.edu.cn



汤瑞东 男,1993 年生于广东汕头,华南理工大学电子与信息学院在读研究生. 研究方向:视频压缩感知.