Information Theoretic Rotationwise Robust Binary Descriptor Learning

Youssef El Rhabi[2(✉)], Loic Simon[1], Luc Brun[1], Josep Llados Canet[3], and Felipe Lumbreras[3]

[1] Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
{loic.simon,luc.brun}@ensicaen.fr
[2] 44screens, Paris, France
yer@44screens.com
[3] Computer Vision Center Dep. Informàtica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
{josep,felipe}@cvc.uab.es

Abstract. In this paper, we propose a new data-driven approach for binary descriptor selection. In order to draw a clear analysis of common designs, we present a general information-theoretic selection paradigm. It encompasses several standard binary descriptor construction schemes, including a recent state-of-the-art one named BOLD. We pursue the same endeavor to increase the stability of the produced descriptors with respect to rotations. To achieve this goal, we have designed a novel offline selection criterion which is better adapted to the online matching procedure. The effectiveness of our approach is demonstrated on two standard datasets, where our descriptor is compared to BOLD and to several classical descriptors. In particular, it emerges that our approach can reproduce equivalent if not better performance as BOLD while relying on twice shorter descriptors. Such an improvement can be influential for real-time applications.

1 Introduction

Since the advent of SIFT [12], extracting local descriptors has become a common practice in order to assess the similarity of image regions. Applications of local descriptors have been considerable, such as image stitching to build panoramas [5], context-based image retrieval, visual odometry or multi-view 3D reconstruction [15]. As a result of its success, this line of research has greatly impacted our everyday behaviour, be it by our use of efficient exemplar based image search engine, or the pervasive introduction of computer vision in mobile devices. Due to this important economical and societal repercussions, the design of ever improving descriptors has drawn a strong interest [4,14]. One of the main enhancements relates to data-driven construction schemes, where a typical database of image correspondences is leveraged to learn an efficient descriptor [8,21].

A. Robles-Kelly et al. (Eds.): S+SSPR 2016, LNCS 10029, pp. 368–378, 2016.
DOI: 10.1007/978-3-319-49055-7_33

In particular, recent approaches based on deep learning techniques [25] have shown a strong improvement on the state of the art.

However, some kind of "no free lunch" principle applies in that quest. Depending on the targeted application, the desired properties of the descriptor may differ significantly, leading to several trade-offs and design principles. Among others, the following questions are recurrent. Is the computational complexity of paramount importance? Does accuracy matter more than completeness? What class of invariance is required? For instance, in context-based image retrieval, a query image is proposed to a system that should propose several images similar to the query. But often semantic similarity is more crucial than purely visual resemblance. On a different note, perspective or affine invariance is a desirable asset in a multi-view reconstruction system but not in a tracking scenario. These central questions become further more complicated in practice, since descriptors are no more than a brick in a complex pipeline. Therefore, some properties of the descriptors can be destroyed or corrected by other parts of the systems. For instance, invariance can be embedded in the design of a descriptor or provided by detecting an orientation and scale before computing a non invariant descriptor. A more sophisticated case is exposed in [7], where the authors acknowledge the benefit of binary descriptors for real-time applications but claim that in a 3D reconstruction system, typical descriptors like SIFT provide a better compromise between accuracy and run time, in particular when matching is accelerated thanks to adapted data structures [16].

In this article, we intend to improve the state-of-the-art of descriptors with real-time applications in mind (e.g. SLAM). We therefore focus on low-complexity binary descriptors based on image intensity comparisons [1,6]. This active line of research lies at the crossroad of several intertwined areas such as feature selection [18,20] and hashing [9]. Our contributions include a clear exposition of a generic framework encompassing the typical state-of-the-art descriptor pipelines, as well as the design of an elegant information theoretic criterion used in the feature selection process. It yields a consensus between the discriminative power of the descriptor and its resilience to rotations. This contribution is evaluated on classical benchmarks and decreases the time and space complexity by a factor 2 compared to a recent state-of-the-art technique [3].

2 State of the Art

Binary descriptors have been in the spotlight during the past decade. Indeed these descriptors come with two central properties: low memory footprint and efficient computation. As a rule of thumb, binary descriptors require up to 512 bit storage, while full-spectrum descriptors typically involve 512 floating point values (32 times larger memory). To reduce the memory requirement one may apply dimensionality reduction and hashing techniques to a full-spectrum descriptor [10,21]. On the contrary, binary descriptors skip the full-spectrum descriptor and produce directly a reduced number of simple binary features (a.k.a tests). As a result, they are not only cheap to store, but are also faster to compute.

The two key strengths of binary descriptors come at a price, namely a lower distinctness. Therefore the main line of research in this area aims at increasing the expressive power of binary descriptors while maintaining a good trade-off in terms of memory and complexity. Attempts in that direction are numerous, and we will hereby extract a few representative ones. Two early instances of binary descriptors are CENSUS and LBP [17,24]. They are based on the systematic comparison of the pixels in a neighborhood to the central pixel. The two methods differ by the shape of the neighborhood: a full-square for CENSUS and a ring for LBP. Such procedures produce a binary string whose length depends directly on the size of the neighborhood. Therefore, in order to remain short and fast, the descriptor must be computed on a small neighborhood, which in turns restricts its distinctness.

BRIEF [6] is a recent approach to tackle the trade-off between locality and efficiency. It is built upon ideas of Locally Sensitive Hashing [9]. It relies on a random pre-selection of pairs of pixels within a large neighborhood. Afterward, the descriptor of a patch is computed by aggregating the results of the binary comparison applied on each pixel pair. In that way, the size of the neighborhood and the length of the descriptor can be chosen independently (e.g. 32×32 and 512). The authors of ORB [19] argue that the selection mechanism should account for the typical data distribution. They propose a principled scheme to select good binary features. Their approach operates in a greedy manner allowing to select uncorrelated features of maximum variance. In Sect. 3, we will give an interpretation of this procedure as a maximization of the overall information quantity. What matters most is that the variance and correlation are estimated on a representative database. In that way, the trade-off between the descriptor complexity and its expressive power is sought according to the data distribution. In addition, some authors guide the feature selection thanks to other principles. For instance, in BRISK [11], a set of concentric sampling points are selected and pairs are created from any two points that are not too far apart. Similarly, FREAK [1] designs a concentric pattern that mimics the retinal layout by having higher sampling density near the center. Then the greedy selection from ORB is used to select good pairs.

Making binary descriptors invariant to natural transformations represents also an important task. By construction, descriptors based on local differences are invariant to contrast changes. On the contrary, noise is by default badly tackled. This can be compensated by pre-smoothing the image patch before computing the descriptor. More sophisticated binary tests were also designed such as in LDB [23] where pixel data are aggregated on cell grids. As for geometric transformations, their impact can be efficiently neutralized if the main orientation (and scale) of the feature is estimated. This is the case for instance in ORB, or in AKAZE [2]. More recently, the authors of BOLD [3] have proposed an alternative where the robustness is introduced by using an online feature selection. The results are compelling and motivated our own work.

In Sect. 3, we establish the general framework underlying our approach and present it in details. Then, we provide an in-depth analysis of the observations

that have led to our formulation. In Sect. 4, we demonstrate the benefits of our contributions on several standard benchmarks. We first compare our descriptor with BOLD, its most direct challenger. In a nutshell, our contributions allows us to achieve similar performance while using half as many features. In addition, we provide also a comparison with a larger collection of classical descriptors such as SIFT [12], SURF [4], LIOP [22] and BRISK [11].

3 Binary Descriptors Construction Scheme

In this section we describe a generic information theoretic framework that encompasses most binary descriptors based on feature selection. In particular, we illustrate how BOLD descriptor can be recovered as a special case, and we analyse its limitations. Based on this analysis, we lay out a possible extension. In what follows we denote patches by the letter $w \in \mathbb{R}^{s\times s}$. A test is a function $t : \mathbb{R}^{s\times s} \to \{0, 1\}$ that maps any patch to a binary value $t(w)$. A database is a collection of patches $\mathcal{D} = \{w_1, \cdots, w_d\}$ drawn from a common (unknown) distribution. Given N tests $t_1, \cdots, t_N$, we denote $x_{k,i} = t_i(w_k)$ the collection of binary samples obtained by applying each individual test to all the patches. For convenience, we denote generically by an upper case W a random variable following the underlying patch distribution, and $X_t = t(W)$ the Bernoulli random variable induced by the test t.

3.1 Global Framework

Our main purpose is, given two patches w_1, w_2, to decide if they are similar up to some allowed transformations. For that we compute a distance d between w_1 and w_2 and match them based on a hard threshold. In practice, binary descriptors x_1 and x_2 are computed for both patches, and the distance is computed between these descriptors. Learning a good metric boils down to selecting good features which is typically done in an offline procedure. In this phase the main goal is to choose a fixed number N of tests which bring as much information as possible. Ideally those tests will be chosen by optimally learning them on a representative database. For that purpose a greedy approach is convenient, where tests are selected iteratively by maximizing a measure tied to the information quantity of the new test given the previously selected ones. This procedure is detailed in Algorithm 1. Overall, the metric construction scheme relies on an offline and an online procedure. Each of them is driven by a criterion: $J_{\mathcal{D}}(t|\mathcal{S})$ for the offline procedure and $d(x_1, x_2)$ for the online one. In Sect. 3.2 we present some common ways to evaluate those two criteria.

3.2 Usual Selection Mechanisms

In order to proceed to the offline test selection we need to set a criterion $J_{\mathcal{D}}(t|\mathcal{S})$. $J_{\mathcal{D}}(t|\mathcal{S})$ can be an estimate of the conditional entropy $H(X_t|X_{\mathcal{S}})$ and $X_{\mathcal{S}}$ is the collection of random variables induced by the current set of selected tests. However, computing the conditional entropy requires estimates of joint probabilities which can be unreliable. Some conditional independence assumptions (*e.g.*

Algorithm 1. Offline algorithm

input : Image patches dataset $\mathcal{D}$
output : $\mathcal{S}$ selected tests
generate a pool $\mathcal{P} = \{t_1, \cdots, t_M\}$ of M random tests // $M \gg N$
$\mathcal{S} = \varnothing$
for $i = 1..N$ **do**
  $t_i^* = \arg\max_{t \in \mathcal{P}}(J_{\mathcal{D}}(t|\mathcal{S}))$
  $\mathcal{S} = \mathcal{S} \cup \{t_i^*\}$

pairwise dependence) can make this task more scalable but it remains computationally intensive. Practitioners [19] often prefer to fall back to related criteria of the form: $J_{\mathcal{D}}(t|\mathcal{S}) = J_{\mathcal{D}}(t) - \infty \mathbb{1}_{|max_{s \in \mathcal{S}}(\text{corr}(X_t, X_s))| > \tau}$. Such a criterion lends itself to an efficient implementation. One may, for example, maximize the first part $J_{\mathcal{D}}(t)$ among the tests that comply with the hard correlation thresholding constraints. $J_{\mathcal{D}}(t)$ can be the entropy, but other measures exist. For instance $J_{\mathcal{D}}(t) = \text{var}(X_t)$ or $J_{\mathcal{D}}(t) = |\mathbb{E}(X_t) - 0.5|$ are preferred. Besides, with Bernoulli variables all those measures are equivalent in terms of maximisation.

In order to compute the online distance between patches w_1 and w_2, the typical matching procedure starts by computing the Hamming distance, $d_{ham}(x_1, x_2)$, between the test results $x_1, x_2 \in \{0,1\}^N$:

$$d_{ham}(x_1, x_2) = \sum_{i=1}^{N} x_{1,i} \oplus x_{2,i} \tag{1}$$

where $\oplus$ is the *XOR* operator and the sum can be efficiently computed thanks to the *popcount* routine.

BOLD improves the robustness to natural transformations by using an online-selection strategy leading to the derivation of a masked Hamming distance. This is done by computing p transformed versions $w_k^1, \cdots, w_k^p$ of each patch w_k ($k \in \{1,2\}$). Then a mask $y_k \in \{0,1\}^N$ allowing to filter out non robust test bits is built as follows[1]:

$$y_{k,i} = \bigoplus_{j=1}^{p} x_{k,i}^j \text{ with } x_{k,i}^j = t_i(w_k^j) \tag{2}$$

Based on this criterion a masked Hamming distance is constructed taking into account only those tests that are robust to the chosen deformations:

$$d_{masked}(x_1, x_2; y_1, y_2) = \sum_{i=1}^{N} \lambda_1 y_{1,i} \wedge (x_{1,i} \oplus x_{2,i}) + \lambda_2 y_{2,i} \wedge (x_{1,i} \oplus x_{2,i}) \tag{3}$$

where the weights[2] are given by $\lambda_k = \frac{|y_k|}{|y_1|+|y_2|}$ with $|y_k|$ the number of 1's in y_k.

[1] In the formula, $\bigoplus$ denotes the n-ary *XOR* (true when all its arguments are equal).
[2] The formula corresponds to the implementation provided by the authors of BOLD and the weights are different from those exposed in their article.

3.3 Analysis of Information Distribution

Data-driven selection methods implicitly rely on the fact that certain distribution estimates generalize well across databases. In particular, in order to preserve tests carrying much information, it is important that the probability estimates p_{x_t} of success of test t should remain consistent independently of the dataset. In Table 1-(a), we present the linear regression between $p_{x_t}^{\mathcal{D}}$ and $p_{x_t}^{\mathcal{D}'}$ for several pairs of databases $(\mathcal{D}, \mathcal{D}')$ from [8]. It shows with high level of confidence that their relationship is well approximated by the identity function.

Setting the number N of selected tests leads to a trade-off between performance and computation time. Besides after a critical number of tests, a saturation point is typically observed, where performance stalls and eventually worsen. Such a phenomenon is shared by data-driven approaches beyond binary descriptors. As an example, in the dimensionality reduced GLOH descriptor [14], the results are worse for a 272 dimension descriptor than for the 128 alternative. This saturation can be observed in Fig. 1-(a) for an offline selection maximizing the test variance. At the saturation, BOLD gets better performance by ignoring bits that are not resilient to some natural transformations. This feature selection is done entirely online because the resilience of a test depends on the chosen patch. A saturation phenomenon can still be observed, as shown in Fig. 1-(b). In this figure, once $N = 512$ tests are selected then no gain in performance is noticeable.

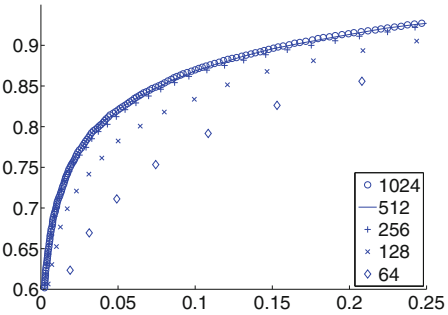
Table 1. Linear regressions (with 95 % confidence intervals) in the form $p^{\mathcal{D}'} = ap^{\mathcal{D}} + b$ with Yosemite as $\mathcal{D}$ and two alternative datasets $\mathcal{D}'$.

$\mathcal{D}'$		**estimate**	**l.b.** 95%	**u.b.** 95%
ND	a	1.083	1.078	1.088
	b	-0.043	-0.046	-0.040
Liberty	a	1.004	0.997	1.011
	b	-0.0002	-0.004	0.004

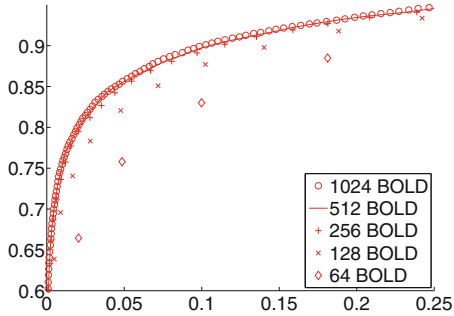
(a) $p_{x_t}^{\mathcal{D}}$ vs $p_{x_t}^{\mathcal{D}'}$

$\mathcal{D}'$		**estimate**	**l.b.** 95%	**u.b.** 95%
ND	a	1.007	1.001	1.014
	b	0.016	0.012	0.020
Liberty	a	0.938	0.930	0.946
	b	0.036	0.031	0.041

(b) $p_{y_t}^{\mathcal{D}}$ vs $p_{y_t}^{\mathcal{D}'}$



(a) Offline selection



(b) BOLD online selection

Fig. 1. Saturation on the ROC curves for the offline selection and BOLD.

3.4 Proposed Approach

Even though the online selection performs on a per patch basis, it is interesting to note that some tests are statistically more robust to geometric transforms than others. This fact becomes manifest when observing the generalization of the probabilities p_{y_t} that a test t is kept by the online selection (see Table 1-(b) for linear regressions across datasets). Since p_{y_t} generalizes well, tests that are robust to geometric transforms can also be learnt offline. We want to take this online filtering into account in the offline selection. Otherwise, information quantity as estimated offline misrepresents the actual information kept during the online phase. We propose a modified information quantity measure $H_{masked}(t)$. It serves as another way to define $J_{\mathcal{D}}(t)$. Therefore $H_{masked}(t)$ is an offline criterion but it is designed to take into account the online selection proposed by BOLD:

$$H_{masked}(t) = -\left[p_{x_t}\log(p_{x_t}) + (1-p_{x_t})\log(1-p_{x_t})\right] \times p_{y_t} \tag{4}$$

where $p_{x_t} = p(X_t = 1)$ is the estimated probability that test t is successful. The interpretation of Eq. 4 is straightforward. On the one hand, $-(p_{x_t}\log(p_{x_t}) + (1-p_{x_t})\log(1-p_{x_t}))$ represents the expected information quantity for bit t irrespectively of the online selection. On the other hand, it is multiplied by p_{y_t} since information will be thrown away with probability $1-p_{y_t}$. A similar definition of a conditional information quantity is possible. Nonetheless, for computational purpose, we choose to rely on a hard decorrelation scheme, and use only the marginal definition which is set as a substitute for $J_{\mathcal{D}}(t)$. This measure is intended to strengthen the overall information flow after the online selection. This asset shall be confirmed by experiments hereafter.

In this section we have set a generic framework that encompasses most of the state of the art test selection based descriptors. In Table 2 we show how online and offline criteria can be combined so as to retrieve some state of the art detectors as well as ours.

Table 2. Description of several methods that derivates from the generic framework

method	offline criterion	online criterion
ORB	$\mathrm{var}(X_t) - \infty\mathbb{1}_{\lvert max_{s\in\mathcal{S}}(\mathrm{corr}(X_t,X_s))\rvert>\tau}$	$d_{ham}(x_1,x_2)$
BOLD	$\mathrm{var}(X_t) - \infty\mathbb{1}_{\lvert max_{s\in\mathcal{S}}(\mathrm{corr}(X_t,X_s))\rvert>\tau}$	$d_{masked}(x_1,x_2;y_1,y_2)$
Proposed method	$H_{masked}(t) - \infty\mathbb{1}_{\lvert max_{s\in\mathcal{S}}(\mathrm{corr}(X_t,X_s))\rvert>\tau}$	$d_{masked}(x_1,x_2;y_1,y_2)$

4 Experiments

In this section, we analyse our descriptor performance on two standard datasets. In both experiments, we use the online selection implementation recommended in BOLD: we draw $p = 3$ rotations (up to 20°) in Eq. 2.

Photo Tourism Dataset: First, we present the evaluation results on the dataset proposed in [8] with the evaluation protocol of [21]. This protocol uses 3

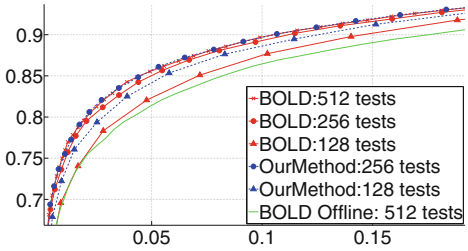


Fig. 2. ROC curves for our descriptor and BOLD under different regimes. (Color figure online)

datasets (Liberty, Notre Dame, Yosemite). The groundtruth on these datasets is encoded through correspondences between pairs of patches. Half of those correspondences are correct matches while the other half correspond to non matches. Interest points are detected with usual detectors (such as differences of Gaussians) and matches are found using a multi-view stereo algorithm as detailed in [21]. In this evaluation We compare our method specifically to the BOLD descriptor (based on their original implementation). In Fig. 2 descriptors were trained on the Yosemite dataset and were tested on 130 k patches of size 32×32 from Notre Dame. This figure highlights that our method with 256 tests, yields as good results as BOLD with 512 tests. On the contrary, BOLD with 256 tests yields lower results. In addition, the first tests selected by our approach are more informative than the ones produced by BOLD. Indeed, we can observe a substantial gap between both approaches when 128 tests are aggregated. Our descriptor performs close to the online saturation level, while BOLD is closer to the offline selection regime (in green).

We tested our approach and BOLD under different configurations (training/testing combinations). We reported area under ROC curves in Tables 3a, b and c. In all configurations, we obtain as good if not better results than BOLD. In particular, with short descriptors our approach is significantly superior to BOLD. Also BOLD reaches the saturation at a slower rate. In a nutshell, saturation occurs around 256 tests for our descriptor against 512 for BOLD.

Table 3. Area under PR curves (values are rounded at 3 decimals).

	Notre Dame		Liberty	
bits	**Bold**	**Us**	**Bold**	**Us**
1024	**0.959**	**0.959**	**0.941**	**0.941**
512	0.958	**0.959**	0.941	**0.942**
256	0.957	**0.959**	0.941	**0.942**
128	0.950	**0.954**	0.934	**0.939**
64	0.932	**0.943**	0.916	**0.929**

(a) Training on Yosemite

Yosemite		Liberty	
Bold	**Us**	**Bold**	**Us**
0.951	**0.952**	0.939	**0.940**
0.952	**0.954**	0.939	**0.943**
0.951	**0.955**	0.939	**0.943**
0.946	**0.951**	0.933	**0.940**
0.930	**0.941**	0.916	**0.931**

(b) Training on Notre Dame

Notre Dame		Yosemite	
Bold	**Us**	**Bold**	**Us**
0.957	**0.959**	0.952	**0.954**
0.956	**0.960**	0.953	**0.955**
0.955	**0.958**	0.952	**0.954**
0.947	**0.953**	0.944	**0.949**
0.933	**0.940**	0.931	**0.938**

(c) Training on Liberty

Vgg Dataset: Here we evaluate our descriptor on the benchmark proposed in [14]. This dataset offers varying testing conditions such as illumination changes, rotations, zooms, blur and jpeg compression. We also compare our descriptor with standard and recent descriptors. We have used the vlfeat implementation of SIFT and LIOP [22], and the Matlab computer vision toolbox implementation of BRISK and SURF [4]. To obtain meaningful comparisons, all the descriptors are extracted from the same key-points computed with the multi-scale Harris-Laplace detector [13]. Table 4 shows the area under ROC curve for several descriptors and image pairs in a nearest neighbor matching scenario. Since our contributions relate to binary feature selection and additional robustness with respect to rotations, we have organised the table as follows. Vertically, columns are ordered according to the level of orientation change in the image pair[3]. We have extracted 3 characteristic rotation regimes. Then horizontally, the binary descriptors are separated from full-spectrum ones. Among the binary descriptors, we consider a recent handcrafted method (BRISK), as well as a few data-driven variants lying within our framework. The chosen variants implement several mechanisms to handle rotations. The first two variations correspond to our descriptor (with 512 or 256 tests) and to BOLD. Then the two remaining variants rest on the offline selection only. The first one, compensates for the orientation of the patch while the other is applied directly on the patch. All full-spectrum descriptors rely by design on explicit rotation compensation. We have highlighted in bold the best results among binary and full-spectrum descriptors.

In the first regime (very small orientation changes), the online selection (BOLD, Us512, Us256) decreases but slightly the performance as compared to the offline-only selection. Explicit compensation of the rotation exacerbates the result deterioration. This observation echoes the fact that enforcing unneeded invariance can be harmful. The second regime (medium angles roughly below

Table 4. Area under ROC curves with a nearest neighbour matching scenario.

pair	ubc 1:5	bikes 1:5	leuven 1:5	boat 1:5	boat 1:2	bark 1:2	boat 1:4	bark 1:4
angle	0°	6°	7°	8°	14°	31°	79°	120°
Us512	0.880	0.828	0.820	**0.619**	**0.720**	0.179	0	0
Us256	0.877	0.832	0.827	0.597	0.703	0.194	0	0
Bold	0.878	0.827	**0.840**	0.605	0.719	0.157	0	0
offline oriented	0.859	0.808	0.796	0.592	0.676	**0.705**	**0.592**	**0.677**
offline	**0.883**	**0.839**	0.830	0.479	0.359	0.012	0	0
BRISK	0.779	0.648	0.647	0.406	0.594	0.668	0.382	0.574
SIFT	**0.865**	**0.859**	**0.880**	**0.749**	**0.698**	0.801	**0.707**	**0.807**
SURF	0.711	0.604	0.553	0.404	0.516	0.582	0.381	0.418
LIOP	0.815	0.792	0.754	0.588	0.662	**0.804**	0.539	0.695

[3] Apart from JPEG experiments (UBC pair), all the image pairs correspond to two independent camera shots and present varying degrees of geometric changes.

20°) is the one for which the online selection was designed. As a matter of fact, this is the mode where BOLD and our descriptors excel among binary ones. They even compete favorably with full-spectrum descriptors. Looking more closely, the orientation compensation is less efficient than the online selection. Besides, our method in this regime performs better than BOLD thanks to the modified entropy proposed in Eq. 4. In the third regime, the online selection cannot tackle the intensity of the underlying rotations. Here only explicit rotation compensation is fruitful, with a large advantage for SIFT. Apart from a single exception, the comparison with BOLD is uniformly at our advantage. This was confirmed by synthetic experiments we carried out on pure rotations, especially for ones between 15° and 30°.

5 Conclusion

In this article we have developed a novel binary image descriptor that finds its roots in the context of real time applications. To construct this descriptor, we have laid a common foundation based on feature selection. This framework covers most of the recent data-driven binary descriptors including a recent one called BOLD. We have also complemented the online selection mechanism proposed in BOLD by an adapted offline criterion applied beforehand. This new mechanism presents an elegant information theoretic interpretation and above all a perceptible practical influence. The immediate comparison to BOLD conveys that in most cases, our descriptor carries as much useful information while being twice more compact. Such an asset is an important benefit in the considered applications. Comparisons to a few other classical descriptor show that our approach obtains favorable results under mild geometric transforms. This situation arises easily in applications on mobile devices where guessing a rough estimate is often possible thanks to additional sensors. Our descriptor is therefore a perfect fit for real-time applications.

References

1. Alahi, A., Ortiz, R., Vandergheynst, P.: Freak: fast retina keypoint. In: CVPR, pp. 510–517. IEEE (2012)
2. Alcantarilla, P.F., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: BMVC (2013)
3. Balntas, V., Tang, L., Mikolajczyk, K.: Bold-binary online learned descriptor for efficient image matching. In: CVPR, pp. 2367–2375 (2015)
4. Bay, H., Tuytelaars, T., Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). doi:10.1007/11744023_32
5. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. IJCV **74**(1), 59–73 (2007)
6. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_56

7. Fan, B., Kong, Q., Sui, W., Wang, Z., Wang, X., Xiang, S., Pan, C., Fua, P.: Do we need binary features for 3D reconstruction? arXiv:1602.04502 (2016)
8. Hua, G., Brown, M., Winder, S.: Discriminant embedding for local image descriptors. In: ICCV, pp. 1–8. IEEE (2007)
9. Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: Proceedings of 30th Annual ACM Symposium on Theory of Computing, pp. 604–613. ACM (1998)
10. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: CVPR, vol. 2, pp. II-506. IEEE (2004)
11. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: binary robust invariant scalable keypoints. In: ICCV, pp. 2548–2555. IEEE (2011)
12. Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV, vol. 2, pp. 1150–1157. IEEE (1999)
13. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. IJCV **60**(1), 63–86 (2004)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI **27**(10), 1615–1630 (2005)
15. Moulon, P., Monasse, P., Marlet, R.: Global fusion of relative motions for robust, accurate and scalable structure from motion. In: ICCV, pp. 3248–3255 (2013)
16. Muja, M., Lowe, D.G.: Fast matching of binary features. In: CRV. IEEE (2012)
17. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)
18. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. PAMI **27**(8), 1226–1238 (2005)
19. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: ICCV, pp. 2564–2571. IEEE (2011)
20. Sechidis, K., Nikolaou, N., Brown, G.: Information theoretic feature selection in multi-label data through composite likelihood. In: S+SSPR, pp. 143–152 (2014)
21. Trzcinski, T., Christoudias, M., Fua, P., Lepetit, V.: Boosting binary keypoint descriptors. In: CVPR, pp. 2874–2881 (2013)
22. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: ICCV, pp. 603–610. IEEE (2011)
23. Yang, X., Cheng, K.T.: Ldb: An ultra-fast feature for scalable augmented reality on mobile devices. In: ISMAR, pp. 49–57. IEEE (2012)
24. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Eklundh, J.-O. (ed.) ECCV 1994. LNCS, vol. 801, pp. 151–158. Springer, Heidelberg (1994). doi:10.1007/BFb0028345
25. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks. In: CVPR, pp. 4353–4361 (2015)