

基于 MAP 的多信息流梯度更新与聚合 视频压缩感知重构算法

杨 鑫, 杨春玲

(华南理工大学电子与信息学院, 广东广州 510640)

摘 要: 现有优秀的基于深度学习的分布式视频压缩感知(Distributed Compressed Video Sensing, DCVS)重构算法利用测量值和参考帧顺序更新非关键帧, 获得了较好的重构性能, 但由于缺乏较严格的理论指导, 无法充分结合这两类信息, 限制了非关键帧重构质量的进一步提升. 针对该问题, 本文首先利用贝叶斯理论及最大后验概率(Maximum A Posteriori, MAP)估计推导出 DCVS 中非关键帧重构的优化方程, 再基于近端梯度算法推导出优化方程的求解框架, 包含多信息流梯度更新聚合方程. 基于此, 本文设计了多信息流梯度更新及聚合模块(Multi-Information flow Gradient update and Aggregation, MIGA), 并构建了深度多信息流梯度更新与聚合网络(Deep Multi-Information flow Gradient update and Aggregation Network, DMIGAN)用于 DCVS 非关键帧重构. MIGA 利用测量值与多参考帧对当前非关键帧进行并行梯度更新, 再做信息交互融合, 从而充分结合多种信息流更新重构帧. 本文级联 MIGA 与去噪子网络用于模拟近端梯度算法的迭代, 作为基础模块(phase), 并通过级联多个 phase 构造深度重构网络 DMIGAN, 实现帧重构的深度优化过程. 实验表明, DMIGAN 与具有代表性的传统迭代优化算法结构相似的帧间组稀疏表示重构算法(Structural SIMilarity based Inter-Frame Group Sparse Representation, SSIM-Inter F-GSR)相比, 在低采样率与高采样率下性能分别提升了 8.8 dB 和 7.36 dB; 和具有代表性的深度学习重构算法 VCSNet-2 相比, 在低采样率和高采样率下性能分别提升了 7.09 dB 和 8.78 dB.

关键词: 分布式视频压缩感知; 最大后验概率估计; 多信息流梯度更新; 信息聚合; 迭代优化; 光流估计

基金项目: 广东省自然科学基金(No.2019A1515011949)

中图分类号: TN919.8

文献标识码: A

文章编号: 0372-2112(2023)11-3320-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220958

MAP-Based Multi-Information Flow Gradient Update and Aggregation for Video Compressed Sensing Reconstruction

YANG Xin, YANG Chun-ling

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

Abstract: Due to the lack of guidance from the parallel update theoretical solver framework, existing deep learning-based distributed compressed video sensing (DCVS) algorithms alternately use measurement values and reference frames to optimize the reconstructed non-key frame, resulting in the inability to fully combine the two types of information and limiting the quality of reconstruction. In order to solve this problem, this paper firstly uses Bayesian theory and maximum a posteriori estimation (MAP) to derive the optimization equation of non-key frame reconstruction in DCVS, and then derives the solution framework of the optimization equation based on the proximal gradient algorithm, including multi-information flow gradient update and aggregation equation. Based on it, this paper designs a multi-information flow gradient update and aggregation neural network module (MIGA), and constructs a deep multi-information flow gradient update and aggregation network (DMIGAN) for DCVS non-key frame reconstruction. MIGA uses the measurement values and multiple reference frames to update the current non-key frame by parallel gradients, and then performs information interaction and fusion, so as to fully combine multiple information flows to update and reconstruct the frame. In this paper, the MIGA and the denoising sub-network are cascaded to simulate a single iteration of the proximal gradient algorithm as the basic phase. The deep

reconstruction network DMIGAN is constructed by cascading multiple phases to realize the deep optimization process of frame reconstruction. Experiments show that, compared with the representative traditional iterative optimization algorithm structural similarity based inter-frame group sparse representation (SSIM-InterF-GSR), the performance of DMIGAN is improved by 8.8 dB and 7.36 dB at low sampling rate and high sampling rate respectively; and compared with the representative deep learning reconstruction algorithm VCSNet-2, the performance is improved by 7.09 dB and 8.78 dB at low sampling rate and high sampling rate, respectively.

Key words: distributed compressed video sensing; maximum a posteriori estimation; multi-information flow gradient update; information aggregation; iterative optimization; optical flow estimation

Foundation Item(s): Natural Science Foundation of Guangdong Province (No.2019A1515011949)

1 引言

分布式视频压缩感知 (Distributed Compressed Video Sensing, DCVS)^[1]将压缩感知理论 (Compressed Sensing, CS)^[2]应用在视频信号采集与重构,以较低采样率(远低于 Nyquist 采样率)对各帧视频进行线性观测,在采样时完成压缩,在重构端利用复杂重构算法将压缩测量值恢复回原始帧.在 DCVS 中,整个视频被划分为多个图像组 (Groups Of Pictures, GOPs),每个图像组包括一个关键帧和 $T-1$ 个非关键帧, T 表示 GOP 大小.其中关键帧与非关键帧分别使用高采样率与低采样率进行线性采样.为了降低采样复杂度, Gan^[3]提出分块压缩感知 (Block Compressed Sensing, BCS) 替换全图采样,该方法成为 DCVS 领域的通用采样方法.与混合视频编码相比^[4-6],DCVS 编码端仅需很少的传感器资源、计算与存储资源,非常适合应用在一些编码端资源受限、采样代价昂贵的特殊场景,例如终端到云端的视频传输系统^[7]、无线视频传感网络^[8]等.然而,DCVS 重构属于高度不适定线性逆向问题,从线性压缩测量值恢复回高质量重构帧面临巨大挑战.

当前针对 DCVS 重构问题的主流算法可以分为两大类:一类为传统迭代优化算法^[9-12],另一类为深度学习重构算法^[13-19].传统迭代优化算法首先通过图像压缩感知重构算法进行帧内独立重构,然后利用高质量参考帧对非关键帧进行帧间联合重构.如何充分利用帧间相关性是 DCVS 重构的关键问题.其中文献[9]通过基于块匹配的运动估计/运动补偿 (Motion-Estimation and Motion-Compensation, MC/ME) 算法实现帧间图像块的运动估计,但是无法准确处理真实场景中光照变化、物体遮挡等情况.文献[10]提出多假设 (Multi-Hypothesis, MH) 算法使用多个假设块增强当前图像块质量,进一步改善帧间运动估计的准确性.基于 MH,文献[11]提出重加权残差稀疏重构算法 (Reweighted Residual Sparsity, RRS) 对多假设预测的残差进行加权稀疏建模,在高采样率下相比 MH,该方法可以进一步提升重构质量.受图像压缩感知重构算法组稀疏表示

(Group-based Sparse Representation, GSR) 的启发,文献[12]提出结构相似的帧间组稀疏表示重构算法 (Structural SIMilarity based Inter-Frame Group Sparse Representation, SSIM-Inter F-GSR),基于参考帧与当前帧的图像块组构造稀疏字典,实现了帧间稀疏建模,得到当前传统迭代优化算法最优重构结果.上述算法需要经过多次迭代,时间复杂度高,同时低采样率下重构质量差,限制了 DCVS 的应用前景.

近年来,随着深度学习研究的发展,其快速并行计算与数据驱动优化的双重优点引起了广泛关注,研究者将其应用在 DCVS,不仅提升了低采样率下的重构质量,而且实现了实时重构.根据帧间对齐方式的不同,深度学习重构算法又可细分为隐式对齐深度重构算法^[13-15]和显式对齐深度重构算法^[16-19].在隐式对齐深度重构算法中,文献[13]与文献[14, 15]分别基于长短期记忆模型 (Long Short-Term Memory, LSTM) 和卷积神经网络 (Convolutional Neural Network, CNN) 进行帧间特征融合,实现隐式对齐与运动补偿.这类算法将神经网络视作黑盒模型,缺乏模型可解释性.并且,这类方法没有充分利用各帧测量值导致重构质量不高.为了解决这些问题,结合深度学习与传统迭代优化算法,研究学者提出基于光流估计网络的显式对齐深度重构算法.文献[16]、文献[17]和文献[18]在显式对齐的基础上分别构造帧间组稀疏表示网络、自适应融合网络以及自适应掩码估计与多帧融合网络进行重构帧质量增强.在文献[16]的基础上,文献[19]提出递归结构网络改进重构顺序与参考帧选取,进一步提升了重构质量.这类算法基于传统迭代优化算法设计网络结构,通过交替利用测量值和参考帧信息顺序更新重构帧.这种串行信息利用方式的明显不足是,利用两种信息交替更新过程中产生的噪声可能会相互影响,从而影响最终重构性能.原因是,重构网络结构设计中直接借鉴传统的迭代优化思路,没有较严格的理论指导,所设计视频压缩感知重构网络不能充分结合观测值和参考帧信息.目前的视频压缩感知重构网络结构设计中都存在缺乏理

论指导的问题。

为了推进上述问题的解决,本文首先基于最大后验概率推导优化方程来建模 DCVS 帧重构问题,然后推导出求解该优化方程的多信息流梯度更新公式,最后设计了深度多信息流梯度更新与聚合网络用于 DCVS 重构. 本文创新点总结如下:

(1) 根据贝叶斯公式,本文首次推导了在给定测量值和多个参考帧条件下当前帧的最大后验概率(Maximum A Posteriori, MAP)估计,得到针对 DCVS 重构的优化方程.

(2) 基于近端梯度算法(Proximal Gradient algorithm, PG)推导出求解 DCVS 重构的迭代框架,每次迭代包含一次多信息流梯度更新聚合过程和一次近端映射过程. 该迭代框架为网络结构的设计提供了理论依据,从而增强模型可解释性,提升重构性能.

(3) 根据多信息流梯度更新聚合公式,设计了多信息流梯度更新聚合模块(Multi-Information flow Gradient update and Aggregation, MIGA). 该模块采用并行更新、交互融合的方式充分利用测量值与参考帧信息,与现有算法的串行交替利用相比,可以提升多种信息的交互程度.

(4) 根据推导的迭代框架设计了用于 DCVS 采样与重构的深度多信息流梯度更新与聚合网络(Deep Multi-Information flow Gradient update and Aggregation Network, DMIGAN),与现有算法相比,重构性能得到了大幅提升.

2 DCVS 重构方程的推导

在 DCVS 的编码端,各帧信号由一个采样矩阵进行帧内独立采样,采样过程定义为

$$\mathbf{y} = \Phi \mathbf{x} \quad (1)$$

式(1)中, $\mathbf{y} \in \mathbf{R}^M$ 为测量值, $\mathbf{x} \in \mathbf{R}^N$ 表示向量化原始帧信号, $\Phi \in \mathbf{R}^{M \times N}$ 为采样矩阵且 $M \ll N$. 帧内独立采样与重构等效于图像压缩感知与重构^[20-24],各帧信号在给定测量值 \mathbf{y} 的条件下,后验概率表示为 $p(\mathbf{x}|\mathbf{y})$. 由于视频各邻近帧存在高度相关性,时间维度邻近的高质量参考帧可以作为已知观测变量用于后验概率建模. DCVS 中各帧联合重构的后验概率重新定义为 $p(\mathbf{x}|\mathbf{y}, \{\mathbf{x}_i^{\text{ref}}\})$,其中 $\{\mathbf{x}_i^{\text{ref}}\}$ 表示参考帧集合, $\mathbf{x}_i^{\text{ref}}$ 为参考帧集合中的第 i 个参考帧. 由此,DCVS 的重构问题定义为最大后验估计问题:

$$\operatorname{argmax}_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}, \{\mathbf{x}_i^{\text{ref}}\}) \quad (2)$$

根据贝叶斯公式,后验概率可以转化为似然概率和先验概率的乘积,所以式(2)可以改写为

$$\begin{aligned} & \operatorname{argmax}_{\mathbf{x}} \log \left(\frac{p(\mathbf{y}, \{\mathbf{x}_i^{\text{ref}}\}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{y}, \{\mathbf{x}_i^{\text{ref}}\})} \right) \\ &= \operatorname{argmax}_{\mathbf{x}} \left(\log p(\mathbf{y}, \{\mathbf{x}_i^{\text{ref}}\}|\mathbf{x}) + \log p(\mathbf{x}) \right) \\ &= \operatorname{argmax}_{\mathbf{x}} \left(\begin{array}{l} \log p(\mathbf{y}|\mathbf{x}) + \\ \log p(\{\mathbf{x}_i^{\text{ref}}\}|\mathbf{y}, \mathbf{x}) + \log p(\mathbf{x}) \end{array} \right) \\ &= \operatorname{argmax}_{\mathbf{x}} \left(\begin{array}{l} \log p(\mathbf{y}|\mathbf{x}) + \\ \sum_i \log p(\mathbf{x}_i^{\text{ref}}|\mathbf{y}, \mathbf{x}) + \log p(\mathbf{x}) \end{array} \right) \end{aligned} \quad (3)$$

式(3)中的第一项 $\log p(\mathbf{y}|\mathbf{x})$ 表示对应测量值 \mathbf{y} 的数据保真项,通常由 l_2 范数定义,即 $\operatorname{argmax}_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x})$ 等效于 $\operatorname{argmin}_{\mathbf{x}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2$. 式(3)中的最后一项, $p(\mathbf{x})$ 定义了真实图像数据分布,在优化方程中 $\log p(\mathbf{x})$ 被转化为由真实图像属性确定的正则项,用 $R(\mathbf{x})$ 表示. 将式(3)转化为最小化最优方程的关键是如何准确定义第二项. 因为参考帧 $\mathbf{x}_i^{\text{ref}}$ 由对应的测量值 $\mathbf{y}_i^{\text{ref}}$ 确定,与当前重构帧的测量值 \mathbf{y} 无关,所以 $p(\{\mathbf{x}_i^{\text{ref}}\}|\mathbf{y}, \mathbf{x})$ 可以改写为 $p(\{\mathbf{x}_i^{\text{ref}}\}|\mathbf{x})$,表示相邻帧之间的像素相关性,可以通过光流对齐算子表达,即

$$\mathbf{x}_i^{\text{ref}} = f_i^{\text{warp}}(\mathbf{x}) + \boldsymbol{\eta} \quad (4)$$

式(4)中 f_i^{warp} 表示由光流确定的当前帧与第 i 个参考帧的线性对齐函数,实现当前重构帧到参考帧的对齐; $\boldsymbol{\eta}$ 表示对齐帧 $f_i^{\text{warp}}(\mathbf{x})$ 与参考帧 $\mathbf{x}_i^{\text{ref}}$ 之间的误差,服从均值为 0 方差为 δ^2 的高斯分布. 因此有

$$p(\mathbf{x}_i^{\text{ref}}|\mathbf{x}) \propto \exp(-\|\mathbf{x}_i^{\text{ref}} - f_i^{\text{warp}}(\mathbf{x})\|_2^2 / \delta^2) \quad (5)$$

根据式(5)可得以下等式成立:

$$\operatorname{argmax}_{\mathbf{x}} \log p(\mathbf{x}_i^{\text{ref}}|\mathbf{x}) = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{x}_i^{\text{ref}} - f_i^{\text{warp}}(\mathbf{x})\|_2^2 \quad (6)$$

其中, $\|\mathbf{x}_i^{\text{ref}} - f_i^{\text{warp}}(\mathbf{x})\|_2^2$ 命名为对应于参考帧 $\mathbf{x}_i^{\text{ref}}$ 的数据保真项.

基于以上分析,式(3)可以转化为式(7),即获得了 DCVS 重构问题的通用优化方程:

$$\operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{2} \sum_i \|\mathbf{x}_i^{\text{ref}} - f_i^{\text{warp}}(\mathbf{x})\|_2^2 + R(\mathbf{x}) \quad (7)$$

本文将使用 PG 算法求解上式优化问题,以下给出详细分析.

3 DCVS 重构问题的近端梯度求解

由于 Φ 与 f_i^{warp} 均为线性前向算子,所以式(7)中对应测量值与对应参考帧的保真项均为凸函数,则式(7)可采用 PG 算法求解,包含多次迭代. 每次迭代包括利用测量值 \mathbf{y} 的梯度更新、利用参考帧的梯度更新和一次近端映射,分别用于优化式(7)中的多个数据保真项和正则项. PG 算法中梯度更新由式(8)和式(9)所示,为

为了得到融合梯度更新,使更新值 $\mathbf{z}^{(k)}$ 可以包含测量值 \mathbf{y} 与参考帧 $\{\mathbf{x}_i^{\text{ref}}\}$ 之间的交互信息,每次迭代需要对多个梯度更新进行非线性聚合,如式(10)所示. 根据 PG 算法,近端映射表示为式(11).

$$\mathbf{z}_0^{(k)} = \mathbf{x}_0^{(k)} + \Phi^T(\mathbf{y} - \Phi \mathbf{x}_0^{(k)}) \quad (8)$$

$$\mathbf{z}_i^{(k)} = \mathbf{x}_i^{(k)} + (f_i^{\text{warp}})^{-1}(\mathbf{x}_i^{\text{ref}} - f_i^{\text{warp}}(\mathbf{x}_i^{(k)})) \quad (9)$$

$$\mathbf{z}^{(k)} = f_{\text{agga}}^{(k)}(\mathbf{z}_0^{(k)}, \{\mathbf{z}_i^{(k)}\}) \quad (10)$$

$$\mathbf{x}^{(k+1)} = \text{prox}_R = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2} \|\mathbf{z}^{(k)} - \mathbf{x}\|_2^2 + R(\mathbf{x}) \quad (11)$$

式(8)表示利用测量值的梯度更新,获得更新图像 $\mathbf{z}_0^{(k)}$; 式(9)表示利用第 i 个参考帧的梯度更新,获得更新图像 $\mathbf{z}_i^{(k)}$,索引值 $i \neq 0$. 基于光流的像素对齐过程等效于像素置换过程,因为置换矩阵的导数为其转置,定义了反向置换,所以 $f_i^{\text{warp}}(\mathbf{x})$ 的导数 $(f_i^{\text{warp}})^{-1}$ 表示参考帧到当前帧的对齐. 式(8)和式(9)并行利用测量值与多个参考帧,得到对应不同信息流的梯度更新图像. 式(10)中 $f_{\text{agga}}^{(k)}$ 表示第 k 次迭代的梯度更新聚合函数, $\mathbf{z}^{(k)}$ 为聚合梯度更新. 由此,式(8)、式(9)与式(10)实现了测量值与参考帧这两类信息的并行更新,交互利用. 根据式(3),式(11)中的 $R(\mathbf{x})$ 由数据自身分布决定. 由于式(8)和式(9)的梯度更新在降低测量误差和参考帧对齐误差

时会引入一定的噪声,近端映射用于将重构图像投影到更加符合视觉效果的流行空间,所以式(11)等效于图像去噪过程. 综上,本文得到用于求解 DCVS 重构问题的 MAP 估计框架及每次迭代的所有步骤,包含一次多信息流的梯度更新聚合过程和一次去噪过程. 在式(8)中的输入 $\mathbf{x}^{(0)}$ 为单通道图像,后续 $\mathbf{x}^{(k)}$ 具有 $1+N$ 个通道,其中 1 表示当前帧维度, N 表示 N 个参考帧的多组逆对齐更新图像. 同理, $\mathbf{z}^{(k)}$ 也具有 $1+N$ 个通道.

4 针对 DCVS 重构问题的深度网络实现

目前深度重构算法分开利用测量值和参考帧,不能充分结合这两类信息,导致重构质量不佳. 本文针对该问题,在上述推导的 MAP 迭代框架指导下,设计了深度多信息流梯度更新与聚合网络 DMIGAN,网络结构如图 1 所示. 该网络包含一个编码器和一个解码器. 编码器为一个 BCS 采样矩阵,解码器由一个初始重构子网络和一个深度重构子网络组成. 非关键帧首先通过编码器得到线性测量 \mathbf{y} ,然后由解码器的初始重构与深度重构子网络重构回高质量图像. 本文选择图像压缩感知网络 HSGANet^[25] 完成关键帧的帧内独立重构. HSGANet 在编码端通过混合 BCS 采样实现关键帧的线性观测,在解码端通过多个级联的梯度注意力网络和去噪子网络实现深度重构,得到高质量重构帧.

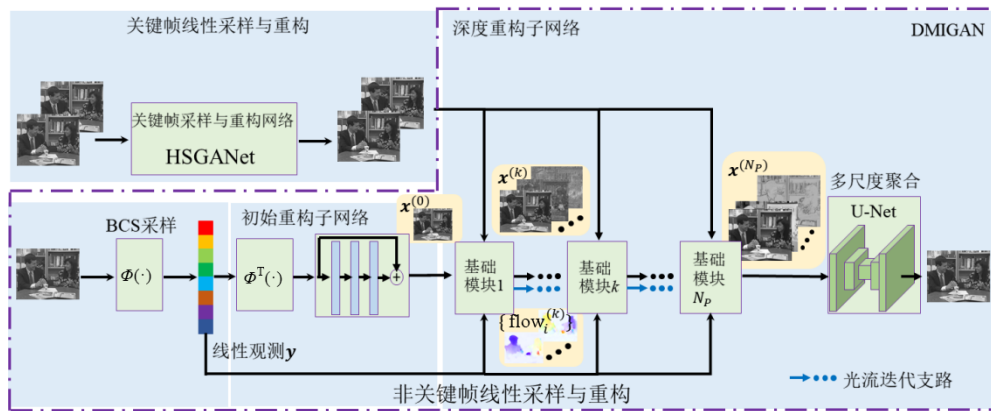


图 1 DMIGAN 结构框架

4.1 BCS 采样编码与初始重构解码

与现有 DCVS 重构网络相同^[14,15,17,18],DMIGAN 的编码器采用单个卷积层实现 BCS 采样. 该卷积层具有 m 个尺寸为 $B \times B \times c$ 的卷积核,其步长为 $B \times B$. 其中 B 表示采样块的大小, c 为图像通道数,采样率由 $m/(B^2c)$ 确定.

DMIGAN 通过采样矩阵的转置 Φ^T 叠加一个残差重构网络得到 MAP 迭代框架的初始重构 $\mathbf{x}^{(0)}$,作为公式(8)与公式(9)的初始值. Φ^T 用于将各个采样块由测量域投影到像素域,得到块内独立重构. 残差重构网络

用于融合各个独立重构的图像块得到全图初始重构. 整个初始重构可表示为

$$\mathbf{x}^{(0)} = f_{\text{ps}}(\{\Phi^T \mathbf{y}_j\}) + f_{\text{RC}}(f_{\text{ps}}(\{\Phi^T \mathbf{y}_j\})) \quad (12)$$

式(12)中, j 为采样块索引, f_{ps} 表示 PixelShuffle 层,用于将图像块拼接回全图, f_{RC} 表示一个全卷积神经网络用于残差重构,既可以去除块效应,又实现了各图像块之间的信息融合.

4.2 深度重构解码

深度重构子网络基于推导的 MAP 迭代框架搭建,

由多个级联的基础模块(phase)组成,用于模拟迭代过程,整体结构如图1所示. phase1的输入为初始重构子网络的输出 $\mathbf{x}^{(0)}$. 由式(8)到式(11)分析可知,每个phase包含一次多信息流梯度更新聚合过程和一个去噪过程. 本文提出多信息流梯度更新聚合模块MIGA实现式(8)到式(11)的多信息流梯度更新与聚合,使用一个残差块作为去噪模块实现去噪过程. MIGA、去噪模块的结构细节详见第4.3小节. 由于单个phase的输出包括对应测量值与参考帧集合的多个重构帧,需要对多个重构帧进行融合. 深度重构子网络使用一个U-Net进行多尺度聚合与信息交互,得到重构输出. U-Net作为多帧融合网络在DCVS^[17,18]与视频帧插值^[26-28]等任务中得到了广泛应用,本文将多尺度融合过程表示为式(13),其中 f_u 表示融合网络.

$$\hat{\mathbf{x}} = f_u(\mathbf{x}^{(N_p)}) \quad (13)$$

4.3 单个phase的结构细节

MIGA的整体结构如图2所示. 当前phase中MIGA的输入包括测量值 \mathbf{y} ,参考帧集合 $\{\mathbf{x}_i^{\text{ref}}\}$,前一个phase输出的多信息流重构帧和多组初始光流. 受目前光流网络^[29-31]采用的coarse-to-fine结构启发,DMIGAN通过phase-by-phase的方式进行逐phase光流估计,前一个phase估计的光流会传到下一个phase作为初始光流. 位于浅层的phase重构帧质量较低,对应光流粗估计;位于深层的phase重构帧质量更高,对应光流细估计. 其中第一个phase的初始光流为0,使用一个空间金字塔网络(Spatial Pyramid Network, SpyNet)^[29]输出多个参

考帧对应的多组光流. SPyNet通过从粗到细的图像金字塔结构来估计每个级别的光流残差,整个金字塔结构包括5个分辨率级别. 通过多级下采样得到的低分辨率图像可以增加各个像素的感受野,进而可以捕获更大范围的运动信息,提升光流估计网络对大范围运动信息的建模准确性. 为了减少DMIGAN模型参数,后续phase只采用SpyNet金字塔最底层对应的全卷积光流微调网络估计原图尺寸的光流残差. 该网络的输入包含当前重构帧、多个参考帧、多组初始光流以及多组初始对齐帧. 输出的光流残差会与初始光流相加得到更新光流. 由于光流法以两帧像素值不变为前提建立帧间一对一的像素对应关系,而现实场景中存在环境光线变化、物体遮挡、物体形变等问题,单个光流无法准确建模帧间相关性. 为了提升帧间相关性建模的准确性,MIGA中的光流估计网络会同时输出多个光流,从而扩展一个光流偏移到一组光流偏移集合. MIGA通过多个光流对齐使单个参考帧对应一组对齐帧,从而产生一组梯度. 因为DCVS重构属于高度不适定线性逆向问题,增加梯度更新的多样性可以提高重构解的准确性. 光流的粗估计的过程如式(14)所示,光流细估计的过程如式(15)所示. 其中 f_{spy} 表示SPyNet, f_{ft} 表示光流微调网络, f^{wrap} 表示使用 $\text{flow}^{(k)}$ 对当前重构帧进行对齐的函数.

$$\{\text{flow}_i^{(0)}\} = f_{\text{spy}}(\{\mathbf{x}_i^{\text{ref}}\}, \mathbf{x}^{(0)}) \quad (14)$$

$$\{\text{flow}_i^{(k+1)}\} = f_{\text{ft}}\left(\begin{matrix} \{\mathbf{x}_i^{\text{ref}}\}, \mathbf{x}^{(k)}, \{\text{flow}_i^{(k)}\}, \\ \{f_i^{\text{wrap}}(\mathbf{x}_i^{(k)})\} \end{matrix}\right) + \{\text{flow}_i^{(k)}\} \quad (15)$$

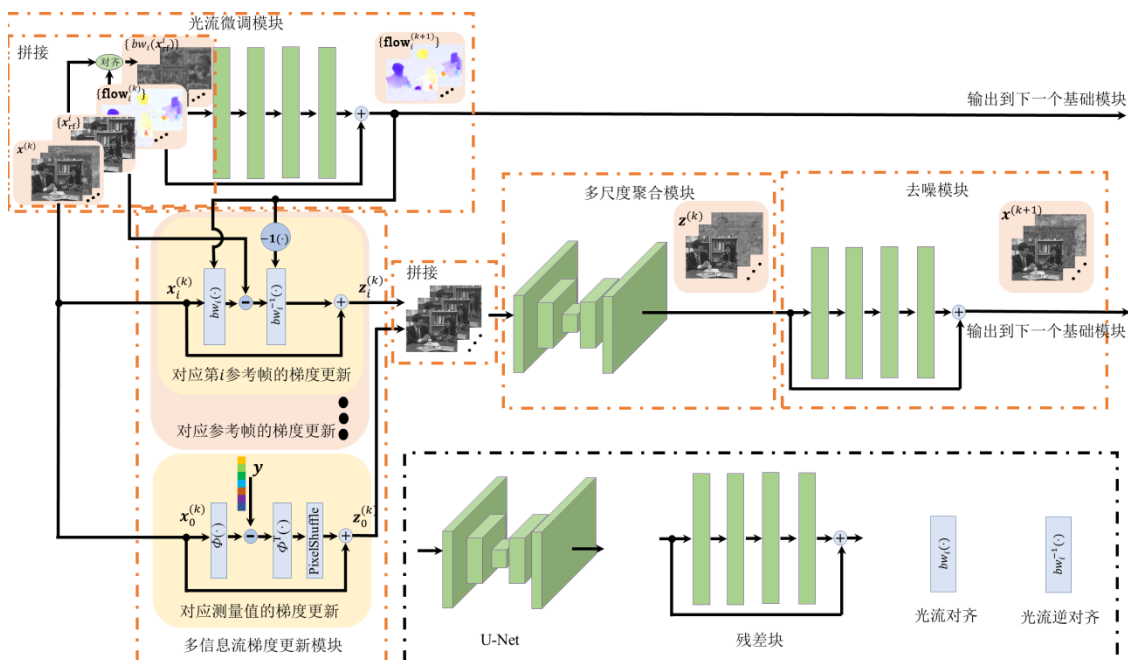


图2 单个phase结构图

MIGA 中对应测量值的梯度更新由式(8)定义,等效于利用测量值对当前帧进行残差重构.具体地,当前重构帧首先通过采样矩阵由像素域变换到测量域,得到重构帧在测量域的线性投影,并与测量值 y 作差求出测量误差.然后测量误差通过采样矩阵的转置变换回像素域,即为重构残差.由于编码端采用 BCS 采样,各残差图像块需由 PixelShuffle 层拼接回全图尺寸.最后当前重构与残差重构相加得到对应测量值的更新重构,与当前重构帧相比,其具有更小的测量误差.

MIGA 中对应参考帧的梯度更新由公式(9)定义,等效于利用参考帧对当前帧进行残差重构.在 DCVS 中,因为关键帧使用高采样率进行帧内独立重构,具有高重构质量,本文采用两个相邻关键帧作为当前非关键帧的参考帧.具体地,当前重构帧首先由更新光流对齐到参考帧并求出对齐误差,接着通过逆光流将对齐误差变换回重构残差.其中逆光流由更新光流乘以 -1 得到.最后当前重构与残差重构相加得到对应参考帧的更新重构,其具有更小的对齐误差.

上述多信息流梯度更新并行利用了测量值和参考帧.为了充分结合这两类信息,MIGA 采用一个 U-Net 对多个更新重构帧进行多尺度信息聚合,得到聚合梯度更新重构,该聚合过程对应公式(10).聚合梯度更新包含多个输出,对应不同信息流的重构帧.综上,MIGA 首先使用并行更新、互相融合的方式结合测量值与参考帧信息,与现有算法中对不同信息串行交替利用相比,提升了多种信息的交互程度,从而增强重构质量.其次,MIGA 根据推导的近端梯度求解公式构建,重构过程具有更强的理论可解释性.

根据式(11),在完成多信息流梯度更新与聚合后需要经过一个去噪模块对近端映射进行模拟.在图 2 中去噪模块通过一个残差块实现,由一个全卷积神经网络重构出当前输入的噪声抑制图,然后将其与输入相加得到去噪图像.去噪过程为

$$\mathbf{x}^{(k+1)} = f_d(\mathbf{z}^{(k)}) + \mathbf{z}^{(k)} \quad (16)$$

4.4 损失函数

DMIGAN 通过端到端学习完成所有网络参数的训练,包含 BCS 采样矩阵、初始重构子网络与深度重构子网络.本文采用均方误差(Mean Square Error, MSE)作为损失函数对网络参数进行优化.因为非关键帧重构能同时利用帧测量和参考帧信息,本文主要针对 DCVS 中的非关键帧重构.DMIGAN 的损失函数具体如下:

$$L(\Phi, \theta_{\text{ini}}, \theta_{\text{deep}}) = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 \quad (17)$$

其中, Φ 和 θ_{deep} 分别表示初始重构子网络与深度重构子网络参数, N 为 batchsize 的样本数, $\hat{\mathbf{x}}_i$ 和 \mathbf{x}_i 分别表示

DMIGAN 的最终重构帧和真实帧.

5 仿真结果及分析

为了公平对比,与现有 DCVS 重构网络相同^[13,15-19],本文使用 UCF-101 数据集^[32]对模型进行训练和测试,包括用于关键帧重构的 HSGANet 和本文提出的用于非关键帧重构的 DMIGAN. UCF-101 包含 101 个视频类别,总共 13 320 个视频序列.该数据集被随机划分为 80% 的训练集,10% 的验证集和 10% 的测试集.所有帧图像被中心裁剪为 160 × 160 分辨率,同时仅保留亮度分量.本文首先训练 HSGANet,接着训练 DMIGAN.训练 HSGANet 时,每个样本为各视频中随机选取帧.训练 DMIGAN 时,每个样本对包含三帧图像,分别为两个相邻关键帧和其间的任意非关键帧.其中两个关键帧使用训练好的 HSGANet 进行帧内重构,得到高质量参考帧.这种训练方式可以在有限计算资源下训练 DMIGAN,并适用于整个 GOP 的任意非关键帧重构.DMIGAN 中深度重构子网络 phase 数设为 13,本文选择 Adam^[33]作为模型的优化器, batchsize 设为 12,整个模型总共训练 390 个 epochs.初始学习率设为 1×10^{-4} ,在第 350 和 375 个 epoch 分别下降为 1×10^{-5} 和 1×10^{-6} .整体网络使用 PyTorch 深度学习框架搭建,并在单张 GeForce RTX 2080 Ti GPU 完成训练.本文选择峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structural SIMilarity, SSIM)^[34]两个指标来评估所有方法的重构性能.

5.1 与传统迭代优化算法结果对比及分析

本文采用 6 个 QCIF 格式(176 × 144)基准序列的前 96 帧与 6 个 CIF 格式(352 × 288)基准序列的前 16 帧测试 DMIGAN 与 4 个具有开源代码的代表性传统迭代优化方法(MC/ME, MH, RRS 和 SSIM-Inter F-GSR),其中 6 个 QCIF 序列分别为 foreman, hall, silent, ice, soccer 和 mobile, 6 个 CIF 序列分别为 coastguard, foreman, mobile, mother-daughter, paris, soccer.这些序列分别代表了不同类型的视频,包括慢速序列(hall, silent, mother-daughter)、背景静止的快速序列(foreman, ice, paris)以及背景运动的快速序列(soccer, mobile, coastguard).关键帧采样率为 0.5,非关键帧采样率为 0.1, 0.05, 0.01, GOP 大小为 8.不同采样率下各模型在 QCIF 序列上的测试结果如表 1 所示,在 CIF 序列上的测试结果如表 2 所示.

由表 1 与表 2 可知,针对 QCIF 格式视频,在非关键帧采样率为 0.1, 0.05, 0.01 条件下,6 个视频序列的平均重构 PSNR 分别提升了 7.36 dB, 7.79 dB, 8.80 dB;针对 CIF 格式视频,6 个视频序列的平均重构 PSNR 分别提升了 5.80 dB, 6.68 dB, 8.30 dB.这些数据表明,对于不

表1 各采样率下DMIGAN与各传统迭代优化算法在QCIF序列上重构结果对比(PSNR(dB)/SSIM), $r_k=0.5$

r_{nk}	算法	foreman	hall	silent	ice	soccer	mobile
0.1	MC/ME ^[9]	29.32/0.702 0	27.54/0.704 8	30.78/0.772 4	28.13/0.550 7	26.39/0.484 7	22.45/0.667 2
	MH ^[10]	34.32/0.946 6	31.99/0.948 4	34.03/0.934 0	30.20/0.926 8	28.42/0.832 6	24.52/0.821 5
	RRS ^[11]	35.17/0.946 0	28.53/0.918 5	33.95/0.925 6	31.09/0.939 6	29.74/0.828 6	20.00/0.552 6
	SSIM-Inter F-GSR ^[12]	35.77/0.953 7	34.44/0.969 2	35.11/0.940 1	31.75/0.953 9	30.25/0.869 1	24.35/0.833 9
	DMIGAN	41.78/0.985 0	42.28/0.988 0	42.23/0.986 2	39.64/0.986 3	36.64/0.947 4	33.31/0.974 5
0.05	MC/ME	27.17/0.628 6	25.63/0.633 6	28.25/0.710 3	25.70/0.437 7	24.19/0.357 6	21.34/0.604 7
	MH	32.31/0.924 2	31.11/0.939 6	33.08/0.921 8	28.57/0.903 8	26.84/0.776 7	23.03/0.761 1
	RRS	26.84/0.846 4	22.05/0.719 9	25.76/0.722 6	26.15/0.858 3	26.80/0.739 7	16.68/0.349 7
	SSIM-Inter F-GSR	33.22/0.933 0	33.21/0.962 1	33.65/0.921 8	28.85/0.922 1	27.49/0.792 9	22.99/0.779 1
	DMIGAN	39.84/0.979 2	40.95/0.985 6	40.91/0.981 8	37.75/0.981 1	34.13/0.918 3	32.56/0.971 4
0.01	MC/ME	16.27/0.259 4	15.33/0.238 4	18.62/0.352 7	17.78/0.168 2	17.16/0.160 2	16.41/0.326 2
	MH	23.61/0.696 7	23.54/0.791 6	25.73/0.740 2	21.87/0.736 3	21.27/0.527 7	16.38/0.303 6
	RRS	18.51/0.468 2	18.18/0.492 7	21.25/0.509 5	20.72/0.646 4	21.42/0.544 4	15.24/0.242 7
	SSIM-Inter F-GSR	26.96/0.837 2	28.52/0.928 7	24.66/0.782 4	24.60/0.824 2	23.43/0.624 9	21.94/0.718 8
	DMIGAN	35.18/0.949 6	38.14/0.979 5	37.52/0.966 3	32.65/0.950 6	28.86/0.798 4	30.58/0.959 9

注:粗体数据表示在各采样率(第1列)下,不同迭代优化算法(第2列)对6个QCIF序列(第1行)进行重构的最佳结果。

表2 各采样率下DMIGAN与各传统迭代优化算法在CIF序列上重构结果对比(PSNR(dB)/SSIM), $r_k=0.5$

r_{nk}	算法	coastguard	foreman	mobile	mother-daughter	paris	soccer
0.1	MC/ME ^[9]	27.10/0.700 3	31.21/0.871 6	20.85/0.626 2	37.55/0.946 7	24.79/0.803 9	26.81/0.715 6
	MH ^[10]	28.57/0.769 1	33.34/0.893 2	21.63/0.645 0	39.19/0.953 7	25.43/0.777 6	27.40/0.740 5
	RRS ^[11]	28.99/0.773 0	34.61/0.928 5	20.12/0.613 5	42.33/0.974 4	24.01/0.817 9	29.17/0.789 4
	SSIM-Inter F-GSR ^[12]	30.70/0.835 7	37.04/0.934 3	22.54/0.775 3	41.70/0.968 2	31.26/0.939 8	30.35/0.809 2
	DMIGAN	33.64/0.902 4	40.97/0.972 2	32.91/0.974 0	47.04/0.987 7	37.56/0.981 1	36.21/0.949 0
0.05	MC/ME	26.00/0.643 7	29.06/0.829 2	19.49/0.538 2	35.93/0.933 1	23.32/0.756 2	25.03/0.644 0
	MH	27.43/0.711 5	31.79/0.869 1	20.37/0.577 2	37.79/0.942 2	24.23/0.745 9	26.09/0.686 4
	RRS	24.70/0.589 2	30.65/0.892 1	16.62/0.372 7	38.14/0.942 9	18.84/0.546 2	26.18/0.681 0
	SSIM-Inter F-GSR	29.07/0.780 9	35.00/0.913 9	20.71/0.664 0	39.94/0.955 2	28.88/0.904 2	27.67/0.731 8
	DMIGAN	32.47/0.867 3	39.72/0.965 9	31.95/0.970 1	46.26/0.986 0	36.28/0.976 6	34.64/0.933 3
0.01	MC/ME	15.24/0.292 1	13.03/0.310 4	13.93/0.277 7	31.20/0.878 5	16.05/0.439 5	13.76/0.314 5
	MH	24.98/0.553 3	27.88/0.806 5	18.08/0.416 0	32.93/0.882 3	20.58/0.600 6	23.06/0.573 1
	RRS	22.03/0.428 3	20.39/0.643 3	15.16/0.285 9	25.33/0.716 0	16.99/0.410 6	21.51/0.516 1
	SSIM-Inter F-GSR	26.20/0.649 6	28.03/0.828 9	19.17/0.555 0	32.27/0.885 0	25.56/0.833 4	22.76/0.581 5
	DMIGAN	31.03/0.817 4	35.70/0.931 2	29.39/0.954 7	44.32/0.981 6	33.11/0.958 8	30.24/0.829 1

注:粗体数据表示在各采样率(第1列)下,不同迭代优化算法(第2列)对6个CIF序列(第1行)进行重构的最佳结果。

同分辨率与不同运动类型的视频序列,DMIGAN在不同采样率下均实现了显著的性能提升,尤其在极低采样率下(0.01采样率),提升效果更为明显。

图3展示了不同算法对CIF格式的mobile序列在关键帧采样率为0.5,非关键帧采样率为0.1的条件下,第13帧的视觉效果。针对运动的背景与前景,MC/ME与MH算法全图均存在明显的噪声,视觉质量极度模糊;RRS与SSIM-InterF-GSR算法对运动的背景丢失过多的信息,图中日历的文字与图案细节丢失尤为明显。本文提出的DMIGAN对运动的背景(画布与日历)、运动的前景(小球与火车)都实现了纹理清晰的高质量重

构,视觉效果提升明显。在低采样率下,传统迭代优化算法帧内重构严重依赖数据稀疏性,帧间重构依赖帧内重构质量,这两点原因导致传统迭代优化算法在低采样率下重构质量差,在极低采样率(0.01)条件下尤为明显。同时,交替利用测量值与参考帧信息产生的噪声会彼此影响,从而影响最终重构。DMIGAN根据本文推导的并行梯度更新聚合公式设计,可以同时测量值与参考帧等多种信息流进行优势互补利用,实现重构帧的迭代更新。并且,采用端到端训练的DMIGAN可以进行数据驱动的参数优化。上述特性使DMIGAN在不同采样率下,对不同类型的视频序列均实现高质量重构。

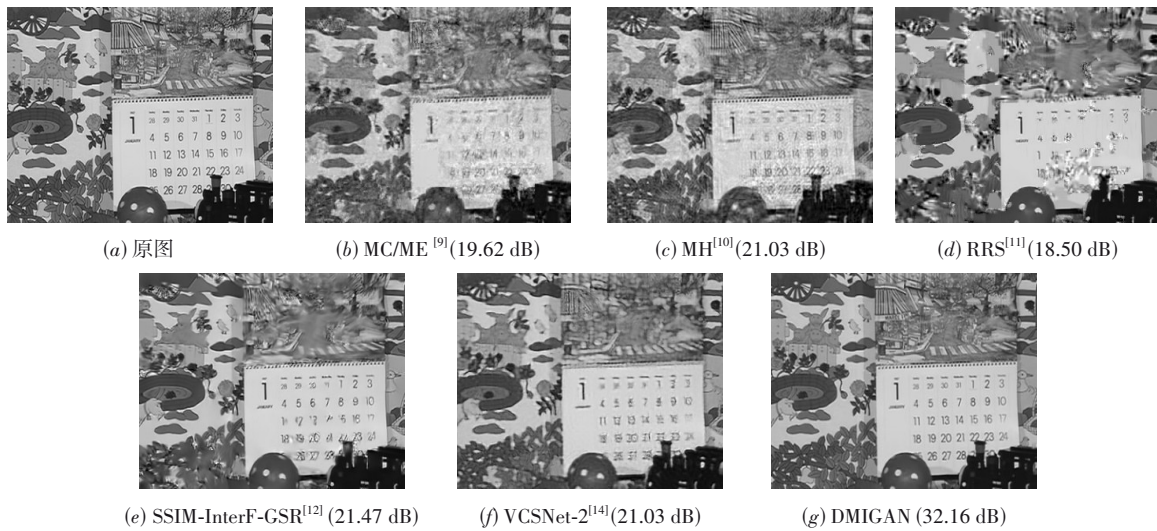


图3 CIF格式mobile序列第13帧重构效果图及对应的PSNR($r_k=0.5, r_{nk}=0.1, GOP=8$)

为了验证 DMIGAN 基于 GPU 并行计算的重构速度优势, 本文将 DMIGAN 与传统迭代优化算法运行时间进行对比. 表 3 显示了不同算法在关键帧采样率为 0.5 非关键帧采样率为 0.1 条件下对单帧 CIF 格式视频的平均重构时间、代码运行框架以及运行代

码的 CPU 或 GPU 的具体型号. 由表 3 可知, 传统迭代优化算法需要多次迭代完成单帧重构, 重构时间长, 无法实时应用. DMIGAN 通过 GPU 的强大并行计算能力使重构速度得到显著提升, 实现实时重构.

表 3 不同算法对单帧 CIF 格式视频的平均重构时间

方法	MC/ME ^[9]	MH ^[10]	RRS ^[11]	SSIM-InterF-GSR ^[12]	DMIGAN
重构时间/s	22.34	4.66	720.25	578.16	0.09
代码运行框架	Matlab 2016a	Matlab 2016a	Matlab 2016a	Matlab 2016a	PyTorch1.10
CPU/GPU	Inter Core i5-9400 CPU	Inter Core i5-9400 CPU	Inter Core i5-9400 CPU	Inter Core i5-9400 CPU	NVIDIA GeForce RTX 2080 Ti GPU

5.2 与深度学习重构算法结果对比及分析

本文采用 UCF-101 的测试集对 DMIGAN 与深度学习重构算法进行性能评估, 具体结果如表 4, 其中 CS-VideoNet, JsrNet 为隐式对齐深度重构算法, 2s-RER-VGSR-Net, STM-Net, ImrNet 为显式对齐深度重构算法. 所有方法采用相同的方式划分训练集与测试集, 对比方法的结果均取自文献. 与目前主流深度学习重构算法相同 (STM-Net, ImrNet, JsrNet, VCSNet-2), 本文在编码端使用了可学习的采样矩阵, 保证多数对比实验

的公平性. 由于 VCSNet-2 训练集与测试集不相同, 因此不参与比较. 由表 4 可知, DMIGAN 在不同 GOP 大小与不同采样率的条件下均取得了显著的性能提升, 具有很好的泛化性能. 具体地, DMIGAN 的平均 PSNR 分别比 CSVideoNet, 2s-RER-VGSR-Net, STM-Net, ImrNet, JsrNet 提高了 8.85 dB, 4.25 dB, 3.04 dB, 2.47 dB, 6.72 dB. 与上述方法相比, DMIGAN 在每个 phase 并行利用测量值与参考帧进行梯度更新, 通过多种信息约束重构解空间以提升解的准确性, 从而实现高质量的帧重构.

表 4 各采样率下不同算法在 UCF-101 测试集上重构 PSNR 对比

单位: dB

GOP 大小	关键帧采样率	非关键帧采样率	CSVideoNet ^[13]	2s-RER-VGSR-Net ^[19]	STM-Net ^[18]	ImrNet ^[17]	JsrNet ^[15]	DMIGAN
10	0.20	0.037	26.87	31.52	32.50	33.40	—	35.86
		0.018	25.09	29.87	31.14	31.90	—	34.23
		0.009	24.23	28.60	29.98	30.54	—	32.65
4	0.25	0.100	—	—	—	37.75	34.15	40.51
		0.040	—	—	—	36.14	31.99	38.82
		0.010	—	—	—	34.37	29.81	36.78

注: 粗体数据表示各采样率下不同算法在 UCF-101 测试集上重构指标 PSNR 的最优数据.

为了进一步验证 DMIGAN 的优越性, 本文将其与开源深度学习重构算法 VCSNet-2 进行深入对比. 表 5 与表 6 显示了两个算法分别在 QCIF 与 CIF 序列上不同采样率的性能表现. 针对 QCIF 格式视频, 在非关键帧采样率为 0.1, 0.05, 0.01 条件下, 6 个视频序列的平均重构 PSNR 与 SSIM 分别提升了 8.78 dB 和 0.078 1, 8.18 dB 和 0.100 6, 7.09 dB 和 0.140 8; 针对 CIF 格式视频, 6 个视频序列的平均重构 PSNR 与 SSIM 分别提升了 7.24 dB 和 0.120 5, 7.14 dB 和 0.150 4, 6.18 dB 和 0.177 4. 图 3 展示了两种方法的视觉重构效果对比, DMIGAN 对于运动的背景(画布与日历)与运动的前景(小球与火车)重构更加精确.

5.3 多信息流梯度更新聚合模块消融实验

多信息流梯度更新聚合模块包含利用测量值的梯度更新(记为 y_{gd})、利用参考帧的梯度更新(记为 r_{gd})与多尺度聚合模块(记为 $agga$), 分别对应式(8)~(10). 本文设计了对应的消融实验验证各个模块的有

效性, 具体结果如表 7, 其中最后一行是本文所提完整算法的重构性能, 其他各行是消融一种或两种技术的重构性能, 无 r_{gd} 情况下使用未对齐的参考帧替换梯度更新对齐帧. 由表 7 可知, 当同时去除 y_{gd} 与 r_{gd} 时, 重构质量下降明显, 原因是非关键帧重构属于高度线性欠定问题, 缺少已知信息对解空间的约束会极大降低解的准确性. 当仅保留 y_{gd} 或 r_{gd} 时会有 1 dB 左右的性能下降, 说明对应测量值与参考帧的并行梯度更新可以实现优势互补的作用, 通过交互融合的方式更加充分地利用测量值与参考帧信息. 多尺度聚合模块通过融合两类梯度更新可以进一步提升重构性能. 利用参考帧的梯度更新需要使用式(14)中的 SPyNet 和式(15)中的光流微调网络进行光流粗估计和细估计, 多尺度聚合模块使用 U-Net 实现. 所有网络均采用 3×3 尺寸的卷积核, 光流微调网络中各层的特征图数目分别为 $\{32, 64, 32, 16, 12\}$, SPyNet 由 5 个光流微调网络组成, U-Net 中各层的特征图数目分别为 $\{32, 32, 50, 50, 64, 64, 50, 50, 32, 1+12\}$.

表 5 各采样率下 DMIGAN 与 VCSNet-2 在 QCIF 序列上重构结果对比 (PSNR(dB)/SSIM), $r_k=0.5$

r_{nk}	算法	foreman	hall	silent	ice	soccer	mobile
0.1	VCSNet-2 ^[14]	31.14/0.918 5	32.06/0.959 3	34.92/0.955 1	30.95/0.939 4	30.51/0.846 6	23.62/0.779 8
	DMIGAN	41.78/0.985 0	42.28/0.988 0	42.23/0.986 2	39.64/0.986 3	36.64/0.947 4	33.31/0.974 5
0.05	VCSNet-2	29.75/0.883 4	31.90/0.959 0	34.52/0.947 0	29.51/0.915 0	28.62/0.774 7	22.79/0.734 9
	DMIGAN	39.84/0.979 2	40.95/0.985 6	40.91/0.981 8	37.75/0.981 1	34.13/0.918 3	32.56/0.971 4
0.01	VCSNet-2	26.07/0.770 2	30.57/0.949 8	31.94/0.921 2	25.77/0.843 0	24.62/0.631 3	21.42/0.644 1
	DMIGAN	35.18/0.949 6	38.14/0.979 5	37.52/0.966 3	32.65/0.950 6	28.86/0.798 4	30.58/0.959 9

注: 粗体数据表示各采样率下 DMIGAN 与 VCSNet-2 两种算法在 QCIF 序列上重构结果对比 (PSNR(dB)/SSIM) 的最优数据.

表 6 各采样率下 DMIGAN 与 VCSNet-2 在 CIF 序列上重构结果对比 (PSNR(dB)/SSIM), $r_k=0.5$

r_{nk}	算法	coastguard	foreman	mobile	mother-daughter	paris	soccer
0.1	VCSNet-2 ^[14]	29.60/0.765 0	33.52/0.911 3	22.39/0.723 3	41.67/0.969 5	27.87/0.896 5	29.87/0.778 1
	DMIGAN	33.64/0.902 4	40.97/0.972 2	32.91/0.974 0	47.04/0.987 7	37.56/0.981 1	36.21/0.949 0
0.05	VCSNet-2	28.48/0.700 2	32.06/0.884 2	20.99/0.638 1	41.04/0.967 2	27.35/0.885 7	28.60/0.721 9
	DMIGAN	32.47/0.867 3	39.72/0.965 9	31.95/0.970 1	46.26/0.986 0	36.28/0.976 6	34.64/0.933 3
0.01	VCSNet-2	27.26/0.643 3	28.13/0.803 1	19.49/0.518 3	40.08/0.964 3	26.03/0.853 0	25.77/0.626 0
	DMIGAN	31.03/0.817 4	35.70/0.931 2	29.39/0.954 7	44.32/0.981 6	33.11/0.958 8	30.24/0.829 1

注: 粗体数据表示各采样率下 DMIGAN 与 VCSNet-2 两种算法在 CIF 序列上重构结果对比 (PSNR(dB)/SSIM) 的最优数据.

表 7 多信息流梯度更新聚合模块消融对比实验

单位: dB

y_{gd}	r_{gd}	agga	QCIF				CIF			
			foreman	mobile	soccer	average	foreman	mobile	soccer	average
无	无	有	27.02	23.11	21.59	23.91	26.96	20.82	22.09	23.29
有	无	有	40.93	32.30	35.89	36.37	40.25	31.38	34.88	35.50
无	有	有	40.31	32.29	35.26	35.95	40.15	31.89	35.25	35.76
有	有	无	41.51	33.03	36.22	36.92	40.60	32.60	35.65	36.28
有	有	有	41.78	33.31	36.64	37.24	40.97	32.91	36.21	36.70

注: 加粗数据是本文所提完整算法的重构性能数据, 其中保留了所有三种多信息流梯度更新聚合模块.

6 结论

本文根据最大后验概率估计推导出了针对视频压缩感知的优化方程,并基于近端梯度算法推导出迭代求解框架.基于此,本文进一步设计了深度多信息流梯度更新与聚合网络(DMIGAN)用于实现该求解框架.DMIGAN通过级联多个 phase 模拟了迭代求解过程,对测量值与参考帧信息进行充分利用.其中每个 phase 包含一个提出的多信息流梯度更新聚合模块(MIGA)与去噪模块,MIGA 可以并行优化测量域误差与多参考帧对齐误差,实现对两类信息的充分结合与互补利用,进而实现高质量帧重构.大量实验证明,与传统迭代优化算法相比,MIGA 极大地提高了重构质量,并且实现了实时重构.与目前主流的串行交替利用两类信息的深度学习重构算法相比,本文提出的 DMIGAN 得到非常显著的性能提升,在低采样率条件下可以准确重构纹理细节.

参考文献

- [1] KANG L W, LU C S. Distributed compressive video sensing[C]//2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2009: 1169-1172.
- [2] DONOHO D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.
- [3] GAN L. Block compressed sensing of natural images[C]//2007 15th International Conference on Digital Signal Processing. Piscataway: IEEE, 2007: 403-406.
- [4] WIEGAND T, SULLIVAN G J, BJONTEGAARD G, et al. Overview of the H. 264/AVC video coding standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 560-576.
- [5] SCHWARZ H, MARPE D, WIEGAND T. Overview of the scalable video coding extension of the H.264/AVC standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(9): 1103-1120.
- [6] PAN Z Q, LEI J J, ZHANG Y, et al. Fast motion estimation based on content property for low-complexity H.265/HEVC encoder[J]. IEEE Transactions on Broadcasting, 2016, 62(3): 675-684.
- [7] ZHENG S, ZHANG X P, CHEN J, et al. A high-efficiency compressed sensing-based terminal-to-cloud video transmission system[J]. IEEE Transactions on Multimedia, 2019, 21(8): 1905-1920.
- [8] XIAO D, LI M, WANG M D, et al. Low-cost and high-efficiency privacy-protection scheme for distributed compressive video sensing in wireless multimedia sensor networks[J]. Journal of Network and Computer Applications, 2020, 161: 102654.
- [9] MUN S, FOWLER J E. Residual reconstruction for block-based compressed sensing of video[C]//2011 Data Compression Conference. Piscataway: IEEE, 2011: 183-192.
- [10] TRAMEL E W, FOWLER J E. Video compressed sensing with multihypothesis[C]//2011 Data Compression Conference. Piscataway: IEEE, 2011: 193-202.
- [11] ZHAO C, MA S W, ZHANG J, et al. Video compressive sensing reconstruction via reweighted residual sparsity[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(6): 1182-1195.
- [12] 和志杰, 杨春玲, 汤瑞东. 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[J]. 电子学报, 2018, 46(3): 544-553
HE Z J, YANG C L, TANG R D. Research on structural similarity based inter-frame group sparse representation for compressed video sensing[J]. Acta Electronica Sinica, 2018, 46(3): 544-553 (in Chinese)
- [13] XU K, REN F B. CSVideoNet: A real-time end-to-end learning framework for high-frame-rate video compressive sensing[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2018: 1680-1688.
- [14] SHI W Z, LIU S H, JIANG F, et al. Video compressed sensing using a convolutional neural network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(2): 425-438.
- [15] CHEN C, WU Y T, ZHOU C, et al. JsrNet: A joint sampling-reconstruction framework for distributed compressive video sensing[J]. Sensors, 2019, 20(1): 206.
- [16] XUAN Y Y, YANG C L. 2Ser-vgsr-net: A two-stage enhancement reconstruction based on video group sparse representation network for compressed video sensing[C]//2020 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2020: 1-6.
- [17] YANG X, YANG C L. Imrnet: An iterative motion compensation and residual reconstruction network for video compressed sensing[C]//2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2021: 2350-2354.
- [18] WEI Z C, YANG C L, XUAN Y Y. Efficient video compressed sensing reconstruction via exploiting spatial-temporal correlation with measurement constraint[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE, 2021: 1-6.

- [19] 禚韵怡, 杨春玲. 基于帧间组稀疏的两阶段递归增强视频压缩感知重构网络[J]. 电子学报, 2021, 49(3): 435-442
XUAN Y Y, YANG C L. Two-stage recursive enhancement reconstruction based on video inter-frame group sparse representation in compressed video sensing[J]. Acta Electronica Sinica, 2021, 49(3): 435-442 (in Chinese)
- [20] MEINHARDT T, MOELLER M, HAZIRBAS C, et al. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 1799-1808.
- [21] LUO G X, ZHAO N, JIANG W H, et al. MRI reconstruction using deep Bayesian estimation[J]. Magnetic Resonance in Medicine, 2020, 84(4): 2246-2261.
- [22] FAN X H, YANG Y, ZHANG J P. Deep geometric distillation network for compressive sensing MRI[C]//2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). Piscataway: IEEE, 2021: 1-4.
- [23] FAN X H, YANG Y, CHEN K, et al. A unifying multi-sampling-ratio CS-MRI framework with two-grid-cycle correction and geometric prior distillation[EB/OL]. [2022-05-14]. <https://arxiv.org/abs/2205.07062>.
- [24] MOU C, WANG Q, ZHANG J. Deep generalized unfolding networks for image restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 17399-17410.
- [25] YANG X, YANG C L, CHEN W J. A hybrid sampling and gradient attention network for compressed image sensing[J]. The Visual Computer, 2022: 1-14.
- [26] JIANG H Z, SUN D Q, JAMPANI V, et al. Super SloMo: High quality estimation of multiple intermediate frames for video interpolation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 9000-9008.
- [27] BAO W B, LAI W S, MA C, et al. Depth-aware video frame interpolation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 3698-3707.
- [28] BAO W B, LAI W S, ZHANG X Y, et al. MEMC-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(3): 933-948.
- [29] RANJAN A, BLACK M J. Optical flow estimation using a spatial pyramid network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2720-2729.
- [30] SUN D Q, YANG X D, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8934-8943.
- [31] HUI T W, TANG X O, LOY C C. A lightweight optical flow CNN-revisiting data fidelity and regularization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2555-2569.
- [32] SOOMRO K, ZAMIR A R, SHAH M. UCF101: A dataset of 101 human actions classes from videos in the wild[EB/OL]. (2012-12-03)[2022-08-13]. <https://arxiv.org/abs/1212.0402>.
- [33] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22) [2022-08-13]. <https://arxiv.org/abs/1412.6980>.
- [34] 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: A Publication of the IEEE Signal Processing Society, 2004, 13(4): 600-612.

作者简介



杨鑫 男, 1991年出生于福建连城. 华南理工大学电子与信息学院博士生. 主要研究方向为图像/视频压缩感知.
E-mail: eexyang@mail.scut.edu.cn



杨春玲(通讯作者) 女, 1970年出生于河南新乡. 华南理工大学电子与信息学院博士生导师. 主要研究方向为图像/视频压缩编码、图像质量评价.
E-mail: eeclyang@scut.edu.cn