A Color-Based Interest Operator

Marta Penas[1] and Linda G. Shapiro[2]

[1] Dpt. of Computer Science, University of A Coruña. 15071, A Coruña. Spain
mpenas@udc.es
[2] Dpt. of Computer Science and Engineering, University of Washington. US
shapiro@cs.washington.edu

Abstract. In this paper we propose a novel interest operator robust to photometric and geometric transformations. Our operator is closely related to the grayscale MSER but it works on the HSV color space, as opposed to the most popular operators in the literature, which are intensity based. It combines a fine and a coarse overlapped quantization of the HSV color space to find maximally stable extremal regions on each of its components and combine them into a final set of regions that are useful in images where intensity does not discriminate well. We evaluate the performance of our operator on two different applications: wide-baseline stereo matching and image annotation.

Keywords: interest operators, feature matching, HSV color space, wide-baseline stereo, image annotation.

1 Introduction

Feature matching is an important challenge in Computer Vision, with broad applications to image retrieval [1,2], video analysis [3,4] and motion tracking [5,6]. Local regions are well suited for matching, since they are robust to occlusion and background clutter. For this reason, several papers in the recent literature describe the use of interest operators [7,8,9,10] to detect image regions suitable for feature matching purposes. Such regions must exhibit some desirable properties such as: distinctiveness and invariability to scaling, rotation, 3D camera viewpoint or changes in illumination.

The interest operators in the literature can be divided in two main categories: those based on edges and corners that detect structured and highly textured regions and those based on intensity that detect blob like features. Examples of the first category are the Harris affine [8] and EBR [7] detectors. The MSER [9], Kadir [10] SIFT [11] and IBR [7] detectors fall in the second category. In many cases, the operators detect complementary image regions and, thus, it is common to find applications that combine their outputs [7,12].

The Harris affine operator [8] detects the interest points in the image through the Harris corner detector, then selects an adequate scale through the LoG kernel and estimates the affine shape of the interest region in the point neighborhood. The Edge Based Regions (EBR) [7] operator starts from interest points, also

P. Foggia, C. Sansone, and M. Vento (Eds.): ICIAP 2009, LNCS 5716, pp. 965–974, 2009.

detected using the Harris detector, and the edges that meet at each point to build an interest region that, as opposed to the Harris affine regions, is not centered at the interest point. The Kadir operator [10] defines the interest regions as those exhibiting unpredictability, in terms of the Shannon entropy, both in their local attributes and their spatial scale. Finally, the Scale Invariant Feature Transform (SIFT) operator [11] detects interest points as those locations invariant to scale changes, defined as the scale-space extrema in the DoG kernel. All the operators just mentioned detect ellipse or circle-shaped regions starting from the initial interest points. In contrast, the following interest operators detect arbitrary-shaped regions through the analysis of the image intensity.

The IBR (Intensity Based Regions) operator [7] defines the interest points as local extrema of the image intensity. The algorithm analyzes the intensity along several rays emanating from the initial interest points. On each of these rays, the point where the intensity changes significantly is selected and, by linking these points, the interest region is determined.

The MSER (Maximally Stable Extremal Regions) operator [9] detects a set of regions that exhibit some desirable properties for feature matching: covariance to adjacency preserving transformations and invariance to scale changes and affine transformations of image intensities. The detection of MSER regions starts with the thresholding of the image intensity at all the 256 gray values. At each threshold level, the pixels below the threshold are colored in black, while the pixels over the threshold are colored in white. The set of connected white components are the maximal regions of the level. Those maximal regions that are stable over a range of thresholdings constitute the maximally stable extremal regions of the image. The algorithm also computes the MSER regions in the inverted image. The performance of the operator is governed by three parameters: the minimum size of a region ms, the maximum size of a region as a percentage of the image size per and the minimum margin of the region mm, this is, the minimum stability of the maximal regions to be considered MSER regions.

A color variant of the MSER that operates in the RGB space was introduced in [13]. The extension to color is made by looking at successive time steps of an agglomerative clustering of image pixels. The selection of time steps is performed by analyzing the evolution of the color differences between neighboring image pixels, with the aim to process a uniform number of pixels per time step. This is not intuitive and not necessarily optimal in the general case, which is why our operator follows a different approach. It works on the HSV color space and imposes hard constraints on the color differences that largely increase the robustness of the method to parameter changes, as demonstrated in sec. 2.

Several color descriptors [14,15] have also been proposed recently. Though the use of color information in the region description could add value to our results, this analysis is beyond the scope of the current paper, where we have used the widely known SIFT descriptor [11] in our experiments.

We are mainly interested in the interest operators that are able to detect regions of arbitrary shape and, to the best of our knowledge, only [13] analyzes the image color, even though it contains important information that could be

of great use for feature matching. For this reason, in sec. 2, we introduce a new interest operator, closely related to the MSER detector, that operates on color and improves the results of the grayscale detector on those images where the intensity is not discriminative enough. Sec. 3 describes some practical applications that test the performance and robustness of our operator, and sec. 4 summarizes the conclusions from our work.

2 Color Based MSER

The MSER operator designed by J. Matas [9] operates on intensity, which is not always discriminative enough, as shown in figure 1. MSER was run on the grayscale version of the input color image and it could only find some of the green areas, which stand out well in the color image, but not in grayscale. Also, the stability of the regions is fragile and very dependent on the input parameters. In this paper we present a color interest operator, called COLOR, based on the same principles as the MSER. Our operator detects interest regions in the HSV color space and is less sensitive to changes in the input parameters and small interconnections among regions, as will be demonstrated.

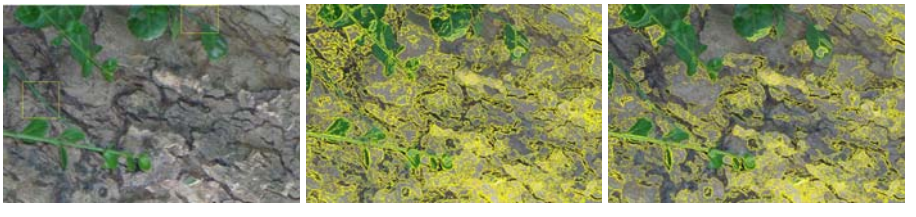

Fig. 1. From left to right: input image; output MSER regions with $mm = 5$ and 7, respectively. MSER was run on the grayscale version of the input.

The HSV color model has been chosen, since it is quite similar to the way humans perceive color. It defines color in terms of three components: the *hue* that ranges from 0 to 360 degrees and defines the color type, the *saturation* that defines the percentage of gray, and the *value* that defines the brightness of the color, both ranging from 0 to 1.

In order to detect interest regions in the HSV color space, an adequate quantization of its components must be first chosen. To do so, two important principles must be taken into account: a value of 0 represents the black color regardless of the hue and the saturation, and a saturation of 0 represents a gray color defined by the value regardless of the hue. Based on these principles, several quantizations can be found in the literature. Smith [16] designed a uniform quantization scheme of the color space into 166 colors. Zhang et al [17] suggest a non-uniform quantization into 36 colors. Huang et al. [18] also suggest a non-uniform quantization into 166 colors.

In this work, we have combined a fine and a coarse overlapped quantization of the HSV diagram. The COLOR operator produces sets of regions from all

three bands: value V, saturation S and hue H. The V-band is quantized into 125 bins and processed with a variant of the MSER operator that does not join pixels if their hues or saturations are very different. These additional constraints have been added in order to avoid joining regions with clear differences in color. We have found that a maximum difference of 0.125 in saturation and 15° in hue, which is a rather coarse quantization of these components, produces good results. Fig. 2 depicts a simplified pseudocode of the interest region detection in the V-band, which yields *value regions*.

```
1. Quantize the V-space
2. Sort pixels according to value.
3. Initialize an empty image I.
4. For each V-bin
   4.1. Place the pixels in the V-bin into the image I
   4.2. Compute the extremal regions in I, connecting two neighboring
        pixels only if their difference in hue is below 15 degrees and
        their difference in saturation is below 0.125.
5. Find minima of the rate of change of the area function
```

Fig. 2. Pseudocode for interest region detection in value

A pixel is analyzed according to its hue and saturation when its value is above 0.2 and its saturation above 0.1, since below these thresholds, the pixel can be considered gray or black and its hue and saturation are undefined. The S-band and the V-band are also quantized into 180 and 125 bins, respectively, and processed with a variant of the MSER that does not join pixels with similar saturation if they have very different hues and vice-versa. Again, the maximum difference in saturation is 0.125 and the maximum difference in hue is 15°. Processing the S-band yields *saturation regions*, while processing the H-band yields *hue regions*. Finally, a postprocessing stage produces the final set of COLOR regions, which is a union of the value, saturation and hue regions in which regions that are approximately the same have been combined; if two regions of similar size overlap by more than 95%, the larger is used.

Fig. 3 depicts a simplified pseudocode of the interest region detection in saturation. The pseudocode for interest region detection in hue is very similar to the one depicted in fig. 3 and, for this reason, it has not been included. As in [9], the current implementation details include the use of the efficient union-find algorithm to store and merge the regions in each bin.

In [9], the author states that the MSER regions can be defined on pixels that come from a totally ordered set, which is not the case with the hue, since it has circular continuity. Despite this fact, the regions detected on hue have proven to contain useful information. Also, since the regions are computed in two directions as in the MSER operator, the effect of the circular continuity can be mitigated by analyzing the hue clockwise starting from 0° and anti-clockwise starting from 180°.

```
1. Quantize the S-space
2. Sort pixels according to saturation (only if value > 0.1 and
   saturation > 0.2)
3. Initialize an empty image I.
4. For each S-bin
   4.1. Place the pixels in the S-bin into the image I
   4.2. Compute the extremal regions in I, connecting two neighboring
        pixels only if their difference in hue is below 15 degrees
5. Find minima of the rate of change of the area function
```

Fig. 3. Pseudocode for interest region detection in saturation

We have previously stated that the regions detected by our operator are more stable to changes in the input parameters than the regions detected by MSER. Figure 4 demonstrates this point since our operator detects the most relevant features in the image regardless of a large variation in the input parameters. The regions detected in the intensity are in fact less stable to parameter changes than the regions detected in the HSV, due to the additional restrictions we have applied on its components.

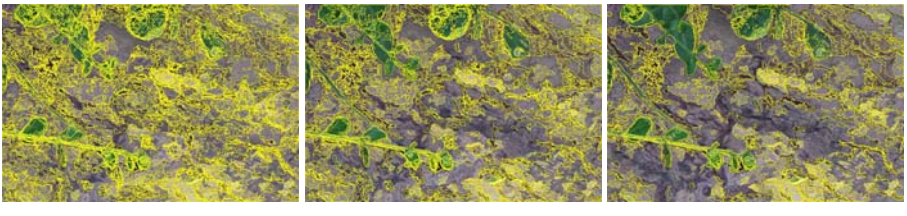

Fig. 4. Interest regions detected by our operator with different input parameters. $ms = 100$, $per = 0.05$, $mm_H = 6, 12, 18$, $mm_S = 5, 10, 15$ and $mm_S = 3, 5, 8$, respectively.

Our interest operator has been tested on several applications and compared with the operators described in sec. 1. The results of these tests have been summarized on the next section.

3 Evaluation of the Operator

In order to test the robustness of our interest operator to several image transformations in a real application, we have implemented a wide-baseline stereo matching system very similar to that in [7]. The goal of such a system is to find the homography that defines the transformation between two scenes. To this end, first a set of interest regions are extracted from the input images and then a descriptor is assigned to each of these regions. We have used the SIFT descriptor [11] in our experiments. Since the SIFT descriptor applies to elliptical regions, and not all the interest operators produce ellipse-shaped regions, the outputs of such operators have been approximated by the best-fitting ellipse.

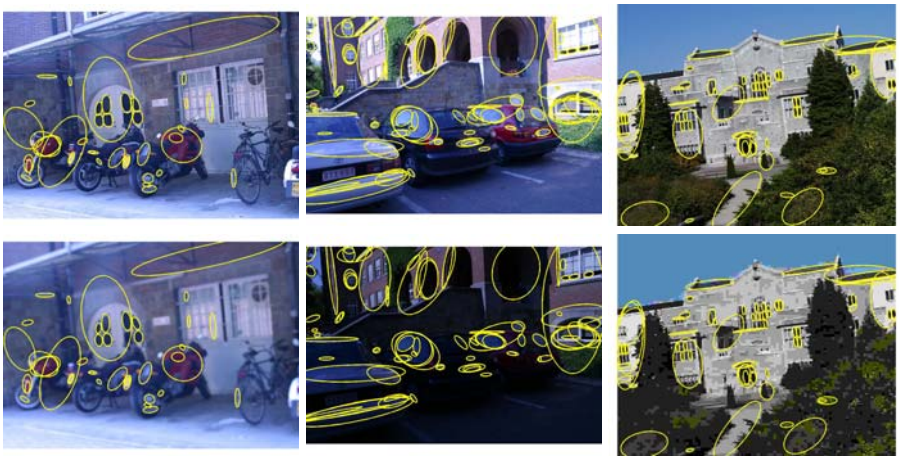

Fig. 5. Top row: reference images for the 'bikes', 'leuven' and 'ubc' image series, respectively. Bottom row: last image of each series.

In order to establish correspondences between regions, a nearest-neighbor based scheme has been used. A tentative match is established between two regions if they are mutual nearest neighbors in terms of Euclidean distance, and this distance is below a predefined threshold. In order to reject false initial matches, a geometric constraint is applied next. The homography transforming two image patches is computed and the match is retained only if its homography is compatible with that defined by at least min_c other matches. We have found that, as stated in [7], $min_c = 8$ produces good results. Finally, the homography between images is computed applying RANSAC on the retained matches.

The wide-baseline stereo matching has been tested on the set of images listed in `http://www.robots.ox.ac.uk/~vgg/research/affine/` since these images are subject to several common transformations like blur or viewpoint change, and the groundtruth for such transformations is available. In total, eight series of six images are available, each composed of a reference image and its successive transformations from easiest to hardest. The homographies between the reference image and the other images in the series are available.

Fig. 5 shows two images of three series subject to photometric transformations with the matching regions detected by our operator superimposed. Our operator is able to solve the correspondences between all the images in each of these series, which proves its robustness to transformations such as blur, illumination changes, and JPEG compression.

Fig. 6 shows two images of three series subject to geometric transformations. In this case, our operator is not always able to find the correspondences between images; for this reason its performance has been compared with the operators described in sec. 1 in order to determine if these operators are more robust than ours to geometric transformations. The results of this comparison are summarized in table 1.

Fig. 6. Top row: reference images for the 'boat', 'graf' and 'wall' image series, respectively. Bottom row: second image of each series.

On the 'bark' series, our operator is not able to find any correspondence, in fact, only the Harris affine operator is able to find one of the five correspondences. On the 'boat' series, our operator is able to find two correspondences, the same number found by all the other operators. On the 'graph' series, our operator finds three correspondences, the same as the MSER and the Harris affine operator, while the Kadir, IBR and EBR operators find only two correspondences. Finally, on the 'wall' series, our operator finds all the correspondences, the same as the other operators except the IBR, which is only able to find three correspondences.

These results show that the proposed operator is robust to both, geometric and photometric transformation, which is of great importance when its outputs are used for feature matching, as will be the case with the following example.

Our operator has also been tested on an image annotation framework. To this end, it has been used as the base region detector in a Spatial Pyramid Matching annotation system [19]. Such a system operates in several stages: first, the

Table 1. Scene matching results for images subject to geometric transformations

	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
	Bark					Boat					Graf					Wall				
Harris	✗	✗	✗	✓	✗	✓	✗	✗	✓	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓
Kadir	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
EBR	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
IBR	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗
MSER	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓
COLOR	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓

Table 2. Results obtained by different interest operators on a subset of the Caltech 256 dataset

	252	**253**	**251**	**145**	**129**	**182**	**140**	**130**	**Mean**
Kadir	**98.36**	92.29	85.64	98.42	**99.65**	81.14	79.10	73.30	**72.71**
Harris	98.11	93.07	**90.54**	95.29	99.17	76.39	**79.68**	70.90	70.48
MSER	97.47	92.80	87.33	**99.14**	93.73	**81.30**	75.63	67.80	69.67
COLOR	98.35	**94.13**	86.41	98.69	96.59	79.86	79.44	**74.07**	71.53

interesting regions of an image are detected and a descriptor is assigned to each of these regions; again the SIFT descriptor has been used. Then, a vocabulary is generated based on a subset of the training images. Each image region is assigned to a dictionary bin, and the image signature is generated by the concatenation of a set of histograms counting the number of regions per bin in several recursive partitions of the input image. Finally, a spatial pyramid matching kernel that weights the differences between histogram bins according to their partition level and position, is used to compute the differences between signatures, which are finally used as the input kernel for an SVM based annotation.

This annotation framework has been applied to a subset of the Caltech 256 dataset. Concretely, the 50 object categories that produce the best results, according to [20], have been used to test the performance of our operator and compare it with the operators introduced in sec. 1. It should be noted that these operators retrieve different image regions, and the best results would probably be obtained through a combination of their results, as is common in annotation frameworks nowadays [12]. This application is just a comparison to determine if our operator is competitive with those most widely used in the literature. Two disjoint sets of 30 and 50 images have been used for training and testing, respectively. The accuracy of the system has been measured in terms of the percentage average precision, which ranges from 0 to 100, and reflects the performance on the relevant images, rewarding systems that retrieve the relevant images quickly. Table 2 shows the results obtained by the operators, Kadir, Harris affine, MSER and COLOR operator, on a subset of the selected categories. The last column in table 2 is the global mean result on the 50 categories analyzed.

The first thing to note is that, on average, the results produced by the four operators are very similar. The Kadir operator has the highest mean average precision, followed by our operator, though it is not robust to affine changes. This could be the case because the "easiest" categories on the Caltech 256 dataset do not contain large viewpoint changes that could affect its performance. Each operator works best on different categories, for example, our color operator produces the best results on 13 categories, the second best result on 20, the third best on 11 and the worst on 6 categories.

Since our operator was designed to improve the original grayscale MSER operator, the comparison to MSER is the most important. Our operator outperforms the MSER operator on 37 of the 50 selected categories, and it achieves a 6.13% gain in performance measured in terms of mean average precision.

4 Discussion and Conclusions

In this paper we have introduced an interest operator based on the MSER operator designed by Matas [9] that operates on color, concretely on the HSV space, where it combines a fine and a coarse quantization in order to find the maximally stable regions present in each component. It detects the most important and distinctive features in the input images and produces high quality results in both structured and non-structured images.

We have tested our operator on two different application. First, a wide-baseline stereo matching applied to scenes undergoing different photometric and geometric deformations has been used to test the robustness of our operator to such deformations. Then, an image annotation application, where our operator produced the best results on some categories and achieved a comparable performance to the most widely used operators in the literature. Our operator outperformed the MSER operator in most of the categories and in terms of average performance. Both applications assess the suitability of our interest operator for feature matching.

Acknowledgements

This work has been funded by the Xunta de Galicia and the *Secretaría de Estado de Universidades e Investigación* of the Ministry of Science and Education of Spain, and partially supported by the U. S. National Science Foundation under grant No. IIS-0705765.

References

1. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(5), 530–535 (1997)
2. Tuytelaars, T., Gool, L.V.: Content-based image retrieval based on local affinely invariant regions. In: International Conference on Visual Information Systems, pp. 493–500 (1999)
3. Sivic, J., Zisserman, A.: Video google: a test retrieval approach to object matching in videos. In: International Conference on Computer Vision (2003)
4. Schaffalitzky, F., Zisserman, A.: Automated location matching in movies. Computer Vision and Image Understanding 92, 236–264 (2003)
5. Harris, C.: Geometry from visual motion. Active Vision, 263–284 (1992)
6. Torr, P.: Motion segmentation and outlier detection. PhD thesis, University of Oxford (1995)
7. Tuytelaars, T., Gool, L.V.: Matching widely separated views based on affine invariant regions. International Journal on Computer Vision 59(1), 61–85 (2004)
8. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal on Computer Vision 60 (2004)
9. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Video Conference, pp. 384–393 (2002)

10. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: European Conference on Computer Vision, pp. 404–416 (2004)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
12. Marszałek, M., Schmid, C., Harzallah, H., van de Weijer, J.: Learning object representations for visual object class recognition. In: Visual Recognition Challenge Workshop, in conjunction with ICCV (October 2007)
13. Forssen, P.: Maximally stable colour regions for recognition and matching. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
14. Van de Weijer, J., Schmid, C.: Coloring local feature extraction. In: European Conference on Computer Vision, pp. 334–348 (2006)
15. Van de Sande, K., Gevers, T., Snoek, C.: Evaluation of color descriptors for object and scene recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
16. Smith, J.R.: Integrated spatial and feature image system: retrieval, analysis and compression. PhD thesis, Columbia University (1997)
17. Zhang, L., Lin, F., Zang, B.: A CBIR method based on color-spatial feature. In: IEEE Region 10 International Conference TENCON, pp. 166–169 (1999)
18. Huang, C., Yu, S., Zhou, J., Lu, H.: Image retrieval using both color and local spatial feature histograms. In: Int. Conference on Communications, Circuits and Systems, pp. 927–931 (2004)
19. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178 (2006)
20. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical report 7694, California Institute of Technology (2007)