STOL: Spatio-Temporal Online Dictionary Learning for low bit-rate video coding

Xin Tang, Hongkai Xiong

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

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

To speed up the convergence rate of learning dictionary, this paper proposes a spatio-temporal online dictionary learning (STOL) algorithm to improve the original adaptive regularized dictionary learning with K-SVD. Experiments show the super-resolution reconstruction based on STOL obviously reduces the computational complexity to 40% to 50% of the K-SVD learning-based schemes with a guaranteed accuracy.

A video sequence is decomposed into a selected high-resolution (HR) key frames (**KF**) X_h and the down-sampled low-resolution (LR) non-key frames (**NKF**) Z_l from Z_h. The high-resolution version $\hat{Z}_h$ would be recovered from $\hat{Z}_l$ by the learning-based super-resolution reconstruction via sparse representation. In training each series of 2-D subdictionaries, the primitives is of low dimensionality. The non-primitive volumes are supposed to be consistent along the motion trajectory with little structure. Correspondingly, their sparse representations over a learned 3-D spatio-temporal dictionary are designed by the online dictionary learning[1] to update the atoms for optimal sparse representation and convergence. Instead of classical first-order stochastic gradient descent on the constraint set, the online algorithm would exploit the structure of sparse coding in the design of an optimization procedure in terms of stochastic approximations. Through drawing a cubic from i.i.d. samples of a distribution in each inner loop and alternating classical sparse coding steps for the decomposition coefficient of the cubic over the previous dictionary, the dictionary update problem is converted to solve the *expected* cost instead of the *empirical* cost. It has been shown that stochastic gradient descent algorithm in online learning is more attractive than standard primal or dual algorithms. For dynamic training data over time, online dictionary learning behaves faster than second-order iteration batch alternatives.

Assume that each patch can be represented as a linear combination of a small subset of patches. Taking temporal redundancy into account, the super-resolution task is defined as an energy minimization as:

$$f^r_{Video}(\{\alpha_{ij}\}_{ij}, X^r_h, F^r_h) = \arg\min_{X_h,\{\alpha_{ij}\}} \{\frac{1}{2}\sum_{i,j}\|Z_l - T_l^r\alpha_{ij}\|_2^2 + \sum_{i,j}\lambda\|\alpha_{ij}\|_0 + \sum_{i,j}\|T_h^r\alpha_{ijt} - R_{ijr}X_h\|_2^2\} \quad (1)$$

where Z_l is the burred and down-sampled version of the high-resolution X_h, T_L is the low-frequency subdictionaries, α_{ij} denotes the sparse solution of T_l under dictionary F_L, and R_{ij} is a projection matrix that selects the $(i,j)_{th}$ patch from X_h. Considering the K-SVD algorithm involves heavy computational burden, an online dictionary learning algorithm is adapted instead.

Experiments of 3-D dictionary learning by K-SVD and STOL on standard video datasets show that the number of STOL iteration would be several times larger than K-SVD within the same duration. It means that the values of empirical and expected cost function are different, and it is obvious that the convergence speed of STOL is significantly faster than K-SVD. For video coding, the proposed STOL algorithm could achieve better performance (PSNR and SSIM) than H. 264/AVC and comparable quality as the K-SVD based learning scheme.

Reference

[1] J. Mairal, F. Bach, J. Ponce, G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," to appear in *Journal of Machine Learning Research*.

The work has been partially supported by the NSFC, under Grants U1201255, 61271218, 61271211, and 61228101.

1068-0314/13 $26.00 © 2013 IEEE
DOI 10.1109/DCC.2013.101