Extending the sGLOH descriptor

Fabio Bellavia and Carlo Colombo

Computational Vision Group, University of Florence, Italy,
{fabio.bellavia, carlo.colombo}@unifi.it

Abstract. This paper proposes an extension of the sGLOH keypoint descriptor which improves its robustness and discriminability. The sGLOH descriptor can handle discrete rotations by a cyclic shift of its element thanks to its circular structure but its performances can decrease when the keypoint relative rotation is in between two sGLOH discrete rotations. The proposed extension couples together two sGLOH descriptor of the same patch with different rotations in order to cope with this issue and it can be also applied straightly to the sCOr and sGOr matching strategies of sGLOH. Experimental results show a consistent improvement of the descriptor discriminability while different setups can be used to reduce the running time according to the desidered task.

Keywords: SGLOH, SIFT, keypoint descriptor, matching

1 Introduction

Keypoint descriptors are valid and useful tools in many computer vision tasks, such as recognition, tracking and 3D reconstruction [10]. In particular descriptors encode properties of image portions extracted by a keypoint detector [11] into numerical vectors in order to evaluate their similarities.

Modern descriptors are mainly based on histogram concatenation such as SIFT and its variants [3,7], MROGH [6] or LIOP [12]. Recently, although they still cannot fully reach the discriminability power of histogram based descriptors, descriptors based on multiple binary comparisons [4], learning [5] and alternative subspace representations [13] received a great effort too due to their interesting properties such as really fast computation and high compactness.

Among the SIFT variants, the sGLOH descriptor [3] coupled with specified matching strategies achieved results comparable with the state-of-the-art LIOP and MROGH descriptors. In particular, the sGLOH descriptor uses a circular grid to incorporate more descriptor instances of the same patch at different orientations into a single feature vector, accessible by a simple cyclic shift of the feature vectors. The matching distance between two features is obtained as the minimum distance among all descriptors for the possible discrete orientations handled by sGLOH. Further improvements can be obtained by limiting the range of allowable orientations according to the scene context. Two matching strategies are derived, namely sCOr and sGOr, respectively when the

orientation range constraint is defined a priori by the user or obtained without user intervention by inspecting the data.

Although robust and efficient, the sGOR descriptor performances can degrade in the case the true relative rotation between the two keypoint patches is equal to half of the sGLOH handled discrete rotation. In order to cope with this issue, we proposed to couple together two sGLOH descriptors for the same patch, with a rotation shift equal to half sGLOH discrete rotation.

A brief introduction to the sGLOH descriptor is given in Sect. 2, also providing details about the matching strategies and the sGLOH worst case. Next, Sect. 3 describes the proposed solution together with the different matching strategies which can rise from the proposed sGLOH extension. Evaluation results on the comparison of the proposed method against the sGLOH and other descriptors is given in Sect. 4, while an analysis of the running times in provided in Sect. 5. Conclusions and future work are discussed on Sect. 6.

2   sGLOH Description

Given a normalized image patch, so that its intensity values are affine normalized and its shape equals the unit circle, the sGLOH descriptor [3] grid is made up of $n \times m$ regions, obtained by splitting n circular rings into m sectors (see Fig. 1).
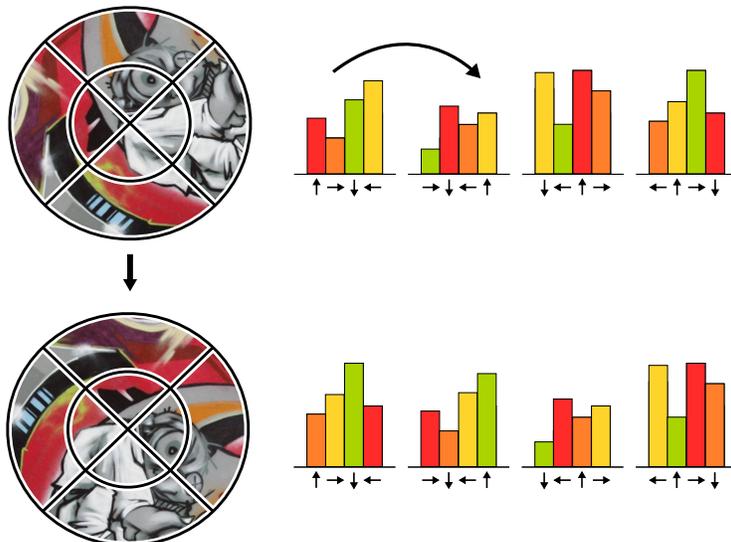

Fig. 1: Rotation of an image patch with the superimposed sGLOH grid by a factor $\frac{2\pi}{m}$ (left), which corresponds to a cyclic shift of the block histogram for each ring (right). In the example $n = 2$ and $m = 4$ (best viewed in color)

For each region the histogram of the m quantized orientations $d_k = k\frac{2\pi}{m}$, $k = 0, \ldots, m-1$ weighted by the gradient magnitude is computed. Gaussian kernel density estimation is used for interpolation and a block histogram H_{ij} with $i = 0, \ldots, n-1$ and $j = 0, \ldots, m-1$, associated to the each sGLOH region is defined by the ordering the computed gradient histogram so that its first bin corresponds to the normal direction of the region and the others follow in clockwise order. The concatenation H of the block histogram, normalized to the unit length on L_1, gives the final sGLOH descriptor.

As shown in Fig. 1, the rotation of the descriptor by a factor d_k, is obtained by k cyclic shifts of the block histogram for each ring, without recomputing the descriptor vector. In this sense, the sGLOH descriptor packs m different descriptors of the same patch with for several discrete dominant orientations. The distance between two sGLOH features H and $\overline{H}$ is then given by

$$\widehat{\mathcal{D}}(H, \overline{H}) = \min_{k=0,\ldots,m-1} \mathcal{D}(H, \overline{H}_k) \tag{1}$$

where $\mathcal{D}(\cdot, \cdot)$ is a generic distance measure and $\overline{H}_k$ is the cyclic shift applied to the descriptor $\overline{H}$ to get the discrete rotation d_k. According to [3] setting $n = 2$ and $m = 8$, which implies that the descriptor dimension is $l = nm^2 = 128$ and the discrete orientation step is $45°$, gives the best sGLOH setup.

A further improvement on the sGLOH can be obtained by limiting the range of the discrete orientations so that part of wrong matches are discarded and cannot be selected by chance. In the sCOr approach one sets a priori $k = 0, 1, m-1$ in (1), i.e. the range of allowable orientations is limited to the first clockwise and counter-clockwise discrete rotations. Although sCOr handles rotations of up to $\pm 67.5°$ only, the method is general enough to be employed in a lot of common practical applications, such as SLAM [1] and sparse matching [9], since transformations are relatively continuous for close images.

The sGOr approach uses instead the scene context to provide a global reference orientation, under the reasonable assumption that all keypoints of the scene undergo roughly the same approximated discrete rotation g, not known a priori. The range of discrete orientations in (1) is modified to $k = g-1, g, g+1$, where $g \in \{0, 1, \ldots, m-1\}$ and the computations are done by using the modulo m arithmetic to take into the account the cyclic nature of k. The value of g is estimated according to the most probable relative orientation among all matches, see [3] for more details.

3 The sGLOH2 Extension

The sGLOH descriptor, especially if coupled with the sCOr and sGOr matching strategies, obtains results comparable with state-of-the-art descriptors [3], but can suffer to performance degradations when the relative rotation between the patches approaches that between two discrete sGLOH rotations, i.e. it is of the form $k\frac{2\pi}{m} + \frac{\pi}{m}$ for $k = 0, \ldots, m-1$.

In order to solve this issue, we define a novel sGLOH2 descriptor $H^\star$, obtained by concatenating two sGLOH descriptors H^1 and H^2 of the

same patch, where H^1 is the standard sGLOH descriptor of the patch, while H^2 is is obtained after applying a rotation of $\frac{\pi}{m}$ to the patch. This corresponds to double the number of possible discrete directions m, leaving the same histogram dimension. Note that this is not equal to double m in the sGLOH parameters, as in this case a smaller and less discriminant descriptor regions would be obtained, since the number of circular sectors will be doubled.

Assuming the sequence $\{0, \frac{\pi}{m}, \frac{2\pi}{m}, \frac{3\pi}{m}, \ldots\}$ of the $2m$ successive discrete rotations by step $\frac{\pi}{m}$, the corresponding ordered set of cyclic shifted descriptors is given by

$$Q(H^\star) = \{H_0^1, H_0^2, H_1^1, H_1^2, \ldots, H_{m-1}^1, H_{m-1}^2\} \tag{2}$$

where H_k^z is the cyclic shift applied to the descriptor H^z, $z \in \{1, 2\}$, to get the discrete rotation d_k as in Sect. 2. The distance between two sGLOH2 features $H^\star$ and $\overline{H}^\star$ becomes

$$\widehat{\mathcal{D}}_2(H^\star, \overline{H}^\star) = \min_{K \in Q(\overline{H}^\star)} \mathcal{D}(H_0^1, K) \tag{3}$$

The sGLOH2 descriptor length is $l^\star = 2l = 256$ and different matching strategies can be obtained in analogy with the sCOr and sGOr matching. In particular, similar to the sCOr strategy we can define more sCOr2 strategies. By limiting the rotations up to $\pm\frac{\pi}{m}$, i.e. using the subset $\{\overline{H}_0^1, \overline{H}_2^0, \overline{H}_{m-1}^2\}$ instead of $Q(\overline{H}^\star)$ in (3) we get the sCOr2.1 strategy. A wider rotation range, up to $\pm 2\frac{\pi}{m}$, could be also used defining the sCOr2.2 matching strategy, i.e. using the subset $\{\overline{H}_0^1, \overline{H}_2^0, \overline{H}_1^1, \overline{H}_{m-1}^1, \overline{H}_{m-1}^2\}$ in (3).

Similar to sGOr, the estimation of the global reference orientation g cab be achieved using all the $2m$ descriptors in Q (sGOr2a strategy) or only the m descriptor belonging to the first concatenated sGLOH descriptor H^1 (sGOr2h strategy), while the relative rotation window could be constrained as above up to $\pm\frac{\pi}{m}$ or $\pm 2\frac{\pi}{m}$, obtaining 4 possible final different matching strategies (sGOr2a.1, sGOr2a.2, sGOr2h.1, sGOr2h.2).

4 Evaluation

We first evaluated sGLOH2 and the corresponding matching strategies on rotation on the same setup described in [3]. A dataset of 16 images, rotating from $0°$ to $90°$ only (due to the cyclic nature of the sGLOH descriptor) is used to measure the average percentage of correct matches for intermediate rotations. The L_1 distance is used since according to [3] it obtains better results than L_2. Plots are reported in Fig. 2.

As it can be noted, both the proposed sGLOH2 and its matching strategy improve on the original sGLOH versions when rotations are close to the worst case. In particular the results of the proposed extensions are better then SIFT and very close to the results of LIOP and MROGH. Only a difference less than 2% with LIOP and MROGH exists for rotations of the form $k\frac{\pi}{m} + \frac{\pi}{2m}$ for $k = 0, \ldots, 2m - 1$, i.e. between rotations of $\frac{\pi}{m}$, in analogy with the original worst case, that is negligible in almost all cases.
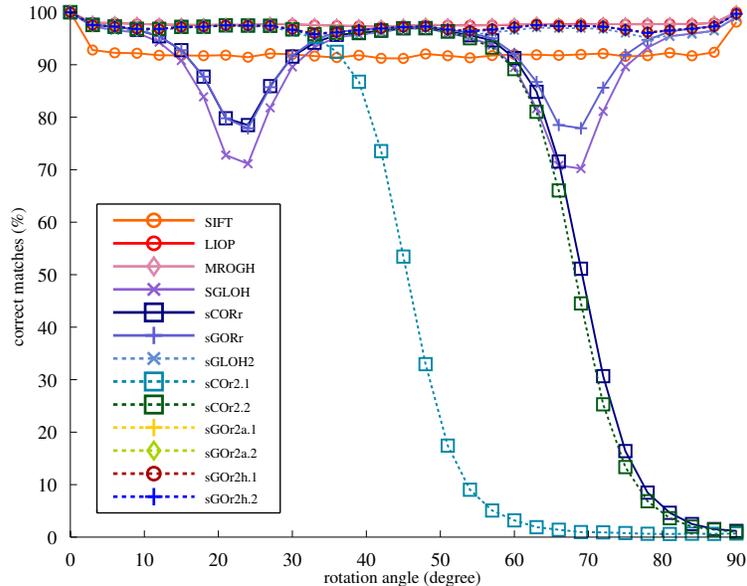
Fig. 2: Descriptor average percentage of correct matches for different rotation degrees (best viewed in color)

Furthermore, by comparing the sCOr and sCOr2 strategies it can be observed that the sCOr2.2 and sCOr matching can handle both angles up to $\pm 67.5°$ while sCOr2.1 works with rotations up to $\pm 45°$ only since it allows less rotation shifts, according to their constructions given in Sect. 2.

No relevant differences can be noted between the sGOr2a and sGOr2h matching strategies, while if we compare the sGLOH2 descriptor against sGOr2a and sGOr2h, these last two work slightly better as they operate on a limited number of possible rotations, reducing the probability that two non-corresponding descriptors are matched by chance. This observation also holds for sCOr and sCOr2.2 for rotation greater than $45°$ since 3 rotations are checked for sCOr while 5 for sCOr2.2, so that in the first case there is, although minimal, less probability of wrong matches by chance.

The sGLOH2 extensions were also tested on the Oxford dataset [8], which offers a more challenging scenarios since different image transformations are applied to the images, not only rotations. In order to get a fair comparison, we used the same setup described in [3] even in this case. The keypoints where extracted using the HarrisZ detector [2], while for matching the NN strategy was used for sGLOH and its variants and the NNR matching for the others, in both cases using the L_1 distance, since these parameters provide the best setup for each descriptor. Results for the first and fourth image pair for each sequence are reported in Figs. 3-4 for varying the precision, both in terms of recall (number of correct matches/number of total correct matches) and correspondences (number

of correct matches/number of total matches). The maximum correspondence (number of total correct matches/number of total matches) for each image pair is reported too. The sCOr and sCOr2 descriptors are not included in this evaluation as there are image pairs with more than their allowable rotations.
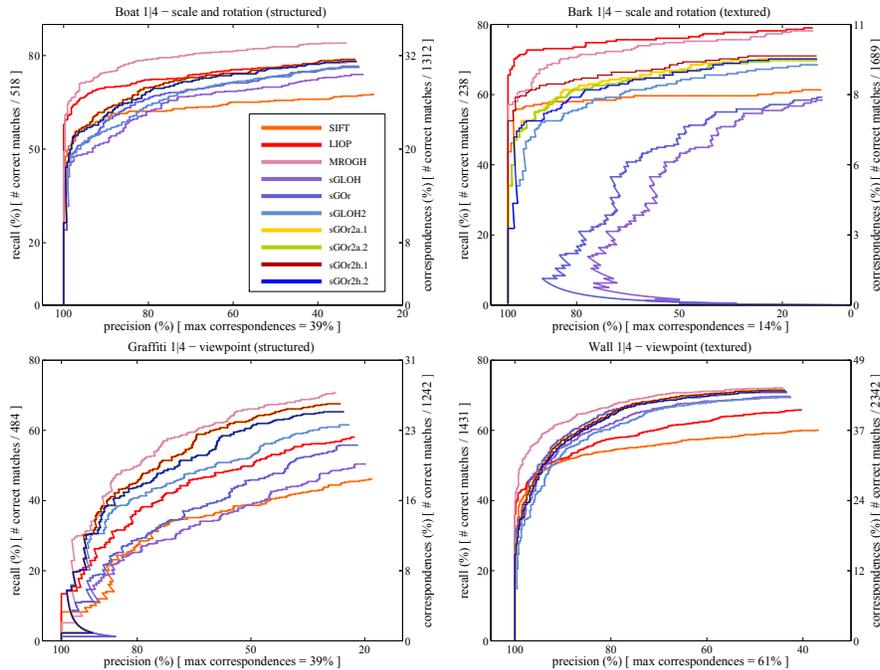

Fig. 3: Precision/recall curves for the Oxford dataset. Image pairs for the first and fourth images in the case of scale, rotation and viewpoint changes are shown (best viewed in color)

The sGLOH2, sGOr2a and sGOr2h provide a relevant improvement upon sGLOH and sGOr, especially in the case of scale, rotation and viewpoint changes, while in the case of variations of blur, luminosity and compression the improvements are less relevant. The most notable case is for the textured image rotation and scale change (Bark sequence), on which one can also note that the sGOr2a matching strategy slightly improves on the sGOr2h, while in the other cases results are the same.

Note that MROGH is introduced in the comparison as a sort of upper bound since, as noted in [3], its support regions is three times those of the other descriptors This leads to a misleading boosting in performances, especially in case of planar image transformations such as for the Oxford dataset. With respect to sGLOH and its matching strategies, the sGLOH2 extensions obtains similar or better results against LIOP in the

structured image sequences with geometric transformations (the Graffiti and Boat sequences); SIFT results seem in general worst than others.
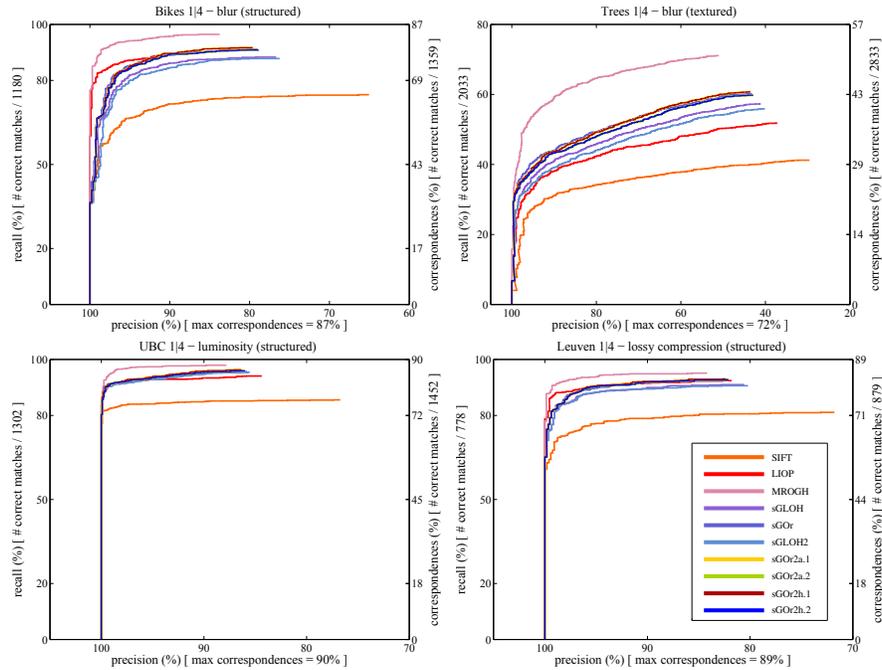

Fig. 4: Precision/recall curves for the Oxford dataset. Image pairs for the first and fourth images in the case of blur, luminosity and image compression changes are shown (best viewed in color)

Plot in Fig. 5 shows the precision/recall curves with the previous setup on the first and fifth image pair of the sequences with geometrical transformations, which represent a more challenging pairs. As it can be noted the improvements given by the sGLOH2 extensions are still remarkable, especially in the case of scale and rotation changes for textured images (the Bark sequence). Furthermore, it can be noted that for these image pairs there are no relevant difference between MROGH and the proposed descriptors, while LIOP and SIFT results seem very varying.

5 Running Time

Figure 6 reports the estimated running times for the evaluated methods, together with a details of the time spent by each computational steps fora particular example. The estimated curves are obtained by quadratic fitting on the image pairs of the Oxford and rotation datasets, denoted by crosses. It can be noted that sGOr2a and sGLOH2 are very time
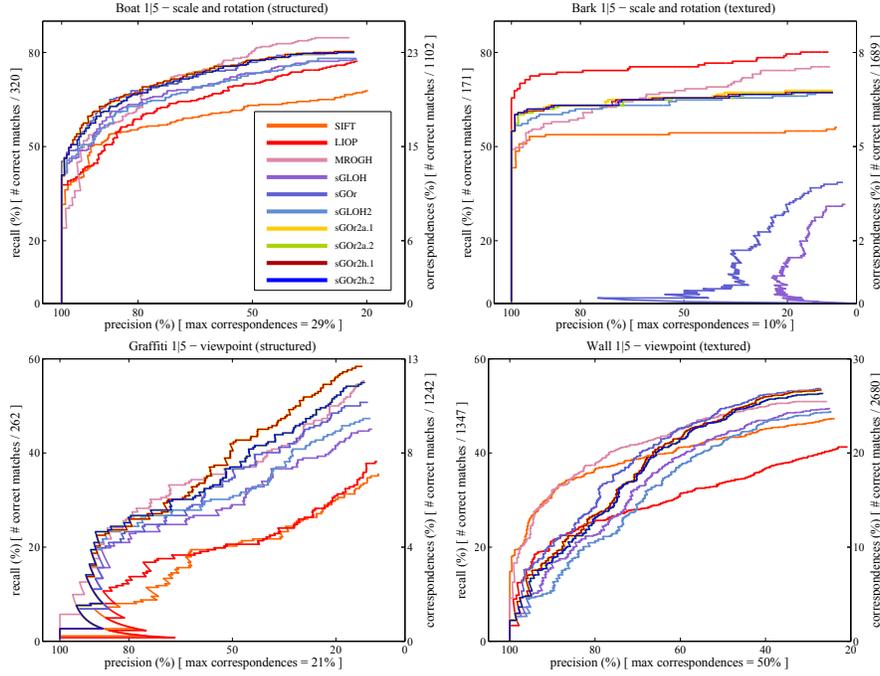
Fig. 5: Precision/recall curves for the Oxford dataset. Image pairs for the first and fifth images in the case of scale, rotation and viewpoint changes are shown (best viewed in color)

consuming, while sGOr2h and sCOr2 running times are still reasonable considering the final matching results, especially up to 2000 keypoints per image. When more keypoints are used and rotations are constrained up to ±45° only the sCOr2.1 strategy is very favourable in terms of computational speed. Note however that according to the authors' experience in general the number of keypoints for matching are bounded to about 1500 for reducing memory issues as well as the probability of wrong matches by chance.

By inspecting the cumulative time histogram on Fig. 6(right), it can be noted that the time for computing the descriptor in the case of LIOP and sGLOH are the lowest, followed by sGLOH2, and more computational cost is required by MROGH and SIFT. The quadratic matching cost (for n features there are n^2 possible matches) makes the difference between the sGLOH and sGLOH2 methods and the others.

In particular, setting to t the time required to compute all the matches between two images using a generic distance measure $\mathcal{D}$, which can be assumed almost equal for SIFT, LIOP and MROGH, a multiplication factor equal to the number of rotation to check should be added according to (1) so that the original sGLOH, sCOr and sGOr require respectively about mt, $3t$ and mt times for matching. Analogously, according to (3), the sGLOH2 match required time is about $2mt$, for sCOr2.1 and

sCOr2.2 respectively $3t$ and $5t$, while for sGOr2a.1, sGOr2a.2, sGOr2h.1, sGOr2h.2 the times are about $2mt$, $2mt$, mt, mt respectively, in accordance with the histogram bars.
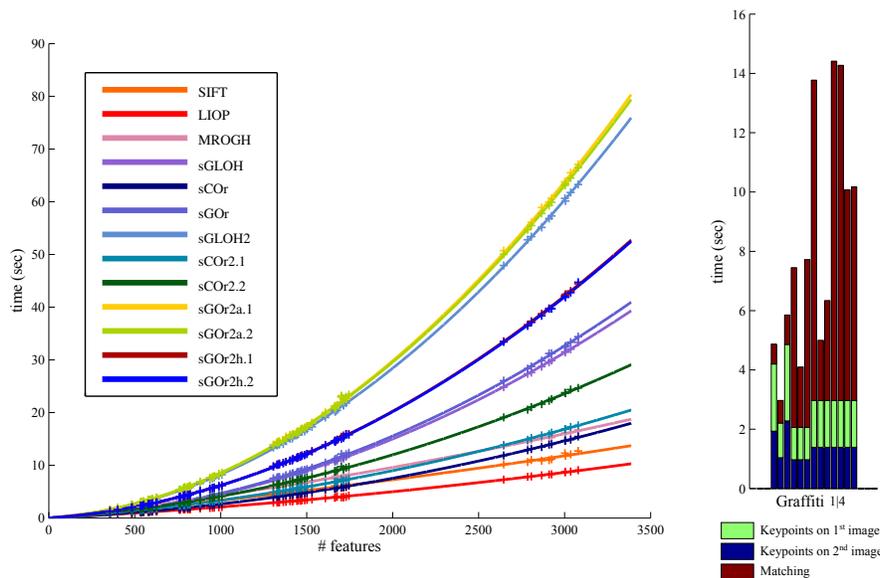

Fig. 6: Estimated running time for the evaluated descriptor (left) and a detailed cumulative time example (right). Descriptor order follows that of the legend (best viewed in color)

6   Conclusions and Future Work

This paper presents an extension to the sGLOH descriptor and the corresponding matching strategies sCOr and sGOr. According to the obtained results, these improvements are effective, through a more robust rotation handling, especially for the sGLOH worst case, but also in the general matching between descriptors. Although the computational time increases, it is still reasonable in a lot of applications and it can be adjusted according to the required task by choosing the right matching strategy. In particular, the sCOr2.1 is the fastest among these and can handle rotation up to $\pm 45°$, followed by sCOR2.2 with constrained rotation up to $\pm 67.5°$, and the sGOr2h.1 matching for general tasks, which differs only slightly from the more costly sGOr2a.2 matching strategy.

Future work will include more experimental evaluations, tests on descriptor binarization to reduce its dimension and to improve its running time as well as fast distance check to discard very probable bad matches.

Acknowledgement

References

1. Bellavia, F., Fanfani, M., Pazzaglia, F., Colombo, C.: Robust selective stereo SLAM without loop closure and bundle adjustment. In: International Conference on Image Analysis and Processing. pp. 462–471 (2013)
2. Bellavia, F., Tegolo, D., Valenti, C.: Improving Harris corner selection strategy. IET Computer Vision 5(2), 86–96 (2011)
3. Bellavia, F., Tegolo, D., Valenti, C.: Keypoint descriptor matching with context-based orientation estimation. Image Vision Computing 32(9), 559–567 (2014)
4. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7), 1281–1298 (2012)
5. Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., Fua, P.: Receptive fields selection for binary feature description. IEEE Transactions on Image Processing 26(6), 2583–2595 (2014)
6. Fan, B., Wu, F., Hu, Z.: Rotationally invariant descriptors using intensity order pooling. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(10), 2031–2045 (2012)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
8. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision 65(1-2), 43–72 (2005)
9. Snavely, N., Seitz, S., Szeliski, R.: Modeling the world from internet photo collections. International Journal of Computer Vision 80(2), 189–210 (2008)
10. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer (2010)
11. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Foundations and Trends in Computer Graphics and Vision 3(3), 177–280 (2008)
12. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: Proc. of the International Conference on Computer Vision. pp. 603–610 (2011)
13. Wang, Z., Fan, B., Wu, F.: Affine subspace representation for feature description. In: Proc. of the European Conference on Computer Vision. pp. 94–108 (2014)