

Optimizing Binary Fisher Codes for Visual Search

Zhe Wang, Ling-Yu Duan, Jie Lin, Jie Chen, Tiejun Huang, and Wen Gao

The Institute of Digital Media, Peking University, Beijing, China

{zhew, lingyu, linjie, cjie, tjhuang, wgao}@pku.edu.cn

Fisher vectors (FV), a global representation obtained by aggregating local invariant features (e.g., SIFT), generates the state-of-the-art descriptor for visual search, due to highly discriminative power and small visual vocabulary. Nevertheless, a high-dimensional raw FV can be further compressed to reduce feature storage and improve search efficiency. In this paper, we formulate the FV compression as a resource-constrained optimization problem. Let $A(\cdot)$ denote search accuracy, $R(\cdot)$ descriptor size, and $C(\cdot)$ compression complexity. Our goal is to design an optimal quantizer $q(\cdot)$ to compress Fisher vector g by maximizing search accuracy $A(q(g))$ subject to the constraints of descriptor length R_{budget} , and computational complexity C_{budget} :

$$\max_q A(q(g)) \quad s.t. \quad R(q) \leq R_{budget} \quad \text{and} \quad C(q) \leq C_{budget}. \quad (1)$$

Accordingly, we present selective binary Fisher codes (SBFC) to compress the raw FV. Given a raw Fisher vector $[X_1, X_2, \dots, X_M]$ with M Gaussian functions, where $X_i (1 \leq i \leq M)$ denotes the i -th sub-vector. Firstly, to fulfill the constraint of compression complexity, we binarize the FV by a sign function; accordingly, we get a binarized Fisher code (BFC) $B = [B_1, B_2, \dots, B_m]$, where each binary sub-vector code $B_i = \text{sgn}(X_i)$. Secondly, we propose to select discriminative bits from the binarized FV (BFC) to maximize search performance, subject to the constraint of descriptor length. We introduce two measurements of “local certainty” and “global informativity” to filter in discriminative bits towards high performance and low complexity, as well as sufficient descriptor compactness. In our work, “local certainty” is defined as the variance of each sub-vector X_i of the raw FV, and “global informativity” the bit-wise entropy of each dimension in BFC derived from an independent set of training images. By “local certainty”, we filter in part of binary sub-vector codes $B_{i_j} (1 \leq j \leq M', M' \leq M)$ from BFC B , while by “global informativity”, we further select a subset of bits with high entropy from each binary sub-vector B_{i_j} .

Extensive experiments over the MPEG Compact Descriptor for Visual Search (CDVS) benchmark datasets have shown that SBFC can achieve a high compression ratio of 128:1 at extremely low complexity of 0.015 MB memory usage and 1 ms (tested on Intel *i5* – 3470) with a minor mAP drop of less than 1%. Compared with typical compression schemes such as Product Quantization(PQ) and Hashing algorithms, SBFC may incur much less memory and time cost. Note that PQ often requires tens of sub-codebooks while Hashing algorithms involve thousands of hash functions. In particular, a simplified version of SBFC (by omitting “global informativity”), SCFV, has been adopted by the MPEG CDVS standard. In the CDVS evaluation framework, SCFV has achieved promising performance with the mean Average Precision (mAP) 85% and the success rate of Top Match 91% on average at the memory cost of 40KB.

Acknowledgements: This work was supported by the National Natural Science Foundation of China under grants 61271311, 61390515, 61421062.