

BOLD - Binary Online Learned Descriptor For Efficient Image Matching

Vassileios Balntas, Lilian Tang and Krystian Mikolajczyk
University of Surrey, UK

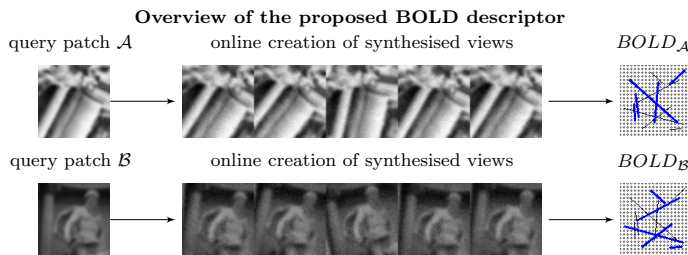


Figure 1: In contrast to typical approaches that use the same measurements for all patches, we adapt the descriptor online to each patch. The blue line ends indicate the selected binary tests from a common superset, based on the measurements from the synthesized views of a common patch. Note that although the final descriptor is different for each patch, it consists of a subset of a fixed set of dimensions. This allows efficient sequential matching and common database storage.

In this paper we propose an approach which combines the advantages of efficient binary descriptors and improved performance of learning based descriptors. The approach is inspired by the linear discriminant embedding [1] that simultaneously increases inter and decreases intra class distances. We demonstrate that there is no single set of measurements that is globally optimal for all patches in a dataset and significant improvement can be gained by adapting the binary tests to the content of each patch. The measurements are first designed to maximize globally the inter-class distances and then a subset is selected online for each patch to minimize the intra-class distances. This is done efficiently such that the extraction time is comparable to other binary descriptors.

1

Global offline optimization is based on a large set of N diverse image patches of normalized size which is different from the datasets used in our evaluation. Our features are sets of binary tests within the patch similar to [2]. For a grid of $P \times P$ locations within a patch (eg. $P = 32$) the total number of tests is $M = \binom{P^2}{2}$. The goal is to identify the subset of discriminative features, which consists of finding features that give a large variance across inter-class examples. The next step is to select a subset of uncorrelated features. We follow the greedy approach from [3] which starts by selecting the first high variance tests from the ranked list and then the process continues by verifying at each iteration the correlation between the new candidate and all selected tests.

Local online learning is performed by intra-class distance minimization to fully benefit from the LDA like optimization. We consider each patch as a separate class, therefore this optimization has to be performed online during descriptor extraction. Given that a patch is a single instance from a class, additional examples have to be synthetically generated to estimate intra-class variance by affine projections of the patch [1]. Having identified the sets that are to be included in the BOLD descriptor, each patch is represented by the string of the adapted binary tests and a second binary string of the same length where 1s indicate which tests are valid for the patch.

Matching of locally adapted descriptors After global and local online selection of discriminative tests during descriptor calculation each patch is represented by a binary string \mathbf{x}_n and a binary mask \mathbf{y}_n . The matching of two descriptors is done using the following symmetric masked Hamming

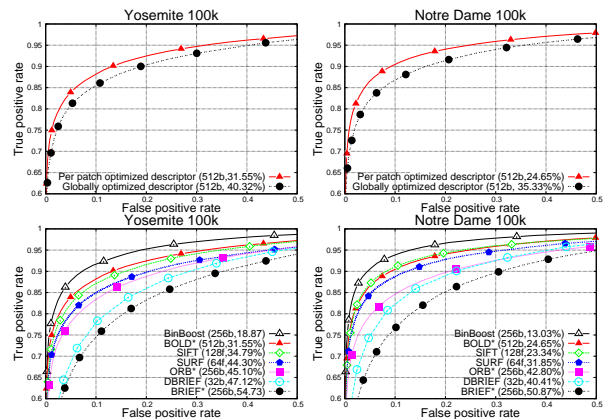


Figure 2: Top: Globally vs. locally optimized features. Bottom: BOLD compared to several state of the art descriptors. Descriptors with * are based on simple intensity tests. Using our per-patch optimization framework, performance of SIFT can be matched with simple intensity tests instead of gradient statistics.

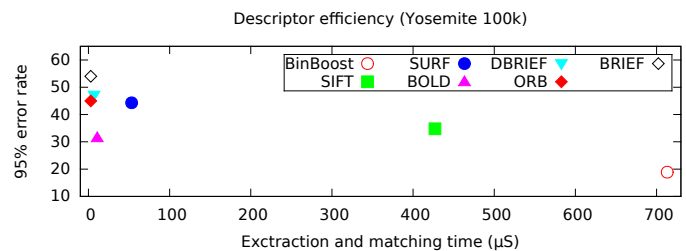


Figure 3: The proposed BOLD descriptor has good properties of low error rates and high computational efficiency. It matches the performance of SIFT, and the efficiency of BRIEF.

distance between the descriptors and their masks:

$$\mathbb{H}(\mathbf{x}_m, \mathbf{x}_n) = \mathbf{y}_m \wedge \mathbf{x}_m \oplus \mathbf{x}_n + \mathbf{y}_n \wedge \mathbf{x}_n \oplus \mathbf{x}_m \quad (1)$$

The results from several experiments (eg. Fig 2, Fig 3) on different datasets show that using a local optimization leads to significant improvements over a global one. Furthermore the efficiency of the proposed implementation is comparable to other binary descriptors and significantly better than real-valued descriptors. Our approach is the first attempt to use a binary per-patch descriptor with successful results in terms of matching performance and speed for typical computer vision applications.

Acknowledgement. This work has been supported by EU Chist-Era and EPSRC EP/K01904X/1 Visual Sense project.

- [1] H. Cai, K. Mikolajczyk, and J. Matas. Learning linear discriminant projections for dimensionality reduction of image descriptors. *Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- [2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: binary robust independent elementary features. In *European conference on Computer vision (ECCV)*, 2010.
- [3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.