Query-Adaptive Late Fusion for Image Search and Person Re-identification

Liang Zheng[1], Shengjin Wang[1], Lu Tian[1], Fei He[1], Ziqiong Liu[1], Qi Tian[2]
[1]Department of Electronic Engineering, Tsinghua University. [2]University of Texas at San Antonio.

Feature fusion has been proven effective in image search and related tasks. Typically, it is assumed that for a given query, a to-be-fused feature works well when used alone and is complementary to existing features. Then, it is expected that a higher search accuracy can be achieved. However, in realistic settings, the problem is that one does not necessarily know in advance whether a heterogeneous feature is good for the query.

Failure in predicting feature effectiveness might result in undesirable search quality. On one hand, the failure of identifying good features may under-utilize features' discriminative power. On the other hand, bad features that escape unpunished may lead to worse consequences: accuracy gets even lower after fusion. This problem is not trivial: some state-of-the-art fusion methods [4, 5], as will be shown, suffer from the fusion of black sheep features. As a result, it is of great importance to identify feature effectiveness in a query-adaptive manner.

Towards this goal, this paper suggests that three guidelines be paid special attention to on effective feature fusion for image search.

- **Query-adaptive.** Given a query image, the effectiveness of a to-be-fused feature should be automatically evaluated, so that good features are used, while bad features are ignored.
- **Unsupervised.** Since we consider generic image search, in which no prior knowledge on the topic of the query image is provided, it is important that we estimate the effectiveness of a feature through unlabeled data.
- **Database independent.** We should keep in mind that the test database keeps growing. Then, it is required that the offline steps be independent on it, so that the fusion scheme is amenable for database update.

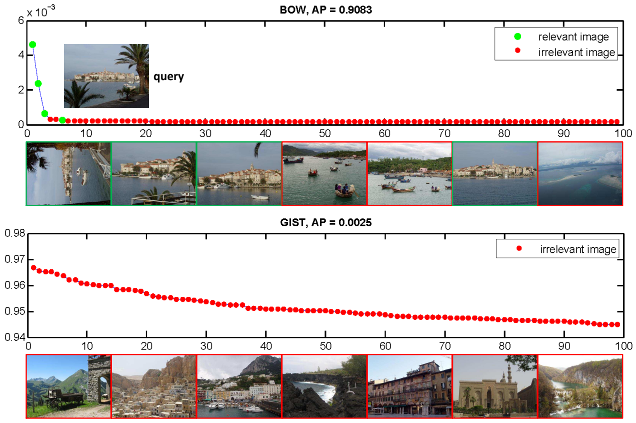

Figure 1: Example of a two-feature system. For a query in the Holidays [1] dataset, the BoW (upper) and GIST (bottom) features are employed to obtain two score lists respectively. There are four relevant images for this query, where BoW produces good performance (AP = 90.83%), but GIST fails (AP = 0.25%). We plot the sorted scores for rank 1-99, and the corresponding 7 top-ranked images. Relevant images are in marked in green, and irrelevant ones red. Note that the sorted score curve is L-shaped for BoW, but gradually descending for GIST.

In light of the above discussions, a query-adaptive late fusion scheme is introduced. Our motivation is illustrated in Fig. 1. We find that the sorted score curve of a good feature is "L" shaped, while that of a bad feature is gradually dropping. In our method, the score curves are firstly normalized by reference curves trained on irrelevant data, which are expected to approximate the tails of the initial score curves. Then, feature effectiveness is estimated as negatively related to the area under the normalized score curve. In our method, the offline operation is independent on the test database, making it well suited to dynamic systems. More importantly, our method identifies "good" and "bad" features on-the-fly, and the results are competitive to the state-of-the-arts on three datasets.

Key Results. On Holidays dataset, apart from the BoW feature, we inject a number of distractor features. Under such circumstance, it is desirable that fusion result not be influenced too much. In our experiment, 20 random projection matrices are generated, so that we are provided with 20 random projection features [3].

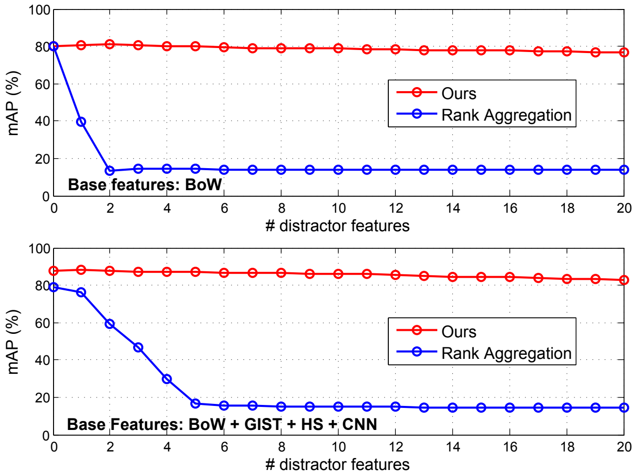

Figure 2: Impact of bad features on Holidays dataset. We plot mAP against a increasing number of random features. **Top:** random features are fused with BoW. **Bottom:** BoW + GIST + HS + CNN is used as baseline. We compare with Rank Aggregation [2].

We evaluate this property in Fig. 2. We compare our method with Rank Aggregation (RA) [2]. In RA, we compute the median rank of each candidate image over all rank lists obtained by different features. We can see that when the number of random features increases, mAP of our method drops very slowly, but that of RA decreases dramatically. When as many as 20 "bad" features are used, mAP of our method drops from 80.16% to 76.58%, and from 87.98% to 82.91% for the two base-feature settings, respectively. In comparison, RA yields an mAP of 13.85% and 14.29%, respectively. Therefore, our method is very robust to "bad" features.

[1] Herve Jegou, Matthijs Douze, and Cordelia Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*. 2008.

[2] Herve Jegou, Cordelia Schmid, Hedi Harzallah, and Jakob Verbeek. Accurate image search using the contextual dissimilarity measure. *PAMI*, 32(1):2–11, 2010.

[3] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Y-i Ma. Robust face recognition via sparse representation. *PAMI*, 31(2): 210–227, 2009.

[4] Shaoting Zhang, Ming Yang, Timothee Cour, Kai Yu, and Dimitris N Metaxas. Query specific fusion for image retrieval. In *ECCV*, 2012.

[5] Shiliang Zhang, Ming Yang, Xiaoyu Wang, Yuanqing Lin, and Qi Tian. Semantic-aware co-indexing for near-duplicate image retrieval. In *ICCV*, 2013.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.