Local Feature Binary Coding for Approximate Nearest Neighbor Search

Li Liu
li2.liu@northumbria.ac.uk
Mengyang Yu
m.y.yu@ieee.org
Ling Shao
ling.shao@ieee.org

Computer Vision and Artificial Intelligence Group,
Department of Computer Science and Digital Technologies,
Northumbria University, UK

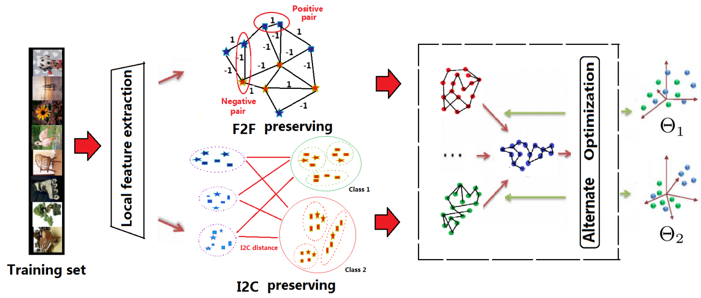


Figure 1: The illustration of the working flow of LFBC learning. The algorithm intends to preserve the pairwise F2F structure and the I2C distances and outputs the optimal bilinear projection matrices Θ_1 and Θ_2.

The potential value of hashing techniques has led to it becoming one of the most active research areas in computer vision and multimedia. However, most existing hashing methods for image search and retrieval are based on global representations, e.g., GIST [3], which lack the analysis of the intrinsic geometric property of local features and heavily limit the effectiveness of the hash code. In this paper, we propose an supervised local feature hashing framework, i.e., Local Feature Binary Coding (LFBC), for visual similarity search, in which the feature-to-feature (F2F) and image-to-class (I2C) structures are successfully preserved and combined together. Specifically, the F2F structure considers the pairwise relationship between local features in the original feature space. While, from a higher-level aspect, I2C structure reflects the connection between images and their corresponding classes, which is derived from [1]. The outline of the proposed method is illustrated in Fig. 1. It is worthwhile to highlight several properties of the proposed method: (1) Different with global representation based hashing, LFBC directly learns hashing function from local features and simultaneously preserves pairwise F2F and I2C structure, which is proved to be more effective for accurate retrieval. (2) Inspired by [2, 4], bilinear projection based hashing function is adopted in our method. Thus, the complexity of the eigen-decomposition, which is the cubic form of the dimensionality, will be significantly reduced. The corresponding integrated LFBC algorithm is depicted in Algorithm 1.

Algorithm 1 Local Feature Binary Coding (LFBC)

Input: Local feature set of each training image $\mathcal{X}_i = \{X_{i1}, \cdots, X_{im_i}\}$ in matrix form, $i = 1, \cdots, n$, the whole local feature set $\mathcal{F} = \bigcup \mathcal{X}_i$, the parameter k for pairwise structure preserving, the number of centroids K in K-means and the label information function $C(\cdot) : \mathcal{F} \to \{1, \cdots, C\}$.

Output: The bilinear projection matrices Θ_1 and Θ_2.

1: Construct local feature pairing set $\mathcal{P} = \{(i,j)|X_i, X_j \in \mathcal{F}\}$ and their corresponding pairwise labels $\ell_{ij} = \{-1, +1\}$, where $\ell_{ij} = +1$ if $X_i \in \mathrm{NN}_k(X_j)$ or $X_j \in \mathrm{NN}_k(X_i)$, and $\ell_{ij} = -1$ otherwise;
2: Employ the K-means clustering algorithm on the set of local features of each class $\bigcup_{C(\mathcal{X}_i)=c} \mathcal{X}_i$, $c = 1, \cdots, C$;
3: Compute pairwise weight W^F_{ij} and I2C similarity W^{I2C}_{ic};
4: Initialize $\Theta_2^{(0)}$ randomly;
5: **repeat**
6: Optimize Θ_1 and Θ_2 alternately;
7: **until** the objective function $\mathcal{L}(\Theta_1, \Theta_2)$ converges.

In the retrieval phase, considering that our method is specifically designed for local features, the original Hamming Ranking and Hamming Table cannot be directly applied to local features for visual indexing. Thus, in this paper, we also introduce an indexing/searching scheme called Local Hashing Voting (LHV) as shown in Fig. 2, which has been demonstrated to be efficient and accurate for image similarity search in our experiments.

For instance, given a bucket with hash code [1,1,-1,-1,1,-1,1,-1,-1,1] of a local feature, we store the indices of the images, which contain the same local feature hash code with this bucket. In this way, we search the hash code $H(\mathbf{q}_i)$ for each local feature $\mathbf{q}_k \in \mathcal{Q}$ in the query image $\mathcal{Q} = \{\mathbf{q}_1, \cdots, \mathbf{q}_m\}$ over the Hamming lookup table within Hamming radius r and return the possible images' indices. Finally, we vote and accumulate the times of each image's indices appearing in relevant buckets and then rank them in decreasing order. The final retrieved samples are returned according to the relevant ranking generated by LHV.

The experimental results on the Caltech-256, SUN397 and NUS-WIDE datasets are demonstrated in Fig. 3, Fig. 4 and Table 1.

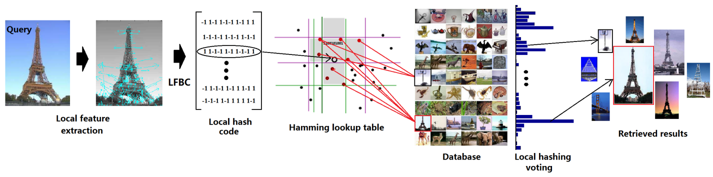
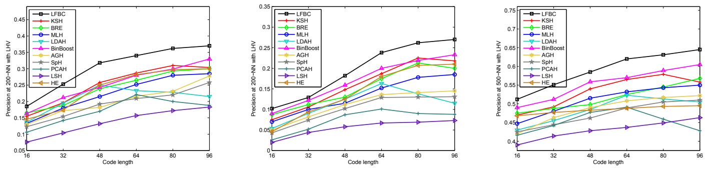
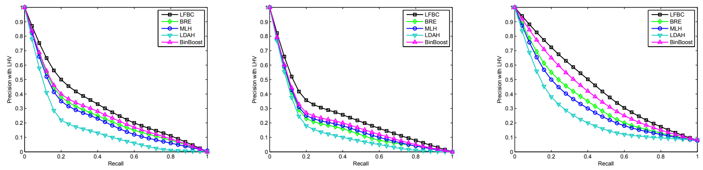


Figure 2: The illustration for the proposed LHV scheme.



(a) Caltech-256 (b) SUN397 (c) NUS-WIDE

Figure 3: Performance comparison with different numbers of bits.



(a) Caltech-256 (b) SUN397 (c) NUS-WIDE

Figure 4: The comparison of precision-recall curves of the supervised algorithms on the three datasets with the code length of 96 bits.

Table 1: Result comparison (32 bits) with/without F2F and I2C term.

Methods	Caltech-256	SUN397	NUS-WIDE
Only F2F preserving	0.189	0.065	0.387
Only I2C preserving	0.227	0.104	0.432
F2F+I2C preserving (LFBC)	**0.253**	**0.129**	**0.551**

[1] Oren Boiman, Eli Shechtman, and Michal Irani. In defense of nearest-neighbor based image classification. In *CVPR*, 2008.

[2] Yunchao Gong, Sanjiv Kumar, Henry A. Rowley, and Svetlana Lazebnik. Learning binary codes for high-dimensional data using bilinear projections. In *CVPR*, 2013.

[3] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[4] Jieping Ye, Ravi Janardan, Qi Li, et al. Two-dimensional linear discriminant analysis. In *NIPS*, 2004.