Progressive Dictionary Learning with Hierarchical Structure for Scalable Video Coding

Xin Tang, Wenrui Dai, Hongkai Xiong

Department of Electronic Engineering, Shanghai Jiao Tong Univ., Shanghai, 200240, China,

Email: {xint14, daiwenrui, xionghongkai}@sjtu.edu.cn

To enable learning-based video coding for transmission over heterogenous networks, this paper proposes a scalable video coding framework by progressive dictionary learning. With the hierarchical B-picture prediction structure, the inter-predicted frames would be reconstructed in terms of the spatio-temporal dictionary in a successive sense. Within the progressive dictionary learning, the training set is enriched with the samples from the reconstructed frames in the coarse layer. Through minimizing the expected cost, the stochastic gradient descent is leveraged to update the dictionary for practical coding. It is demonstrated that the learning-based scalable framework can effectively guarantee the consistency of motion trajectory with the well-designed spatio-temporal dictionary.

Here, scalability is enabled by restricting motion-compensated prediction to reference frames with a temporal layer identifier of the frame to be predicted. With hierarchical B-frame, dyadic enhancement layer frames are coded as B pictures which are predicted by the previous referenced frames and regarded as the referenced frames for the next enhancement layer. The temporal layers are identified by T that starts from 0 for the base layer and increase by 1 from one temporal layer to the next. The enhancement layer pictures shall be decoded as B-pictures. Each set of layers $T_0,\dots,T_k$ can be decoded independently of all layers with identifier $T > k$. Since the training examples are extracted as "cube" by concatenating the patches from both two I-frames in the same location of successive I-frames in one GOP to exploit the motion within sparse representation, a couple of successive frames can be always predicted at once. In each layer T_k, the low-resolution frames $\hat{Z}_l$ are reconstructed by super-resolution in terms of a learned dictionary.

Taking advantage of the property that online gradient descent operates without reference to the training set, we can directly use the examples observed in the current layer. For layer T_1 to target the enhancement layer T_K, the initial dictionary for layer T_k is $\mathbf{D}^{k-1}$ and the current dictionary is updated based on the input sample and previous dictionary. Instead of averaging the gradient of the loss over the complete training batch, each iteration of the online gradient descent consists of choosing an example $\mathbf{x}_t^k$ at the distribution $p(\mathbf{x})$, here we consider it as uniform distribution, and updating the dictionary $\mathbf{D}_L^k$ as Eq. (1).

$$\mathbf{D}_L^k = \mathbf{D}_L^{k-1} - \gamma_t \nabla_{\mathbf{D}} l(\mathbf{D}_L^{k-1}, \mathbf{x}_t^{k-1}) = \mathbf{D}_L^{k-1} - \gamma_t \nabla_{\mathbf{D}} \left(\frac{1}{2}\|\mathbf{x}_i^{k-1} - \mathbf{D}_L^{k-1}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1\right) \tag{1}$$

The initial dictionary of the progressive dictionary learning for the process of one GOP is trained by high-resolution frames. The sparse coding progress is an ℓ_1-regularized least-squares problem, which can be solved by Cholesky decomposition for the classical LARS-Lasso algorithm. The observed training set can be split as a mini-batch to speed up the convergence speed. The dictionary $\mathbf{D}_t$ is updated by block-coordinate descent with warm restarts, which minimizes the expected cost without tuning learning rate.

Experiments on CIF(352×288) sequences show that the progressive dictionary learning based coding scheme could achieve better performance (PSNR) than H.264/AVC both in sequential coding and hierarchical coding. Moreover, it is witnessed that the proposed framework obtains a significant PSNR gain over H.264 on the rate-distortion performance, specially in low bit-rate range.

The work was supported in part by the NSFC under Grants 61425011, 61271218, and U1201255, and in part by the Shu Guang project under Grants 13SG13.

1068-0314/15 $31.00 © 2015 IEEE
DOI 10.1109/DCC.2015.28