Semantic-Context-Based Augmented Descriptor For Image Feature Matching

Samir Khoualed[1], Thierry Chateau[1], and Umberto Castellani[2]

[1] Institut Pascal, CNRS/University of Blaise Pascal
Clermont-Ferrand, France
{samir.khoualed,thierry.chateau}@lasmea.univ-bpclermont.fr
[2] VIPS, University of Verona
Verona, Italy
umberto.castellani@univr.it

Abstract. This paper proposes an augmented version of local features that enhances the discriminative power of the feature without affecting its invariance to image deformations. The idea is about learning local features, aiming to estimate its semantic, which is then exploited in conjunction with the bag of words paradigm to build an augmented feature descriptor. Basically, any local descriptor can be casted in the proposed context, and thus the approach can be easy generalized to fit in with any local approach. The semantic-context signature is a 2D histogram which accumulates the spatial distribution of the visual words around each local feature. The obtained semantic-context component is concatenated with the local feature to generate our proposed feature descriptor. This is expected to handle ambiguities occurring in images with multiple similar motifs and depicting slight complicated non-affine distortions, outliers, and detector errors. The approach is evaluated for two data sets. The first one is intentionally selected with images containing multiple similar regions and depicting slight non-affine distortions. The second is the standard data set of *Mikolajczyk*. The evaluation results showed our approach performs significantly better than expected results as well as in comparison with other methods.

1 Introduction

Image description is an important task for many applications in computer vision. It is based on computing image feature descriptors, which are distinctive and invariant to different geometric image transformations and imaging conditions. For this purpose, many approaches have been proposed. The successful among them provide descriptors with high discriminative power, invariant (constant) under aggressive image degradations.

In the seminal-work of Mikolajczyk and Schmid [1], a number of promising approaches are evaluated and compared for image feature matching. The evaluation results suggest SIFT [2], PCA-SIFT [3] and GLOH [1] as the most successful descriptors whereas for object recognition, the evaluation conducted by Bay et al. [4] shows that SURF descriptor performs better than SIFT, PCA-SIFT and GLOH.

Despite their apparent usefulness, it seems no approach was found to perform best in general. The main problem with these approaches (mostly local information-based) is the sensitivity to scenes exhibiting multiple similar motifs, like those of homogeneous-structured and highly-textured environments as well as for those present complicated

non-affine distortions. Hence, it becomes difficult for matching features within images obtained from these scenes assuming affine warps or 2D-rigid transformations. By allowing non-affine image transformations, the local descriptors may fail to address the matching problem.

These constraints can be harmful to the applications involving much accuracy and precision. To overcome these limitations, we propose the approach of the semantic-context-based descriptor. This is an extended 2D version (for images) of the approach introduced in [5], which is developed for the registration of multiple partial 3D views. Our approach is based on concatenating both *local* and *semantic-context* information to obtain an augmented feature descriptor. The semantic-context information is built around the local information using clustering of feature descriptors collected on different images.

The paper is organized as follows: In Section 2, a short overview of the previous work related to image feature description is given. The section 3 details the principle of the semantic-context approach and describes different steps in computing the semantic-context-based descriptor. In Section 4, we outline the evaluation scheme while presenting data set, performance criteria, descriptors, support regions, and matching strategies. In Section 5, we show the evaluation results. The conclusion is given in Section 6.

2 Related Work

Many computer vision tasks rely heavily on image feature selection and description. Video tracking, object recognition, and features matching are examples. The latter is a typical task in which image descriptors are collected to obtain support for recognizing similar features on different images. In this context, different approaches have been proposed. These can be divided into two categories, *global* and *local* methods.

Although considered to be less successful than the local descriptors, the global descriptors (or context) are still well-suited for particular scenes, like those containing large similar motifs. A basic global descriptor is a two-dimensional histogram representing the distribution of interest-points across an uniform square-grid.

Based on a similar idea, instead of interest points, Shape-Context [6] is computed as log-polar histograms for spatial distribution of edges, which are extracted with Canny detector [7]. This technique has been successfully evaluated in shape recognition, where edges are reliable features.

A 2D version of spin images was developed by Lazebnik et al. [8]. The proposed *intensity-domain spin images* (abbreviated here as Spin-Image) descriptor is inspired by the standard spin images [9], in which the traditional coordinates are replaced by the spatial point position and brightness. The Spin-Image descriptor has a high degree of invariance for representing affine normalized patches.

In contrast with the global approaches, the local descriptors have been paid more attention, for feature matching in particular. This is due to their best performances. The early approaches of *differential invariants* [10] and *steerable filters* [11] use derivatives computed by convolution with an approximated Gaussian. Differential-invariants is based on the property that the derivatives of the blurred illumination are equal to the convolution of the original image with certain filters. To build the descriptor, a set of special filters are concatenated to obtain higher order derivatives at lower resolution. The steerable-filters technique is similar to the differential invariants. It uses oriented

(steered) filters to compute derivatives in an arbitrary direction. The oriented filter is a linear combination of basis filter. The responses of the basis filters are used to determine the direction of the derivatives. These derivatives are invariant to rotation if they are computed in the direction of gradient.

Moment-invariant [12] is a method developed in the context of viewpoint invariant recognition of planar pattern. It uses the traditional geometric moments as the basic features. The moment invariants are functions of both shape moments and intensity moments. The descriptor based on moment invariants performs well with color images.

Cross-correlation is a basic descriptor represented by a vector of image pixels. In the framework of first template matching [13], the unnormalized cross-correlation is normalized using pre-computed tables containing the integral of the image and its square over the search window. This descriptor is well suited for special effects feature tracking.

Complex filters [14] is a descriptor developed for multi-view matching within large number of images where no ordering information is provided. The invariance of complex filters to image transformation is achieved by mapping the neighborhood onto an unit disk. Thus, the rotational invariants is computed from responses of a bank of linear filters.

SIFT [2] is the most popular descriptor, which has been proven to be very successful in many applications. It is a scale invariant descriptor based on the distribution of gradient magnitudes. Among its variants, there are *gradient location and orientation histogram* (GLOH) [1] and PCA-SIFT [15]. Both are built around the standard SIFT, then processed through the principal components analysis.

Combining local and global information is a promising technique, though only few methods use this concept. To the best of our knowledge, there are only those of Carneiro and Jepson [16] and Mortensen et al. [17]. The latter was proposed in the context of insects classification for which the authors claimed that the method is robust to local appearance ambiguity and non-rigid transformations without showing convincing evaluations and results. Other interesting approaches exploit a similar concept of bag of words are proposed in [18–20] for multiple class segmentation using a unified framework over mean-shift patches, large scale partial-duplicate web image search, and linear spatial pyramid matching using sparse coding, respectively.

Adopting the term of "*semantic context*" with bag-of-words is not new since it is already used by [5] and [21] in different manners to address two different problems. The first is related to registration of multiple range images while the second to visual words disambiguation.

Besides, much more emphasis has been placed on image feature description and many other approaches have been introduced despite the methods mentioned above still gaining more interests.

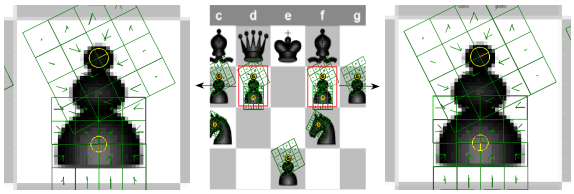
3 Semantic-Context-Based Augmented Descriptor

The proposed approach is an extended 2D version (*i.e.*, for images) of [5], which is developed for the registration of multiple partial 3D views. The concept of the *semantic-context-based* (SSC) [3] descriptor is based on integrating both *context* and *semantic*

[3] Abbreviated SSC in reference to Semantic-Shape-Context used interchangeably with Semantic-Context. This is to remove any confusion with SC used as abbreviation of Shape-Context descriptor [6].

informations into a local descriptor. The *context* information is built around a *semantic* vocabulary generated for describing relationships between different images of a same scene. SSC improves the performance of local descriptors as follows:

- The context information (or global) helps to reduce ambiguities occurring locally between similar regions. This has an effect to increase the discriminative power of descriptors, for scenes of multiple similar regions in particular. The illustration given in Fig. 1 shows how local approaches such as SIFT may fail to correctly match features where the context information is well-suited.



(a) (b) (c)

FIG. 1: A typical case to demonstrate that multiple similar motifs inside an image, can result in highly ambiguous local descriptors. Though features selected inside both red boxes of (b) are spatially different, it seems by comparing (a) and (c) they have similar local presentations with SIFT descriptors. This illustration uses features obtained with SIFT detector, whereas in our experiments we adopt the approach of `harris-affine`, which will be presented later.

- There are mainly two types of errors that influence the descriptors computation: *intrinsic* and *extrinsic* errors. The intrinsic errors are related to the descriptor algorithm itself whereas the extrinsic are caused by imprecision in support region errors (detectors) as well as errors arise from common local image defects. These are often resulting from imperfection of image sensors and uncontrollable imaging conditions. When a semantic information is used, these errors have less impact on descriptor performance. This is because, the clustering-based strategy for generating the semantic vocabulary helps to recover resemblance between deficient descriptors. Thus, different inaccurate descriptors computed for the same feature on different images is represented with the same visual semantic feature within all images.
- Furthermore, combing context and semantic informations seem quite appealing for particular scenes where obtained images present complicated non-affine distortions and non-rigid movements. These often caused by non-stationarity of objects inside images. That is to say, objects move independently during image capturing or deformation. For instance, the *giraffe* and *leaves* objects shown in Fig. 3c and Fig. 3j page 7, can move while the scenes are under different viewpoints, *e.g.*, camera angle of view changes. Hence, it becomes difficult for matching features within images obtained from these scenes assuming affine warps or 2D-rigid transformations. By allowing non-affine image transformations, the local descriptors may fail to address the matching problem.

3.1 Descriptor components

The proposed SSC-based descriptor is a composite of *local* and *semantic-context* signatures. The SSC-based descriptor vector, **D**, is a weighted concatenation of the normalized local and the semantic-context components, **L** and **S**, respectively:

$$\mathbf{D} = \begin{bmatrix} w\mathbf{S} & (1-w)\mathbf{L} \end{bmatrix}, \tag{1}$$

where w is a weighting factor (fixed to 0.5 in our experiments) introduced in order to compromise the contribution of the two components. More details on this parameter are given in Section 5.1.

The local component, **L**, can be any of local descriptor such as SIFT, SURF etc. The computation of the semantic-context component, **S**, is based on the local descriptor and it is a three-step process, as described in the following paragraph.

Semantic-context component. Given a set of images taken, as example, from a same scene under different angles of camera viewpoint, the semantic-context information is computed as follows:

1. First, local descriptors are computed for support regions collected on different images. These are reassembled to obtain one set of descriptor vectors.
2. The obtained set is then partitioned based on *k-means* clustering. The resulting clusters (fixed to 25 in our experiments) constitute what we called semantic vocabulary (or visual features). Thus, each feature on image has an assigned cluster which is visible on other images. It is qualified as semantic because it expresses the connection existing between images related to or having dealing with each other (*e.g.*, obtained from the same scene).
3. Finally, each feature in each image is assigned to its corresponding semantic word, then, the semantic-context component is computed as illustrated in Fig. 2.

In this figure, we assume for an instructive purpose, that the chess pieces replace semantic features (words) resulting from clustering stage. There are five semantic features: *black-pawn*, *bishop*, *rook*, *knight*, and *white-pawn*. Thus the following steps compute the semantic-context component for the feature-point p.

- The surrounding space of p (*i.e.*, around the spatial position of the local feature p) is partitioned into log-radial concentric shells (fixed to 12 in our experiments).
- For each shell, the repeated occurrences of each semantic feature (each chess piece) are accumulated to give one two-dimensional histogram per shell.
- The obtained histograms are concatenated, providing thus, the semantic-context component of the descriptor computed for the feature-point p.

Performance. In addition to the above-mentioned reasons, the SSC-based descriptor is inherently rotation invariant, since the accumulated occurrences of each visual feature inside each concentric shell are nearly unchanged under rotation. The resulting histograms are notably insensitive to arbitrary rotations applied to images. It is more intuitive to

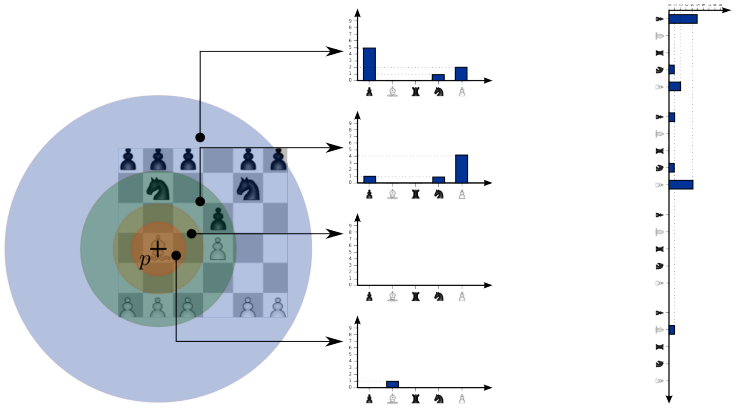


FIG. 2: An instructive example for computing semantic-shape-context component. For the purpose of simplicity, we suppose that the chess pieces replace semantic features (words) resulting from clustering stage. There are five semantic features: *black-pawn*, *bishop*, *rook*, *knight*, and *white-pawn*. The semantic-context component for the feature point p, is computed as follows: (a) The surrounding space of p is partitioned into log-radial concentric shells. (b) For each shell, the repeated occurrences of each semantic feature (each chess piece) are accumulated to give one two-dimensional histogram per shell. (c) The obtained histograms are concatenated, providing thus, the semantic-context component of the descriptor computed for the feature-point p.

consider the invariance to translation, since all histograms are built relative to image points. The robustness against scale changes is also increased, since the radii of the inner and outer concentric-shells are set according to the normalized distances between all images features. Furthermore, we expect low outlier rates, since the semantic words simulate features collected on image overlapping areas. Thus, the descriptor robustness is enhanced by getting rid of no longer representative features.

4 Performance Evaluation

The performance of SSC-based descriptor is evaluated using the standard benchmark of Mikolajczyk [1] available on-line[4]. It contains programs and dataset for evaluating and comparing the descriptor performances for image feature matching. To compute the semantic-context component, we wrote a basic c/c++ program, which is compiled on `Intel(R)Core(TM)2 Duo CPU P8700 @2.53GHz` machine model, running under `Linux-2.6.33.7-x86-64` environment.

Data set. The data set consists of a set of scenes depicting different geometric transformations and imaging conditions. In addition, we also include other scenes mainly selected from homogeneous environments of multiple similar regions, as shown along the first row of Fig. 3.

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

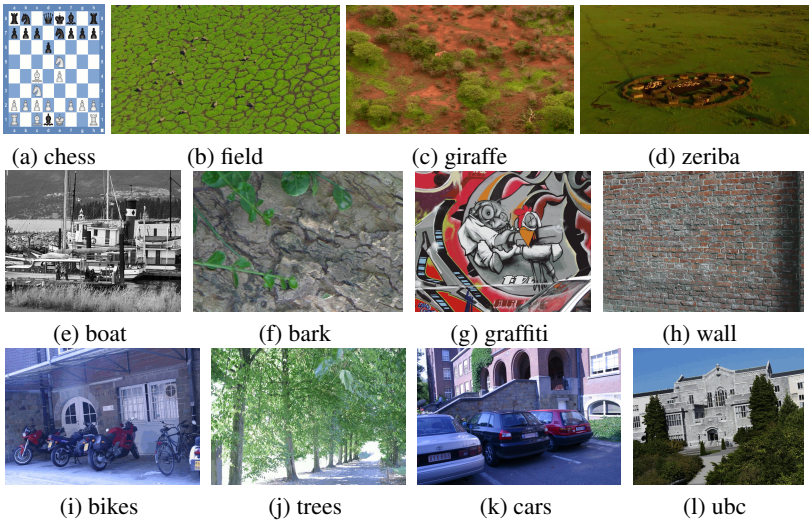


FIG. 3: The data set used in our evaluations. There are: (a)(b) rotation, (c) scale change, (e)(f) combined rotation-scale, (d)(g)(h) viewpoint change, (i)(j) image blur, (k) illumination change, and (l) JPEG compression.

Evaluation criteria. The descriptor performances are evaluated according to both discriminative power and invariance criteria. The discriminative power evaluates the ability of a descriptor to distinguish between different image features. It is given by ROC[5] curves showing *recall* score as function of *1-precision* score. The invariance (or robustness) measures the *constancy* of the descriptor performance under gradual increase in image degradation.

Descriptors. The performance of three variants of SSC-based descriptors, SIFT-Based-SSC, SPIN-Based-SSC and CC-Based-SSC, are evaluated and compared to ten state-of-the-art approaches described in Section 2.

These include the local descriptors of SIFT [2], spin images (SPIN) [8], complex filters (CF) [14], differential invariants (KOEN) [10], steerable filters (JLA) [11], moment invariant (MOM)[12], normalized cross-correlation (CC) [13], GLOH [1], and PCA-SIFT [15]. In addition, we incorporate the global approach of shape context (SC) [22].

Support regions. The descriptors are computed on regions obtained with *harris-affine* [23] approach, which is an affine covariant regions detector. That is, covariant with respect to translation, image rotation, scale change, and image shearing.

Matching strategies. Similar to SIFT matching approach, we adopt one-to-one feature correspondences by using the *nearest-neighbor* strategy. This is mostly correct (*i.e.*, with high precision), since it selects only one match per feature –the best below a threshold– while discarding all the rest. It also well suited for image feature matching.

[5] ROC as an abbreviation of Receiver Operating Characteristics

5 Experimental Results

Based on the experimental setup described in the previous section (§4), following are the results of different conducted evaluations.

Before going into more details, we show in Fig. 4 an example of image region matching based on SIFT-Based-SSC descriptor. Besides, we present in Fig. 5 a preview of the experimental results, in which the performance of the evaluated descriptors are compared (in terms of recall percentages) with respect to different image deformations. This shows how well SIFT-Based-SSC outperforms the other descriptors on both structured and textured scenes as well as for all types of image deformations.

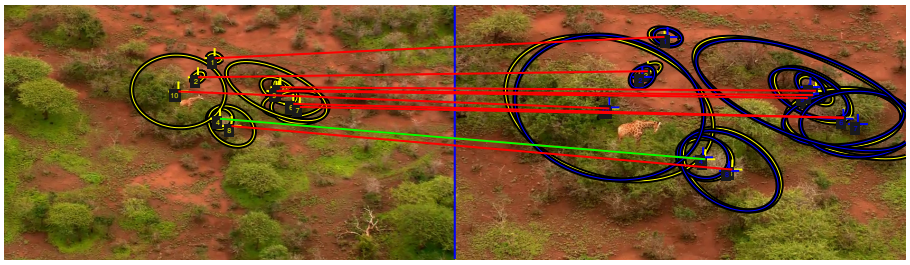


FIG. 4: An example of image region matching based on SIFT-Based-SSC descriptor. This is obtained on the textured scene of `giraffe` which reflects images subjected to scale change of factor 3. The detected regions are in yellow while their correspondences –transformed from the reference image (*left*) to the second image (*right*) using ground truth– are colored blue. The region correspondences computed based on ground truths and region overlap errors are highlighted with blue lines, whereas matches identified as correct using descriptors are highlighted with green lines. For the purpose of clarity, only reduced numbers of correspondences and matches are displayed.

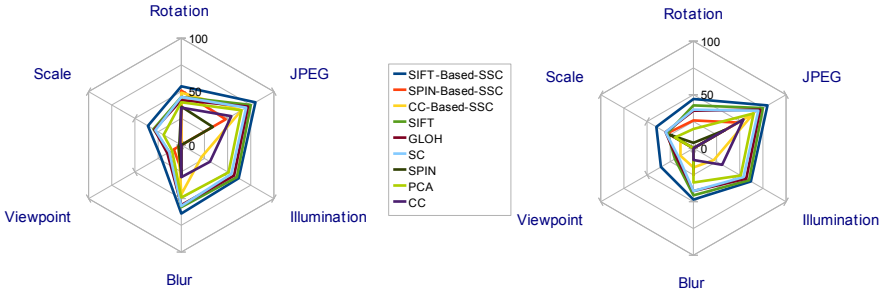
For the computational times (in terms of wall-clock time), we reported 1.9 *ms*/per-feature for computing the SSC component of SIFT-Based-SSC. This is approximately 41% of SIFT computational time (4.8 *ms*/per-feature). The recorded times are obtained on the highly textured scene of `zeriba` for which the numbers of cluster and images are reduced to 5 and 2, respectively.

5.1 Parameter setting

The SSC-based descriptor has three main adjustable parameters. These include: weighting factor, w, number of words, k, and number of concentric shells, s.

In addition, two other parameters related to radii of the inner and outer shells are set. The robustness of SSC components against scale changes depends on the values of these parameters.

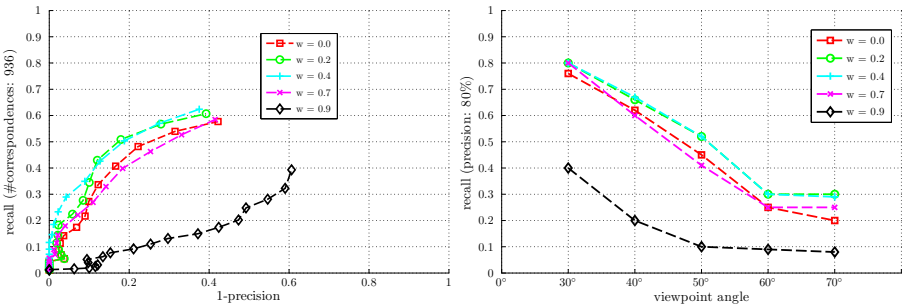
To minimize the effect of scale on the SSC robustness, we adopt an approach similar to [22]. The idea is setting the inner and outer radii to $1/8$ and 2 respectively, after normalizing (*e.g.*, by the mean) of the pairwise euclidean distances between the spatial locations of all image features.



(a) Structured scenes (b) Textured scenes

FIG. 5: A preview of the experimental results showing a comparison of the descriptor performances (in terms of the percentages of recall scores) for different types of images deformations. The recall percentages are computed for precision values vary in $80\% - 95\%$.

The heuristic evaluation illustrated in Fig. 6 suggests that the reasonable range of w is $0.4 - 0.7$. Hence, we set $w = 0.5$ in all our experiments. Similarly, we obtained values of $k = 25$ and $s = 12$, reasonably enough to ensure good performances.

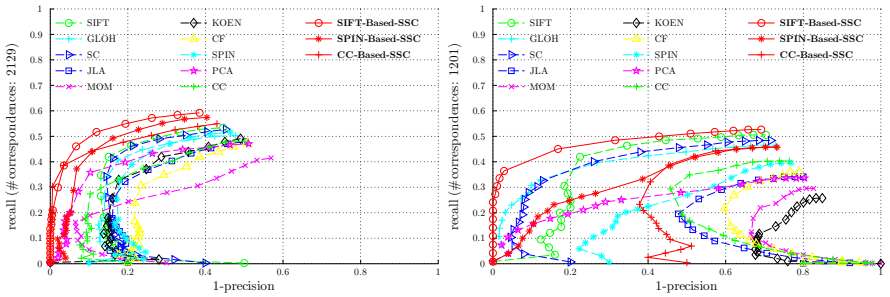


(a) Discriminative power (b) Invariance

FIG. 6: Effect of weighting factor, w, on (a) discriminative power and (b) invariance. The results are with respect to viewpoint changes using the textured scene of `zeriba`. The discriminative power is evaluated according to a viewpoint angle of $50°$.

5.2 Image rotation

The performance evaluations for image rotation are conducted on the structured and textured scenes of `chessboard` and `field`. These intentionally selected to present a large number of similar regions. The results of the discriminative power evaluations are reported in Fig. 7.



(a) Structured scene (b) Textured scene

FIG. 7: Results of discriminative power evaluations for image rotation, obtained on the structured and textured scenes of (a) `chessboard` and (b) `field` with rotation angles of 45° and 25°, respectively.

As we can see, the curves illustrate how much the distinctiveness power of all SSC-based descriptors are highly augmented compared to those of the others. Thus, SIFT-Based-SSC is recording the best performances on both scenes, and the discriminative power of SPIN and the simple CC are considerably improved after adding SSC components. For instance, CC-based-SSC becomes more effective than the competitive SIFT, as shown in Fig. 7a.

For the structured scene of `chessboard` which contains multiple similar regions, we observe the ROC curves (see Fig. 7a) of different SSC descriptors go above those of the others. This means that all SSC variants are more discriminative.

Fig. 7b (obtained on the textured `field`), shows also SIFT-Based-SSC performs largely as the best. Besides, it illustrates that a number of descriptors (*e.g.*, JLA, KOEN, and SPIN) are completely unusable when requiring high precisions (*1-precision* considerably close to zero).

5.3 Scale change

The evaluations against scale changes are performed with respect to discriminative power and invariance using `giraffe`'s scene.

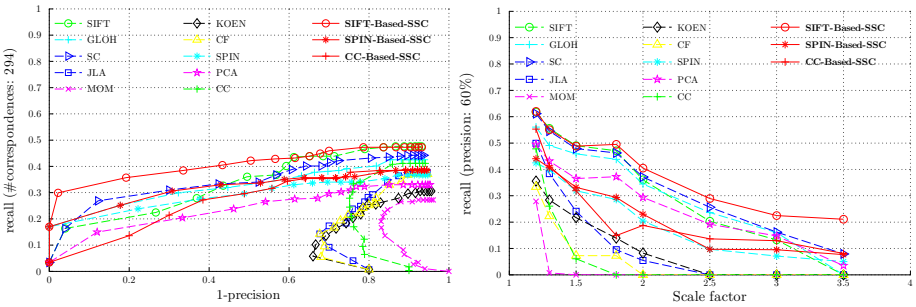
The results of discriminative power evaluations are displayed in Fig. 8a.

This shows the ROC curve of SIFT-Based-SSC being far above all descriptors. We also notice that the distinctiveness of CC-Based-SSC is extremely increased, yet it outperforms that of SIFT. This shows how well the performance of a simple descriptor like CC can be improved when adding the SSC information.

Though it is agreed that SC performs well on the textured scene like `giraffe`, we observe it to be fully outperformed by SIFT-Based-SSC. This is illustrated in Fig. 8a, in which SIFT-Based-SSC is ranked first while SC is second.

It is interesting to notice the SIFT ranking before and after incorporating the SSC information. It jumped from the third to the first spot.

It is easy to remark from Fig. 8b that SIFT invariance is highly enhanced when SSC component is added. This can checked through comparing the constancy of SIFT and



(a) Discriminative power (b) Invariance

FIG. 8: Results of performance evaluations under scale changes for the textured scene of `giraffe`. The discriminative power is evaluated for a scale factor of 3.5. The invariance evaluation uses 8 images with different scale factors ranged from 1.2 to 3.5.

SIFT-Based-SSC curves. It appears that SIFT-Based-SSC curve drops down more slowly than those of the other descriptors.

Similar to the discriminative power evaluation, SC is again completely outperformed by SIFT-Based-SSC. Tab. 1 summarizes the degradations of recall and precision scores when the scale factor is increased from 1.2 to 3.5.

TAB. 1: Degradation of (a) recall and (b) precision scores under scale changes. These are obtained on the scene of `giraffe`. The degradations are computed as the differences between the scores recorded at low and high scale factors of 1.2 and 3.5, respectively.

(a) Recall

Descriptor/Scale	Low	High	*Degradation*↓
SIFT	**0.62**	0.00	0.62
SC	**0.62**	0.09	0.53
GLOH	0.58	0.00	0.58
SIFT-Based-SSC	**0.62**	**0.20**	**0.42**

(b) Precision

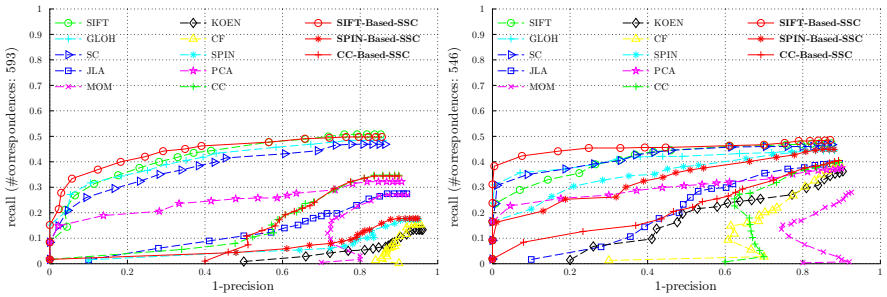
Descriptor/Scale	Low	High	*Degradation*↓
SIFT	**1.00**	0.30	0.70
SC	**1.00**	0.34	0.66
GLOH	**1.00**	0.15	0.85
SIFT-Based-SSC	**1.00**	**0.70**	**0.30**

These results reveal the impact of SSC component on increasing the invariance under scale changes.

5.4 Rotation-enlargement transformation

Rotation-enlargement transformation is a particular image deformation combining rotation and scale change. The descriptors are evaluated with respect to this transformation, and the results are highlighted in Fig. 9.

These illustrate SIFT-Based-SSC obtaining the best discriminative power. Though it performs differently, it still ranked first on both of scenes. However and as expected we obtain the best performance, as we can see in Fig. 9b, on the textured scene.



(a) Structured scene (b) Textured scene

FIG. 9: Results of discriminative power evaluations under rotation-enlargement for the structured and textured scenes of (a) `boat` and (b) `bark`, using the 6^{th} and 3^{rd} images, respectively.

In addition, the distinctiveness of SIFT is unexpectedly enhanced on the structured scene when it is coupled with SSC component, as illustrated in Fig. 9a. This is motivating in a sense that the scene contains only a few number of similar regions.

5.5 Viewpoint change

The most challenging geometric transformation that images can be subjected to, is that related to viewpoint changes (*i.e.*, out-of-plane rotation). The results of performance evaluations in terms of discriminative power, are depicted in Fig. 10.

First, it is worth noting that the number of correspondences in Fig. 10a is much smaller than that obtained in the same experiment of [1]. This is because the matching strategies are different. We use a one-to-one `nearest-neighbor` approach, while the other adopts a `threshold-based` matching strategy, in which a region can have several correspondances. Therefore the number of correspondances is generally too high with the latter.

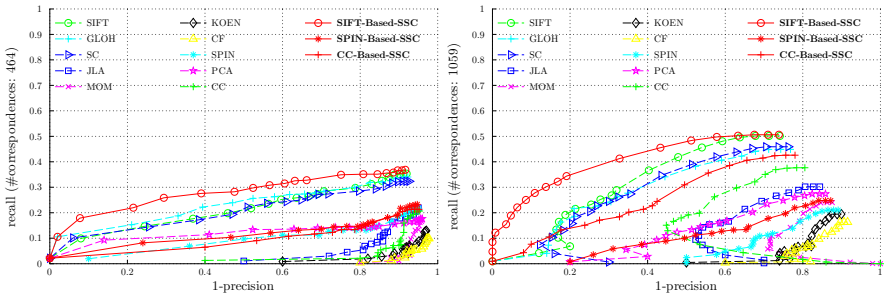
Fig. 10 shows without a doubt that SIFT-Based-SSC is winning largely the other approaches, especially on the textured scene as can be seen from Fig. 10b.

In addition, this illustrates other descriptors, like, GLOH, PCA , SC, and SIFT to perform poorly for high precision scores. It also demonstrates, JLA, CF, CC, KOEN, MOM becoming completely unusable. In contrast with these latter, CC-Based-SSC built on a simple CC pixel vector, shows to obtain high recall scores with high precisions compared to CC and some other descriptors, such as PCA and SPIN.

Moreover, we observe the discriminative powers for all SSC descriptors significantly increased on the structured scene, as shown in Fig. 10a.

6 Conclusion

This paper proposes an augmented version of local feature. The proposed descriptor consists of local feature concatenated with a feature context descriptor that uses bag-of-words (BoW) representation. The words, we called *semantic features*, are generated



(a) Structured scene (b) Textured scene

FIG. 10: Results of discriminative power evaluations under viewpoint changes for the structured and textured scenes of (a) `graffiti` and (b) `zeriba`, using images of viewpoint angles of 60° and 70°, respectively.

using local features collected on different images, then accumulated around the position of the local feature using log-concentric shells (weighted w.r.t the initial position of the feature). The idea is to use both context and semantic information around the local feature to improve the discriminative power of the feature without affecting its robustness to different image deformations.

The evaluation results bring out the effectiveness of the approach to address the problem of image feature matching. It is showed to perform significantly above the expected performances, thereby it can be a solution for many feature matching related tasks, like object recognition.

References

1. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence (2005) 1615–1630
2. Lowe, D.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60** (2004) 91–110
3. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. (2004)
4. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. Computer Vision–ECCV 2006 (2006) 404–417
5. Khoualed, S., Castellani, U., Bartoli, A.: Semantic Shape Context for the Registration of Multiple Partial 3D Views. IEEE Transactions on pattern analysis and machine intelligence **14** (2009) 239–256
6. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (2002) 509–522
7. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (1986) 679–698
8. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Transactions on Pattern Analysis and Machine Intelligence (2005) 1265–1278
9. Johnson, A.: Spin-images: a representation for 3-D surface matching. (1997)
10. Koenderink, J., Van Doorn, A.: Representation of local geometry in the visual system. Biological cybernetics **55** (1987) 367–375

11. Freeman, W., Adelson, E., of Technology. Media Laboratory. Vision, M.I., Group, M.: The design and use of steerable filters. IEEE Transactions on Pattern analysis and machine intelligence **13** (1991) 891–906
12. Van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns. Computer VisionECCV'96 (1996) 642–651
13. Lewis, J.: Fast normalized cross-correlation. In: Vision Interface. Volume 10., Citeseer (1995) 120–123
14. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets. Computer VisionECCV 2002 (2002) 414–431
15. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. (2004)
16. Carneiro, G., Jepson, A.: Pruning local feature correspondences using shape context. In: Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Volume 3., IEEE (2004) 16–19
17. Mortensen, E., Deng, H., Shapiro, L.: A sift descriptor with global context. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1., IEEE (2005) 184–190
18. Yang, L., Meer, P., Foran, D.: Multiple class segmentation using a unified framework over mean-shift patches. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
19. Wu, Z., Ke, Q., Isard, M., Sun, J.: Bundling features for large scale partial-duplicate web image search. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 25–32
20. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1794–1801
21. Su, Y., Jurie, F.: Visual word disambiguation by semantic contexts. In: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE (2011) 311–318
22. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (2002) 509–522
23. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International journal of computer vision **60** (2004) 63–86