SIViP (2017) 11:1261–1269
DOI 10.1007/s11760-017-1083-x

ORIGINAL PAPER

An approach to vision-based localisation with binary features for partially sighted people

Mariusz Oszust[1] · **Jaroslaw Padjasek**[1] · **Przemyslaw Kasprzyk**[1]

Received: 9 June 2016 / Accepted: 20 March 2017 / Published online: 1 April 2017

Abstract In this paper, an approach to the development of a localisation system for supporting visually impaired people is proposed. Instead of using unique visual markers or radio tags, this approach relies on image recognition with local feature descriptors. In order to provide fast and robust keypoint description, a new binary descriptor is introduced. The descriptor computation pipeline selects four image patches with scale-dependent sizes around the keypoint and then places five square pixel blocks within each patch. The binary string is obtained in pairwise tests between directional gradients obtained for blocks. In contrary to other binary descriptors, tests take into account gradient values obtained for blocks from all patches. The proposed approach is extensively tested using six demanding image datasets. Some of them contain labelled indoor and outdoor images under different real-world transformations, as well as challenging illumination conditions. Two datasets were prepared for the needs of this research. Experimental evaluation reveals that the introduced binary descriptor is more robust and achieves shorter computation time than state-of-the-art floating-point and binary descriptors. Furthermore, the approach outperforms other techniques in image recognition tasks, making it more suitable for the vision-based localisation.

Keywords Assistive technology · Image recognition · Keypoint matching · Binary descriptor

✉ Mariusz Oszust
marosz@kia.prz.edu.pl

[1] Department of Computer and Control Engineering, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszow, Poland

1 Introduction

Assistive technology [13,30] helps people with different disabilities to overcome their daily-life problems [7]. In the literature, there are many attempts to support blind or visually impaired people using for this purpose: social applications [7], text readers [10], Global Positioning System (GPS) [41], radio-frequency identification (RFID) tags [15,27], radio beacons [37], QR codes [14], visual markers [20,33], LED markers [26], ultra-wideband (UWB) technology [22], infrared (IR) cameras [18], or ultrasonic sensors [33]. Most of these approaches aim to provide an accurate position of a blind person relying on a device attached to the location of interest. RFID tags are often used for this purpose, as it can be seen in [15]. However, due to their short range, other types of tags are preferred. For example, Martinez-Sala et al. in [22] introduced UWB positioning technique, used to obtain a path to the destination, taking into account obstacles, walkable areas, or places of interest. In [33], a wearable system was introduced in which images captured by RGB camera were processed to find visual markers using Haar classifiers. In that work, ultrasonic sensor was also used for detecting obstacles. In [18], IR camera was used instead of ultrasonic sensor for obstacle detection. Such solution was able to provide a 3D map of observed environment, making it superior over approaches relying on proximity sensors. RGB cameras provide more information about the environment, and thus, their application can be found in systems directed to the visually impaired people. In a one of such applications [20], pie-shaped, large colour markers were recognised using a mobile device with a camera. Tapu et al. used bag of visual words with HOG descriptor [34] for detection of obstacles and category classification of observed objects. A computer vision approach with image matching based on Scale-Invariant Feature Transform (SIFT) [19] to localisa-

tion was proposed in [23]. Another descriptor, Speeded Up Robust Features (SURF) [5], was used in a banknote recognition system for the blind by Hasanuzzaman et al. [11].

In this paper, an approach to object recognition based on image matching is proposed. In order to provide higher recognition accuracy in shorter time than it can be achieved with popular floating-point and binary descriptors, a new binary descriptor is proposed. The descriptor performs binary tests between directional gradients of a small number of pixel blocks which are placed on four, scale-dependent image patches centred on the keypoint. The main novelty of the approach lies in the placement of pixel blocks within the patch, as well as in an arrangement of binary tests, which are performed between pixel blocks that belong to all selected patches. Since in this paper a localisation of a person is determined based on the labels of recognised images, two real-world, demanding datasets which contain labelled indoor scenes are introduced.

The rest of the paper is organised as follows. In Sect. 2, local keypoint descriptors which are often used in image matching tasks are presented, as well as the proposed binary descriptor. Section 3 covers evaluation of the approach on typical image benchmarks. The section also contains its comparison with state-of-the-art descriptors in image recognition tasks, having in mind their possible use in vision-based assistive technology for partially sighted people. Section 4 concludes the paper.

2 Proposed method

Local feature descriptors are often used in vision-based object recognition [11,21], retrieval [6], or scene categorisation [38]. However, there is still a place for faster and more robust techniques, able to successfully describe and match images despite various transformations, distortions, or illumination conditions [6,12,24,25].

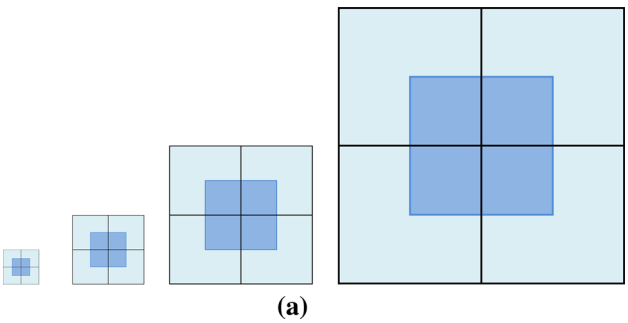
SIFT [19] and SURF [5] descriptors are among the most commonly used floating-point techniques. They are also interest point detectors, using extrema of difference of Gaussians (SIFT) or the determinant of the Hessian (SURF). For keypoint description, SIFT uses spatial histogram of the image gradients, while SURF introduced many approximations to this approach, using Haar wavelet responses determined for a scale-dependent window, or integral images to speed up computations. Despite high-quality description provided by SIFT [16], this technique, and also SURF, suffers from long computation and matching time. Therefore, binary descriptors have been developed. Here, information carried by an image patch around the keypoint is transformed into a binary string using pairwise binary tests between some image regions, pixel blocks, or raw pixels, according to a sampling pattern. Such binary strings can be compared using Hamming distance implemented as fast bitwise XOR operation followed by a bit count. In Binary Robust Independent Elementary Features (BRIEF) [8], pairs of pixels are selected from uniform distribution. In Oriented FAST and Rotated BRIEF (ORB) [32], in turn, a machine learning approach determined the sampling pattern for BRIEF features. Here, rotation invariance is achieved using intensity centroid [32], and keypoints are determined with FAST detector [31]. Another descriptor, Binary Robust Invariant Scalable Keypoints (BRISK) [17], uses AGAST [17] for interest point detection and incorporates a circular sampling pattern. A retinal sampling pattern is used in Fast Retina Keypoint (FREAK) [2] technique. All these binary descriptors rely on intensity comparisons; therefore, having in mind well-performing, floating-point techniques which use image gradients, several new binary descriptors have been introduced. In Ordinal and Spatial information of Regional Invariants (OSRI) [39], binary tests on intensities and gradients of regional invariants are performed. However, OSRI suffers from long computation time of its 21576-bit string, which additionally has to be reduced. In BinBoost [35], gradient-based image features are used for training AdaBoost classifier. Binary tests are replaced by learned binary hash functions. Among recently introduced techniques, Binary Online Learned Descriptor (BOLD) [4] is independently optimised for each image patch, and Receptive Fields Descriptor (RFD) [9] thresholds fields' responses of rectangular or Gaussian pooling regions. In Optimised Binary Robust fAst Features (OBRAF) [28], up to 12 image patches with different scale-dependent sizes are divided into 3×3 pixel blocks and then pairwise tests on intensities and directional gradients are performed. In that solution, the binary string is reduced using a simulated annealing algorithm, or only four patches are used, leaving intensity tests in a simplified version of this descriptor [29]. Local Difference Binary (LDB) [40] descriptor uses comparison of pixel blocks. There is one image patch with fixed size, divided into 4, 9, 16, and 25 blocks. LDB and OBRAF were coupled with SURF keypoints. Further extension of LDB, Accelerated-KAZE AKAZE [3], introduced scale invariance, using the keypoint's scale for calculation of the size of the patch. In AKAZE, interest points are detected using Fast Explicit Diffusion [3].

Well-performing binary descriptors often require dimensionality reduction [28,39] or learning which can be prone to the overfitting, e.g. BinBoost showed outstanding performance in patch-based benchmarks, while obtaining mediocre results in typical image matching tests [4,9]. Furthermore, their computation time is close or longer than floating-point techniques, as it can be seen for AKAZE, LDB, BinBoost, or OSRI.

In this paper, a novel binary descriptor is proposed which allows fast, robust, and scale- and rotation-invariant keypoint description by: (1) selection of scale-dependent patches

around keypoint, (2) calculation of keypoint's dominant orientation, (3) using a small number of pixel blocks per patch, and (4) performing binary tests on directional gradients that belong to different patches. The first two properties are present in many known solutions. The usage of scale-dependent patches seems to be an intuitive way of description of keypoints detected at different scales, which was confirmed in AKAZE or OBRAF. Interestingly, AKAZE and SURF share the size of the patch, which is equal to 20σ, where σ denotes the scale of the interest point. Estimation of the dominant orientation is often achieved using sums of Haar wavelet responses (SURF), rotation of the integral image or the grid (LDB, AKAZE), or using intensity moments approach (ORB). The proposed descriptor uses five pixel blocks per patch (20 blocks in total), OBRAF uses 99 blocks of pixels, BRAF 36, and AKAZE with LDB 54. It can be seen that the amount of information required to create the binary string is significantly smaller for the proposed technique than for other block-based descriptors. Furthermore, in contrary to them, the proposed descriptor, namely Simple Binary Descriptor (SBD), divides each patch into four disjunctive blocks (2×2) and adds one centre block of the same size. Figure 1 presents partitioning strategy introduced in SBD. In AKAZE, OBRAF, or LDB, all-against-all binary tests are performed between blocks that are placed on the same patch. SBD, in turn, performs binary tests on values obtained for blocks that belong to all selected patches. The values, i.e. gradients, are normalised in respect to the size of their blocks.

The creation pipeline of SBD can be described as follows. For each keypoint, $n \in N$, detected on the image, four square image patches ($P_i, i = 1, \ldots, 4$) are selected around it. The size of i-th patch, A_i, is determined by the scale of the interest point (σ), i.e. $A_i = M_i \times M_i$, where $M_i = \{6\sigma, 12\sigma, 24\sigma, 48\sigma\}$. Then, i-th patch is divided into four square pixel blocks, $B^i_j, j = 1, \ldots, 4$ and one additional block is placed in the centre ($B^i_{j=5}$). Blocks are characterised by directional gradients, D_x and D_y. Here, information on intensity present in most binary descriptors is not used to ensure shorter binary string. Directional gradients are obtained using integral images [5] and Haar-like box filters calculated for each block [3,5]. The dominant orientation in SBD is calculated with the half of wavelet responses in horizontal and vertical directions used by SURF [5]. The computation of the binary string can be written as:



(a)

Fig. 1 Patch partitioning strategy used in SBD, each patch contains five pixel blocks

$$\mathrm{SBD} = \sum_{1 \le o \le 190} 2^{o-1} T_{D_x} + \sum_{1 \le o \le 190} 2^{o-1} T_{D_y}, \tag{1}$$

where o denotes the pair of compared blocks ($B^i_j(o)$ and $B^l_k(o)$, $j \neq k \wedge i \neq l$; $j, k = 1, \ldots, 5$; $i, l = 1, \ldots, 4$), and the test is defined as:

$$T_D = \begin{cases} 1, & \text{if } \frac{D(B^i_j(o))}{\mathrm{size}(B^i_j(o))} < \frac{D(B^l_k(o))}{\mathrm{size}(B^l_k(o))} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

3 Experiments

In this section, the influence of the parameters of the SBD on its performance in matching tests is presented. Then, the proposed binary descriptor is compared with state-of-the-art binary and floating-point descriptors on popular image benchmarks. Finally, three demanding, real-world datasets are used to assess a possible usage of compared descriptors in vision-based localisation approach for partially sighted people.

3.1 Influence of the parameters

There are two main parameters used in SBD creation pipeline: (1) the number of image patches and (2) the size of each image patch. In order to show how they influence the performance of SBD, matching tests on two popular image datasets were performed. In such a test, detected and described keypoints from two images are compared. Two keypoints are considered to be matched if the distance ratio between the first and the second closest keypoint is smaller than 0.8, taking into account three pixel localisation errors and 40% overlap [12,24]. The area under *Recall versus 1-Precision* curve was used as the performance index. *Precision* expresses the number of verified matches to the returned matches, and *Recall* counts how many verified matches were found out of possible correct matches. In matching tests, 500 keypoints per image were detected using SURF and described with SBD, and then, threshold-based similarity matching was applied [5].

Oxford [24] and Heinly et al. [12] datasets were used in experiments. These popular benchmarks contain base images, as well as sequences of transformed images with known homographies between them. In datasets, there are

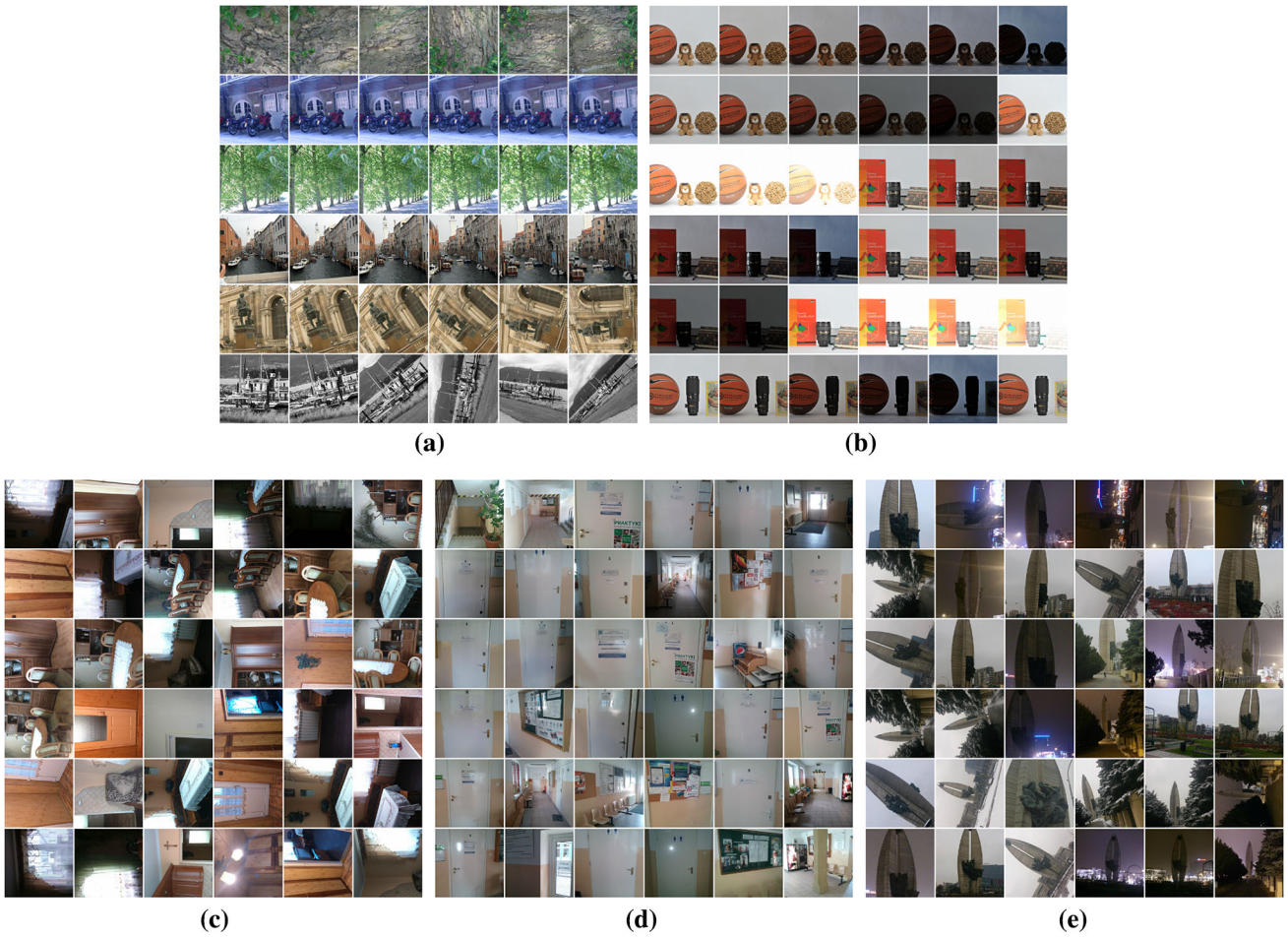


Fig. 2 Exemplary images from datasets: **a** Oxford [24] and Heinly et al. [12], **b** Phos [36], **c** the AH, **d** the DC, and **e** the BR [29]

images that exhibit a large amount of scaling, rotation, viewpoint change, blur, illumination changes, exposure, or compression. Figure 2a contains some images from these datasets. For each image pair, the area under *Recall versus 1-Precision* curve was calculated, and then, the mean value for all sequences from both datasets was provided as the measure of performance of the given set of SBD's parameters. There are many possible combinations of the number of image patches centred on the keypoint and the relation of their sizes to the keypoint's scale. In experiments, the number of patches was in the range [1, 4] and their sizes, expressed as the length of the patch's side multiplied by the keypoint's scale, was in [5, 50] range. The size of one patch was changed, while other patches were not used or the size of smaller patch was two times larger than the size of its predecessor starting from 5, e.g. in the case of three patches, M is equal to 5 and 10, for the first and the second patch, respectively. Since in this paper a new concept of binary tests between values that belong to different patches is introduced, this experiment was divided into two parts. At first, binary tests were performed only between pixel blocks which belong to the same patch, as in a typical block-based descriptor, and, in the second part, binary tests covered all blocks. Obtained results are presented in Fig. 3. It can be seen that the proposed approach was able to provide stable results disregarding the growing size of the examined patch. The number of patches had a positive influence on the performance of the resulted descriptor. Interestingly, the usage of binary tests performed on blocks from all patches led to the considerably better performance of the descriptor, which is shown in Fig. 3b. Such a gain in the performance was achieved due to the comparison of the areas that contain different amount of information. Furthermore, the results for more than two patches were better than results for compared state-of-the-art binary descriptors (see Sect. 3.2).

3.2 Comparative evaluation

Image matching benchmarks were also used to provide comparative evaluation of the SBD with state-of-the-art binary descriptors. SBD is implemented in Java, and thus, all available binary descriptors from *BoofCV* (http://boofcv.org/) [1]

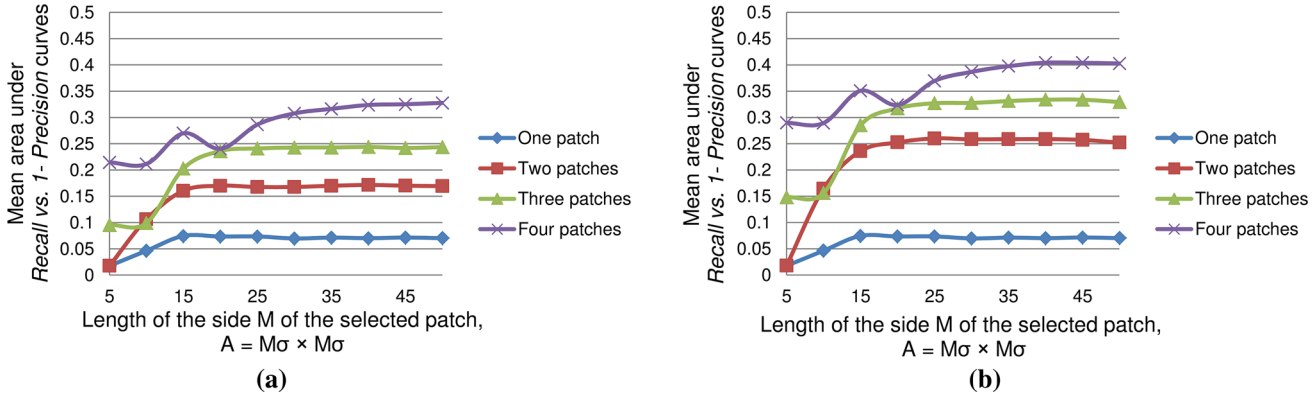


Fig. 3 Influence of the number of image patches and their sizes on the mean area under *Recall versus 1-Precision* curves calculated for image sequences from Oxford [24] and Heinly et al. [12] datasets. The experiments were divided into two parts in which binary tests were performed between pixel blocks that belong to: **a** the same image patch and **b** to different patches. The second approach to the arrangement of binary tests is introduced in this paper

and *javaCV* (https://github.com/bytedeco/javacv) libraries were used, i.e. BRIEF, BRISK, ORB, and AKAZE; *javaCV* is a Java wrapper for widely used OpenCV library (C++). Furthermore, two floating-point descriptors, SURF and SIFT, were also used in tests, in order to show that SBD can outperform them running in a fraction of their description time. Binary descriptors were compared using Hamming distance, while floating-point counterparts were compared using Euclidean distance. SBD and BRIEF described SURF keypoints, SBD was also run with FAST keypoints, which were used by ORB, since fast interest point detection can be desired in some applications.

Three datasets were used in these experiments. Oxford and Heinly et. al datasets contain mostly rotated and scaled images, and in order to provide more thorough evaluation of the descriptors against various illumination conditions, Phos dataset [36] was used. Phos contains 15 scenes captured changing the strength of uniform and degrees of non-uniform illumination. There are underexposed (−) and overexposed images (+) in this dataset, and a strong directional light source was used for capturing non-uniform images.

The mean area under *Recall versus 1—Precision* curves was used to compare descriptors. Obtained results are presented in Table 1, and they reveal that the introduced binary descriptor, SBD, outperformed compared binary counterparts by a large margin, i.e. overall means reported for Oxford and Heinly et al. datasets for SBD with SURF and FAST keypoints were 1.5 and 1.36 times better than the result obtained by the best other binary descriptor (AKAZE). Mean values for each dataset bring similar observation. Taking into account floating-point, heavy solutions, SURF obtained the best overall mean, but was worse than SBD on Heinly et al. dataset. SIFT was better than other descriptors on this dataset; however, in general, it was outperformed by SBD with SURF keypoints. It can be seen that this version of SBD was only worse than BRIEF for image sequence with exposure (*Leuven*) and worse than AKAZE for two image sequences that contain rotated images (*Bikes* and *Ceiling*). For other image sequences, SBD with SURF keypoints clearly outperformed other descriptors, and for *Boat Graffiti*, *Wall*, and *Day and Night* sequences it was better than floating-point techniques. For Phos, BRIEF showed good performance, since here test images are not rotated, and applied binary tests, which are also present in other descriptors, were able to compensate illumination changes. SBD using FAST and SURF keypoints outperformed all other compared descriptors on this dataset. Comparing these two interest point detectors, it seems that SURF keypoints are less stable against illumination changes.

The implementation of SBD was run as single threaded on a CPU with Intel Core i5-5200u 2.2 GHz processor using 8 GB RAM, Java 8.0, and Microsoft Windows 7. The obtained description time, measured per keypoint, for the first image from *Bikes* sequence, was as follows: SURF 0.1403 ms, SIFT 0.7517 ms, BRIEF 0.0276 ms, ORB 0.0225 ms, AKAZE 0.1406 ms, BRISK 0.0425 ms, and SBD 0.029 ms. SBD was slightly slower than ORB and BRIEF, but it considerably outperformed them in tests. It was faster than BRISK and almost five times faster than AKAZE or SURF. SIFT was the slowest competing descriptor. Matching time depends on the length of the binary string and the number of detected keypoints, which was constant for all techniques. Here, only ORB with its 256-bit string was faster than SBD. They were followed by AKAZE (486 bits), BRISK, and BRIEF (512 bits), and by floating-point descriptors, for which matching time limits the number of their possible applications. The upright version of SBD, in which the dominant orientation of the keypoint is not used, was computed in 0.007 ms, which is almost four times faster than BRIEF, which also does not contain this step.

Table 1 Comparison of the approach with state-of-the-art binary and floating-point descriptors in matching tests on Oxford, Heinly et al., and Phos datasets, in terms of the mean area under *Recall versus 1-Precision* curves

Image sequence, transformation	Floating-point descriptors		Binary descriptors					
	SURF [5]	SIFT [19]	BRIEF [8]	AKAZE [3]	BRISK [17]	ORB [32]	SBD (SURF)	SBD (FAST)
Oxford dataset								
Bark, rotation	**0.2780**	0.1436	0.0029	0.1766	0.0571	0.0328	0.2314	0.0316
Bikes, blur	**0.5784**	0.5344	0.3199	0.3683	0.3245	0.0604	0.3365	0.3118
Boat, rotation	0.3863	0.2123	0.0531	0.1990	0.0345	0.0426	**0.5254**	0.3740
Graffiti, viewpoint	0.1800	0.1774	0.0400	0.1215	0.0644	0.0251	**0.2403**	0.1486
Leuven, exposure	0.5875	**0.6421**	0.4268	0.3833	0.1318	0.1328	0.3480	0.3128
Ubc, JPEG compression	**0.7464**	0.6038	0.5561	0.4659	0.3331	0.1805	0.6986	0.5815
Wall, viewpoint	0.4380	0.4540	0.3029	0.2455	0.0906	0.0897	**0.6403**	0.5664
Trees, blur	0.3822	**0.3905**	0.2724	0.2037	0.1290	0.0769	0.2862	0.3853
Mean	**0.4471**	0.3948	0.2468	0.2705	0.1456	0.0801	0.4133	0.3390
Heinly et al. dataset								
Ceiling, rotation	0.4583	**0.6163**	0.0077	0.4640	0.3204	0.0827	0.3804	0.4493
Day and night, illumination	0.0606	0.1580	0.1028	0.0383	0.0497	0.0188	0.4007	**0.4800**
Rome, rotation	0.5632	**0.6585**	0.0026	0.3542	0.3550	0.1255	0.3759	0.3710
Semper, rotation	0.3179	**0.6196**	0.0346	0.2667	0.3947	0.1188	0.3443	0.4436
Venice, scaling	**0.6524**	0.1514	0.0655	0.2252	0.2437	0.1347	0.5560	0.3300
Mean	0.4105	**0.4407**	0.0426	0.2697	0.2727	0.0961	0.4115	0.4148
Overall mean	**0.4330**	0.4124	0.1683	0.2702	0.1945	0.0862	0.4126	0.3682
Phos dataset								
Directional + uniform	0.5415	0.5610	0.6122	0.4173	0.4148	0.1780	0.6486	**0.7471**
Directional + 0.8 uni.	0.4781	0.5034	0.5902	0.4020	0.3494	0.1377	0.6201	**0.7049**
Directional + 0.6 uni.	0.3836	0.4792	0.5379	0.3641	0.2146	0.1110	0.5587	**0.6243**
Directional + 0.4 uni.	0.2900	0.4268	0.5022	0.3349	0.1392	0.0865	0.4895	**0.5585**
Directional + 0.2 uni.	0.2274	0.4090	0.4500	0.3237	0.0955	0.0806	0.4510	**0.5117**
Directional	0.1410	0.3039	0.3427	0.2355	0.0380	0.0695	0.3615	**0.4167**
Underexposed (−1)	0.6176	0.6043	0.6947	0.4411	0.4676	0.1607	0.7124	**0.8166**
Underexposed (−2)	0.5317	0.6075	0.6716	0.4217	0.3131	0.1265	0.6889	**0.7795**
Underexposed (−3)	0.4131	0.5654	0.5834	0.4304	0.1821	0.1077	0.6334	**0.7275**
Underexposed (−4)	0.3571	0.5548	0.5354	0.4697	0.0241	0.1039	0.6072	**0.6963**
Overexposed (1)	0.6370	0.5935	0.6901	0.4276	0.5727	0.1764	0.7181	**0.8124**
Overexposed (2)	0.5587	0.5688	0.6310	0.4181	0.4698	0.1526	0.6796	**0.7824**
Overexposed (3)	0.4271	0.4947	0.5894	0.3465	0.3237	0.1176	0.6111	**0.7158**
Overexposed (4)	0.2149	0.3390	0.5287	0.3041	0.1614	0.0891	0.4836	**0.5887**
Mean	0.4156	0.5008	0.5685	0.3812	0.2690	0.1213	0.5903	**0.6773**

The results for the best descriptor are written in bold, and underlined results indicate the two best binary descriptors
The name of each image sequence is written in italics

Table 2 Comparison of the approach with state-of-the-art binary and floating-point descriptors in object recognition tests on the AH, the DC, and the BR datasets, in terms of the number of correctly recognised objects or places

No. of keypoints per image	SURF [5]	SIFT [19]	BRIEF [8]	AKAZE [3]	BRISK [17]	ORB [32]	SBD (SURF)	SBD (FAST)
The AH dataset, 60 test images								
20	9	6	12	**32**	11	22	16	**31**
50	32	28	32	**40**	13	37	39	**42**
100	40	36	41	41	24	**44**	**48**	43
200	42	**47**	45	42	29	46	**51**	43
300	45	48	46	40	35	**49**	**51**	42
400	47	**48**	47	42	39	**48**	**50**	42
500	**49**	47	45	42	40	47	**51**	44
The DC dataset, 126 test images								
20	8	13	3	**36**	5	25	6	**28**
50	32	31	13	**53**	16	**42**	31	38
100	42	44	14	**54**	29	40	**45**	**45**
200	57	**62**	14	57	45	**62**	57	44
300	63	65	16	58	54	**68**	**67**	45
400	64	67	13	58	55	**73**	**70**	44
500	66	**70**	12	58	56	**71**	**70**	44
Sequence	The BR dataset, 500 test images, 500 keypoints per image							
A. day	173	**284**	106	161	13	81	**263**	163
A. night	203	**227**	100	204	12	80	**316**	170
W. day	161	**312**	92	144	22	58	**218**	162
W. night	237	**268**	100	220	12	80	**297**	193
S. day	202	**325**	105	203	20	90	**238**	186
S. night	209	**250**	102	290	43	182	**313**	202

A, *W*, and *S* denote autumn, winter, and summer, respectively. The results for the two best descriptors for each case are written in bold, and the results for the best binary descriptor are underlined

3.3 Application to vision-based localisation

In order to evaluate the usability of the developed binary descriptor in a vision-based assistive technology for supporting partially sighted people, a specific image datasets are required. They should contain images of building interiors, as well as images of outdoor objects or scenes. It can be assumed that images of such places or objects are labelled, and, upon recognition, their labels can be pronounced using text-to-speech technology. Therefore, three labelled image datasets were used in this paper. Two of them, the At Home (AH) and the Doors and Corridors (DC) datasets, were created for the needs of this study. They can be downloaded at http://www.marosz.kia.prz.edu.pl/datasets.html. The AH dataset contains 250 images taken at an apartment. There are 190 learning images and 60 test images; most of them are rotated (90°). Here, labels refer to the part of the apartment and observed objects. The DC dataset, in turn, contains labelled images of corridors and doors captured at the Department of Computer and Control Engineering, at the Rzeszow University of Technology, Poland. There are 111 learning examples and 126 test images in this dataset. The third image collection, the Beautiful Rzeszow (BR) dataset [29], is much larger and contains 3000 images depicting 50 tourist attractions in Rzeszow, Poland. They were photographed varying the time of the day (day and night) and season (spring, autumn, and winter). The dataset is particularly challenging, since it covers many image transformations such as scale, viewpoint, and rotation. There are also difficult illumination changes and occlusions. Images captured at a different time of the day were used for testing. Exemplary images from these three datasets are shown in Fig. 2c–e.

SBD descriptor was compared with other state-of-the-art binary and floating-point techniques. The test images were recognised using k-nearest neighbour classifier ($k = 1$) working on the number of returned matched descriptor pairs. Since all recognised images are indicating the localisation of a person, as well as seen objects, such image recognition approach can be used for supporting partially sighted people.

Obtained recognition results on three datasets are presented in Table 2. The number of used keypoints per image varied from 20 to 500 for the first two datasets and set to 500

for the BR dataset. For the AH dataset, SBD's version using SURF keypoints was better than other descriptors. However, AKAZE achieved good performance with small number of detected keypoints, close to the results obtained by SBD with FAST interest points. Since only a part of images in this dataset are rotated, and scale change is small, BRIEF's performance is worth noticing. Floating-point descriptors were better than SBD (FAST), AKAZE, and BRISK. The second dataset, the DC, turned out to be more difficult, since door images are very similar. Furthermore, it can be seen that matching-based recognition has difficulties in case of repetitive patterns. For this dataset, SBD on SURF keypoints, ORB and SIFT outperformed other descriptors. The recognition results obtained for the BR dataset show outstanding performance of SBD with SURF keypoints. Here, SBD recognised similar number of images as it is reported for SIFT, in a fraction of its description and matching time. Also, SBD with FAST keypoints performed as well as AKAZE, and better than other binary descriptors. Due to high robustness of the presented binary descriptor against illumination conditions, recognition results for images taken at night are much better than for other techniques. In general, SBD using SURF keypoints presented the best recognition accuracy.

4 Conclusion

In this paper, an approach to image recognition with binary features for the localisation purposes for supporting visually impaired people was considered. Since the matching-based image recognition performance with widely used binary descriptors is not satisfactory, as well as the computation and matching time of their floating-point, heavy counterparts, a new binary descriptor was introduced. SBD achieves fast computation time and is more robust to different image transformations than compared techniques. Its creation pipeline selects four scale-dependent image patches centred on a keypoint, covers them with five pixel blocks, and then performs binary tests on directional gradients calculated for blocks. In contrary to other block-based descriptors, the binary tests are also performed between values determined for blocks from different patches. The descriptor was evaluated and compared with state-of-the-art using three popular image benchmarks, as well as three real-world image collections with labelled images. Obtained results are promising; they confirm the usability of SBD for vision-based recognition and localisation.

Compliance with ethical standards

Author's contribution MO conceived and designed the experiments, analysed the data and wrote the paper. MO, JP, and PK performed the experiments and contributed analysis tools (JP, PK—SURF descriptor, MO—SBD and other descriptors). JP and PK prepared the datasets.

References

1. Abeles, P.: Speeding up SURF. In: Proceedings of the International Symposium on Advances in Visual Computing (ISVC), pp. 454–464. Springer (2013)
2. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012. pp. 510–517 (2012). doi:10.1109/CVPR.2012.6247715
3. Alcantarilla, P.F., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: British Machine Vision Conference (BMVC) (2013)
4. Balntas, V., Tang, L., Mikolajczyk, K.: BOLD - Binary online learned descriptor for efficient image matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. pp. 2367–2375 (2015). doi:10.1109/CVPR.2015.7298850
5. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 404–417. Springer (2006). doi:10.1007/11744023_32
6. Bianco, S., Mazzini, D., Pau, D.P., Schettini, R.: Local detectors and compact descriptors for visual search: a quantitative comparison. Digit. Signal Process. **44**, 1–13 (2015). doi:10.1016/j.dsp.2015.06.001
7. Brady, E., Morris, M.R., Zhong, Y., White, S., Bigham, J.P.: Visual challenges in the everyday lives of blind people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2117–2126. ACM (2013)
8. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision ECCV 2010, Lecture Notes in Computer Science, **6314**,778–792. Springer, Berlin (2010). doi:10.1007/978-3-642-15561-1_56
9. Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., Fua, P.: Receptive fields selection for binary feature description. IEEE Trans. Image Process. **23**(6), 2583–2595 (2014). doi:10.1109/TIP.2014.2317981
10. Guerreiro, J., Gonçalves, D.: Text-to-speeches: evaluating the perception of concurrent speech by blind people. In: Proceedings of the 16th international ACM SIGACCESS Conference on Computers and Accessibility, pp. 169–176. ACM (2014)
11. Hasanuzzaman, F.M., Yang, X., Tian, Y.: Robust and effective component-based banknote recognition for the blind. IEEE T. Syst. Man Cybern. C **42**(6), 1021–1030 (2012)
12. Heinly, J., Dunn, E., Frahm, J.M.: Comparative evaluation of binary features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 759–773. Springer (2012)
13. Hersh, M., Johnson, M.A.: Assistive Technology for Visually Impaired and Blind People. Springer, Berlin (2010)
14. Idrees, A., Iqbal, Z., Ishfaq, M.: An efficient indoor navigation technique to find optimal route for blinds using qr codes. In: IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), 2015. pp. 690–695 (2015). doi:10.1109/ICIEA.2015.7334197
15. Ivanov, R.: Indoor navigation system for visually impaired. In: Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in

Computing on International Conference on Computer Systems and Technologies, pp. 143–149. ACM (2010)
16. Khan, N., McCane, B., Mills, S.: Better than SIFT? Mach. Vis. Appl. **26**(6), 819–836 (2015). doi:10.1007/s00138-015-0689-7
17. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: Binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision (ICCV), 2011. pp. 2548–2555 (2011). doi:10.1109/ICCV.2011.6126542
18. Li, B., Zhang, X., Muñoz, J.P., Xiao, J., Rong, X., Tian, Y.: Assisting blind people to avoid obstacles: an wearable obstacle stereo feedback system based on 3d detection. In: 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 2307–2311. IEEE (2015)
19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). doi:10.1023/b:visi.0000029664.99615.94
20. Manduchi, R., Coughlan, J., Ivanchenko, V.: Search Strategies of Visually Impaired Persons Using a Camera Phone Wayfinding System. Springer, Berlin (2008)
21. Manipoonchelvi, P., Muneeswaran, K.: Significant region-based image retrieval. Signal Image Video Process. **9**(8), 1795–1804 (2015). doi:10.1007/s11760-014-0657-0
22. Martinez-Sala, A.S., Losilla, F., Sánchez-Aarnoutse, J.C., García-Haro, J.: Design, implementation and evaluation of an indoor navigation system for visually impaired people. Sensors **15**(12), 32168–32187 (2015)
23. Mekhalfi, M.L., Melgani, F., Bazi, Y., Alajlan, N.: Toward an assisted indoor scene perception for blind people with image multilabeling strategies. Expert Syst. Appl. **42**(6), 2907–2918 (2015). doi:10.1016/j.eswa.2014.11.017
24. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005). doi:10.1109/tpami.2005.188
25. Mukherjee, D., Wu, Q.M.J., Wang, G.: A comparative experimental study of image feature detectors and descriptors. Mach. Vision Appl. **26**(4), 443–466 (2015). doi:10.1007/s00138-015-0679-9
26. Nakajima, M., Haruyama, S.: New indoor navigation system for visually impaired people using visible light communication. EURASIP J. Wirel. Comm. **2013**(1), 1–10 (2013)
27. Nassih, M., Cherradi, I., Maghous, Y., Ouriaghli, B., Salih-Alj, Y.: Obstacles recognition system for the blind people using RFID. In: 6th International Conference on Next Generation Mobile Applications, Services and Technologies, 2012. pp. 60–63 (2012). doi:10.1109/NGMAST.2012.28
28. Oszust, M.: An optimisation approach to the design of a fast, compact and distinctive binary descriptor. Signal Image Video Process. pp. 1–8 (2016). doi:10.1007/s11760-016-0907-4
29. Oszust, M.: Towards binary robust fast features using the comparison of pixel blocks. Meas. Sci. Technol. **27**(3), 035,402 (2016). doi:10.1088/0957-0233/27/3/035402
30. Peetoom, K.K., Lexis, M.A., Joore, M., Dirksen, C.D., De Witte, L.P.: Literature review on monitoring technologies and their outcomes in independently living elderly people. Disabil. Rehabil. **10**(4), 271–294 (2015)
31. Rosten, E., Porter, R., Drummond, T.: Faster and better: a machine learning approach to corner detection. IEEE Trans. Pattern Anal. Mach. Intell. **32**(1), 105–119 (2010). doi:10.1109/tpami.2008.275
32. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), 2011. pp. 2564–2571 (2011). doi:10.1109/ICCV.2011.6126544
33. Simoes, W.C.S.S., de Lucena, V.F.: Blind user wearable audio assistance for indoor navigation based on visual markers and ultrasonic obstacle detection. In: IEEE International Conference on Consumer Electronics (ICCE), 2016. pp. 60–63 (2016). doi:10.1109/ICCE.2016.7430522
34. Tapu, R., Mocanu, B., Bursuc, A., Zaharia, T.: A smartphone-based obstacle detection and classification system for assisting visually impaired people. In: The IEEE International Conference on Computer Vision (ICCV) Workshops (2013)
35. Trzcinski, T., Christoudias, M., Fua, P., Lepetit, V.: Boosting binary keypoint descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013. pp. 2874–2881. IEEE (2013). doi:10.1109/CVPR.2013.370
36. Vonikakis, V., Chrysostomou, D., Kouskouridas, R., Gasteratos, A.: A biologically inspired scale-space for illumination invariant feature detection. Meas. Sci. Technol. **24**(7), 074,024 (2013). doi:10.1088/0957-0233/24/7/074024
37. Wawrzyniak, P., Korbel, P.: Wireless indoor positioning system for the visually impaired. In: Federated Conference on Computer Science and Information Systems (FedCSIS), 2013. pp. 907–910. IEEE (2013)
38. Wei, X., Phung, S.L., Bouzerdoum, A.: Visual descriptors for scene categorization: experimental evaluation. Artif. Intell. Rev. **45**(3), 333–368 (2015). doi:10.1007/s10462-015-9448-4
39. Xu, X., Tian, L., Feng, J., Zhou, J.: OSRI: a rotationally invariant binary descriptor. IEEE Trans. Image Process. **23**(7), 2983–2995 (2014). doi:10.1109/TIP.2014.2324824
40. Yang, X., Cheng, K.T.: LDB: an ultra-fast feature for scalable augmented reality on mobile devices. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2012. pp. 49–57 (2012). doi:10.1109/ISMAR.2012.6402537
41. Zenk, S.N., Schulz, A.J., Odoms-Young, A., Wilbur, J., Matthews, S.A., Gamboa, C., Wegrzyn, L.R., Hobson, S., Stokes, C.: Feasibility of using global positioning systems (GPS) with diverse urban adults: before and after data on perceived acceptability, barriers, and ease of use. J. Phys. Activity Health **9**(7), 924 (2012)