RBRIEF: a robust descriptor based on random binary comparisons

Wei Huang¹, Ling-Da Wu¹,², Han-Chen Song¹, Ying-Mei Wei¹

¹School of Information System and Management, National University of Defense Technology, ChangSha 410073, People’s Republic of China
²The Key Laboratory of the Academy of Equipment Command and Technology, Beijing 410073, People’s Republic of China
E-mail: weeihuang@hotmail.com

Abstract: The authors propose a robust descriptor based on BRIEF, called RBRIEF. Unlike the original BRIEF, the proposed descriptor is also robust to scale and in-plane rotation transformations. Furthermore, the authors use first derivative as sample function to do binary comparisons which has proven to be better compared against the function of intensity used in BRIEF. In the feature matching stage, the authors use Hamming distance to evaluate the descriptor similarity. As a result, the performance of the proposed descriptor outperforms SURF, BRIEF and ORB using standard benchmarks. In particular, the experiments demonstrate the proposed descriptor’s superior performance in the presence of image blur, JPEG compression and light changes. Furthermore, the descriptor exhibits robust performance using only relatively few bits compared to other descriptors.

1 Introduction

Local features description have been widely studied as a key technology in many fields of computer vision and pattern recognition, such as wide baseline matching, panorama stitching, object recognition, image retrieval, three-dimensional scene reconstruction and robot localisation [1–6]. Recently, local feature descriptors have received an enormous attention and numerous approaches have been presented. Essentially, the design of local image descriptors is based on image features extracted from local image regions using certain types of spatial pooling (such as histogram [7, 8], random sampling [9, 10]). Real-world applications frequently require local feature descriptors to be robust to illumination changes, perspective distortions, image blur and image zoom.

The state-of-the-art feature descriptors, such as SIFT [7], SURF [8] and GLOH [11], first obtain the gradients information of a region of interest (ROI) densely, then partition the ROI into fine sub-blocks, within which a statistical sub-histogram is computed. Finally, a high-dimensional histogram vector is obtained for the ROI as a feature vector by concatenating all the sub-histograms. Although these methods perform very well, they are quite complex and time-consuming. In contrast, the recently developed BRIEF [9] descriptor, based on random binary comparisons, has been shown to be of low computational complexity. Moreover, its performance has been shown to be better or at least equal to that of SIFT and SURF in many aspects.

In comparison to the SIFT and SURF descriptors, the BRIEF descriptor has the following advantages: First, it randomly samples some sub-blocks of the local patch rather than enumeration of the local patch’s attribute information, leading to computation efficiency. Secondly, the descriptor compares the attribute information of the sub-blocks and stores the results as binary strings, leading to lower storage cost.

Owing to the advantages mentioned previously, the random binary comparison-based BRIEF descriptor has attracted some attention recently. However, the BRIEF descriptor has some serious limitation because of its sensitivity to rotation and scale changes. Therefore in this paper, we propose a robust variant of the BRIEF descriptor, dubbed RBRIEF (robust BRIEF, RBRIEF), which has the advantages of being invariant to in-plane rotation and scale changes. In the proposed RBRIEF descriptor, we first normalise the local ROI according to the estimated orientation and scale. Furthermore, people have found that, in order to obtain good discriminative power, derivative information is more effective than intensity of image pixel [7] because it is somewhat robust to intensity shift or affine brightness changes. As a result, we will sample the derivative information of the ROI, instead of the intensity information, which is used in the BRIEF.

The proposed RBRIEF descriptor employs the advantages of SURF in computing the histogram using the gradient information, and those of BRIEF in computational efficiency and lower storage.

The rest of the paper is organised as follows: In Section 2, we briefly review some related works. In Section 3, we present the details of the proposed local descriptor RBRIEF. In Sections 4 and 5, we give the experimental tests of the proposed descriptor and compare it with other existing descriptors. Section 6 concludes the paper.

2 Related works

With the rapid development of computer vision, local feature descriptors have received great attention in recent years. As a
result, various outstanding feature descriptors have been proposed. For more details, the readers are referred to [11]. Overall, the proposed descriptors can be partitioned into two main classes: The first class is distribution-based. They use the distribution of different image properties such as intensity, colour, gradient and edges to represent the local patches. The important examples include SIFT descriptor, which is proposed by Lowe in his milestone paper [7]. The SIFT descriptor, based on the histogram of oriented gradients within a local image region, has exhibited excellent performance in the matching and recognition due to its invariance to scale, rotations and partially invariant to illumination and viewpoint changes. However, the high computational cost of SIFT descriptor is notable. This has led to an intensive search for efficiency replacements. The SURF descriptor [8] is designed as an efficient ‘speeded up’ alternative to SIFT. It uses simple two-dimensional (2D) box filters that approximate the effects of derivative filters and uses integral image to efficiently compute the box filter. Gradient location orientation histogram (GLOH) [11] is another extension of SIFT descriptor. It uses the log-polar form of concentric circles instead of the SIFT’s 4 × 4 Cartesian location grid. Rotation invariant feature transform [12] is also a rotation invariant descriptor with a 2D histogram, where the dimensions are the distance from the centre point and the intensity of the pixel. Compared to SIFT, the calculation of the main direction of the local features is not needed. The DAISY descriptor [13], which is designed for wide-baseline matching, replaces the weighted sums of gradient norms by convolutions and thus has a much faster computational efficiency. Furthermore, this class includes some descriptors which are built based on the dimensionality reduction method such as principal component analysis [14], compressive sensing theory [15] or other projection methods [16].

The second class describes the ROI based on comparison of the image properties rather than enumerating. In [17], the authors first build the view-sets and then use them to train a classifier, which is called a randomised tree. It is used at run-time to recognise the keypoints using the tree which shifts much of the computational burden to training phase. In [18], the authors formulate the problem in a Naïve Bayesian classification framework. Abandoned the classifier of randomised tree, it directly uses hundreds of simple binary features and models class posterior probabilities which is called fern. The above two methods use binary comparisons in constructing classification tree/fern. BRIEF [9] inherits the properties of the above two methods. It directly uses binary strings which is the result of random binary comparisons to describe the features. During matching, it uses Hamming distance to evaluate the descriptor similarity. Its performance is similar or better than the state-of-the-art SURF descriptor as long as invariance to large in-plane rotations and scale change is not required. In [10], it proposed a descriptor based on BRIEF, called ORB which is rotation-invariant but not scale-invariant. Recently, DBRIEF [19] is proposed which uses learned discriminative projects to build binary descriptor. It has a superior performance but it needs a number of training patches.

3 Our descriptor

BRIEF is inspired by the work that random binary comparisons can be used to train a classifier [17]. It uses the binary string as the descriptor which is the result of randomised binary comparisons between pixels in a local patch. Its performance is similar or better than SURF in many respects. However, it is not designed to be invariant to in-plane rotation and scale change. In this paper, we will add rotational and scale invariance to BRIEF. Otherwise, it uses the intensity of the image pixel as the function of binary comparisons in BRIEF. We will also test other functions of binary comparison to do binary comparisons in our descriptor.

It needs two steps to construct the descriptor of a keypoint. First, we assign the keypoint’s orientation. Using the orientation and scale of the keypoint, we obtain a normalised local patch. Then, we use a given function to sample the pixels in the normalised patch. Finally, we use binary strings resulting from the random binary comparisons to describe the keypoint features.

In the first step, we assign the keypoint’s orientation based on the method used in SURF [8]. Specifically, the orientation of keypoints is identified based on the Haar-wavelet responses in horizontal and vertical direction in a circular neighbourhood of radius 6s around the keypoint, where s is the scale of the keypoint. The sampling step of Haar is chosen to be s, which makes it scale-independent. Calculating the Haar response and weighting it with a Gaussian (σ = 2s) which is centred at the keypoint, we obtain the gradient in horizontal and vertical direction. Then, we use a sliding orientation window to estimate the dominant orientation. After identifying the dominant orientation, we sample a (S∗s) × (S∗s) patch around the keypoint and resize it into S × S (see Fig. 1), where S is a scale parameter. By doing this, we obtain a normalised local patch P with size S × S.

As known, BRIEF descriptor is a binary string description of an image patch which resulted from the random binary comparisons. We use the same flow as BRIEF to obtain our descriptor. Consider the keypoint and its local image patch P with size S × S, then a binary comparison τ is defined as [9]

\[ \tau(p, x, y) := \begin{cases} 1, & \text{if } f(x) > f(y) \\ 0, & \text{otherwise} \end{cases} \]

where \( f(\cdot) \) is the function of binary comparisons. \((x, y)\) is a sample pair. It gets the property of the sample pixel in the patch \( P \) at \( x = (u, v)^T \), where \( u, v \) are the coordination of sample point \( x \). In BRIEF, the function \( f(\cdot) \) gets the intensity of the sample pixel.

![Fig. 1] According to the keypoint’s orientation and scale, we sample a (S∗s) × (S∗s) patch around the keypoint and resize it into S × S. After doing this, we obtain a normalised patch around the keypoint.
Choosing a set of \( n_d \) \((x, y)\) sample pairs to do binary comparisons, we define the descriptor as a vector of \( n_d \) binary comparisons \([9]\)

\[
F_{n_d}(P) := \sum_{1 \leq i \leq n_d} 2^{i - 1} \tau(P, x_i, y_i)
\]

As a result, the length of our descriptor is \( n_d \) bits. Since we store it in bytes, the length is \( n_d/8 \) bytes when we do \( n_d \) binary comparisons. In our algorithm, different types of sample function \( f(\cdot) \) have been considered. If \( f(\cdot) \) is chosen to be a function of first derivative, it has horizontal and vertical entries. Then, the descriptor is built by concatenating the two vectors and has a length of \( n_d/4 \) bytes. In the case of intensity and second derivative, the descriptor has a length of \( n_d/8 \) bytes. Our descriptor consists of binary strings in nature, so we use Hamming as its distance metric.

More specifically, the set of sample coordinate \( x = (u, v)^T \) are generated using Gaussian distribution \((\sigma = (S - 1)/10)\) around the centre of the patch which has proven to perform well. The response \( f(x) \) is computed as a smoothed sum in the sub-window around \( x \). We set the size of the sub-window as \( w \times w \). In our implementation, we compute the response in a \( 3 \times 3 \) sub-window within the \( 47 \times 47 \) normalised window.

4 Evaluation

4.1 Dataset

4.1.1 ‘2D’ Oxford Dataset: The image sequences are available at http://www.robots.ox.ac.uk/~vgg/research/affine. We evaluate our descriptor and compare the performance against other descriptors on real image with different geometric and photometric transformations and for different scene types. Fig. 2 shows the publicly test image sequences used for the evaluation. In detail, the image sequences are used to evaluate: scale and rotation change (Figs. 2a and b), viewpoint change (Figs. 2c and d), image blur (Figs. 2e and f), JPEG compression (Fig. 2g) and illumination change (Fig. 2h). Every image sequence includes six images. The first image is a basic map. The other five images are all transformation versions based on the first one. The respective ground-truth matrix is known. As a result, there are five matching pairs: 1-2, 1-3, 1-4, 1-5 and 1-6 (i.e. from 1-2 to 1-6). There is an increasing baseline that brings negative effects to all the descriptors. To facilitate the display in the figure, we only show the first, the second and the last image in the image sequence and all the images are displayed in a much lower resolution.

4.1.2 Lidar Dataset: The dataset are available at http://cvlab.epfl.ch/~strecha/multiview/. This is a publicly available datasets [20], for which dense ground-truth correspondences are available. This dataset contains real 3D distortions which is different from the well-known 2D Oxford dataset, where the images are related by a single homography. We focus on image sequence fountain-P11 depicted in Fig. 3 in order to give us some idea about the performance of our descriptor under this distortion.

4.2 Evaluation metric

We evaluate the performance of the descriptors using the same metric as in [11]. It is based on the number of correct matches and the number of false matches obtained for an image pair. Recall is the number of correctly matched regions over the number of corresponding regions between two images of the same scene. The number of false
matches relative to the total number of matches is represented by 1-precision. The results are presented with recall against 1-precision

\[
\text{recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}
\]

\[
1\text{-precision} = \frac{\#\text{false matches}}{\#\text{correct matches} + \#\text{false matches}}
\]

where the \#correspondences stands for the ground truth number of matching regions whose distance is below a threshold. The curves are obtained by varying the threshold.

### 4.3 Feature detector

We will compare our descriptor against SURF, BRIEF and ORB in the paper. In order to evaluate the performance of the descriptors fairly, we use keypoints detected by SURF detector in the evaluation of SURF, BRIEF and RBRIEF. Especially, the description vector of ORB is computed based on FAST detector, which is proposed by [10]. In practice, any detector that can provide scale information of the keypoint is proposal. The choice of SURF detector was motivated by the fact that it is computational simplicity.

Fig. 3 Lidar image used for evaluation (a), (b) and (c) Chosen from fountain-P11. We will evaluate the performances of the descriptors using image pairs (a)–(b) and (a)–(c)

Fig. 4 Evaluation for different sample functions of binary comparisons (a) and (b) Performances of descriptors on ‘boat’ and ‘trees’ image sequence, respectively

Fig. 5 Evaluation for descriptors using different lengths. We use first derivative as the sample function of binary comparisons (a) and (b) Performance tested on boat and trees sequence, respectively
Note that the output of the detector is a local patch that includes scale information.

## 5 Experimental result

### 5.1 Parameter selection

In our descriptor, we have a few free parameters: the function of binary comparisons, the length of the descriptor, the sample strategy of random \( n_d (x, y) \) pairs, the scale parameter \( S \) and the size of the sub-window \( w \). In order to select a sample method, we tested different sample strategies such as uniform distribution and Gaussian distribution with different invariance. We have found that Gaussian distribution with invariance \( \sigma = (S - 1)/10 \) is preferable. The scale parameter \( S \) and the size of the sub-window \( w \) are also tested using different values. In our implementation, we compute the response of \( f(\cdot) \) in a \( 3 \times 3 \) grid.
sub-window within the $47 \times 47$ normalised window. The selection of the other two parameters is discussed below.

5.1.1 Function of binary comparisons: In this section, we will investigate the influence of function $f(\cdot)$ on the descriptor performance. We consider three kinds of function $f(\cdot)$: the intensity of the sample pixel, the first derivative and the second derivative. In the experiment, we refer them as RBRIEF-0D, RBRIEF-1D and RBRIEF-2D, respectively. The length of the descriptors is all 32 bytes. We test them on boat and trees sequence and show the recall-precision curve in Fig. 4.3

As can be seen from the figure, RBRIEF-1D has the better performance compared against RBRIEF-0D and RBRIEF-2D in two image sequences. The experimental result validates that first derivative has a better discriminative power than the others. In the following, we will choose first derivative as the sample function of binary comparisons to build our descriptor.

5.1.2 Length of the descriptor: In this section, we will investigate the influence of descriptor length on the descriptor performance. We refer the test descriptor as RBRIEF-$k$, where the tail number stands for the number of bytes of the descriptor. We also test them on boat and trees image sequence and show the recall-precision curve in Fig. 5.

In Fig. 5, different lengths of RBRIEF descriptors are evaluated on boat and trees sequence, respectively. The performance gradually increases with increasing descriptor length. But the improvement in performance is neglectable from RBRIEF-32 to RBRIEF-64. As a result, we will use RBRIEF-32 which has a length of 32 bytes as our final descriptor to compare with the state-of-the-art descriptors.

5.2 Evaluation of the descriptors

Although many descriptors exist, we compare our descriptor against SURF, BRIEF and ORB. The SURF descriptor is a de facto standard because of its accuracy and efficiency. BRIEF is selected since it is close to our approach. The ORB is a variant of BRIEF which is said to be rotation-invariant.

We refer our descriptor as RBRIEF that uses first derivative as the sample function of binary comparisons. As we know, BRIEF is not designed to have orientation and scale invariance. To complete the evaluation, we run BRIEF on normalised patch using the method like RBRIEFs on Figs. 3a–d and 4. We refer it as NBRIEF where ‘N’ stands for the normalisation of orientation and scale. When evaluated on Figs. 3e–h, we will not compute the keypoint’s orientation and refer the descriptor as RBRIEF (U) (upright RBRIEF). In the experiment, the SURF descriptor is a 64-dimensional floating point vector which needs 256 bytes to store. BRIEF, ORB and RBRIEF has a length of 32 bytes. The detectors used in this paper and the SURF, BRIEF, ORB descriptors are all implemented based on OpenCV 2.3.

5.2.1 Evaluation on Oxford dataset: We will evaluate the descriptors on image sequence in Fig. 2, which includes rotation, scale change, viewpoint change, blur, JPEG compression and light change. The performance of our descriptor outperforms the state-of-the-art SURF descriptor. Our descriptor has a similar performance with ORB in the case of viewpoint change in Fig. 5c. In some cases of distortions, the ORB descriptor has a better performance in the condition of high precision, such as Figs. 5b, f and g.

The viewpoint, scale and rotation change is more challenging transformations compared to the other distortions. In this case, it introduces a perspective transformation to the images. The perspective transformation changes the shape of the local patch which makes the coordinates of sample pixel to vary. The distortion gives a great impact to our descriptor which builds descriptor based on sample pixels in the normalised local patch. As a result, the performance of our descriptor is not as good as in other evaluations.

The descriptors based on binary comparisons should be invariant under monotonic deformations of underlying image measurements [21]. In fact, our descriptor performs well in the condition of JPEG compression (Fig. 6g), image blur (Figs. 6e and f) and light change (Fig. 6h). We can observe from the figure that our descriptor and BRIEF outperforms SURF by a large margin. The ORB which takes into account the orientation information when building the descriptor has not performed well. This tells us that the descriptor with more robustness will reduce its discriminability.

![Fig. 7 Comparison of RBRIEF against other descriptors using 'fountain-P11'](image_url)

a Matching performance between Figs. 3a and b
b Matching performance between Figs. 3a and c
5.2.2 Evaluation on Lidar dataset: We focus on two pairs of image in fountain-P11 (Fig. 3). Fig. 7 shows the results obtained from image pairs Figs. 3a–b and 3a–c, respectively. Our descriptor obtains the best score. The comparison of two curves has shown that the performance of all presented descriptors declines as the change of viewpoint increases.

5.2.3 Timing efficiency: The keypoint detection and matching pipeline can be split into feature detection, feature description and feature matching. Our descriptor mainly includes the construction of the descriptor and the evaluation of distances between vectors. Here we evaluate the overall timing efficiency of the descriptors during feature description and matching stages. In the matching stage, we used the exhaustive searching strategy. The SURF uses the L2 distance metric. RBRIEF, SURF and ORB use the Hamming distance metric. We also test the efficiency of RBRIEF with upright direction. The timings are shown in Table 1 which used 7360 corresponding keypoints. It is measured on a standard x86 PC with Intel Core i5 2.7 Hz.

As can be seen from the table, BRIEF has a great timing advantage and our descriptor has a similar timing efficiency as SURF in description stage. Descriptor matching based on Hamming metric uses less time than L2 metric. Meanwhile, the description time of RBRIEF declines greatly when it does not take into account the orientation information of the local patch.

<p>| Table 1 Comparison of calculation time used in description and matching stages (in seconds) |</p>
<table>
<thead>
<tr>
<th>SURF</th>
<th>BRIEF</th>
<th>ORB</th>
<th>RBRIEF</th>
<th>RBRIEF(U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>description</td>
<td>1.616</td>
<td>0.076</td>
<td>0.155</td>
<td>1.467</td>
</tr>
<tr>
<td>matching</td>
<td>2.770</td>
<td>2.323</td>
<td>2.323</td>
<td>2.323</td>
</tr>
</tbody>
</table>

6 Conclusions

In this paper, we have proposed a new descriptor RBRIEF based on the well-known SURF and recently developed BRIEF descriptors. The descriptor computation starts from a scale and orientation normalised patch around the keypoint. Then, it builds the descriptor based on the binary comparisons which has been used in BRIEF. Differing from BRIEF, it uses first derivative as the sample function to do binary comparisons which has been proved its better performance against the original intensity function. As a result, RBRIEF inherits the virtue of SURF and BRIEF. The experimental result shows that it has better or at least similar performance against the state-of-the-art descriptors on standard benchmarks. Moreover, it needs only 32 bytes to store against the 256 bytes of SURF descriptor. Besides, it uses Hamming distance to evaluate the descriptor similarity, which is efficient compared to Euclidian distance.

In the process of building our descriptor, assigning the orientation of keypoints is time-consuming. Using fast orientation estimator, we can speed up the computation of the descriptor. In addition, we can further optimise the computation and matching of our descriptor based on the GPU/ SSE.

7 Acknowledgments

This research has been supported by the National Science Foundation of China under contract nos. 61103081 and 61202336.

8 References