FRIF: Fast Robust Invariant Feature

Zhenhua Wang
wzh@nlpr.ia.ac.cn
Bin Fan
bfan@nlpr.ia.ac.cn
Fuchao Wu
fcwu@nlpr.ia.ac.cn

National Laboratory of Pattern Recognition,
Institute of Automation,Chinese Academy of Sciences,
100190, Beijing, China

Local image feature is a fundamental component of many computer vision applications such as 3D reconstruction, image retrieval, object recognition and object categorization. The main goal is to find salient image points that can be repeatably detected under various image transformations and then construct distinctive and robust representations for them. Many methods have been proposed in the literature [2, 3, 4, 6]. However, it is still very challenging to obtain a high quality feature whilst maintaining a low computational cost.

This paper aims to tackle this problem by developing a novel Fast Robust Invariant Feature (FRIF). For feature detection, scale invariant and stable keypoints are selected in the scale space according to our Fast Approximated LoG (FALoG) filter responses. For feature description, distinctive binary descriptors are constructed by encoding both local pattern and inter-pattern information. By employing factorization and integral image, FALoG can be computed very fast. Meanwhile, the binary descriptor is constructed directly by intensity comparisons. Thus, both the detection and description of FRIF can be done very efficiently, making it suitable for real-time applications.

Fast Approximated LoG Detector We develop a scale-invariant feature detector based on the LoG function since it is found to be more stable than other derivative based functions in characteristic scale selection [5]. The proposed Fast Approximation of LoG (FALoG) filter assigns the surrounding areas with different weights according to their distances to the central point and keeps the DC response to be zero. As shown in Figure 1, a 9×9 FALoG filter is used to approximate LoG with $\sigma = 1.2$. To efficiently compute the response of FALoG filter, it is factorized into four rectangles with different weights. The response of each rectangle can be computed rapidly using integral image. By summing over all the responses of these four rectangle filters, we obtain the response of FALoG filter.

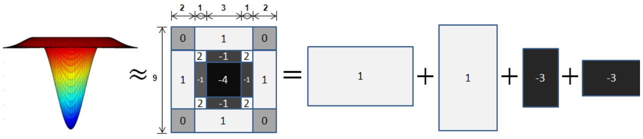

Figure 1: The approximation of LoG kernel.

The scale space is implemented in a way similar to BRISK [3]. To obtain scale invariant keypoints, we first compute the integral image for each octave and intra-octave. This can be done very fast since only one scan over the scale space is required. Next, we compute the 9×9 FALoG filter response for each pixel in the scale space separately, and a threshold T_F is applied to filter out weak responses. Then, for the remaining points, non-maximum suppression is performed so that only the extremas of FALoG responses in a local neighborhood over locations and scales are kept as potential keypoints. Finally, we refine the scale and location of each keypoint via quadratic function fitting.

Similar to LoG, FALoG filter has strong responses along edges, which are poorly localized and sensitive to noise. To filter out these edge responses, we compute the ratio of principal curvatures by the trace and determinant of Harris matrix. Keypoint candidates with ratios larger than T_L are considered as edges and are removed. Note that the Harris matrix is computed on the scale-space image.

Mixed Binary Descriptor The basic idea of our descriptor is to incorporate both the local pattern and inter-pattern information to improve matching performance of current stat-of-the-art binary descriptor, *e.g.* BRISK [3] and FREAK [1]. More specifically, the descriptor is constructed based on a modified BRISK pattern which has more overlapping areas than the original one. The local gradients of pattern locations and the inter-pattern intensity comparisons are combined to create the mixed binary descriptor.

Let $\mathcal{P}$ be the set of all the N pattern locations, for each pattern location $p_i = (x_i, y_i) \in \mathcal{P}$, a set of four points $\mathcal{S}(p_i) = \{s_{i,k}, k = 1,2,3,4\}$ are equally sampled on a circle of radius R centered at p_i. Let $I(x, \sigma)$ be the smoothed intensity of point x with sigma σ and θ be the local dominant orientation estimated by the average local gradients [1, 3]. For each rotated pattern location $p_i{}^\theta$, we compare the pair-wise intensities of its sampling points $s^\theta_{i,k} \in \mathcal{S}(p^\theta_i)$. Then, the descriptor is constructed by assembling all the test results into a binary string, each bit b of which corresponds to:

$$b = \mathrm{sign}(\mathrm{I}(s^\theta_{i,k}, \sigma_{\mathrm{i}}) - \mathrm{I}(s^\theta_{i,t}, \sigma_{\mathrm{i}})), \tag{1}$$
$$\forall p^\theta_i \in \mathcal{P} \wedge s^\theta_{i,k}, s^\theta_{i,t} \in \mathcal{S}(p^\theta_i) \wedge k,t = 1,2,3,4 \wedge k \neq t.$$

As four points are sampled for each pattern location, the dimension of the descriptor is $N \times C^2_4 = 6N$ bits. It is worth noting that the intensity comparisons between sampling points $s_{i,k}$ is closely related to the local gradient operator, both of which consider the intensity differences between pairs of local samplings.

The above descriptor encodes the local pattern information into binary strings. We complement it with more global information, which is encoded by the inter-pattern intensity comparisons. Let the set $\mathcal{A}$ be all pairs of pattern locations:

$$\mathcal{A} = \{(p_i, p_j) \mid p_i, p_j \in \mathcal{P} \wedge i \neq j\}. \tag{2}$$

A binary string can be computed by comparing the intensities of all the rotated pairs (p^θ_i, p^θ_j) in a subset $\mathcal{B}$ of $\mathcal{A}$, each bit of which corresponds to $\mathrm{sign}(\mathrm{I}(p^\theta_i) - \mathrm{I}(p^\theta_j))$. To form the subset $\mathcal{B}$, we tested three different selection criteria: 1) randomly select M pairs from $\mathcal{A}$. 2) select the shortest M pairs of $\mathcal{A}$. 3) select the longest M pairs of $\mathcal{A}$. Experiments show that the shortest pairs are more stable than others. This result is consistent with BRISK, which uses the short-distance pairs to build the descriptor. The difference is that, in our method only the shortest M pairs are used as a complementary part of the previous local gradient based binary descriptor.

By concatenating the two kinds of complementary binary strings together, we obtain our mixed binary descriptor. As the BRISK and FREAK descriptor are both 512 bits, we select $M = 512 - 6N$ to make the mixed binary descriptor the same dimension to them.

[1] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition*, pages 510 –517, 2012.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: speeded up robust features. In *European Conference on Computer Vision*, pages 404–417, 2006.

[3] S. Leutenegger, M. Chli, and R.Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *International Conference on Computer Vision*, pages 2548–2555, 2011.

[4] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[5] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, volume 1, pages 525–531, 2001.

[6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *International Conference on Computer Vision*, pages 2564–2571, 2011.