Supplementary Materials: Spatial-Temporal Weighted Pyramid using Spatial Orthogonal Pooling

Yusuke Mukuta[1], Yoshitaka Ushiku[1], and Tatsuya Harada[1,2]

[1]The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, Japan
[2]RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi Chuo-ku, Tokyo, Japan
{mukuta,ushiku,harada}@mi.t.u-tokyo.ac.jp

A. Comparison to other pooling methods

In this section, we apply the proposed pooling method to with different feature vectors to compare with existing feature pooling methods.

A.1. Application to fisher vector with different parameter

We applied the proposed method on Fisher Vector with the dimensionality of the SIFT reduced to 64 and the number of components 256 on CUB dataset [5] to compare with Generalized Max Pooling (GMP) [3]. The proposed method with kernel parameter a = 0.25 showed 19.47, 19.54, and 22.94 with degree 0, 1, and 2 while GMP showed 17.0. Thus, the proposed method showed better performance with the same pyramid size and we can get higher performance by choosing a larger pyramid size.

A.2. Application to Locality-constrained Linear Coding

Next, we applied the proposed method on locality-constrained linear coding (LLC) [4] on Caltech101 dataset [1]. Caltech-101 dataset consists of 9,144 images with 101 object categories and background class. We randomly sampled 15 training samples per class as training data and the rest as test data. We extracted SIFT features and then encode them to LLC feature with codebook size 2,048 to compare with original LLC and Geometric L_p norm Feature Pooling (GLFP) [2]. We observed that the proposed method with kernel parameter a = 0.25 showed 57.96, 63.50, 69.42, 70.64, 71.68, and 72.04 with degree 0, 1, 2, 3, 4, 5 respectively, while original LLC with pyramid scales $[1 \times 1, 2 \times 2, 4 \times 4]$ showed 65.43 and GLFP with pyramid scales $[1 \times 1, 2 \times 2, 4 \times 4]$ showed 70.34. Thus, we can get comparable performance to GLFP with about half the pyramid size.

References

[1] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *CVIU*, 106(1):59–70, 2007.

[2] J. Feng, B. Ni, Q. Tian, and S. Yan. Geometric lp-norm feature pooling for image classification. In *CVPR*, 2011.

[3] N. Murray and F. Perronnin. Generalized max pooling. In *CVPR*, 2014.

[4] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2010.

[5] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-ucsd birds 200. Technical Report CNS-TR-201, Caltech, 2010.