

A Local Image Descriptor Robust to Illumination Changes

Sebastian Zambanini and Martin Kampel

Computer Vision Lab, Vienna University of Technology, Austria
{zamba, kempel}@caa.tuwien.ac.at

Abstract. In this paper we address the problem of building a local image descriptor that is insensitive to the complex appearance changes induced by illumination variations on non-flat objects. The presented descriptor is based on multi-scale and multi-oriented even Gabor filters and constructed in such a way that typical effects of illumination variations like changes of edge polarity or spatially varying brightness changes are taken into account for illumination insensitivity. For evaluation, a dataset of textured as well as textureless objects is used which introduces a greater challenge towards evaluating the robustness against illumination changes than conventional datasets used in the past. The experiments finally show the superiority of our descriptor compared to existing ones under illumination changes.

1 Introduction

Representing images by sets of local image descriptors has shown to be an effective approach for computer vision tasks like image correspondence search [1] or object recognition [2]. The purpose of local image descriptors is to represent discriminative information of a local image patch. Preferably, the descriptor should be insensitive to certain image transformations like rotation, blurring or brightness change. In this paper, we focus on designing an image descriptor that is insensitive to the illumination conditions in the scene and consequently to the possibly complex appearance variations evoked by changes of the illumination direction.

The computation of image descriptors typically follows a common workflow, where the main components are the extraction of pixel-wise low-level features and their spatial pooling [3]. Image gradient directions are a popular low-level feature, e.g. used by SIFT [4], DAISY [5], WLD [6], GLOH [7] or MROGH [8]. Another type of low-level features are filter bank responses, e.g. Haar wavelets (SURF [9]), higher-order derivatives [10] or steerable filters (the best descriptor reported in [3]). Encoding the local image structures by means of a set of patterns has also been proposed, e.g. local binary patterns (CSLBP [11], MRRID [8]), local ternary patterns [12], local intensity order patterns (LIOP [13]). The OSID descriptor [14] applies an ordinal labeling of pixel intensities.

The role of the spatial pooling stage is to encode the spatial distribution of the per-pixel features. The standard SIFT implementation uses a squared 4×4 grid

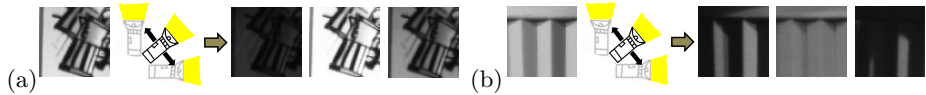


Fig. 1. Image patches showing the same object regions under different lighting conditions; (a) a flat, textured object; (b) a non-flat, textureless object

for the spatial pooling and other descriptors like SURF, OSID or CSLBP adopt this pooling scheme. GLOH uses a log-polar grid for spatial pooling. Similarly, the DAISY descriptor uses circular cells of varying size arranged on concentric rings for spatial pooling and Brown et al. [3] proposed a learning scheme to obtain optimal configurations of such cell arrangements. MROGH, MRRID and LIOP do not use fixed spatial cells but define cell locations by tiling the image patch based on pixel-intensity orders. This has the advantage that the resulting descriptor is inherently invariant to image rotations.

There is also a line of research which rather focuses on the spatial correlation of the features than on their spatial distribution. The GLAC descriptor [15] computes the correlation of gradient directions between pixel pairs. Binary descriptors like FREAK [16] are constructed by comparing image intensities over the local patch.

Common to all descriptors is that the final outcome is a feature vector (e.g. by concatenating all sub-histograms of the spatial cells), to which various optional postprocessing steps can be applied. Normalization of the vector to unit length is done to make the descriptor invariant to global linear brightness changes and to reduce the influence of large descriptor elements [3,4,9,11,15]. Dimensionality reduction techniques and quantization can be applied to improve both the descriptors' performance and storage space [3,7,16].

When it comes to changes of the illumination conditions in a scene, current image descriptors are in general only invariant to linear or monotonic brightness changes which occur on flat, textured objects (i.e. object with varying albedo), as shown in Figure 1(a). However, on non-flat, *textureless* objects (i.e. objects with constant albedo) more challenging appearance changes occur that include locally varying brightness changes and edge polarity changes, as shown in Figure 1(b).

In this paper, we propose a new local descriptor for application areas where strong illumination changes can be expected and texturedness as well as flatness of objects are not necessarily given. This is achieved in the first place by using pixel-wise low-level features that exhibit a higher recognition power under changing lighting conditions than the low-level features commonly used in existing descriptors. Osadchy et al. [17] propose to use oriented second derivative filters of Gaussians as an effective feature for capturing smooth/isotropic as well as non-smooth/anisotropic surface characteristics. However, they use single scale filters for recognizing objects in aligned images. We extend their approach towards local descriptor construction by applying multi-scale filters and incorporating spatial statistics in order to describe local image patches in an illumination-insensitive way. In our work, we use the real part of Gabor filters [18] instead of second

derivatives of Gaussians due to the similarity but higher flexibility of Gabor filters. This makes us able to explore the influence of the Gabor filter parameters on recognition performance.

As a further contribution of the paper, we evaluate the robustness of existing local image descriptors against changes in illumination conditions by using a new image patch database extracted from images of the ALOI dataset [19]. Existing evaluations [7,20,21] have the problem of only marginally considering illumination invariance: Mikolajczyk and Schmid [7] and van de Sande et al. [21] only test image brightness changes and ignore changes of the light source direction. Moreels and Perona [20] use only three different lighting configurations and couple the descriptor performance evaluation with the interest point detection step. Our experiments aim at a broader evaluation: testing the robustness of the small image patch descriptions against changes of the light source direction for textured as well as textureless objects, the latter being the more challenging type of objects.

2 Methodology

2.1 2D Gabor Filters

A 2D Gabor filter [18] is a complex filter that consists of a real/even part G_e and an imaginary/odd part G_o ,

$$G_e(x, y) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\omega}\right), G_o(x, y) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} \sin\left(2\pi \frac{x'}{\omega}\right) \quad (1)$$

with $x' = x \cos \theta + y \sin \theta$ and $y' = -x \sin \theta + y \cos \theta$. The filters have the form of a sinusoidal plane wave with wavelength ω and orientation θ multiplied by a Gaussian envelope with standard deviation σ . In order to ensure equal shape of Gabor filters of different sizes, σ is defined as a linear function of ω by $\sigma = c \cdot \omega$. The parameter γ is the spatial aspect ratio of the filter.

2.2 Local Image Descriptor from Even Gabor Filter Responses

According to Osadchy et al. [17], for illumination insensitivity smooth image regions are best described by a whitening filter, which reduces the autocorrelation within the image signal, whereas non-smooth image regions are best described by a directional filter. Therefore, oriented second derivative filters of Gaussians are a good choice for general cases, as they combine the Laplacian-of-Gaussian whitening effects with a directional filter. The authors also mention the similarity between second derivative filters and even Gabor filters and report similar results for both filters. This similarity is given when the cosine bandwidth is selected such that the Gaussian envelope roughly covers the cosine range of $[-1.5\pi, 1.5\pi]$, i.e., $c \approx 0.4$ (see Figure 3(c) in Section 3.1). In our work, we use the even part of Gabor filters in order to have more flexibility for testing the influence of various filter shapes on the descriptor performance (see Section 3.1).

While the directionality of the filter is given by the rotation parameter θ , the optimal size of the filter depends on the surface properties, as for smoother surface parts a wider filter is needed than for less smooth surface parts. As in a general scenario the surface characteristics are usually unknown and varying, we additionally use a set of different filter scales for creating our descriptor. For given values of the filter shape parameters c and γ , we use N equally spaced orientations $\theta_1 \dots \theta_N$, where $\theta_i \in [0, \pi[$ and $\theta_1 = 0$. The M scales $\omega_1 \dots \omega_M$ are exponentially sampled to achieve homogeneous intervals, i.e. $\omega_j = k^{j-1} \omega_1$.

In order to build our descriptor, we first compute absolute filter responses I^{θ_i, ω_j} by convolving the image patch I with the $N \cdot M$ filters $G_e^{\theta_i, \omega_j}$,

$$I^{\theta_i, \omega_j} = |I \star G_e^{\theta_i, \omega_j}|. \quad (2)$$

These output images are arranged to a feature map F that contains for every image point \mathbf{p} and discrete filter parameters θ_i and ω_j the absolute filter responses,

$$F(\mathbf{p}, \theta_i, \omega_j) = I^{\theta_i, \omega_j}(\mathbf{p}). \quad (3)$$

Obviously, the filter outputs I^{θ_i, ω_j} depend on the image contrast, e.g., stronger ridges produce higher values in $F(\mathbf{p})$. This gives a natural weighting of the local low-level features, in the same manner as, for instance, the gradient magnitude is used to weight the histogram inputs in the SIFT descriptor [4]. Global linear brightness changes on the image patch can then be compensated by normalizing the final histogram to unit length. However, different light source directions can lead to brightness changes that vary locally, as demonstrated in Figure 1(b). Therefore, it is beneficial to normalize F on a per-pixel level,

$$\tilde{F}(\mathbf{p}, \theta_i, \omega_j) = \frac{F(\mathbf{p}, \theta_i, \omega_j)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^M F(\mathbf{p}, \theta_i, \omega_j)^2}}. \quad (4)$$

An option would also be to normalize only over the responses of different orientations, but normalizing over all responses is more robust when parts of the image region are over- or undersaturated. On discrete images, linear brightness changes lead to a clipping of values which are outside the dynamic range of the sensor. The normalized feature map \tilde{F} is not invariant to brightness changes when such effects occur, but normalizing over all orientation and scale responses is more robust in presence of partial over- and undersaturation as wider filters and thus more data samples are included.

The last step of the descriptor construction is to perform a spatial pooling on \tilde{F} to increase the descriptor's discriminative power by adding spatial information. Formally, L cells $C_l(\mathbf{p}), l = 1 \dots L$ are defined that represent the weighting of the spatial location \mathbf{p} for the cell's local sub-histogram. The final descriptor is a 3D joint histogram $H(\theta_i, \omega_j, l)$ of the values in \tilde{F} ,

$$H(\theta_i, \omega_j, l) = \sum_{\mathbf{p} \in \tilde{F}} C_l(\mathbf{p}) \cdot \tilde{F}(\mathbf{p}, \theta_i, \omega_j). \quad (5)$$

The cells C_l can be, for instance, of Gaussian shape to achieve a DAISY-like pooling [3,5] or perform the bilinear weighting between squared cells as in the SIFT descriptor [4]. However, please note that the optimal pooling scheme depends on the application scenario. The goal of spatial pooling is to achieve a optimal trade-off between an maximal descriptor discriminability and a maximal descriptor insensitivity to variations of the features' spatial distribution caused by inaccurate keypoint detection, varying viewpoints and different object instances. Thus, for object matching between viewpoints one can use smaller cells the less viewpoint differences are expected. In our experiments, we use the standard SIFT 4×4 squared cells with bilinear weighting [4] for spatial pooling as it achieved reasonably good results on all datasets.

Although Gabor filters are well known and often used [22], to the best of our knowledge they have never been used before in this manner for local image descriptors. Multi-scale and/or multi-oriented filter banks are used by others for image recognition tasks, e.g. by [23,24] with Gabor filters and by [3] with their single-scale Gaussian derivative counterparts, but the filters are usually used in quadrature, whereas we only use their real part. As shown in our experiments (see Section 3.2), using only the real part achieves similar results under strong illumination variations while saving computation time. Larsen et al. [10] also use filter bank responses to build a local descriptor, but they rely on higher-order derivative filters which are applied to single positions on the patch. In contrast, we establish statistics on the filter responses at all pixel positions of the patch that are well-founded for an illumination-insensitive descriptor by means of measuring the spatially varying frequency of occurrence of locally normalized responses.

3 Experiments

The goal of our experiments is to test a descriptor's ability to describe small image patches in a discriminative way while ignoring the effects arising from different lighting conditions. In contrast to evaluation strategies that combine the descriptor performance evaluation with an evaluation of interest point detection [7,20], we adopt the performance evaluation scheme of Brown et al. [3] which is based on sets of true and false image patch pairs. True patch pairs show the same object patch but from different viewpoints and with different illumination conditions, whereas false patch pairs show different object patches. An optimal descriptor will minimize the distance between the true image patches and maximize the distance between false image patches. Hence, we measure these two groups of distances for a given descriptor and integrate the two histograms to build a ROC curve of which the area under curve (AUC) is computed as performance measure. A highly illumination-insensitive descriptor will have less overlap between the two histograms and thus a higher AUC than a less illumination-insensitive descriptor.

We use three configurations of our descriptor for evaluation: one that uses only single-scale even Gabor filters (SSEG) with dimensionality N , one that uses

multi-scale even Gabor filters (MSEG) with dimensionality $N \cdot M$ and the full descriptor which uses a 4×4 grid for spatial pooling (MSEG4x4) with dimensionality $N \cdot M \cdot L$. We compare these descriptors to several descriptors proposed in literature where implementations are available for download: SIFT [4], SURF [9], DAISY [5], MROGH, MRRID [8], LIOP [13], FREAK [16] and GLAC [15]. Additionally, we implemented the best DAISY descriptor reported in [3] (BEST-DAISY) which uses second order steerable quadrature pair filters, a DAISY-like spatial pooling and a final vector normalization with range clipping. We also tested a modified version of the SIFT descriptor which uses unsigned gradients in the range $[0, \pi[$ to handle the polarity changes of edges on textureless surfaces under opposite lighting directions (UGSIFT).

Experiments are conducted on four datasets of true and false image patch pairs. The datasets *ALOI Textureless Patches* and *ALOI Textured Patches* have been newly generated for our purpose of testing insensitivity against lighting variations. To test the descriptors on scenarios with less lighting variations we also used the existing image patch pair databases *Liberty* and *Virtual World* for evaluation. Examples of true patch pairs contained in the datasets are shown in Figure 2.

ALOI Textureless: patches of this dataset were extracted from the Amsterdam Library of Object Images¹ (ALOI) [19], an image database of 1000 objects that were photographed from three viewpoints and with eight illumination configurations each. We manually identified 80 textureless objects with a wide range of surface smoothness and material BRDF (e.g., a basket, a nut, a sponge, a lemon, a plastic pig...) in the dataset and randomly picked true and false patch pairs of size 64×64 . Correspondences for the true patch pairs were identified by manually estimating the homography between images from the three viewpoints. The viewpoint changes are small enough to describe the image correspondences by a homography and thus errors are also considered as being small enough to just simulate the uncertainty of interest point detection. In total, 60000 true and 60000 false patch pairs were extracted.

ALOI Textured: likewise, we selected 80 textured objects (e.g., labeled boxes, an alarm clock, a calendar, ...) and selected a total set of 120000 true and false patch pairs.

Liberty: the *Liberty* dataset² consists of true and false patches sampled from 3D reconstructions of the Statue of Liberty. This dataset has been used for descriptor learning [3] and represents an appropriate descriptor evaluation dataset for the scenario of multi-view reconstruction of large-scale outdoor objects. Hence, it also includes realistic outdoor lighting variations, although their amount and frequency is unknown. We use 50000 patch pairs for evaluation.

Virtual World: this dataset³ contains 3000 photorealistic images of a virtual city model and has been used by Kaneva et al. [25] for image descriptor evaluation in the same manner as the real image patches in [3]. Likewise to [25],

¹ <http://staff.science.uva.nl/~aloi/>

² <http://www.cs.ubc.ca/~mbrown/patchdata/patchdata.html>

³ <http://people.csail.mit.edu/biliana/projects/iccv2011/>

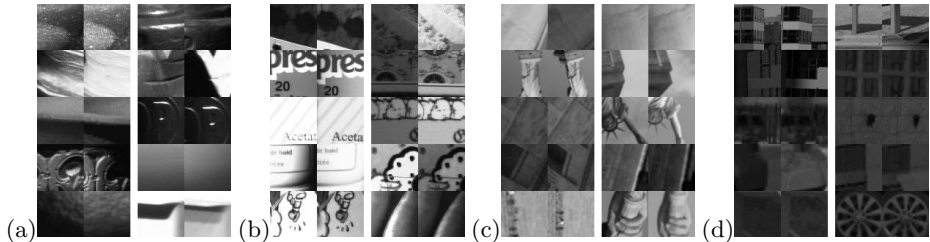


Fig. 2. Examples of true patch pairs in the datasets: (a) *ALOI Textureless*, (b) *ALOI Textured*, (c) *Liberty* and (d) *Virtual World*

we extracted 120000 true patch pairs by identifying corresponding Difference-of-Gaussians keypoints between viewpoints and different times of the day to introduce changing lighting conditions. Finally, we resized the detected patches to a standard size of 64×64 based on the detected scale. The advantage of this dataset over the Liberty dataset is that it exhibits a more controlled and evenly distributed variation of the lighting conditions, as each scene was rendered under five different lighting conditions (different times of the day).

Note that the correct patch pairs of all datasets show no rotation differences. Hence, to allow for a fair comparison, we do not use the rotation-invariant versions of the descriptors, except for MROGH, MRRID and LIOP which are inherently rotation invariant. The other descriptors can be made rotation invariant by determining a canonical orientation per patch and describing the per-pixel features and cells relative to this orientation. The same principle can be used to make our descriptor rotation-invariant, although it is not treated in this paper.

3.1 Gabor Filter Parameter Selection

In order to investigate the relation of our descriptor parameters to the recognition performance, we defined parameter selections in discrete intervals and determined the AUC of all parameter combinations on a selected training set. As our main goal is to achieve an optimal performance under strong lighting variations, we built a mixed dataset by randomly extracting 25000 patches from the representative datasets *ALOI Textureless* and *ALOI Textured*. MSEG4x4 was then modified according to the filter shape parameters c and γ as well as the number of orientations N and tested on the mixed dataset.

Figure 3(a) shows the influence of the number of orientations by plotting the best AUC for a given value of N and all values of c and γ . It can be seen that for $N > 6$ no substantial improvement can be achieved. We therefore chose a value of $N = 6$ for all further experiments. In Figure 3(b) the AUC values for the filter parameters c and γ are shown. It is evident that the best performance is not achieved for filter parameters that make the Gabor filter similar to the second derivative of Gaussian used in [17] ($c \approx 0.4$), but for values of c close to 0.6 where the Gaussian envelope is wider and thus a higher frequency and

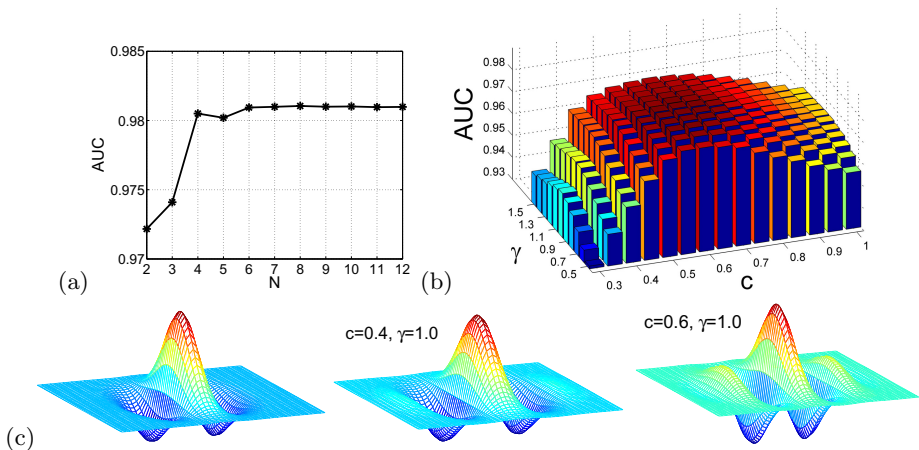


Fig. 3. (a)-(b) Performance of descriptor MSEG4x4 for different values of N , c and γ ; (c) from left to right: second derivative of Gaussian with reversed sign, the corresponding even Gabor filter and the filter used for constructing our descriptor

orientation resolution is provided [22]. This comes at the price of an increased spatial uncertainty of the filter, which, however, seems not to be critical due to the subsequent pooling step. For values of c in this range the aspect ratio γ has only a minor influence on the performance. Based on these results, we used parameter values of $c = 0.6$ and $\gamma = 1$ for our experiments. Figure 3(c) shows the shape of the filter used and demonstrates the difference to the shape of a second derivative of Gaussian filter and its Gabor counterpart. Optimal parameters for the multi-scale spacing have also been determined on this dataset as $\omega_1 = 2$, $k = \sqrt{2}$ and $M = 8$.

3.2 Results and Discussion

Results on the four datasets were achieved by applying the descriptors to all patch pairs but excluding the patches used for parameter selection from the *ALOI Textureless* and *ALOI Textured* datasets. The obtained ROC plots are shown in Figure 4. The corresponding legends list the descriptors sorted by their achieved AUC given in brackets.

For the datasets representing strong illumination variations between patches (*ALOI Textureless* and *ALOI Textured*), all versions of our descriptor outperform the other descriptors, with a larger advance in performance for the more challenging textureless objects. On the ALOI datasets even the descriptor SSEG with a dimensionality of 6 achieves a better recognition performance than the remaining high-dimensional descriptors with dimensionalities of ≥ 64 . MSEG4x4 clearly shows the best performance on these datasets as well as on the *Virtual World* dataset which is assumed to represent more lighting variations than the *Liberty* dataset. On this dataset MSEG4x4 is outperformed by BESTDAISY but shows

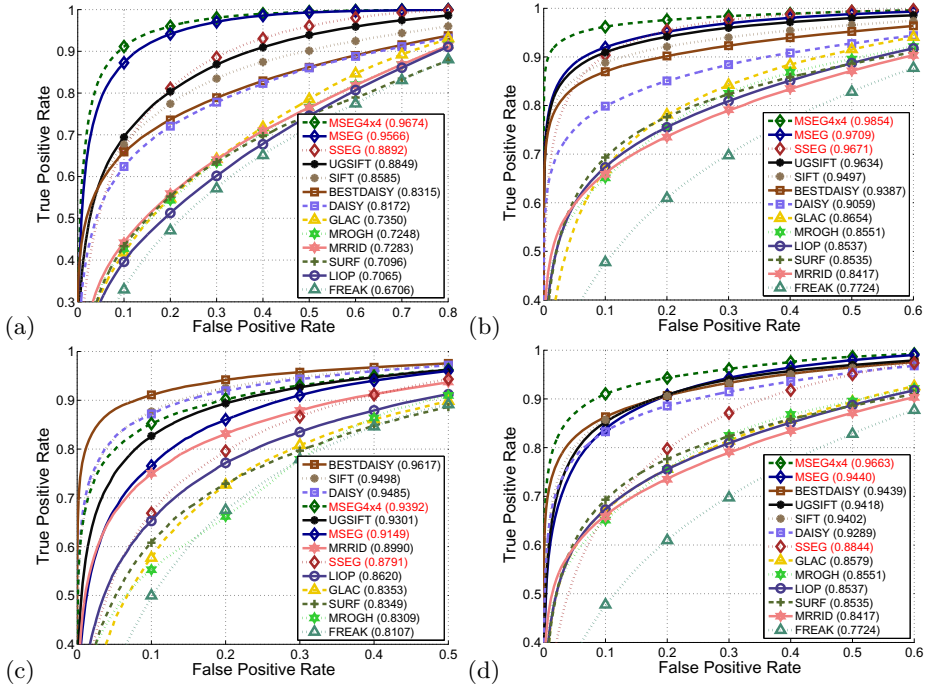


Fig. 4. ROC curves of the descriptors on the datasets (a) *ALOI Textureless*, (b) *ALOI Textured*, (c) *Liberty* and (d) *Virtual World*

nearly the same performance as SIFT and DAISY. However, BESTDAISY has been especially optimized for this dataset. It is worth noting that the best parameters for our descriptor were selected according to the datasets *ALOI Textureless* and *ALOI Textured*, but parameter tuning can also be used to improve the results of our descriptor on the *Liberty* dataset. By using Gabor filters with a shape more similar to a second Gaussian derivative ($c = 0.4$) and $N = 8$, MSEG4x4 is competitive to BESTDAISY on the *Liberty* dataset (AUC=0.9582), while still achieving the best performance on *ALOI Textureless* and *ALOI Textured* (AUC of 0.9563 and 0.9807, respectively). In general, we can conclude that MSEG4x4 shows the best performance under strong illumination changes for a wide range of filter shapes. Even the worst parameter combination with $c = 0.3$ and $\gamma = 0.5$ (see Figure 3(b)) achieves a better performance than the other descriptors on *ALOI Textureless* (AUC=0.9054, not shown in Figure 4).

Among the remaining descriptors, the gradient-based descriptors SIFT and DAISY show the best performance under illumination changes. It is also shown that using unsigned gradients (UGSIFT) is beneficial for the SIFT descriptor by making it invulnerable to edge polarity changes. However, this lowers also the discriminability of the SIFT descriptor and thus its recognition performance is decreased when less illumination changes are present in the data (AUC of 0.9301 compared to 0.9498 on the *Liberty* dataset). Image gradients have

Table 1. Comparison of AUC values achieved when using even or complex Gabor filter responses for our descriptor

	ALOI Textureless	ALOI Textured	Liberty	Virtual World
Even Gabor Filters	0.9674	0.9854	0.9392	0.9663
Complex Gabor Filters	0.9666	0.9855	0.9397	0.9656

previously been mentioned to exhibit illumination-insensitivity properties [17,26] and thus descriptors, which rely on per-pixel features that are not well adapted to the problem of changing lighting conditions (GLAC, MROGH, MRRID, LIOP, SURF, FREAK), have a worse performance compared to SIFT and DAISY.

We have noted before that we do not use Gabor filters in quadrature but only the even part due to the similarity to second derivatives of Gaussians [17]. In Table 1 the results achieved by either using the even Gabor filters or the entire complex filters are compared. It is shown that using the magnitude of both the even and odd filter as feature map does not contribute to considerably better results, while consuming twice the computational power. The advantage of the complex filters is in general that the response is invariant to the phase of the signal, but this does not help to improve illumination insensitivity, as has also been noted by Osadchy et al. [17].

4 Conclusions

Our experiments reveal that current image descriptors, while performing reasonably well in scenarios with textured objects and only low changes of illumination conditions, show a tremendous decrease of performance in scenarios with strong changes of illumination conditions, especially when textureless objects are involved. The absence of texture on objects as well as strong illumination variations makes the recognition more challenging and this scenario has been neglected in descriptor design and evaluation in the past. We proposed a descriptor based on even Gabor filter responses that is more robust against the effects caused by changing lighting conditions on non-flat surfaces, and thus gain a recognition performance boost in such scenarios. Future work will focus on improving the descriptor with respect to computational time, dimensionality and rotation invariance.

Acknowledgement. This research has been supported by the Austrian Science Fund (FWF) under the grant TRP140-N23-2010 (ILAC).

References

1. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: SIGGRAPH, pp. 835–846. ACM, New York (2006)
2. Zhang, J., Marszaek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. IJCV 73, 213–238 (2007)

3. Brown, M., Hua, G., Winder, S.: Discriminative learning of local image descriptors. *PAMI* 33(1), 43–57 (2011)
4. Lowe, D.: Distinctive image features from scale-invariant keypoints. *IJCV* 60(2), 91–110 (2004)
5. Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. *PAMI* 32(5), 815–830 (2010)
6. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: Wld: a robust local image descriptor. *PAMI* 32(9), 1705–1720 (2010)
7. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *PAMI* 27(10), 1615–1630 (2005)
8. Fan, B., Wu, F., Hu, Z.: Rotationally invariant descriptors using intensity order pooling. *PAMI* 34(10), 2031–2045 (2012)
9. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). *CVIU* 110(3), 346–359 (2008)
10. Larsen, A.B.L., Darkner, S., Dahl, A.L., Pedersen, K.S.: Jet-based local image descriptors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part III. LNCS*, vol. 7574, pp. 638–650. Springer, Heidelberg (2012)
11. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. *Pattern Recognition* 42(3), 425–436 (2009)
12. Gupta, R., Patil, H., Mittal, A.: Robust order-based methods for feature description. In: *CVPR*, pp. 334–341 (2010)
13. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: *ICCV*, pp. 603–610 (2011)
14. Tang, F., Lim, S., Chang, N., Tao, H.: A novel feature descriptor invariant to complex brightness changes. In: *CVPR*, pp. 2631–2638 (2009)
15. Kobayashi, T., Otsu, N.: Image feature extraction using gradient local auto-correlations. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part I. LNCS*, vol. 5302, pp. 346–358. Springer, Heidelberg (2008)
16. Alahi, A., Ortiz, R., Vandergheynst, P.: Freak: Fast retina keypoint. In: *CVPR*, pp. 510–517 (2012)
17. Osadchy, M., Jacobs, D., Lindenbaum, M.: Surface dependent representations for illumination insensitive image comparison. *PAMI* 29(1), 98–111 (2007)
18. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2, 1160–1169 (1985)
19. Geusebroek, J., Burghouts, G., Smeulders, A.: The amsterdam library of object images. *IJCV* 61(1), 103–112 (2005)
20. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3d objects. *IJCV* 73(3), 263–284 (2007)
21. Van De Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. *PAMI* 32(9), 1582–1596 (2010)
22. Kamarainen, J., Kyrki, V., Kalviainen, H.: Invariance properties of gabor filter-based features-overview and applications. *TIP* 15(5), 1088–1099 (2006)
23. Ahonen, T., Pietikäinen, M.: Image description using joint distribution of filter bank responses. *PRL* 30(4), 368–376 (2009)
24. Kyrki, V., Kamarainen, J., Kälviäinen, H.: Simple gabor feature space for invariant object recognition. *PRL* 25(3), 311–318 (2004)
25. Kaneva, B., Torralba, A., Freeman, W.: Evaluation of image features using a photorealistic virtual world. In: *ICCV*, pp. 2282–2289 (2011)
26. Chen, H., Belhumeur, P., Jacobs, D.: In search of illumination invariants. In: *CVPR*, pp. 254–261 (2000)