Complementary Keypoint Descriptors

Clark F. Olson(✉), Sam A. Hoover, Jordan L. Soltman, and Siqi Zhang

School of STEM, University of Washington, Bothell, USA
cfolson@uw.edu

Abstract. We examine the use of complementary descriptors for keypoint recognition in digital images. The descriptors combine multiple types of information, including shape, color, and texture. We first review several keypoint descriptors and propose new descriptors that use normalized brightness/color spatial histograms. Individual and combined descriptors are compared on a standard data set that varies blur, viewpoint, zoom, rotation, brightness, and compression. Results indicate that substantially improved results can be achieved without greatly increasing keypoint descriptor length, but that the best results combine information from complementary descriptors.

1 Introduction

We investigate the use of complementary descriptors for keypoint recognition. These keypoints combine multiple unrelated keypoint descriptors to form a longer descriptor that is better able to discriminate between correct and incorrect matches.

Historically, the most popular keypoint descriptors have used image gradients to encode shape information in the local region around keypoint. Lowe's SIFT descriptor [1] has been highly influential and has spawned many competing descriptors, including HOG [2], GLOH [3], SURF [4], BRIEF [5], and ORB [6]. These descriptors are based on pixel intensities (grayscale images).

Color has been incorporated into keypoint descriptors in several ways. One possibility is to run SIFT on each color channel in some color space and stack the results into a longer descriptor. Bosch et al. [7] use the HSV color space. Van de Sande et al. [8] additionally consider RGB and an opponent color space. However, these methods are still based on gradients and thus focus on shape, not color.

Less work has examined the use of color information directly. Van de Weijer and Schmid [9] use a hue histogram (without location information) stacked with the SIFT descriptor. Luke et al. [10] stack SIFT with a separate SIFT-like descriptor that replaces gradient orientation with pixel hue and gradient magnitude with pixel saturation. Olson and Zhang [11] instead use histograms of normalized colors in the grid cells of the keypoint neighborhood. Other color descriptors were considered by van de Sande et al. [8], but were found to have lower performance.

Texture descriptors have been developed by Lazebnik et al. [12]. Their descriptors are rotationally invariant based on the distance from the keypoint

G. Bebis et al. (Eds.): ISVC 2016, Part I, LNCS 10072, pp. 341–352, 2016.
DOI: 10.1007/978-3-319-50835-1_32

center (and not the orientation). We consider additional descriptors that make use of the computed keypoint orientation in a manner similar to SIFT. Both techniques can be additionally extended to color images by stacking the descriptors for each color channel.

Given this rich set of descriptors that use differing image cues, we examine the use of combinations of descriptors in order to improve precision/recall performance. We use a straightforward method of combining the descriptors, concatenating the vectors and then taking the Euclidean distance between the concatenated vectors. In the context of multi-class object classification, Gehler and Nowozin [13] found that such simple methods yield equivalent results to more complicated combination methods, but with much faster results.

Previous work has combined descriptors in a similar fashion. Zhang et al. [14] concatenate the SPIN, SIFT, and RIFT descriptors to generate more effective combinations. Van de Sande et al. [8] found that even by combining a highly correlated set of descriptors (SIFT, OpponentSIFT, rgSIFT, C-SIFT, and RGB-SIFT), all of which rely on gradient (shape) information, a significant improvement in mean average precision is possible. In contrast, we combine highly differing descriptors using information from multiple modalities (shape, color, texture). Bo, Ren, and Fox [15] also concatenate their kernel descriptors for gradient, color, and shape into a single encompassing descriptor that outperforms the individual components.

Our experiments using the Oxford affine covariant regions data set [3,16] demonstrate techniques comparable and superior in performance to SIFT that do not use gradients and that combinations of descriptors outperform any single descriptor. The disadvantage of this technique is the additional computation time and space required. This yields a trade-off in selecting an appropriate descriptor/combination. Fortunately, much of the computation is parallelizable.

The next section describes the descriptors that we consider in this work. Section 3 describes the keypoint recognition process and metrics. Section 4 details our experiments and results. Finally, Sect. 5 gives our conclusions and observations.

2 Descriptors

We examine several keypoint descriptors that encode shape, texture, and color information.

2.1 SIFT

The Scale-Invariant Feature Transform (SIFT) descriptor was described by Lowe [1]. We use the OpenCV contributed implementation based on the code of Hess [17]. The technique generates histograms of the gradient orientation (weighted by the gradient magnitude) in a 4×4 grid around the keypoint center rotated to the keypoint orientation and scaled to the keypoint size. With eight magnitude bins, the SIFT technique yields a 128-dimensional descriptor

for each keypoint. The SIFT descriptor and a variation called GLOH were found to be the best performing local descriptors by Mikolajczyk and Schmid [16] when compared to several grayscale descriptors.

2.2 RGB-SIFT

RGB-SIFT is the concatenation of the SIFT descriptors computed separately for three RGB channels, yielding a 384-dimensional descriptor. This descriptor was considered in the study by van de Sande, Gevers, and Snoek [8] and found to be among the top performers. Since normalization is performed separately on each channel, this descriptor is invariant not only to illumination intensity and shift, but also illumination color.

2.3 OpponentSIFT

OpponentSIFT is similar to RGB-SIFT, except that the color channels are first transformed into an opponent color space [8]:

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} (R-G)/\sqrt{2} \\ (R+G-2B)/\sqrt{6} \\ (R+G+B)/\sqrt{3} \end{bmatrix} \tag{1}$$

In order to prevent channels with little signal from becoming magnified, we first concatenate the channel descriptors and then normalize the values together. This yields significant improvements in our experiments.

Like RGB-SIFT, this yields a 384-dimensional descriptor and was one of the top performing descriptors in the study by van de Sande, Gevers, and Snoek [8].

2.4 SPIN and CSPIN

SPIN is a texture descriptor introduced by Lazebnik, Schmid, and Ponce [12]. Unlike other descriptors described here, it does not require the keypoint orientation; it is invariant to orientation based on its construction using concentric circular bins. We use 8 circular bins with 16 intensity bins each (unlike the 10×10 histogram in the original work) to yield a 128 dimensional descriptor. This circular descriptor is scaled to use the same image area as the other (square) descriptors.

CSPIN is our simple generalization of SPIN to color images by stacking the SPIN descriptors for each color channel.

2.5 HoNI: Histograms of Normalized Intensities

The HoNI descriptor uses the same rotated and scaled 4×4 grid as SIFT, but instead of gradient orientations, the histogram values are normalized image

intensities. Like in the SIFT descriptor, the histogram votes are Gaussian-weighted according to their distance from the keypoint center and spread among spatially adjacent cells. The intensities are normalized such that average weighted intensity within the keypoint boundary is 127.5 and the standard deviation of the intensities is 64. This provides invariance to affine intensity changes (bias and gain). The intensity histograms have 8 buckets per grid cell, yielding a 128-dimensional descriptor. To our knowledge, this descriptor has not been previously studied, although it is similar to the SPIN descriptor [12] on a square grid and the HoNC descriptor [11] without color information.

Color HoNI (CHoNI) is an extension that stacks HoNI descriptors for each color channel.

2.6 HoNC: Histograms of Normalized Colors

Similar to HoNI, the HoNC descriptor [11] uses a 4×4 SIFT-like grid, but for each grid cell a simple 8-bin ($2 \times 2 \times 2$) color histogram is computed. The average color intensity is normalized to 127.5 (over all three color channels, not each channel separately) and the average standard deviation of the color channels is normalized to 48. This yields invariance to changes in illumination intensity, but not changes in illumination color. Since the color histograms have 8 buckets per grid cell, this also yields a 128-dimensional descriptor.

2.7 HoWH: Histograms of Weighted Hues

Luke, Keller, and Chamorro-Martinez [10] have suggested stacking the SIFT descriptor with a similar descriptor that replaces the gradient orientation and magnitude at each pixel with the hue and saturation. We consider here a version that is not (necessarily) stacked with the SIFT descriptor. Since hue is a circular value (like gradient orientation) and the saturation describes the strength of the hue (similar to gradient magnitude), a similarly structured descriptor results. This descriptor has the drawback that it will not work on grayscale images (or others with low saturation), since all grays have undefined hue. As suggested by van de Weijer and Schmid for their hue-based descriptor [9], when combining this descriptor with others, we weight it 60% as much as other descriptors, and this improves the performance.

2.8 CNN3: Deep Convolutional Descriptor

Simo-Serra et al. [18] developed a descriptor learned using a deep convolutional neural network. The input to the network is a 64×64 grayscale image patch. The network generates a 128-dimensional descriptor similar to SIFT and related descriptors. This descriptor was demonstrated to be superior to SIFT and recently developed competitors, including DAISY [19] and a state-of-the-art learned descriptor [20].

This technique doesn't fit clearly into the class of shape descriptors or texture descriptors, but it undoubtedly uses both shape and texture cues. While it is

not invariant to any illumination changes (intensity, shift, or color), it is resilient to such changes and works well in practice. Interestingly, it generates descriptor vectors that are more correlated than the other methods (they have a smaller average angle between them). This is important when combining them with other methods. The vectors require lengthening for them to have equivalent weight in the combined score. We use an empirically determined scale factor of 5. This allows combinations of the descriptors to outperform individual descriptors.

2.9 Summary

In addition to the descriptors discussed below, we considered SURF [4], rgSIFT [8], and C-SIFT [8,21,22] descriptors. However, these performed poorly in our previous work [11] and are not included here.

Table 1 summarizes the characteristics of the descriptors. We classify SIFT, RGB-SIFT, and OpponentSIFT as shape descriptors, since they are primarily based on gradient orientations. We classify SPIN, CSPIN, HoNI, and CHoNI as texture descriptors, since they are based on spatial relationships of intensity or individual color channels. We classify HoWH and HoNC as color descriptors, since they do not separate the color channels in constructing the descriptor. CNN3 incorporates both shape and texture information.

3 Keypoint Recognition

In order to recognize keypoints between images, we first detect the keypoints in the images using the SURF keypoint detector [4]. We have found this detector to be fast and to generate good features for matching.[1] When the detector finds

Table 1. Characteristics of descriptors.

Name	Size	Type	Illumination Invariance
SIFT	128	shape	intensity + shift
RGB-SIFT	384	shape (color)	intensity + shift + color
OpponentSIFT	384	shape (color)	intensity + shift
CNN3	128	shape/texture	none
SPIN	128	texture	intensity + shift
CSPIN	384	texture (color)	intensity + shift + color
HoNI	128	texture	intensity + shift
CHoNI	384	texture (color)	intensity + shift + color
HoNC	128	color	intensity + shift
HoWH	128	color	intensity + shift

[1] Note that this is the SURF keypoint detector, not the descriptor, which has not performed well in our experiments [11].

more than 1000 keypoints, only the top 1000 are retained for each image in order to maintain efficiency and relevance. Descriptors for each keypoint are constructed using the techniques described above. Individual descriptors may be used, but we also consider combinations of descriptors that are concatenated (i.e., stacked) into longer descriptors. When combined, each individual descriptor vector is scaled to have the same length, regardless of size (except as noted above for HoWH and CNN3).

The best match for each keypoint in the reference image is found in the target image using the Euclidean distance between the keypoint descriptors. Matches are considered correct if the projection of the keypoint location into the other image (using a known homography) lies within the computed size of the corresponding keypoint (and in reverse).

We measure the matching performance of each descriptor using the mean average precision as follows. The precision and recall are defined as:

$$\text{precision} = \frac{\#\text{ correct matches detected}}{\#\text{ total matches detected}} \tag{2}$$

$$\text{recall} = \frac{\#\text{ correct matches detected}}{\#\text{ keypoints possible to detect}} \tag{3}$$

In computing the recall, we exclude from the denominator those keypoints from the reference image that do not appear in the target image (because they have moved outside the boundaries of the image). We do not exclude keypoints that appear in the target image, but that are missed by the keypoint detector. As the threshold on descriptor distance varies, the number of matches changes and the precision versus recall can be plotted. The *average precision* is the average of the precision over the interval $r \in [0, 1]$ (the area under the curve). The *mean average precision* computes the mean over multiple plots. The maximum value is one and the minimum is zero. Figure 1 shows an example plot of precision versus recall for one image pair.

We ran experiments with each descriptor and many combinations of descriptors on the Oxford affine covariant regions data set[2] that models variations in viewpoint, rotation, zoom, lighting, blur, and compression. All six images (five pairs with the same reference image) of each of the eight data subsets were used. Some pairs are difficult to match, with no combination of descriptors achieving an average precision above 0.05. Others are straightforward, with most techniques performing well. This tends to compress the MAP differences between descriptors.

4 Results

We tested every possible combination of five (or less) descriptors from the set described in Sect. 2 using the process from Sect. 3. Table 2 shows the top performing combinations sorted first by the number of descriptors (or equivalently the descriptor norm) and then by the maximum average precision.

[2] http://www.robots.ox.ac.uk/~vgg/data/data-aff.html.

Table 2. Top descriptor combinations by number of descriptors (vector norm).

Number	Descriptors	Size	MAP
1	CHoNI	384	.5330
1	CNN3	128	.5267
1	OpponentSIFT	384	.5237
1	RGBSIFT	384	.5185
1	HoNI	128	.5139
1	SIFT	128	.5136
1	HoNC	128	.5043
1	CSPIN	384	.4951
1	SPIN	128	.4403
1	HoWH	128	.3206
2	CNN3+CHoNI	512	.5704
2	CNN3+HoNC	256	.5668
2	CNN3+CSPIN	512	.5650
2	CHoNI+OpponentSIFT	768	.5621
2	CNN3+HoNI	256	.5604
2	CHoNI+RGBSIFT	768	.5599
2	CHoNI+SIFT	512	.5589
2	CNN3+OpponentSIFT	512	.5577
2	HoNC+OpponentSIFT	512	.5550
2	HoNI+OpponentSIFT	512	.5550
3	CNN3+CHoNI+OpponentSIFT	896	.5781
3	CNN3+CHoNI+HoWH	640	.5770
3	CNN3+CHoNI+RGBSIFT	896	.5765
3	CNN3+CHoNI+SIFT	640	.5758
3	CNN3+HoNC+OpponentSIFT	640	.5733
3	CNN3+CSPIN+OpponentSIFT	896	.5728
3	CNN3+CSPIN+HoNC	640	.5727
3	CNN3+HoNC+RGBSIFT	640	.5725
3	CNN3+CHoNI+CSPIN	896	.5724
3	CNN3+HoNI+OpponentSIFT	640	.5722
4	CNN3+CHoNI+HoWH+OpponentSIFT	1024	.5848
4	CNN3+CHoNI+HoWH+RGBSIFT	1024	.5842
4	CNN3+CHoNI+HoWH+SIFT	768	.5839
4	CNN3+CSPIN+HoWH+OpponentSIFT	1024	.5826
4	CNN3+CSPIN+HoWH+RGBSIFT	1024	.5823
4	CNN3+HoNI+HoWH+OpponentSIFT	768	.5818
4	CNN3+CSPIN+HoWH+SIFT	768	.5813
4	CNN3+HoNI+HoWH+RGBSIFT	768	.5810
4	CNN3+CSPIN+HoNC+OpponentSIFT	768	.5804
5	CNN3+CHoNI+HoWH+CSPIN+OpponentSIFT	1408	.5857
5	CNN3+CHoNI+HoWH+CSPIN+RGBSIFT	1408	.5855
5	CNN3+HoNC+HoWH+CSPIN+OpponentSIFT	1152	.5854
5	CNN3+HoNC+HoWH+CSPIN+RGBSIFT	1152	.5854
5	CNN3+HoNC+HoWH+CSPIN+SIFT	896	.5849

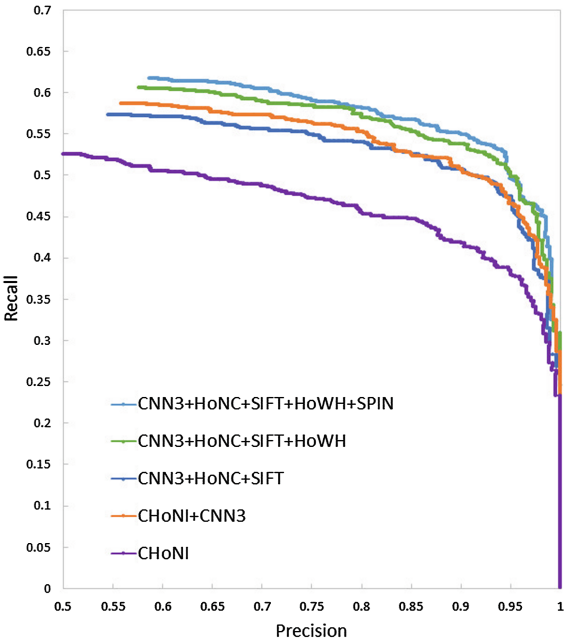

Fig. 1. Example precision/recall plot from the first pair of "bark" images in the Oxford data set.

For individual descriptors, three shape/texture descriptors that incorporate all color channels performed well, with CHoNI surpassing OpponentSIFT and RGB-SIFT. CNN3 was, by far, the strongest descriptor that did not include color information. SIFT is near the middle of these uncombined descriptors. Surprisingly, even a simple histogram of normalized intensities (HoNI) performed equivalently. It is clear that using SIFT as the only baseline for comparison to new descriptors is no longer sufficient to demonstrate the state-of-the-art.

Thirty-three pairs of descriptors (when stacked together) surpassed the best individual descriptor. The top three combined CNN3 with a descriptor that incorporates color. CHoNI also performed well when combined with the shape descriptors.

When triples of descriptors are considered, fifteen combinations surpassed the best pair. Most of these combine a descriptor from each type (shape, color, texture). All of the top performers combined CNN3, CHoNI, and another descriptor. CNN3 was included in the top twelve combinations.

Nineteen quadruples of descriptors surpassed the top triple. The top performers all combined CNN3, a shape-based descriptor, a color-based descriptor, and a texture-based descriptor. This demonstrates that the complementarity of the descriptors is important to improved keypoint recognition performance. While HoWH was a poor performer by itself, it is included in most of the top quadruples, indicating that it includes information that is not redundant with the other descriptors.

Only five quintuples were able to (barely) surpass the best performing quadruple. We have reached the limit of this set of descriptors. To improve performance beyond this point, we would require additional descriptors that incorporate information unused by the current set.

Another way of looking at the descriptor combinations is with respect to the number of elements in the descriptor vector (a minimum of 128 for the smallest vectors used in this work). Table 3 shows the results. The top five for each vector size are shown, unless fewer surpass previous (smaller) vectors. From

Table 3. Top performing descriptor combinations by number of elements (vector size).

Size	Descriptors	Number	MAP
128	CNN3	1	.5267
128	HoNI	1	.5139
128	SIFT	1	.5136
128	HoNC	1	.5043
128	SPIN	1	.4403
256	CNN3+HoNC	2	.5658
256	CNN3+HoNI	2	.5604
256	HoNC+SIFT	2	.5522
256	CNN3+SIFT	2	.5518
256	HoNI+SIFT	2	.5506
384	CNN3+HoNC+SIFT	3	.5715
384	CNN3+HoNI+HoWH	3	.5711
384	CNN3+HoNI+SIFT	3	.5694
384	CNN3+HoNC+SPIN	3	.5684
384	CNN3+HoWH+SIFT	3	.5675
512	CNN3+HoWH+HoNI+SIFT	4	.5800
512	CNN3+HoWH+HoNC+SIFT	4	.5779
512	CNN3+HoWH+SPIN+SIFT	4	.5776
512	CNN3+HoNC+SPIN+SIFT	4	.5763
512	CNN3+HoWH+HoNC+SPIN	4	.5738
640	CNN3+HoNC+HoWH+SPIN+SIFT	5	.5827
640	CNN3+HoNI+HoWH+SPIN+SIFT	5	.5801
768	CNN3+CHoNI+HoWH+SIFT	4	.5839
896	CNN3+HoNC+HoWH+CSPIN+SIFT	5	.5849
1152	CNN3+HoNC+HoWH+CSPIN+OpponentSIFT	5	.5854
1152	CNN3+HoNC+HoWH+CSPIN+RGBSIFT	5	.5854
1408	CNN3+CHoNI+HoWH+CSPIN+OpponentSIFT	5	.5857
1408	CNN3+CHoNI+HoWH+CSPIN+RGBSIFT	5	.5855

Fig. 2. Matching examples. The top matches detected using CNN3+HoNC+SIFT on four example pairs. The top pair shows the 100 best matches. The other pairs show the 200 best matches.

this perspective, CNN3, HoNI, SIFT, and HoNC are the top short descriptors. Even at size 256 (shorter than RGBSIFT, OpponentSIFT, CHoNI and CSPIN), we are able to achieve significant improvements by combining CNN3 and/or SIFT with other short descriptors (CNN3+HoNC performs best).

At 384 elements, we improve noticeably on the longer single descriptors by combining three shorter descriptors. The top performers generally combine CNN3 with a color descriptor and a shape or texture descriptor (one exception combines CNN3 with both a shape and texture descriptor). Combining four short descriptors yields some additional gains (mostly CNN3 combined with a short descriptor of each type). Beyond that, the gains are even smaller, with the best overall descriptor (CNN3+CHoNI+HoWH+CSPIN+OpponentSIFT) besting the top 512-element vector only .5857 to .5800. This suggests that descriptors with size between 256 and 512 elements achieve the best trade-off between vector size (i.e., computation time) and performance.

Overall, CNN3 comes out a big winner, participating in most of the top combinations. The drawback to CNN3 is that it requires significantly higher computational expense when compared to SIFT or similar descriptors. A GPU implementation can improve this, but it is still not competitive with the speed of SIFT [18].

Figure 2 shows four examples of the use of a combination of descriptors. These examples combine CNN3, HoNC, and SIFT, which comprise the top performing descriptor with 384 elements. Despite changes in scale, perspective, and illumination, the descriptor combination finds a large number of correct matches, with very few mismatches.

5 Conclusions

We have demonstrated that improved recall/precision results can be achieved by stacking multiple keypoint descriptors. In particular, a combination of descriptors that use shape, color, and texture information significantly improve upon those that use a single modality (or even two). We believe that this will prove true not just for the descriptors studied here, but also more generally. This would imply that nearly every previously published descriptor could benefit through a combination with additional complementary descriptors.

Acknowledgment. This work was supported, in part, by a Worthington Distinguished Scholar award from the University of Washington Bothell.

References

1. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)

3. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1615–1630 (2005)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: speedup up robust features. Comput. Vis. Image Underst. **110**, 346–359 (2008)
5. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Proceedings of the European Conference on Computer Vision, pp. 778–792(2010)
6. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Proceedings of the International Conference on Computer Vision, pp. 2564–2571 (2011)
7. Bosch, A., Zisserman, A., Muñoz, X.: Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Anal. Mach. Intell. **30**, 712–727 (2008)
8. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1582–1596 (2010)
9. van de Weijer, J., Schmid, C.: Coloring local feature extraction. In: Proceedings of the European Conference on Computer Vision, pp. 334–348 (2006)
10. Luke, R.H., Keller, J.M., Chamorro-Martinez, J.: Extending the scale invariant feature transform descriptor into the color domain. ICGST J. Graph. Vis. Image Process. **8**, 35–43 (2008)
11. Olson, C.F., Zhang, S.: Keypoint recognition with histograms of normalized colors. In: Proceedings of the 13th Conference on Computer and Robot Vision (2016)
12. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1265–1278 (2005)
13. Gehler, P., Nowozin, S.: On feature combination for multiclass object recognition. In: Proceedings of the International Conference on Computer Vision, pp. 221–228 (2009)
14. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. Int. J. Comput. Vis. **73**, 213–238 (2007)
15. Bo, L., Ren, X., Fox, D.: Kernel descriptors for visual recognition. In: Advances in Neural Information Processing Systems 23, pp. 244–252 (2010)
16. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comput. Vis. **65**, 43–72 (2005)
17. Hess, R.: An open-source SIFT library. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 1493–1496 (2010)
18. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative learning of deep convolutional feature point descriptors. In: Proceedings of the International Conference on Computer Vision, pp. 118–126 (2015)
19. Tola, E., Lepetit, V., Fua, P.: DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 815–830 (2010)
20. Simonyan, K., Vedaldi, A., Zisserman, A.: Learning local feature descriptors using convex optimisation. IEEE Trans. Pattern Anal. Mach. Intell. **36**, 1573–1585 (2014)
21. Abdel-Hakim, A.E., Farag, A.A.: CSIFT: a SIFT descriptor with color invariant characteristics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1978–1983 (2006)
22. Burghouts, G.J., Geusebroek, J.M.: Performance evaluation of local colour invariants. Comput. Vis. Image Underst. **113**, 48–62 (2009)