

DOI:10.11918/202105063

视频压缩感知采样率自适应的帧间片匹配重构

陈根龙¹, 刘浩^{1,2}, 周健¹, 黄荣^{1,2}

(1. 东华大学 信息科学与技术学院, 上海 201620; 2. 东华大学 数字化纺织服装技术教育部工程研究中心, 上海 201620)

摘要: 相比于香农采样需要大规模样本, 压缩采样在视频信号的高能效表征方面具有独特优势。当关键帧采样率与非关键帧采样率不一致时, 现有的帧间片匹配算法在多种联合采样率下呈现出稳定的重建质量。为了充分利用帧间的相关性进行视频压缩感知重构, 本文提出了一种采样率自适应的帧间片匹配重构算法, 根据关键帧采样率与非关键帧采样率的相对变化执行差异化的帧间片匹配, 以更好地适配不同联合采样率生成的视频码流。所提算法首先对各帧的观测值分别进行帧内重构, 然后按照关键帧采样率较非关键帧采样率的增长率情形进行自适应的帧间重构: 若增长率较低, 当前帧直接选取最近的关键帧和同向的邻近非关键帧的重构结果, 执行基于同向双参考帧的片匹配重构; 若增长率较高, 当前帧逐步地挑选多个参考帧的重构结果, 完成更为精细的多阶段片匹配重构。相较于现有重构算法, 所提算法在视频压缩感知的典型架构中能够获得更为稳定的重构性能, 视频序列的重建质量得到了一定程度的提升。

关键词: 视频压缩感知; 采样率; 帧间重构; 片匹配; 参考帧

中图分类号: TN919.8 **文献标志码:** A **文章编号:** 0367-6234(2022)05-0146-06

Subrate-adaptive interframe patch matching for video compressive sensing reconstruction

CHEN Genlong¹, LIU Hao^{1,2}, ZHOU Jian¹, HUANG Rong^{1,2}

(1. College of Information Science and Technology, Donghua University, Shanghai 201620, China; 2. Engineering Research Center of Digitized Textile & Apparel Technology, Ministry of Education, Donghua University, Shanghai 201620, China)

Abstract: In contrast with the Shannon sampling which requires large-scale samples, the compressive sampling has unique advantages in energy-efficient representation of video signals. When the sampling subrates of a keyframe and a non-keyframe are inconsistent, the existing interframe patch matching algorithms show unstable recovery quality under different subrate combinations. In order to fully utilize the temporal correlation between frames, a subrate-adaptive interframe patch matching algorithm was proposed for video compressive sensing reconstruction, where the differential interframe matching mechanism was performed according to the relative change between keyframe sampling subrate and non-keyframe sampling subrate for better adaptation of video bitstream generated from different subrate combinations. Firstly, the measurements of each frame were reconstructed to obtain the corresponding intraframe results respectively. Then, the adaptive interframe reconstruction was implemented according to the growth ratio of the keyframe sampling subrate to the non-keyframe sampling subrate. In the case of low growth ratio, the current frame selected the nearest keyframe and the non-keyframe in the same direction as its co-directional double reference frames for patch matching reconstruction. In the case of high growth ratio, the current frame gradually selected the reconstruction results of multiple reference frames for multi-stage patch matching reconstruction. Compared with the existing reconstruction algorithms, the proposed algorithm could achieve more stable reconstruction performance during the typical framework of video compressive sensing, and thus the recovery quality of video sequence was consistently improved.

Keywords: video compressive sensing; sampling subrate; interframe reconstruction; patch matching; reference frame

压缩感知理论旨在以亚奈奎斯特速率采样信号, 高能效地获取信号的稀疏表征。该理论表明, 具

有稀疏性的图像/视频信号可通过少量随机观测值进行恢复^[1]。文献[2]提出一种基于时域帧分层的视频压缩感知编码框架, 该框架实质上是采用三维多尺度变换对视频信号进行表征, 其测量端的实现复杂度过高。分布式压缩感知为视频压缩感知提供了部分理论基础, 每帧单独测量生成的多帧观测向量通常服从某种联合稀疏模型, 根据空时联合信息

收稿日期: 2021-05-14
基金项目: 中央高校基本科研业务费专项(2232021G-09);
国家自然科学基金(62001099)
作者简介: 陈根龙(1982—), 男, 讲师
通信作者: 刘浩, liuhao@dhu.edu.cn

利用程度的不同,视频重构的性能及资源消耗也不一样^[3]。视频压缩感知需要设计一个可以高概率地满足有限等距性质 (restricted isometry property, RIP) 且尽可能保存足够有效信息的观测矩阵^[4-5]。文献[6]将卷积神经网络与压缩感知相结合,形成视频压缩感知网络 (video compressed sensing network, VCSNet), 通过端到端的训练来提升重建视频的保真度,但忽视了设计满足 RIP 的观测矩阵。Gaussian 随机观测矩阵是普适满足 RIP 的常用观测矩阵。此外,分块压缩感知 (block compressive sensing, BCS) 对测量端的传感或存储资源要求较低,不仅可以实现低内存的观测矩阵,并且可以逐块或并行传输^[7-8]。BCS 将每帧分成同等尺寸的子块,对每个子块进行独立观测,得到相应的观测值。

视频码流通常由连续的图像组 (group of pictures, GOP) 构成,每一组 GOP 含有一个关键帧以及后续若干非关键帧。残差信号通常比原始信号具有更大的稀疏性,基于 GOP-BCS 典型架构的残差稀疏重构是目前最具代表性的视频压缩感知重构技术:各帧先进行独立的帧内重构,然后非关键帧利用已重构的参考帧进行片匹配重构。文献[9]提出了多假设 (multi-hypothesis, MH) 帧间片匹配算法,将已重构的关键帧作为当前帧的参考帧,结合 Tikhonov 正则化对当前帧进行残差稀疏重构。在 MH 算法的基础上,文献[10]在分裂布雷格曼迭代中不断更新信号残差的权重,构建重加权残差稀疏 (reweighted residual sparsity, RRS) 模型,实现多假设预测精度和重建帧质量的交替提升。进一步地,文献[11]提出了一种结合长短距离 (long short distance, LSD) 的帧间片匹配算法,通过在一个 GOP 内动态地确定参考帧,在一定程度上改进了重建质量。对于视频压缩感知重构,文献[12]提出了一种结合光流的结构组稀疏 (structural group sparsity, SGS) 算法,其中帧间片匹配交替地实现光流估计和残差稀疏重构,该最新算法适合于非关键帧采样率较低的情况。

在 GOP-BCS 典型架构下,测量端逐帧地采用分块随机观测进行视频信号的特征,关键帧采样率与非关键帧采样率称为联合采样率。现有的帧间片匹配算法仅关注非关键帧采样率较低的情况,此时低质量的多个非关键帧特别适合进行稀疏建模,单一的假设片选择模式有利于结合稀疏模型取得稳定的片匹配性能。对于更多联合采样率的情况,仍有待提出有效的片匹配重构机制。本文的主要创新点在于:改变了以往单一的参考帧选取方式,根据联合采样率的变化自适应地执行双向的帧间片匹配重构,

以更好地应对不同联合采样率导致的码流变化,稳定地改善视频序列的重建质量。

1 残差稀疏重构

视频残差稀疏重构首先对关键帧和非关键帧分别进行帧内重构,每个 GOP 的关键帧通常具有较高的采样率,因此关键帧在帧内重构之后具有较好的恢复效果,可作为最终的重构结果;而非关键帧使用了较低的采样率,在完成帧内重构后,还需要进行帧间重构,使用其他帧的重构结果作为参考帧进行片匹配,非关键帧的参考帧可以是其前一关键帧或后一关键帧,图 1 给出了非关键帧的当前片从前后参考帧中选取假设片的示意图。在获得最终的重构结果前,每个非关键帧均需要执行多次迭代的帧间重构。

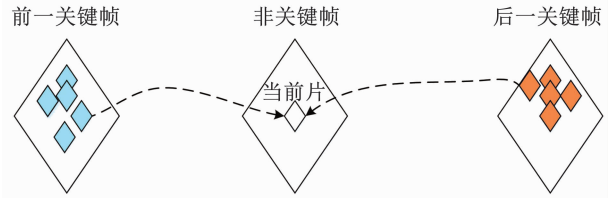


图 1 非关键帧片匹配的示意

Fig. 1 Schematic diagram of non-keyframe patch matching

在每次迭代的帧间重构中,当前帧分成重叠的若干个片,以降低块效应。当前片在以其所处位置为中心的搜索窗内,从参考帧中选取与其具有较高相似度的一组假设片,通过稀疏化当前片与其假设片之间的残差进行视频重构。在参考帧中使用均方误差 (mean square error, MSE) 准则搜索出与当前片近似的一组假设片,选出其中 MSE 较小的 K 个假设片, h_k 表示当前片 x 的第 k 个假设片 ($1 \leq k \leq K$)。一般而言,第 k 个假设片的权重 α_k 与其 MSE 之间有着近似反比例的关系,若某假设片与当前片的相似度不高,则该假设片会被赋予一个较小的权重,即 MSE 大的分配较小权重, MSE 小的分配较大权重,执行如下的权重度量学习:

$$\alpha_k = \frac{e^{\frac{1}{s^2} \|x - h_k\|^2}}{\sum_{1 \leq k \leq K} e^{\frac{1}{s^2} \|x - h_k\|^2}} \quad (1)$$

式中 s^2 表示片大小。当前片 x 的残差 $R(x)$ 是当前片与其假设片线性组合之间的差值,该残差可表示为

$$R(x) = x - \sum_{1 \leq k \leq K} \alpha_k \cdot h_k \quad (2)$$

式中第 k 个假设片的权重 α_k 会影响多假设片匹配的预测精度。当关键帧采样率与非关键帧采样率不一致时,现有的重构算法在变化的联合采样率下呈现

出不稳定的恢复性能。

2 采样率自适应的帧间片匹配

视频码流的每个 GOP 由一个关键帧和若干非关键帧组成。为了应对更多联合采样率的情况,本文提出了一种采样率自适应的帧间片匹配(subrate-adaptive interframe patch matching, SIPM)重构算法,所提 SIPM 算法根据关键帧与非关键帧采样率的变化情况自适应地挑选假设片,并加入了新的参考帧选取机制。图 2 是 SIPM 算法的流程,主要包括以下 3 个步骤:

步骤 1 视频码流由若干个 GOP 组成,每组

GOP 的帧总数为 M ,每一组 GOP 的第一帧是关键帧,后续的 $M-1$ 帧是非关键帧。每帧的观测值独立地进行帧内重构,至此完成关键帧的重构,接下来准备进行非关键帧的帧间重构。

步骤 2 获取关键帧采样率 S_k 、非关键帧采样率 S_n ,计算关键帧采样率较非关键帧采样率的增长率 $G_r = (S_k - S_n)/S_n$,并与阈值 τ 进行比较:若 $G_r < \tau$,表明各帧的重建质量差别不大,邻近非关键帧也具有较好的参考价值,因此使用步骤 3 中的模式 I 对非关键帧进行帧间重构;否则,引入更为精细的参考帧机制,使用步骤 3 中的模式 II 对非关键帧进行多阶段的片匹配重构。

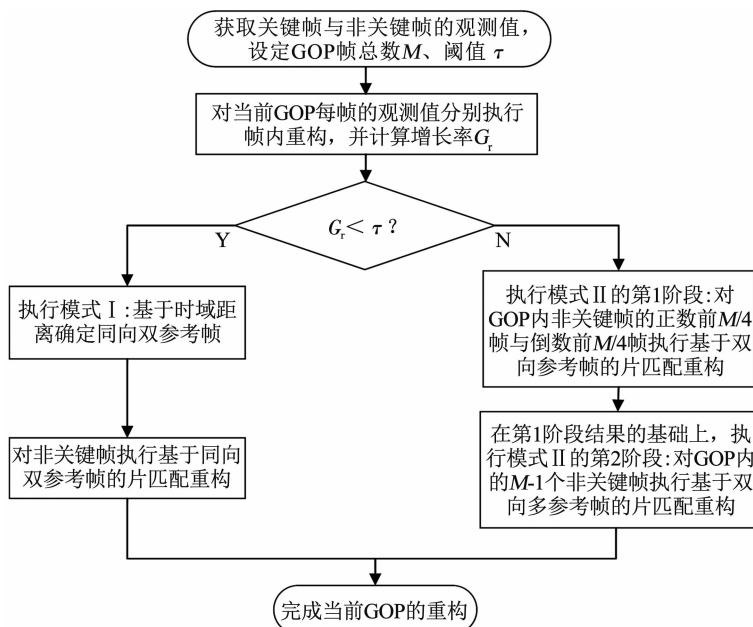


图 2 采样率自适应的帧间片匹配

Fig. 2 Subrate-adaptive interframe patch matching

步骤 3 本步骤只对采样率较低的非关键帧进行帧间重构。根据步骤 2 中的模式选择,对当前 GOP 的非关键帧执行模式 I 或模式 II。

模式 I: 基于帧间的时域距离,当前帧选取最近的关键帧和同向的邻近非关键帧的重构结果作为参考帧,见图 3,图中每个 GOP 的帧总数 $M=8$,帧 1、帧 9 为关键帧,其余帧为非关键帧,箭头指向参考帧。随后,模式 I 使用残差稀疏模型进行基于同向双参考帧的片匹配重构。

模式 II: 为了充分利用关键帧与非关键帧的重构结果,模式 II 分为第 1 阶段与第 2 阶段。如图 4 所示,把各帧重构的结果(包括重构的中间结果:有些帧可能多次进行重构)按照重构的先后次序进行编号,记作 R_i ,其中 i 为某帧在模式 II 中的重构序号。

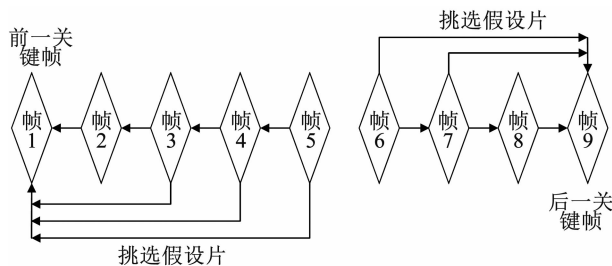


图 3 模式 I 的参考帧示意

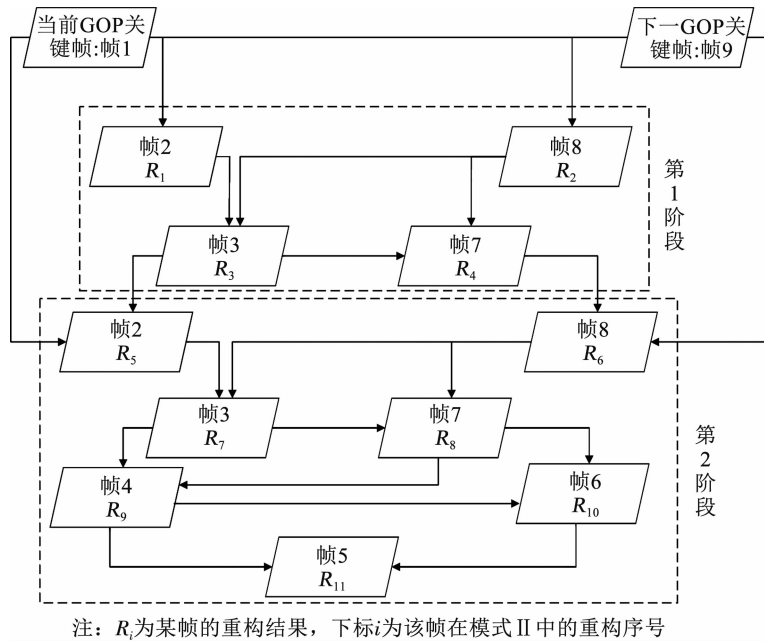
Fig. 3 Schematic diagram of reference frames for mode I

第 1 阶段: 为当前 GOP 内的帧 $M=8$ (最后一个非关键帧)挑选帧间片匹配的参考帧,将当前 GOP 关键帧与下一 GOP 关键帧作为两个参考帧,然后使用残差稀疏模型进行求解,得到序号 1 的重构结果 R_1 。对于当前 GOP 的帧 2(首个非关键帧),其参考帧是当前 GOP 关键帧与下一 GOP 关键帧,同样对其执行残差稀疏模型,得到序号 2 的重构结果 R_2 。

对于当前 GOP 的帧 3(第 2 个非关键帧),其参考帧是 R_1 与 R_2 ,为其运行残差稀疏模型,得到序号 3 的重构结果 R_3 。对于当前 GOP 的帧 $M-1$ (倒数第 2 个非关键帧),其参考帧是 R_3 与 R_2 ,然后进行重构得到序号 4 的重构结果 R_4 。以此类推,直到完成当前 GOP 非关键帧的正数前 $M/4$ 帧与倒数前 $M/4$ 帧的重构。

第 2 阶段:基于当前 GOP 的首个非关键帧在上述第 1 阶段的重构结果,进行再次重构,其参考帧是下一 GOP 关键帧与其相邻非关键帧在上述第 1 阶段的重构结果(当前 GOP 的第 2 个非关键帧在上述

第 1 阶段的重构结果),利用选取的 2 个参考帧运行残差稀疏模型,得到重构结果 $R_{\frac{M}{2}+1}$;对于倒数第 1 个非关键帧进行重构,其参考帧是当前 GOP 的关键帧与当前 GOP 的倒数第 2 个非关键帧在第 1 阶段的重构结果 $R_{\frac{M}{2}+2}$;接下来对当前 GOP 的正数第 2 个关键帧进行重构,其参考帧是 $R_{\frac{M}{2}+1}$ 与 $R_{\frac{M}{2}+2}$,得到其重构结果 $R_{\frac{M}{2}+3}$;接下来,依次从 GOP 的两端向中心交替地进行重构,每个非关键帧的参考帧是在第 2 阶段中先其重构的前两次重构结果,至此完成基于双向多参考帧的片匹配重构。



注: R_i 为某帧的重构结果, 下标 i 为该帧在模式 II 中的重构序号

图 4 模式 II 的参考帧示意

Fig. 4 Schematic diagram of reference frames for mode II

3 仿真实验

仿真实验在装有 MATLAB R2020a 的计算机上进行,其主要配置为: Intel Core i9-10900K CPU、16 GB 内存和 Windows 10 操作系统。针对不同的视频码流,分别选取多组 GOP 进行测试,并为待重构的视频帧从 1 开始按观测顺序进行编号,每个 GOP 的帧总数 $M=8$ 。为验证所提算法在各种联合采样率下的恢复效果,仿真实验基于 GOP-BCS 典型架构,在 5 组联合采样率 (0.7, 0.1)、(0.7, 0.2)、(0.7, 0.4)、(0.7, 0.6)、(0.7, 0.7) 下对视频码流进行残差稀疏重构,括号中的前一个值为关键帧采样率,括号中的后一个值为非关键帧采样率,前 3 组联合采样率为现有算法所采用。所有算法共有的测试参数设置如下:测量端采用 Gaussian 随机观测矩阵,分块尺寸为 16×16 ;在参考帧中,片尺寸 $s^2=8^2$,假设片的搜索窗尺寸 $L=30 \times 30$,当前片挑选的假设

片个数 $K=10$ 。每个视频序列选取前 25 帧进行测试,其中帧 1、帧 9、帧 17、帧 25 为关键帧,其余帧为非关键帧。在本文所提 SIPM 算法中,阈值 τ 选取常用的分界点 0.5,配合采样率自适应的帧间片匹配重构,可取得整体较优的重建质量。

为了便于叙述,文献[9]中的帧间片匹配算法简称 MH,文献[10]中的帧间片匹配算法简称 RRS,文献[11]中的帧间片匹配算法简称 LSD,文献[12]中的帧间片匹配算法简称 SGS。将本文 SIPM 算法与这 4 种代表性的帧间重构算法进行比较。当执行同样的帧内重构算法后,MH 算法主要基于前一关键帧进行片匹配重构,RRS 算法基于前后两个关键帧进行片匹配重构,LSD 算法则在 GOP 内先验地挑选假设片,SGS 算法对帧间残差信号进行复杂的光流估计,这些算法均忽略了联合采样率的差异化,所提 SIPM 算法能根据关键帧与非关键帧采样率的相对变化自适应地执行帧间片匹配。典型地,实验选

择 CIF 格式(分辨率 352×288)的 Akiyo、News、Foreman、Container、Football、Calendar 视频序列,在 5 组联合采样率下依次执行 MH、RRS、LSD、SGS 与 SIPM 算法。

首先,在联合采样率(0.7,0.1)和(0.7,0.6)下,图 5 给出了利用不同算法恢复 Foreman 序列的逐帧结果,峰值信噪比(peak signal noise ratio, PSNR)越大表示重建质量越好,本文以 R_{PSN} 表示。当非关键帧采样率不断降低,即增长率 G_r 越来越大时,在两个相邻的关键帧中各帧的重建质量形成一个不断加深的波谷。因此,联合采样率的差异对于恢复效果具有显著的影响,所提 SIPM 算法根据不同联合采样率执行差异化的帧间片匹配,得到了较好的整体性能。

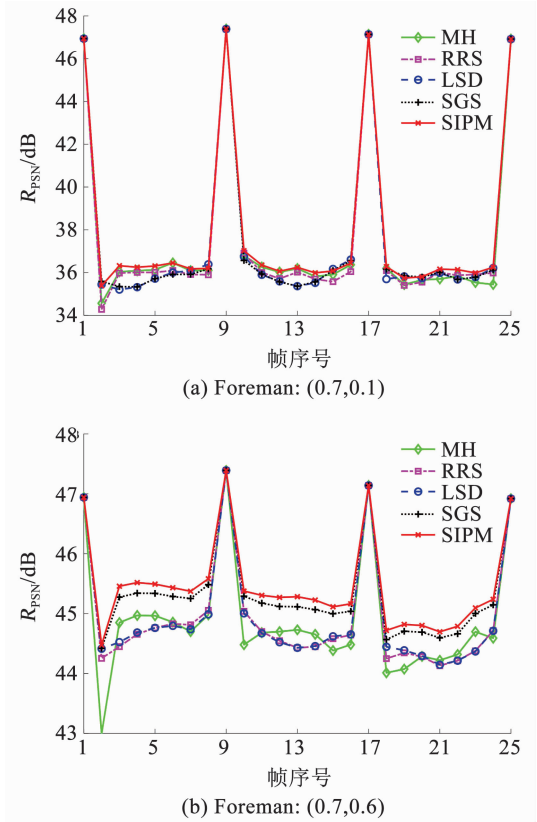


图 5 不同算法的逐帧 R_{PSN} 比较

Fig. 5 Frame-by-frame R_{PSN} comparison of different algorithms

为了直观地比较各种算法的重构性能,图 6 给出了在不同联合采样率下各种算法对 Foreman 序列进行重构的平均结果。从图中可看出,SIPM 算法的重建质量在测试条件下优于现有算法。由于 SIPM 算法学习到了质量更高的参考帧信息,其恢复效果的改善较为明显。

表 1 进一步地给出了各种算法在不同视频序列上重建质量的平均 R_{PSN} 值,在每组联合采样率中,关键帧采样率是逗号前的数值,非关键帧采样率是逗

号后的数值。当增长率 G_r 较低时,帧间质量差别不大,SIPM 算法的帧间重构较好地挖掘了邻近帧的参考价值;而当增长率 G_r 较高时,SIPM 算法通过更为精细的参考帧选取,执行多次迭代的帧间重构,充分利用了质量更好的关键帧信息。

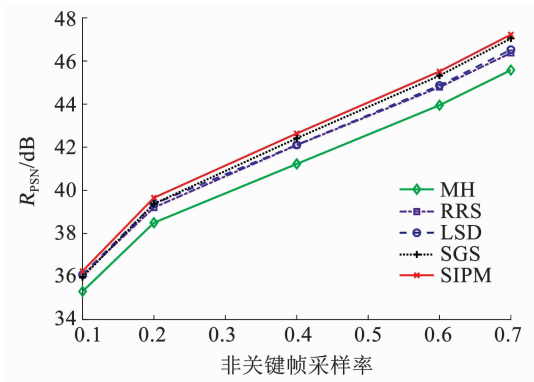


图 6 不同联合采样率的重构结果

Fig. 6 Reconstruction results under different substrate combinations

表 1 各种算法在不同视频序列上的平均 R_{PSN}

Tab. 1 Average R_{PSN} values of different algorithms on different sequences dB

联合采样率	视频序列	MH	RRS	LSD	SGS	SIPM
(0.7,0.1)	Foreman	35.31	36.06	36.09	35.98	36.25
	Container	30.92	31.76	32.25	34.06	33.83
	News	33.89	34.17	35.93	36.07	36.33
	Akiyo	34.27	35.01	35.37	36.29	35.67
	Calendar	28.76	29.36	30.02	30.48	30.35
	Football	25.98	26.41	26.56	26.67	26.73
(0.7,0.2)	Foreman	38.50	39.20	39.37	39.38	39.66
	Container	32.32	33.75	33.95	34.47	34.30
	News	35.42	36.03	36.51	37.39	37.69
	Akiyo	36.46	36.82	37.15	38.26	38.03
	Calendar	31.85	32.39	33.15	34.67	34.46
	Football	28.05	28.67	28.91	29.78	30.19
(0.7,0.4)	Foreman	41.22	42.10	42.1	42.42	42.67
	Container	34.97	34.72	36.11	37.15	36.74
	News	37.11	38.23	38.78	39.47	39.91
	Akiyo	37.52	38.04	38.29	38.89	39.26
	Calendar	34.36	35.27	36.01	37.32	37.63
	Football	29.95	30.65	32.47	32.41	32.56
(0.7,0.6)	Foreman	44.25	44.68	44.77	45.22	45.41
	Container	37.52	38.38	38.85	39.22	39.57
	News	39.72	41.21	42.13	42.31	42.69
	Akiyo	41.69	43.14	43.98	44.19	44.57
	Calendar	36.65	37.18	37.62	38.15	38.48
	Football	32.92	33.56	33.94	34.56	34.76
(0.7,0.7)	Foreman	45.58	46.37	46.52	47.05	47.33
	Container	38.55	39.46	40.72	41.09	41.36
	News	40.58	42.07	43.11	43.34	43.78
	Akiyo	42.37	44.52	44.57	44.82	45.31
	Calendar	37.78	38.31	38.75	39.28	39.61
	Football	34.05	34.69	35.07	35.69	35.89

最后,对各种算法的复杂度进行简要分析。本文 SIPM 算法的复杂度主要由挑选假设片的过程所决定。考虑到上述测试参数的设置,SIPM 算法在极端场景下可能接近 RRS 算法复杂度的 2 倍,但大多数场景下均达不到此上限。在实际测试中,对于 CIF 大小的视频序列,RRS、LSD、SGS 算法重构每帧图像分别平均花费大约 3、4、7 min,而 SIPM 算法重构每帧图像平均花费约 5 min。本文 SIPM 算法通过增加少量的复杂度,相比现有算法取得了更好的平均 R_{PSN} 值。

4 结 论

针对视频压缩感知的联合采样率变化,本文提出了一种采样率自适应的帧间片匹配重构算法,所提算法动态地利用了视频重构过程中的帧间相关性。当关键帧所提供的参考信息较少时,所提算法执行基于同向双参考帧的片匹配重构,反之则采用更为精细的多阶段片匹配重构。在各种联合采样率下,所提算法按照关键帧采样率较非关键帧采样率的增长率情形进行自适应的帧间片匹配重构,有效地减少了帧间变化引入的干扰,相比现有重构算法能够获得更为稳定的重建质量。

参考文献

[1]BARANIUK R G, GOLDSTEIN T, SANKARANARAYANAN A C, et al. Compressive video sensing: Algorithms, architectures, and applications[J]. IEEE Signal Processing Magazine, 2017, 34(1): 52. DOI:10.1109/MSP.2016.2602099

[2]GAO Xinwei, JIANG Feng, LIU Shaohui, et al. Hierarchical frame based spatial-temporal recovery for video compressive sensing coding [J]. Neurocomputing, 2016, 174(part A): 404. DOI:10.1016/j. neucom. 2015. 07. 110

[3]CHEN Can, ZHOU Chao, LIU Pengyuan, et al. Iterative reweighted Tikhonov-regularized multihypothesis prediction scheme for distributed

compressive video sensing[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(1): 1. DOI:10.1109/TCSVT.2018.2886310

[4]ZHENG Shuai, ZHANG Xiaoping, CHEN Jian, et al. A high-efficiency compressed sensing-based terminal-to-cloud video transmission system[J]. IEEE Transactions on Multimedia, 2019, 21(8): 1905. DOI:10.1109/TMM.2019.2891415

[5]TREVISI M, AKBARI A, TROCAN M, et al. Compressive imaging using RIP-compliant CMOS imager architecture and Landweber reconstruction[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(2): 387. DOI:10.1109/TCSVT.2019.2892178

[6]SHI Wuzhen, LIU Shaohui, JIANG Feng, et al. Video compressed sensing using a convolutional neural network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(2): 425. DOI:10.1109/TCSVT.2020.2978703

[7]DADKHAH M, DEEN M J, SHIRANI S. CMOS image sensor with area-efficient block-based compressive sensing [J]. IEEE Sensors Journal, 2015, 15(7): 3699. DOI:10.1109/JSEN.2015.2397874

[8]OBERMEIER R, MARTINEZ-LORENZO J A. Sensing matrix design via capacity maximization for block compressive sensing applications [J]. IEEE Transactions on Computational Imaging, 2019, 5(1): 27. DOI:10.1109/TCI.2018.2884291

[9]FOWLER J E, MUN S, TRAMEL E W. Block-based compressed sensing of images and video[J]. Foundations and Trends in Signal Processing, 2012, 4(4): 297. DOI:10.1561/20000000033

[10]ZHAO Chen, MA Siwei, ZHANG Jian. Video compressive sensing reconstruction via reweighted residual sparsity [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(6): 1182. DOI:10.1109/TCSVT.2016.2527181

[11]ZHOU Jian, LIU Hao. Video compressive sensing reconstruction via long-short-term double-pattern prediction [J]. Optoelectronics Letters, 2020, 16(3): 230. DOI:10.1007/s11801-020-9112-3

[12]CHEN Jian, CHEN Zhifeng, SU Kaixiong, et al. Video compressed sensing reconstruction based on structural group sparsity and successive approximation estimation model [J]. Journal of Visual Communication and Image Representation, 2020, 66: 102734. DOI:10.1016/j. jvcir. 2019. 102734

(编辑 苗秀芝)