Color-based and rotation invariant self-similarities

Xiaohu Song, Damien Muselet, Alain Tremeau

Color-based and rotation invariant self-similarities

Xiaohu Song, Damien Muselet and Alain Tremeau
Univ Lyon, UJM-Saint-Etienne, CNRS, LaHC UMR 5516, F-42023, SAINT-ETIENNE, France
{damien.muselet, alain.tremeau}@univ-st-etienne.fr

Keywords: Color descriptor, self-similarity, classification, invariance.

Abstract: One big challenge in computer vision is to extract robust and discriminative local descriptors. For many applications such as object tracking, image classification or image matching, there exist appearance-based descriptors such as SIFT or learned CNN-features that provide very good results. But for some other applications such as multimodal image comparison (infra-red versus color, color versus depth, ...) these descriptors failed and people resort to using the spatial distribution of self-similarities. The idea is to inform about the similarities between local regions in an image rather than the appearances of these regions at the pixel level. Nevertheless, the classical self-similarities are not invariant to rotation in the image space, so that two rotated versions of a local patch are not considered as similar and we think that many discriminative information is lost because of this weakness. In this paper, we present a method to extract rotation-invariant self similarities. In this aim, we propose to compare color descriptors of the local regions rather than the local regions themselves. Furthermore, since this comparison informs us about the relative orientations of the two local regions, we incorporate this information in the final image descriptor in order to increase the discriminative power of the system. We show that the self similarities extracted by this way are very discriminative.

1 INTRODUCTION

Evaluating self-similarities within an image consists in comparing local patches from this image in order to determine, for example, the patch pairs that look similar. This information is used for super-resolution (Glasner et al., 2009; Chih-Yuan et al., 2011), denoising (Zontak and Irani, 2011), inpainting (Wang et al., 2014), ... The spatial distribution of the self-similarities within each image is also a robust and discriminative descriptor that is very useful in some applications. Indeed, in order to compare images that look very different because of light variations, multi-modality (infrared versus color sensors) or, for example, the images of figure 1, the appearance-based descriptors such as SIFT (Lowe, 1999) or Hue histograms (van de Weijer and Schmid, 2006) completely fail whereas the self-similarities provide accurate information (Kim et al., 2015). We can note that CNN features (Krizhevsky et al., 2012) do not cope with this problem because they are based on learned convolutional filters that can not adapt themselves alone to a new modality. Consequently, these deep-features have been recently mixed with self-similarities in order to improve the results (Wang et al., 2015).

The idea of self-similarity consists in describing the content of the images by informing how similar are some local regions from each other (Shechtman and Irani, 2007; Chatfield et al., 2009; Deselaers and Ferrari, 2010). By this way, when two red and textured regions are similar in an image, their contribution to the final descriptor will be the same as this of two green and homogeneous regions. This representation is also invariant to any illumination condition variations, to changes in the colors of the objects (a red bike will have the same description as a blue one) and to modifications of the textures. For example, the figures {1(c), 1(e), 1(g)}, {1(j), 1(l), 1(n)} and {1(q), 1(s), 1(u)} show some self-similarities we have evaluated from the 3 images 1(a), 1(h) and 1(o) respectively. For this aim, we have extracted 3 patches {1(b), 1(d), 1(f)}, {1(i), 1(k), 1(m)} and {1(p), 1(r), 1(t)} at corresponding positions in each image 1(a), 1(h) and 1(o) repectively and we have evaluated the similarities between each of this patch with all the patches in the corresponding images. We can see that the similarities remain stable across variations in colors and textures and consequently their local (Shechtman and Irani, 2007; Chatfield et al., 2009) or global (Deselaers and Ferrari, 2010) spatial distribution can be used to efficiently describe the contents

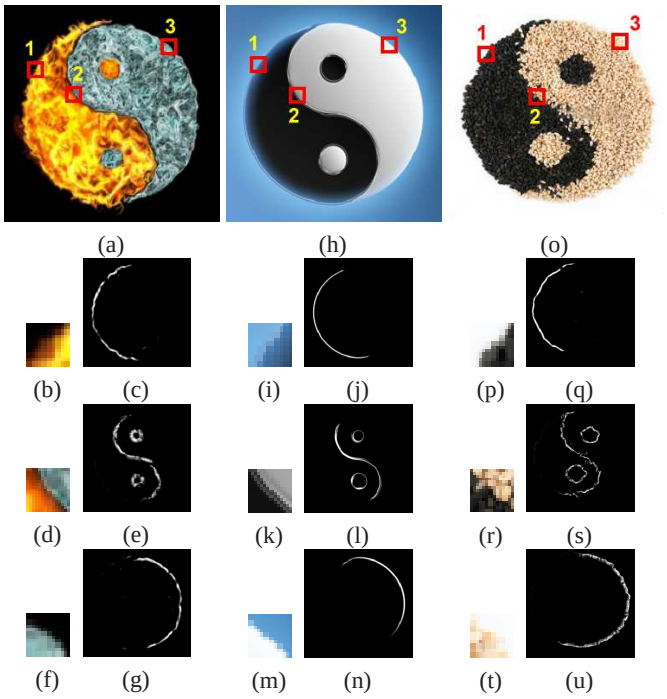


Figure 1: The images 1(a), 1(h) and 1(o) represent the same symbol but do not share any shape or color information at the pixel level. The local patches numbered 1 (1(b), 1(i), 1(p)), 2 (1(d), 1(k), 1(r)) and 3 (1(f), 1(m) and 1(t)) are extracted from similar relative positions in these 3 images. The figures 1(c), 1(j), 1(q), 1(e), 1(l), 1(s), 1(g), 1(n) and 1(u) represent the similarities (evaluated with our method) between these respective patches and the image from where they have been extracted. They represent self-similarities.

of the images.

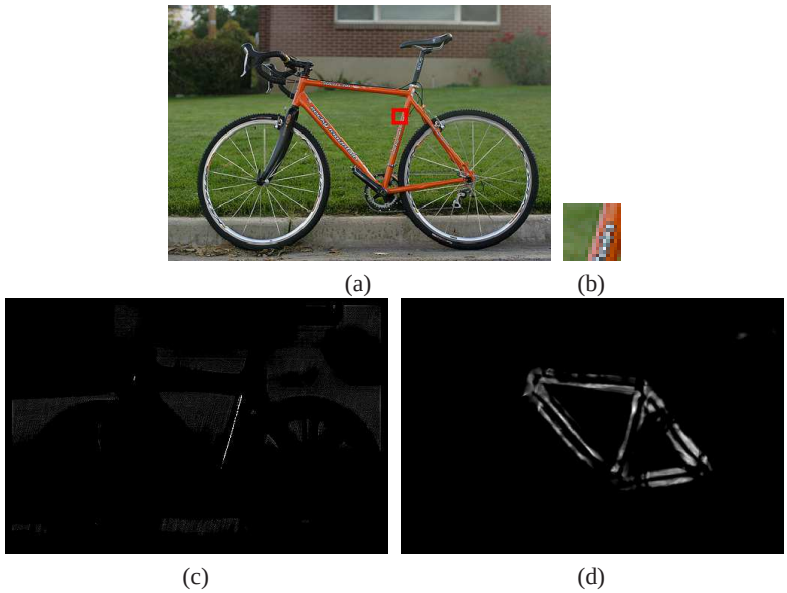
In this context, the classical approaches propose to extract the similarity between two local patches by evaluating their correlation (Shechtman and Irani, 2007; Chatfield et al., 2009). The drawback of the correlation is that it is based on a pixelwise multiplication and hence is not invariant to rotation. Indeed, the similarity between two patches can be very low if the first one of these patches is a rotated version of the second one. The image 2(c) represents the similarities between the patch 2(b) extracted from the image 2(a) and all the patches in the image 2(a) with the correlation-based method. We can see that only a part of the bike frame is detected by this way because the orientations of the other parts of the frame are different.

In this paper, we propose to associate each patch in an image with one particular spatio-colorimetric descriptor and to evaluate the similarity between two patches by comparing their descriptors. The proposed descriptor represent both the colors of the pixels in the patch and their relative spatial positions while being invariant to rotation. In order to design this descriptor we exploit the work from Song *et al.* (Song et al., 2009). The similarities evaluated by this way are displayed in the image 2(d). In this case, we can see that almost the whole frame of the bike can be detected whatever the orientation of each part. Furthermore, we will show that our descriptor-based self-similarity evaluation provides us the information of the angle difference between the orientations of the two compared patches and that this information can be introduced in the representation of the spatial distribution of the self-similarities in an image.

In the second part of this paper, we present how the classical approaches extract the self-similarities from the images and how they represent their spatial distribution. Then, in the third part, we introduce our new local spatio-colorimetric descriptor on which is based our self-similarity evaluation. We propose to represent the spatial distribution of the self-similarities by a 3D structure presented in the fourth part. The fifth part is devoted to the experimental results obtained in object classification task and we conclude in the sixth part.

2 RELATED WORKS

Shechtman *et al.* have designed a local descriptor based on self-similarity (Shechtman and Irani, 2007;



(a) (b) (c) (d)

Figure 2: The images 2(c) and 2(d) represent the similarities between the patch 2(b) extracted from the image 2(a) and all the patches in the image 2(a) with the correlation-based method (2(c)) and with our descriptor-based method (2(d)).

Kim et al., 2015). Considering one patch, the idea consists in measuring the similarity between it and its surrounding patches. To determine the similarity between two patches, they propose first to evaluate the pixelwise sum of square differences (*SSD*) and second they transform this distance to a similarity measure $SM = exp(-\frac{SSD}{\sigma})$ where σ is related to the variance of all the *SSD* locally evaluated. Then, the neighborhood of the considered patch is discretized on a log-polar grid and the maximal value of *SM* is stored within each bin grid. Chatfield *et al.* have also shown that the use of this local description provide better results than appearance-based descriptors such as SIFT for matching non-rigid shape classes (Chatfield et al., 2009).

Deselaers *et al.* argue that in the context of object classification and detection, the self-similarities should be evaluated globally rather than locally (Deselaers and Ferrari, 2010). Thus, they propose to compare each patch with all the patches in the image. Since this approach is highly time consuming, they propose an alternative to the classical pixelwise similarity evaluation that is based on the bag-of-words approach. Indeed, they associate each patch with a vector which can be its discrete cosine transform (DCT) or the patch itself (reshaped into a vector) and apply a k-means clustering in this vector space in order to get the most frequent vectors. Then, giving a patch, they evaluate its distance with all the cluster representatives and associate it with the nearest neighbor. By this way the similarity measure between two patches is a binary value, 1 if the patches are associated with the same cluster representatives and 0 if not. In order to design their global descriptor, called *SSH* for Self-Similarity Hypercube, Deselaers *et al.* first propose to project a regular D_1xD_2 grid onto the image. Then, they evaluate the similarity between all the patches in one grid cell GC_i and all the patches in the image. This returns a correlation surface whose size is the same as the image size. Then, they sub-sample this correlation surface to the size D_1xD_2 and put it in the grid cell GC_i. By doing that for all the grid cells GC_i, $i = 1,...,D_1$xD_2, they obtain their 4D *SSH* of size D_1xD_2xD_1xD_2 which represent the global repartition of the self-similarities in the image.

In these main works, the self-similarities are evaluated either by using pixelwise correlation between patches or by comparing the indexes of the associated cluster representatives of the two patches. In the first case, the extraction of the self-similarities can be highly time consuming and in the second case, the similarity between two patches is a binary value, 0 or 1. Furthermore, the results provided by the second approach are highly dependent on the quality of the clustering step. Consequently, we propose another approach that speed up the self-similarity evaluation while giving a real value as similarity measure between two patches. Therefore, we propose to evaluate a color descriptor of each patch and then to evaluate a similarity measure between these descriptors. Since these descriptors are not dependent on the orientations of the patches, our self similarities are rotation invariant. This is illustrated in Fig.2, where we can see that the classical correlation-based approaches (image

2(c)) can not detect all the self-similarities in the image whereas the self-similarities detected by our approach in image 2(d) clearly underline the discriminative power of the proposed color descriptor as well as its rotation invariance. Our local descriptor is introduced in the next section.

3 ROTATION INVARIANT DESCRIPTOR

In this section, we present our spatio-colorimetric descriptor used to evaluate the self-similarities. Therefore, we propose to exploit the paper from Song *et al.* (Song et al., 2009). The main idea of this paper, presented in the next paragraph, consists in applying an affine transform from image space to color space in order to design a local descriptor. Nevertheless, this descriptor requires a rotation invariant local region detection in order to be invariant to rotation. In the context of self similarity extraction, the image (or a part of the image) is dense sampled in order to extract patches at regular positions and to evaluate the similarity between each pair of patches. If the descriptor of each patch is not stable across rotation in the image, the extracted self-similarities will not be invariant to rotation. Consequently, in the second paragraph we propose to extend the approach of Song *et al.* in order to design a new rotation invariant descriptor.

3.1 The spatio-colorimetric descriptor from Song *et al.* (Song et al., 2009)

Song *et al.* have proposed a way to extract a descriptor from each detected local region (patch) in an image. The main idea consists in applying an affine transform to the pixels of the considered patch, from the image space to a color space. Thus, considering a pixel P_i characterized by the 3D-position $\{x_i^{Co}, y_i^{Co}, z_i^{Co}\} = \{c_i^R, c_i^G, c_i^B\}$ in the color space *Co* and the 2D-position $\{x_i^{Pa}, y_i^{Pa}\}$ in the patch-relative space *Pa* so that the center of the patch has $\{0,0\}$ for coordinates. Song *et al.* propose to define an affine transform from the patch space to the color space as:

$$\begin{bmatrix} m_1 & m_2 & t_x \\ m_3 & m_4 & t_y \\ m_5 & m_6 & t_z \end{bmatrix} \begin{bmatrix} x_i^{Pa} \\ y_i^{Pa} \\ 1 \end{bmatrix} = \begin{bmatrix} x_i^{Co} \\ y_i^{Co} \\ z_i^{Co} \end{bmatrix} \quad (1)$$

where t_x, t_y and t_z are the translation parameters and the m_i are the rotation, scale and stretch parameters.

This equation can be re-written as follows: $[Af] \times [Pa] = [Co]$.

It is based on the coordinates of one pixel but all the pixels of the patch can be accounted by adding columns in the matrices $[Pa]$ and $[Co]$. Since there exists no such affine transform from the image space to a color space, Song *et al.* propose to estimate the best one as the least-squares solution:

$$Af = Co[Pa^T Pa]^{-1} Pa^T. \quad (2)$$

Once the transform parameters have been determined, they apply the transform to the 4 pixels corresponding to the corners of the patch and whose coordinates in the patch space are $\{-sx/2, -sy/2\}$, $\{sx/2, -sy/2\}$, $\{-sx/2, sy/2\}$ and $\{sx/2, sy/2\}$, where *sx* and *sy* are the width and height of the patch, respectively. The positions of these corners in the color space after applying the affine transform constitute the descriptor:

$$Descriptor = Af \times \begin{bmatrix} -sx/2 & sx/2 & -sx/2 & sx/2 \\ -sy/2 & -sy/2 & sy/2 & sy/2 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \quad (3)$$

The use of the least-squares solution method provides the discriminating power of the descriptor. Indeed, the resulted destination position of a pixel depends not only on its color but also on the colors and the relative positions of the other pixels of the local region. Thus, considering two patches characterized by the same colors but by different spatial color arrangements, the resulted descriptors will be different. This characteristic is very interesting in the context of object recognition.

Nevertheless, in the context of self-similarity evaluation, the patches are dense sampled without any information about their orientation. So, the *x* and *y* axis of the patches are all oriented along the horizontal and vertical directions respectively. In this case, this descriptor is not stable across rotation in the image space and we show in the next paragraph the way to reach this invariance.

3.2 The proposed rotation invariant descriptor

We consider two patches Pa_1 and Pa_2 so that the second one is a rotated version of the first one:

$$\begin{bmatrix} x_i^{Pa_2} \\ y_i^{Pa_2} \end{bmatrix} = Rot \times \begin{bmatrix} x_i^{Pa_1} \\ y_i^{Pa_1} \end{bmatrix}. \quad (4)$$

By applying the approach of Song *et al.* on these two patches, we can determine one affine transform for the first patch $Af_1 = Co[Pa_1^T Pa_1]^{-1} Pa_1^T$ and one for the second patch $Af_2 = Co[Pa_2^T Pa_2]^{-1} Pa_2^T$. From

equation (4), we have $Pa_2 = Rot.Pa_1$ and then:

$$\begin{aligned} Af_2 &= Co[Pa_2^T Pa_2]^{-1} Pa_2^T \\ Af_2 &= Co[(Rot.Pa_1)^T (Rot.Pa_1)]^{-1} (Rot.Pa_1)^T \\ Af_2 &= Co[Pa_1^T.Rot^T.Rot.Pa_1]^{-1} Pa_1^T.Rot^T \\ Af_2 &= Co[Pa_1^T.Pa_1]^{-1} Pa_1^T.Rot^T \\ Af_2 &= Af_1.Rot^T, \end{aligned} \tag{5}$$

because for any rotation matrix Rot, $Rot^T.Rot = Identity$.

By using these two transforms Af_1 and Af_2 in the equation (3), we can see that the descriptors of the two patches Pa_1 and Pa_2 are different. So these descriptors can not be used directly in order to find rotation invariant self similarities in images.

However, the equation (5) shows that the transforms obtained for two patches, the second patch being a rotated version of the first, are related to each other by the rotation applied in the patch space. Our intuition is to use this transform itself as a descriptor after removing the rotation from it. Since this rotation is applied in the patch space, it is a 2D rotation and can be represented by a 2x2 matrix Rot_{2x2}. Thus, we propose to rewrite the transform Af_k of the patch Pa_k, $k = 1$ or 2, as follows:

$$\begin{aligned} Af_k \begin{bmatrix} x_i^{Pa_k} \\ y_i^{Pa_k} \\ 1 \end{bmatrix} &= \begin{bmatrix} mk_1 & mk_2 & tk_x \\ mk_3 & mk_4 & tk_y \\ mk_5 & mk_6 & tk_z \end{bmatrix} \begin{bmatrix} x_i^{Pa_k} \\ y_i^{Pa_k} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} mk_1 & mk_2 \\ mk_3 & mk_4 \\ mk_5 & mk_6 \end{bmatrix} \begin{bmatrix} x_i^{Pa_k} \\ y_i^{Pa_k} \end{bmatrix} + \begin{bmatrix} tk_x \\ tk_y \\ tk_z \end{bmatrix} \\ &= Mk_{3x2} \begin{bmatrix} x_i^{Pa_k} \\ y_i^{Pa_k} \end{bmatrix} + Tk_{3x1}. \end{aligned} \tag{6}$$

Since each transform Af_k is evaluated in the corresponding patch Pa_k relative coordinates system (where the center of the patch has $\{0,0\}$ for coordinates), the translation parameters in $T1_{3x1}$ and $T2_{3x1}$ are the same and the only variation between Af_1 and Af_2 holds in the Mk_{3x2} matrices. Furthermore, from equation (5), we can deduce that $M2_{3x2} = M1_{3x2}Rot_{2x2}^T$. Our aim is to remove the rotation term from the Mk_{3x2} matrices so that they become identical. Therefore, we propose to use the QR factorization tool. This factorization decomposes a matrix Mk_{3x2} into a product of a rotation matrix Qk_{3x3} and a triangular upper right matrix Rk_{3x2} so that $Mk_{3x2} = Qk_{3x3}Rk_{3x2}$. Since the rotation matrix is applied in the 2D patch space, it is a 2D rotation and so can be represented by a 2x2 matrix. So we rather propose to apply the QR factorization on the transpose of the Mk_{3x2} matrix. By this way, we will obtain $Mk_{3x2}^T = Qk_{2x2}Rk_{2x3}$ and hence $Mk_{3x2} = Rk_{2x3}^T Qk_{2x2}^T$ where the matrix Qk_{2x2} contains the rotation part of the matrix Mk_{3x2}. Since the matrix Rk_{2x3} is not sensitive to rotation variation we have $R1_{2x3} = R2_{2x3}$.

To summarize, considering two patches, we propose to:

- evaluate the affine transforms (from image space to color space) Af_1 and Af_2 of the patches by using the Song *et al.* method (Song et al., 2009) (equation (2)),
- decompose each transform Af_k into two transforms Mk_{3x2} (rotation, scale, stretch) and Tk_{3x1} (translation) (equation (6)),
- apply the QR factorization on the transposes of the Mk_{3x2} matrices giving two matrices Rk_{2x3} and Qk_{2x2}.

Previously, we have shown that if the second patch Pa_2 is a rotated version of the first one Pa_1, we have $T1_{3x1} = T2_{3x1}$ and $R1_{2x3} = R2_{2x3}$. Consequently, we propose to take these two matrices Tk_{3x1} and Rk_{2x3} as the rotation invariant descriptor for the patch Pa_k. Since the matrix Rk_{2x3} is a triangular upper right matrix, we consider only the 5 non-zero values among its 6 values. Thus, the descriptor we propose in this paper is constituted by only $3+5=8$ values.

Thus, the advantages of our descriptor is threefold. First since it represents both the colors and their relative spatial distribution in the patch space, it is highly discriminative and so can determine if two patches are similar or not. Second, the time processing required to evaluate the similarity between two patches is very low since each descriptor is constituted by only 8 values. Third, this descriptor is fully invariant to rotation in the patch space. The discriminative power and the rotation invariance property of our descriptor can be checked in the figures 1 and 2.

4 REPRESENTATION OF THE SPATIAL DISTRIBUTION OF THE ROTATION INVARIANT SELF-SIMILARITIES

We consider one patch Pa_0 in an image and we want to represent the spatial distribution of the self-similarities around this particular patch. If in its surrounding, one other patch Pa_1 has similar colors spatially arranged in a similar way as Pa_0, the similarity between their descriptors will be high, even if there exists a rotation between them in the image space. In this case, 3 values can be used to represent the

relative position and orientation of these patches: 2 values Δx and Δy represent the translation along the x and the y axis respectively, and 1 value $\Delta\theta$ represents the rotation angle between these two patches. The translation values are easily obtained by evaluating the differences between the x and y coordinates of the centers of the patches. The $\Delta\theta$ can be obtained from their respective Qk_{2x2} matrix introduced in the previous section. Indeed, these matrices represent the rotation in the image space applied to each patch so that they match the same position in the color space. Consequently if the patch Pa_0 is rotated by an angle θ_0 and the patch Pa_1 by an angle θ_1 in order to match the same position, the $\Delta\theta$ is just the difference between these two angles. Consequently, we propose to create a 3D structure around the patch Pa_0 whose axis are Δx, Δy and $\Delta\theta$ and to put the value of the similarity between Pa_0 and Pa_1 in the cell whose coordinates are $\{x_1 - x_0, y_1 - y_0, \theta_1 - \theta_0\}$, where x_k and y_k are the position in the image space of the center of the patch Pa_k, $k = 0$ or 1. Likewise, we can do the same for all the patches Pa_i, $i > 0$, around Pa_0 in order to represent the spatial distribution of the self-similarities around Pa_0. For this, the neighborhood of the patch Pa_0 is discretized into a grid of size 5x5 and the angle axis is discretized into 4 values. The maximal similarity is stored within one cell if several similarities are falling in the same position. This representation is similar to this proposed by Shechtman *et al.* (Shechtman and Irani, 2007) but since we can find similarities between patches with different orientations, we have added a third dimension for the angle. The dimension of the feature of Shechtman *et al.* was 4 radial intervals x 20 angles = 80 while our is 100 (5 Δx intervals x 5 Δy intervals x 4 angles).

5 EXPERIMENTS

5.1 Experimental approach

The rotation invariance of our self-similarity having been theoretically demonstrated, we propose to assess the discriminative power of the final descriptor and to compare it with the other self-similarities that are classically used in many applications, as mentioned in the introduction. For this purpose, we consider the context of object classification by using the PASCAL VOC 2007 dataset (Everingham et al.,). This dataset contains 9963 images representing 20 classes. The aim of this experiment is not to get state-of-the-art classification score on the considered dataset, but rather to fairly compare the discriminative powers of the different self-similarity descriptors. In order to test our self-similarity, we propose to use the Bag-of-words approach which is based on the following successive steps:

- keypoint detection (dense sampling is used for all tested methods),
- local descriptor extraction around each keypoint (we use the 3D structures presented in the previous section for our method),
- clustering in the descriptor space, the cluster representatives are called visual words (k-means is used with 400 words for all tested methods),
- in each image, each local descriptor is associated with the nearest visual word,
- each image is characterized by the histogram of visual word,
- learning on the train images and classification of the test images (linear SVM is used for all tested methods).

Furthermore, we propose to compare our results with the local self-similarity descriptor (Shechtman and Irani, 2007) and with the global self-similarity descriptor (Deselaers and Ferrari, 2010). For both, we use the codes provided by the authors. For the global self-similarity descriptor, we have constructed the SSH (with $D_1 = D_2 = 10$) from the image and consider each grid cell as a feature. So, this feature is also of dimension 100. For all the approaches, we reduce the dimension of the features to 32 by applying PCA.

5.2 Results

The results are shown in figure 3. In this figure, BOCS stands for Bag Of Correlation Surfaces (Deselaers and Ferrari, 2010), BOLSS stands for Bag Of Local Self Similarities and BORISS is our proposed approach and means Bag Of Rotation Invariant Self Similarities.

In this figure, we can see that our approach outperforms the two other ones for most of the 20 classes. Furthermore, the Mean Average Precision (MAP) provided by our BORISS is 25% while it is around 18% for the two other approaches. This experiment shows that the self similarities are more efficient to characterize the content of an image if they are invariant to the rotation. Of course, these results are not competitive with the ones provided by SIFT or CNN features, but the aim of these experiments was to show that adding color and rotation invariance into the self-similarity descriptors improves the discriminating power of the final features. These results clearly show that our self-similarity representation is a good candidate to be used as complementary information with the appearance-based features.

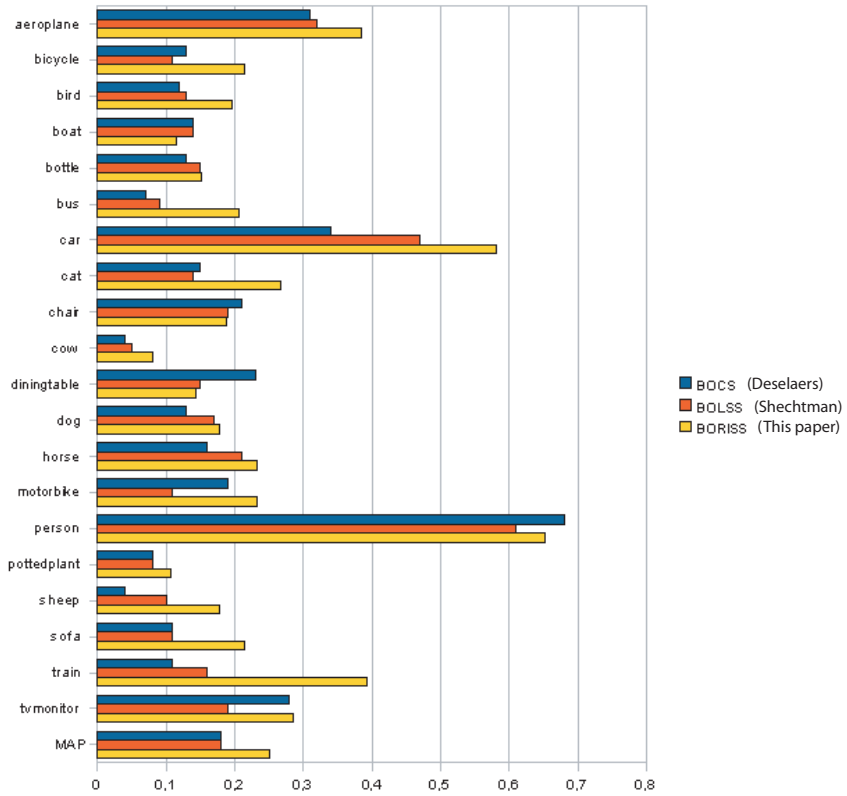


Figure 3: Mean average precision obtained by the three tested approaches on the VOC 2007 database.

6 CONCLUSION

In this paper, we have presented a new method to represent the spatial distribution of the self-similarities in an image. First, we have proposed to extract rotation invariant self-similarities. This extraction is based on a comparison of new spatio-colorimetric descriptors. We have shown that these descriptors extract discriminative information from local regions while being very compact and invariant to rotation. Then, we have proposed a 3D structure to represent the spatial distribution of these self-similarities. This structure informs about the translation and rotation there exist between two similar local regions. The experimental results provided by this method outperform those of the classical self-similarity based approaches. In this work, we found a way to represent translation and rotation that occur between self-similar regions and as future works we are trying to add the other possible transformation such as scale variation or stretch. Finally, the discriminative color descriptors introduced in this paper could be used as a color texture descriptor since it is representing both the colors and their spatial distributions within the local neighborhood.

REFERENCES

Chatfield, K., Philbin, J., and Zisserman, A. (2009). Efficient retrieval of deformable shape classes using local self-similarities. In *NORDIA workshop in conjunction with ICCV*.

Chih-Yuan, Y., Jia-Bin, H., and Ming-Hsuan, Y. (2011). Exploiting self-similarities for single frame super-resolution. In *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part III*, pages 497–510, Berlin, Heidelberg. Springer-Verlag.

Deselaers, T. and Ferrari, V. (2010). Global and efficient self-similarity for object classification and detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, DC, USA. IEEE Computer Society.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

Glasner, D., Bagon, S., and Irani, M. (2009). Super-resolution from a single image. In *ICCV*.

Kim, S., Min, D., Ham, B., Ryu, S., Do, M. N., and Sohn, K. (2015). Dasc: Dense adaptive self-correlation descriptor for multi-modal and multi-spectral correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2103–2112.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157 vol.2. IEEE Computer Society.

Shechtman, E. and Irani, M. (2007). Matching local self-similarities across images and videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Song, X., Muselet, D., and Tremeau, A. (2009). Local color descriptor for object recognition across illumination changes. In *ACIVS09*, pages 598–605, Bordeaux (France).

van de Weijer, J. and Schmid, C. (2006). Coloring local feature extraction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3952 of *Lecture Notes in Computer Science*, pages 334–348.

Wang, J., Lu, K., Pan, D., He, N., and kun Bao, B. (2014). Robust object removal with an exemplar-based image inpainting approach. *Neurocomputing*, 123:150 – 155.

Wang, Z., Yang, Y., Wang, Z., Chang, S., Han, W., Yang, J., and Huang, T. S. (2015). Self-tuned deep super resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

Zontak, M. and Irani, M. (2011). Internal statistics of a single natural image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.