Analysis of spatial partitioning approaches for image classification

Aleksej Avramović, *Student Member, IEEE,* Vladimir Risojević, *Member, IEEE*

***Abstract*—Spatial partitioning is proven to be beneficial for the tasks of image classification, scene categorization and object recognition. The most popular method to capture rough spatial structure of the scene is spatial pyramid matching. However, spatial pyramid matching results in an image representation that is sensitive to rotations. In this research we investigate the influence of upright and rotated partitions on image classification regardless of the image filtering step. We show that simple combination of rotated spatial partitions improves classification accuracy up to 10% compared to single spatial partition commonly used in spatial pyramid matching.**

***Index Terms*—Spatial pyramid matching, rotation invariance, image classification.**

I. Introduction

SPATIAL partitioning is proven to benefit the accuracy in the tasks of image classification, scene categorization and event recognition. In their seminal work, Oliva and Torralba [1] considered images depicting different scenes and concluded that the spatial envelope of a scene can be captured by dividing the image using 4×4 spatial grid, and representing each block using the statistics of Gabor coefficients.

Lazebnik *et al.* introduced the spatial information into the bag-of-words model [2]. They partitioned the image into subregions on different levels of its pyramidal decomposition and locally pooled codewords obtained by vector quantizing SIFT descriptors. When comparing the obtained histograms of codeword appearances they are weighted according to the corresponding pyramid levels. The authors termed this approach *spatial pyramid matching* and showed that it considerably improves the scene classification performance compared to the original bag-of-words implementation.

Following the same line of research in [3] CENsus TRansform hISTogram (CENTRIST) descriptor is proposed for scene classification. CENTRIST is also based on partitioning an image into subregions and integrating the filtering results in these subregions. More specifically, CENTRIST uses the information from 31 spatial blocks on three levels of the spatial pyramid. The same spatial partitioning scheme is used in mCENTRIST descriptor [4] which is a multichannel extension of CENTRIST. Although this kind of spatial partition helps in encoding rough global structure of an image thus improving the accuracy on the scene classification tasks, it is sensitive to rotation, which can limit discriminative power, e.g. in the cases of texture classification and aerial image classification.

This paper addresses the described shortcoming of the approaches based on spatial pyramid matching. In order to improve robustness to rotation, as well as classification accuracy we propose partitioning the image into rotated subregions in addition to upright partitioning commonly used in spatial pyramid matching. In this way we examine the effect of upright and rotated partitions on classification accuracy. Our approach is independent on the image filtering step and can be used in the cases where spatial pyramid matching has been used. We performed our experiments on three publicly available image datasets and showed consistent improvements in classification accuracies when rotated blocks are included in descriptor extraction.

The recent paper [5] deals with the same problem by using different approach. The classifiers are trained and validated using randomly generated spatial partitions. Two approaches are then tested. The first one tries to find the best performing partition for each image category, and then uses that partition for test images. The second one uses boosting to assign weights to different partitions for each category. The authors show that this approach leads to performance improvements compared to the basic spatial pyramid matching model on different datasets. This approach can be regarded as the late fusion of information obtained using different spatial partitions. On the other hand, our approach corresponds to the early fusion since we do not train classifiers using different partitions but incorporate the information from different spatial partitions into a single image descriptor and leave to the classifier to weigh them appropriately.

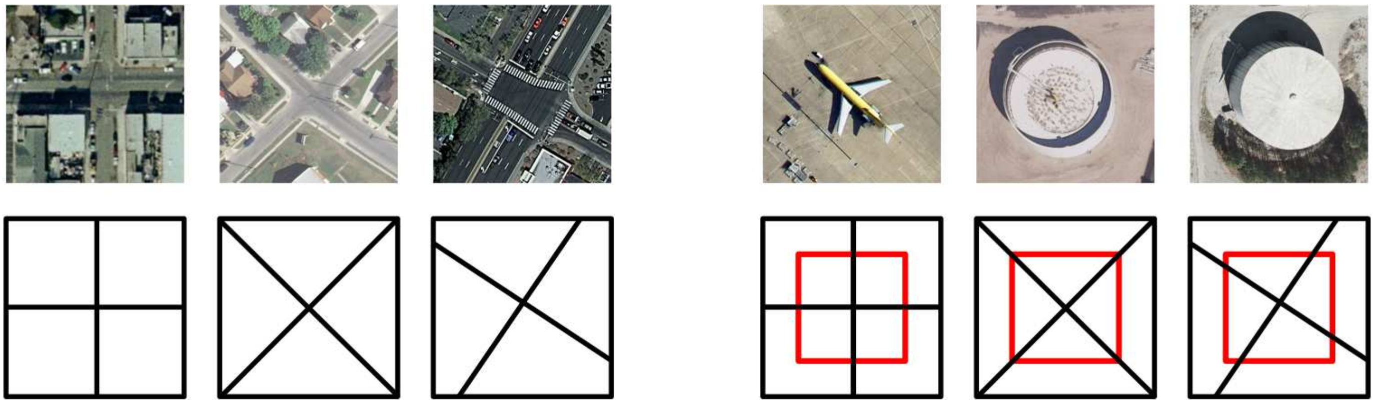
The rest of the paper is organized as follows. In Section II a detailed motivation for this research is given. Section III briefly describes used data and methodology, while Section IV gives description and results of the performed experiments. Section V gives concluding remarks.

II. Motivation

Although dividing the image space using the upright spatial grid on multiple pyramid levels improves image classification and scene categorization accuracies, various inter-class variations of the spatial layout of an image cannot be fully represented using simple upright grids. For example, let us consider the aerial images showing intersections given in the first row of Fig. 1a. We can notice that the intersections are differently rotated which implies that the usage of the upright rectangular grid, as in the spatial pyramid matching, may not

A. Avramović is with the School of Electrical Engineering, University of Belgrade Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia and the Faculty of Electrical Engineering, University of Banja Luka (e-mail: aleksej@etfbl.net)

V. Risojević is with the Faculty of Electrical Engineering, University of Banja Luka, (e-mail: vlado@etfbl.net)



(a) Examples of aerial images of intersections with different rotations. A rotated spatial grid which depicts the outline of intersection is given below each image.

(b) Additional areas in the center of image can be used to exploit information on objects given in images.

Fig. 1. Different spatial layouts in aerial images and corresponding spatial partitions.

completely capture the spatial layout of the image. On the other hand, the rotated spatial partitions, shown in the second row of Fig. 1a much better correspond to the actual spatial layouts in the images.

In this research we will examine the effect of using four differently rotated image partitions, given in Fig. 2, on classification accuracy. In each of the cases, the image filtering results are integrated in each of the shown subregions and the obtained descriptors are concatenated. The first case, shown in Fig. 2a corresponds to the partition used in spatial pyramid matching, while the others are its rotated variants. We also examine the descriptors obtained by concatenation of descriptors computed for differently rotated partitions, as well as descriptors computed on different levels of the pyramid.

Another interesting aspect of exploiting the spatial layout of images arises from the task of object recognition. For example, CENTRIST descriptor [3] assumes that the object of interest is usually close to the center of the image so subregions around the center of the image are used to extract the information about the shown object. The examples are given in the second row of Fig. 1b where we can notice additional areas marked with red, which are chosen to extract information about the objects shown in the images in the first row of Fig. 1b.

III. USED DATA AND METHODOLOGY

For the purpose of testing of the effect of different spatial partitions on image classification accuracy, we used three publicly available datasets. The first one is the UC Merced (UCM) dataset which contains 2100 high-resolution color aerial images of size 256×256 pixels and spatial resolution of 30 cm taken from USGS National Map [6][1]. These images have been manually classified into 21 semantic categories. As in the previous papers that report experiments on this dataset, we randomly split 100 images from each category to 80 training and 20 test images.

The second dataset is Scene15 dataset, which is focused on the task of scene recognition and contains 15 categories of different natural scenes[2]. Each category contains between 216 and 400 grayscale images of different sizes. In this case, we randomly select 100 images for classifier training and the remaining images are used for testing.

The third dataset is UIUC dataset and it contains 25 categories of texture images[3]. Each category contains 40 grayscale images of textures with different scales and rotations. In the experiments we used 30 images from each category for training and the rest for testing.

The spatial partitioning schemes shown in Fig. 2 can be used with different image filtering steps. In this paper we chose two filters which have shown good performance on scene recognition and texture classification tasks. The first one is Census transform which is used as a filtering step in computing CENTRIST descriptor [3]. After the filtering, histograms of the responses are computed for each of the subregions and concatenated. The original descriptor uses only upright image partition, marked as **Partition1** in Fig. 2a, for the first level partitioning. We compute the descriptor using multiple spatial partitions shown in Fig. 2, and concatenate the obtained histograms. Therefore the resulting dimensionality is $\#Subregions \times 254$.

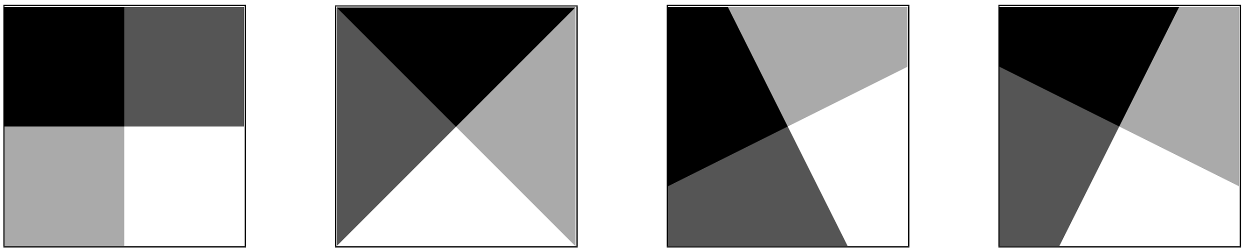
The second image filter used in the experiments is Local Binary Pattern (LBP) operator used for computation of LBPriu2 descriptor [7]. In the original descriptor the histogram of operator outputs is computed for the entire image, while in this paper we compute separate histograms for each of the subregions given in Fig. 2 and concatenate them. We compute the operator outputs using three resolutions with 8, 16 and 24 points and radii 1, 2 and 3, respectively, which yields histogram with 54 bins for one subregion. The dimensionality of the resulting descriptor is $\#Subregions \times 54$.

In all experiments we use support vector machine (SVM) with linear kernel [8] for classification. Multiclass classification is obtained by training binary classifiers for each class separately in one-vs-all manner and classifying the test

[1] This dataset is publicly available at: http://vision.ucmerced.edu/datasets

[2] This dataset is publicly available at: http://www-cvr.ai.uiuc.edu/ponce_grp/data/scene_categories/scene_categories.zip

[3] This dataset is publicly available at: http://www-cvr.ai.uiuc.edu/ponce_grp/data/texture_database/



(a) **Partition1** Non-rotated (upright) spatial partition.

(b) **Partition2** Spatial partition rotated 45 degrees.

(c) **Partition3** Spatial partition rotated 22 degrees counterclockwise.

(d) **Partition4** Spatial partition rotated 22 degrees clockwise.

Fig. 2. Rotated spatial partitions.

sample into the class which corresponds to the maximum SVM response. The penalty factor is set to $C = 2^{-5}$ as in [4].

In order to avoid the bias of a specific training/test data split we performed the experiments with five different splits and averaged the results. To ensure fair comparison, all experiments with different descriptors or parameters use the same training/test data split in each of the five runs.

IV. Experimental results and discussion

A. *Effect of rotated spatial partitions*

In the first experiment we use only the rotated spatial partitions given in Fig. 2 for the descriptor extraction along with the feature vector extracted from the original image. The original image is previously resized to contain the same number of pixels as each partition. Descriptors from each subregion are concatenated into the final image descriptor.

We consider the following six cases:

i **Case1** Includes resized original image and *Partition1* set,
ii **Case2** includes resized original image and *Partition2* set,
iii **Case3** includes resized original image and *Partition3* set,
iv **Case4** includes resized original image and *Partition4* set,
v **Case5** includes resized original image, *Partition1*, *Partition2*, *Partition3* and *Partition4* set,
vi **Case6** includes resized original image, *Partition1*, *Partition2*, *Partition3*, *Partition4* and additional central area as in Fig. 1b.

In the first four cases we use rotated partitions, thus it is enough to extract local features once from the original image and to pool them into appropriate histograms according to their spatial properties. No additional features were extracted, but the existing features are pooled into four additional histograms. The final descriptor is concatenation of histograms extracted from the resized image and four histograms obtained from the rotated partitions. **Case5** uses descriptors that contain histograms of partitions from all the previous cases. Furthermore, **Case6** uses an additional histogram extracted from the central area.

Since UCM dataset contains color images, we conducted experiments for the six cases described above using both grayscale and color images. For color images we used RGB and Opponent color spaces. Images are converted from RGB to the Opponent color space according to the equation given in [9]:

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

Descriptors are computed for each color channel separately and concatenated.

TABLE I
Average classification accuracies and standard deviations on the UCM dataset, for six spatial partitions. The dimensionalities of descriptors per color channel are given in brackets next to the specific cases.

Case		Spectral Band		
		Grayscale	RGB	Opponent
CENTRIST	Case1 (1270)	57.33±(3.21)	70.10±(2.94)	71.52±(2.55)
	Case2 (1270)	61.19±(2.11)	73.14±(1.66)	75.05±(1.78)
	Case3 (1270)	61.29±(2.77)	73.33±(1.31)	75.81±(1.39)
	Case4 (1270)	61.52±(2.55)	72.19±(3.28)	74.62±(1.74)
	Case5 (4318)	67.90±(2.56)	77.48±(1.85)	80.05±(1.71)
	Case6 (4572)	70.48±(2.83)	79.00±(2.18)	81.33±(2.10)
LBPriu2	Case1 (270)	44.33±(2.42)	59.00±(3.36)	60.81±(2.85)
	Case2 (270)	44.33±(2.63)	58.57±(3.26)	60.29±(3.13)
	Case3 (270)	43.71±(2.48)	58.52±(3.16)	61.10±(3.61)
	Case4 (270)	44.29±(2.46)	58.71±(2.98)	60.67±(2.62)
	Case5 (918)	54.71±(3.14)	67.57±(2.56)	70.24±(2.85)
	Case6 (972)	57.62±(3.11)	69.62±(3.02)	72.10±(2.59)

For the UCM dataset the experimental results are given in Table I. If we compare the first case with the second three cases, for CENTRIST descriptor, we can notice that rotated spatial partitions result in increased classification accuracy. That might indicate that rotated image layouts are more common in this dataset than the layout given by **Partition1** in Fig. 2a. However, this behaviour is not observed in the case of LBPriu2 descriptor.

Furthermore, the results are around 10% better for both descriptors in the **Case5** when all four partitions are included in the descriptor computation. This is due to the fact that images in the same category of UCM dataset often have different orientations and inclusion of different spatial partitions into the descriptor computation helps in achieving robustness to rotations.

We can also notice that additional information extracted from the central area in the **Case6** generally improves classification accuracy up to 3%. This happens primarily because it increases the classification accuracy on the classes which contain characteristic objects located near the center of the image. This property can be also useful in the task of object recognition.

The same set of experiments are conducted on UIUC texture dataset as well as on Scene15 dataset and the results are given in Table II. In this case, however, the results for **Case1–Case4** are pretty much the same for both descriptors, which indicates that there is no preferred partition orientation in these cases. However, for both datasets concatenation of descriptors computed using different spatial partitions (**Case5**) is beneficial. This is especially prominent in the case of UIUC dataset which contains rotated images and the classification accuracy can be improved 10% to 15% when combination of differently rotated partitions is used. Increase in the classification accuracy in **Case5** for Scene15 dataset is somewhat smaller because in natural scenes there are no significant variations in orientation. Finally, inclusion of the central area in **Case6** results only in marginal improvements in the classification accuracies because neither texture images in UIUC dataset nor scene images in Scene15 dataset do not contain salient objects near the center of the image.

TABLE II
AVERAGE CLASSIFICATION ACCURACIES AND STANDARD DEVIATIONS ON UIUC AND SCENE15 DATASETS, FOR SIX SPATIAL PARTITIONS. THE DIMENSIONALITIES OF DESCRIPTORS PER BAND ARE GIVEN IN BRACKETS NEXT TO THE SPECIFIC CASES.

Case		Dataset	
		UIUC	Scene15
CENTRIST	Case1 (1270)	59.04±(2.99)	72.92±(0.56)
	Case2 (1270)	58.96±(4.05)	73.38±(0.61)
	Case3 (1270)	59.28±(3.65)	73.80±(0.58)
	Case4 (1270)	59.92±(3.77)	73.64±(0.60)
	Case5 (4318)	72.80±(2.10)	76.47±(0.85)
	Case6 (4572)	73.44±(2.54)	76.84±(0.84)
LBPriu2	Case1 (270)	45.28±(1.12)	53.09±(0.90)
	Case2 (270)	46.64±(2.44)	55.34±(0.61)
	Case3 (270)	46.24±(2.34)	55.18±(1.16)
	Case4 (270)	46.96±(2.07)	54.71±(0.85)
	Case5 (918)	61.20±(2.45)	62.77±(0.51)
	Case6 (972)	62.16±(2.27)	63.26±(0.51)

B. Pyramidal image partitioning

In the first set of experiments we computed the descriptors using the entire image as well as different spatial partitions shown in Fig. 2. These correspond to pyramid levels 0 and 1 in the terminology of spatial pyramid matching. Here we examine the possibility to include further pyramid levels in the computation of the descriptor. However, while including further pyramid levels can improve classification accuracy it also significantly increases descriptor dimensionality. Generally, if we use L levels, the total number of upright and rotated subregions from which we extract features can be calculated as

$$Nsubreg = 1 + 16 \times \frac{(4^L - 1)}{3} + 1 + 4 \times \frac{(4^{L-1} - 1)}{3} \quad (2)$$

So, if we take into consideration that every subregion gives additional 254 bins (for CENTRIST) and 54 bins (for LBPriu2), total dimensionality of descriptor increases rapidly with the increase of the number of pyramid levels. To overcome this problem, we use descriptor dimensionality reduction technique based on Principal Component Analysis to reduce dimensionality of every block to 32.

In this experiment we add the second level of the pyramid to the computation of the descriptor and compute subregion descriptors as in **Case6** for each of the subregions from the first level. All subregions are resized to have the same size and prior to concatenation normalization of histograms extracted from each subregion is done. The classifier and the experimental setup are the same as in the first experiment. The obtained results are given in Table III. Since UIUC and Scene15 datasets do not contain color images the results are given only for the grayscale case.

TABLE III
CLASSIFICATION ACCURACIES FOR DIFFERENT DATASETS IN THE CASE WHEN THE SECOND LEVEL PARTITIONING IS USED. NUMBERS IN BRACKETS, NEXT TO THE CLASSIFICATION ACCURACIES ARE STANDARD DEVIATION CALCULATED FOR 5-FOLD CROSS VALIDATION. THE DIMENSIONALITY OF DESCRIPTORS IS 2592 FOR EACH CASE.

Dataset		Accuracy (%)
CENTRIST	UCM	74.48 ± (2.13)
	UIUC	81.60 ± (1.65)
	Scene15	79.77 ± (0.19)
LBPriu2	UCM	60.00 ± (1.74)
	UIUC	81.92 ± (1.78)
	Scene15	70.79 ± (0.37)

Comparing the results from the Tables I and II and Table III we can notice additional improvement of classification accuracy in the cases when the second level partitioning is included. If we compare classification rates with the results reported in the literature, we can notice that this simple approach can achieve results better or comparable to the state-of-the-art. Classification rate of 74.5% on UCM database obtained using CENTRIST is better than 73.4% reported in [4]. For the Scene15 dataset obtained classification accuracy is 79.8% which is lower than 83.9 % reported in [3], but that result is obtained using SVM with radial basis kernel, which is computationally more expensive.

V. CONCLUSION

In this paper we examined the influence of rotated image partitioning schemes on classification accuracy. We noticed that usage of rotated blocks on the first level of partitioning can be beneficial in the cases when dataset contains rotated images, such as aerial images. Also, we showed that concatenation of features extracted from the partitions with different rotation angles can improve classification rates up to 10%, compared to using single partition. The observed improvement is obtained with both CENTRIST and LBPriu2 descriptors on three datasets. Moreover, adding the second level of the pyramid into the partitioning scheme improves the classification accuracies even further. The main drawback of the proposed scheme is large dimenstionality of the obtained descriptors. We plan to further investigate this issue in the future work.

References

[1] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vision*, vol. 42, no. 3, pp. 145–175, May 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1011139631724

[2] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, vol. 2, 2006, pp. 2169–2178.

[3] J. Wu and J. Rehg, "CENTRIST: A visual descriptor for scene categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, Aug 2011.

[4] Y. Xiao, J. Wu, and J. Yuan, "mCENTRIST: A multi-channel feature generation mechanism for scene categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 823–836, Feb 2014.

[5] Y. Jiang, J. Yuan, and G. Yu, "Randomized spatial partition for scene recognition," in *ECCV*, ser. Lecture Notes in Computer Science, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Springer Berlin Heidelberg, 2012, pp. 730–743.

[6] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *ACM SIGSPATIAL GIS*, 2010, pp. 270–279.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul 2002.

[8] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. and Technol*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[9] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, sep 2010.