Compass Radius Estimation for improved Image Classification using Edge-SIFT

Compass Radius Estimation for improved Image Classification using Edge-SIFT

E. Fidalgo[1], E. Alegre[2], V. González-Castro[3], L. Fernández-Robles[4]

[1,2] Department of Electrical, Systems and Automatics Engineering. University of León, Campus de Vegazana S/N 24071, León, Spain.
[3] Neuroimaging Sciences. Centre for Clinical Brain Sciences. University of Edinburgh. 49 Little France Crescent. Edinburgh EH16 4SB (United Kingdom)
[4] Department of Mechanical, Informatics and Aerospace Engineering. University of León, Campus de Vegazana S/N 24071, León, Spain.

[1] **efidaf00@estudiantes.unileon.es – Corresponding author.**
[2] enrique.alegre@unileon.es
[3] victor.gonzalez@ed.ac.uk
[4] l.fernandez@unileon.es

Abstract:
The combination of SIFT descriptors with other features usually improves image classification, like Edge-SIFT, which extracts keypoints from an edge image obtained after applying the compass operator to a colour image. We evaluate for the first time, how the use of different radii in the compass operator affects the classification performance. We demonstrate that the value proposed in the literature, radius = 4.00, is not the optimum from an image classification point of view. We also put in evidence that in ideal conditions, choosing an appropriate radius for each image yields accuracy values even higher than 95%. Finally, we propose a new method to estimate the best radius for the compass operator in each dataset. Using a training subset selected on the basis of a minimum dispersion criterion of edges density, we construct a richer dictionary for each dataset in our Bag of Words pipeline. From that dictionary it is selected a radius for the whole dataset that yields higher accuracy than using the value proposed in the literature. Using this method, we obtained improvements in the accuracy up to 24.4% in Soccer, 6.77% in COIL-RWTH-2, 4.46% in Birds, 3.82% in ImageNet_Dogs, 2.75% in ImageNet_Birds, 2.02% in Flowers and 1.75% in Caltech101 datasets.

Keywords: Image Classification, Bag of Words, dense SIFT, Edge-SIFT, Edge Compass Operator, Radius estimation, Support Vector Machine.

1. Introduction

Feature extraction is one of the most critical steps in image classification and, for this reason, the performance of different kinds of descriptors has been widely studied in the literature. One very well-known example are SIFT descriptors [1], which have mainly been used to obtain intensity-based information related to the appearance of objects [2]. Other works describe objects and scenes using different characteristics based either on colour features [3], histograms [4], adding chromatic information to the Bag of Words descriptor [5], employing Colour Name features [6] or proposing adaptive colour mathematical-morphology-based descriptors [7] among others. There are also approaches that characterize objects of interest by means of their texture [8]. Among the last ones, Local Binary Patterns (LBP) [9] are

probably the most popular in the literature, either some of the original proposals from Ojala et al. [9] or any other of the several improvements that have been proposed more recently [10][11][12].

The information is relevant when the descriptors are extracted just from the region of interest, or foreground. The data extracted from the background do not provide relevant information to the image description, so extracting features from the whole image may result in suboptimal classifications [13].

There are several ways to deal with this drawback and select an appropriate region of interest to work with. One of them is visual saliency, which highlight the most relevant objects or regions by means of different methods. A complete revision on recent advances in Visual Saliency can be found in [13]. Other efficient methods for image segmentation are the active contour models [14], which extract the contour of the object of interest using energy minimization. The success of these methods is demonstrated in [14], [15] and [16].

Following a different perspective, Xie et al. [17] proposed the extraction of a new set of descriptors, called Edge-SIFT. This method extracts the dense SIFT descriptors from the edges of the image, which are obtained by means of the compass operator proposed by Ruzon et al. [18]. The information extracted from these edges complements the one provided by the dense SIFT descriptors extracted from the original image. Henceforth, these images will be called “edge images”.

The information contained in the edge image depends on the radius parameter used to calculate it. Although compass operator has been widely used [19][20][21][22], the parameter to obtain the edge images always remains fixed. In the experiments carried out by Xie et al. the value of the compass operator radius is fixed as proposed in literature [17] for all images. The use of a fixed radius will not always guarantee that the edges of the objects of interest are contained in the resulting edge image but also might contain noise from the background.

In this work we demonstrate how the radius of the compass operator affects the type of information extracted from the image. It influences the number of features extracted from both the background and foreground and, thus, the quality of the description and the classification results. The objective will be to create edge images that mainly contain descriptors from the object of interest in order to obtain a more accurate object description and classification. Moreover, it is empirically demonstrated that if the best radius parameter to create an edge image is manually chosen for each image on a dataset, the accuracy obtained would outperform the results obtained when only one radius is used for all the dataset.

As we have discussed, the radius selected for the compass operator influences the amount of information that the edge image provides. So we propose an algorithm that predicts what it will be the best parameter to use in the complete dataset. This estimation is carried out by means of a richer dictionary

built using a subset of images selected upon a minimum dispersion criterion. The edge-based descriptors computed using this radius yield better accuracy when they are combined with the dense SIFT descriptors. In the experiments carried out using several datasets it is demonstrated that the accuracy obtained when using the estimated radius is superior than the one obtained when using the fixed radius proposed in the literature [17].

The rest of the paper is organized as follows. In Section 2 the image classification pipeline is explained. Then, Section 3 describes the performance of using different radius parameters into the compass operator and makes an overview of the resulting classification with the best radius parameter manually chosen for each image. Section 4 describes the radius estimation method and its effectiveness in the datasets described in Section 2. Results on six publicly available datasets and two subsets from ImageNet are discussed in Section 5. Finally the conclusions and future perspectives are presented in Section 6.

2. The image classification process

In the next subsections, the image classification pipeline followed in this paper is briefly described.

2.1 Feature extraction

An image dataset can be represented as a set of pairs of images and their corresponding labels, as follows:

$$\boldsymbol{S} = \{(\boldsymbol{I_n}, \boldsymbol{y_n})\} \quad , \tag{1}$$

where $\boldsymbol{I_n}$ represents the $\boldsymbol{n}$-**th** image of the dataset, $\boldsymbol{y_n}$ stands for its label, $\boldsymbol{n} = \boldsymbol{1} \ldots \boldsymbol{m}$ and $\boldsymbol{m}$ is the dataset size.

In the first step we extract the dense SIFT descriptors following the same procedure as Xie et al. [17]. In this way, an image can be represented by the set $\boldsymbol{d_I}$:

$$\boldsymbol{d_I} = \{(\boldsymbol{d_1}, \boldsymbol{k_1}), (\boldsymbol{d_2}, \boldsymbol{k_2}), \ldots (\boldsymbol{d_s}, \boldsymbol{k_s})\} \quad , \tag{2}$$

where d_i is the descriptor extracted at the coordinates of the keypoint k_i and $\boldsymbol{s}$ the number of descriptors computed from the image.

In this work, dense SIFT features are extracted from each original image and Edge-SIFT descriptors are extracted from the edge image, i.e. the image containing object's contours as a result of applying the compass operator to each original image in the dataset. We use the values of the parameters

proposed by Xie [17]. These values are sampling density (step) 7 and scale of the extracted descriptors (size of the window) 7, for the original images and sampling 7 and scale 12 for the edge images. To compute the dense SIFT descriptors, the VLFeat library [23][24] has been used.

2.2 Visual dictionary and Bag of Words

Once all features are extracted for each image, the dataset **S** can be represented as a set of descriptors $\boldsymbol{D}$:

$$\boldsymbol{D} = \{\boldsymbol{d_{I1}}, \boldsymbol{d_{I2}}, \dots, \boldsymbol{d_{Im}}\} \quad , \tag{3}$$

where d_{In} is the set of descriptors of an image $\boldsymbol{I_n}$. The size of D is 128 x $\boldsymbol{s}$ x $\boldsymbol{m}$, being 128 the standard size for a SIFT descriptor, $\boldsymbol{s}$ the number of descriptors computed for each image and $\boldsymbol{m}$ is again the size of the dataset.

Subsequently, a visual dictionary is built by clustering D. Each cluster is called a visual word, or a codeword, which is a vector representative of similar patches within the image. The most usual clustering method to create the dictionary is K-means [25] and, thus, the visual words will be the centroids of the resulting clusters. Since we are going to use large dictionaries, the clustering process has been carried out by means of the approximate nearest neighbour algorithm proposed by Lloyd [26] because it is computationally more efficient.

For the dataset represented in (3), the following visual dictionary is built:

$$\boldsymbol{V} = [\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_c}] \quad , \tag{4}$$

where $\boldsymbol{V}$ is the Visual Dictionary of size 128 x $\boldsymbol{c}$, where $\boldsymbol{c}$ is the number of visual words. The number of words, 2048 in our experiments, is the number of centroids, $\boldsymbol{c}$, that are selected with the K-means method [25] implemented in the VLFeat library [23]. No more than 2 million of descriptors were used, as in Xie did [17].

Bag of Words feature vectors [27] are built upon this visual dictionary. To do that, each image is represented by a histogram that counts the number of times that each visual word is present in the dictionary. In our case, each image will be represented by the frequency of 2048 visual words.

$$\boldsymbol{B} = [\boldsymbol{H_1}, \boldsymbol{H_2}, \dots, \boldsymbol{H_m}] \quad , \tag{5}$$

where $\boldsymbol{B}$ is the Bag of Words matrix of size $\boldsymbol{m}$ x $\boldsymbol{c}$ and $\boldsymbol{H_n}$ is the histogram for the image $\boldsymbol{I_n}$.

The recognition of the visual words and the encoding process has been carried out using Euclidean distance.

2.3 Classification

The BoW feature vectors were used to create a model with a Support Vector Machine (SVM) [28][29] with a linear kernel and using *LibLinear* library [30]. Five different training and test sets have been randomly chosen for accuracy repeatability evaluation. Results are measured as the average of these five classification values. To understand better how the sets were chosen, some examples of these sets of images can be seen in **Table 1** and **Table 2**. Information about the size of each training and test set used is presented in Section 2.4.

	Images_class1									Images_class2									Images_class3			
Train_set_1	16	17	18	19	…	32	33	34	35	56	57	58	59	…	77	78	79	80	96	97	98	…
Train_set_2	3	17	20	9	…	28	11	32	22	61	41	50	52	…	53	80	48	51	91	94	110	…
…					…									…								…
Train_set_5	34	23	9	17	…	26	7	35	28	72	71	69	59	…	56	44	58	48	85	113	111	…

Table 1. *Sample of the training sets. Selected images for 3 out of 10 training sets on Soccer Dataset.*

	Images_class1									Images_class2									Images_class3			
Test_set_1	1	2	3	4	…	12	13	14	15	41	42	43	44	…	52	53	54	55	81	82	83	…
Test_set_2	25	23	2	7	…	1	8	37	5	57	43	63	74	…	42	55	78	45	86	115	84	…
…					…									…								…
Test_set_5	25	15	4	29	…	22	3	39	40	51	45	54	73	…	67	70	65	47	89	97	120	…

Table 2. *Sample of the test sets. Selected images for 3 out of the 10 tests used on Soccer Dataset.*

2.4 Datasets

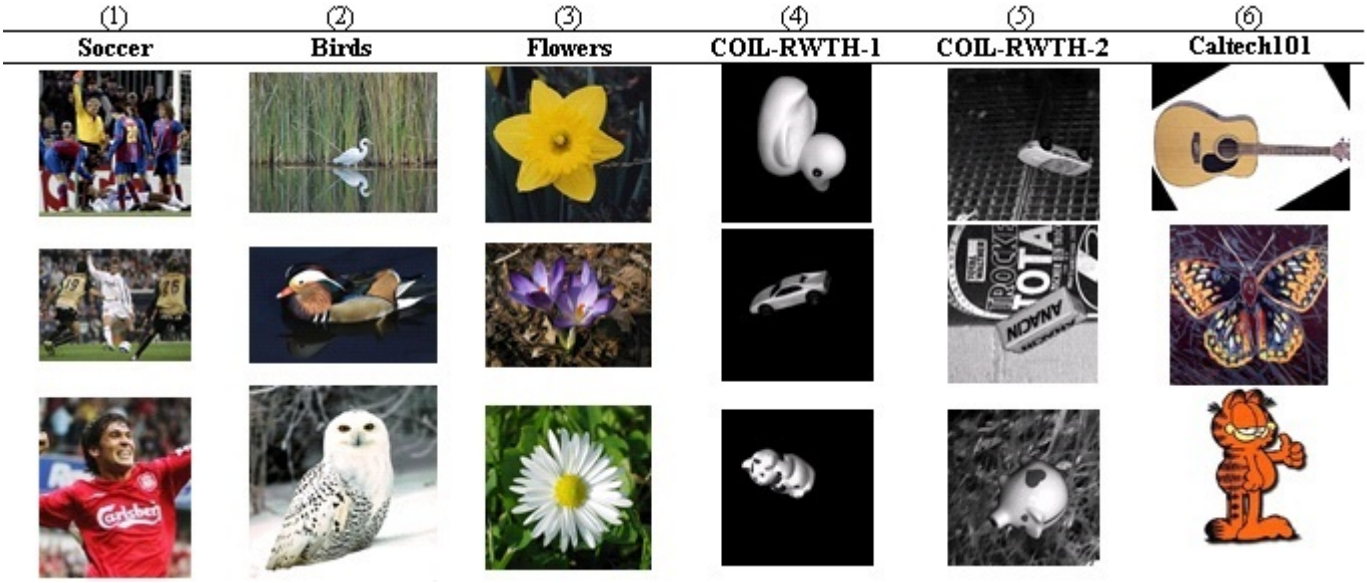


Fig. 1 presents samples from the six datasets used to validate our method. Soccer contains 280 images that show 7 football teams (40 images per team) [4]. Birds dataset is formed by 600 images of 6 different species of birds (100 images per specie) [33]. 1360 images grouped in 17 classes (80 images per class) make up Flowers dataset [34]. COIL-RWTH-1 contains 5760 images coming from 20 different objects placed on a homogeneous black background [35]. This dataset is based on COIL-20 [36]. COIL-RWTH-2 consists of 5760 images of 20 objects with heterogeneous real-world background [35]. The 20 objects are the same as the ones on COIL-RTWH-1. Unlike the three first datasets, in both COIL-RWTH-1 and COIL-RWTH-2 the images are not uniformly distributed on each class. Finally, Caltech101 [37] is composed by 9136 images divided in 101 categories with 40 to 800 images per class. Most of the classes have around 50 images.

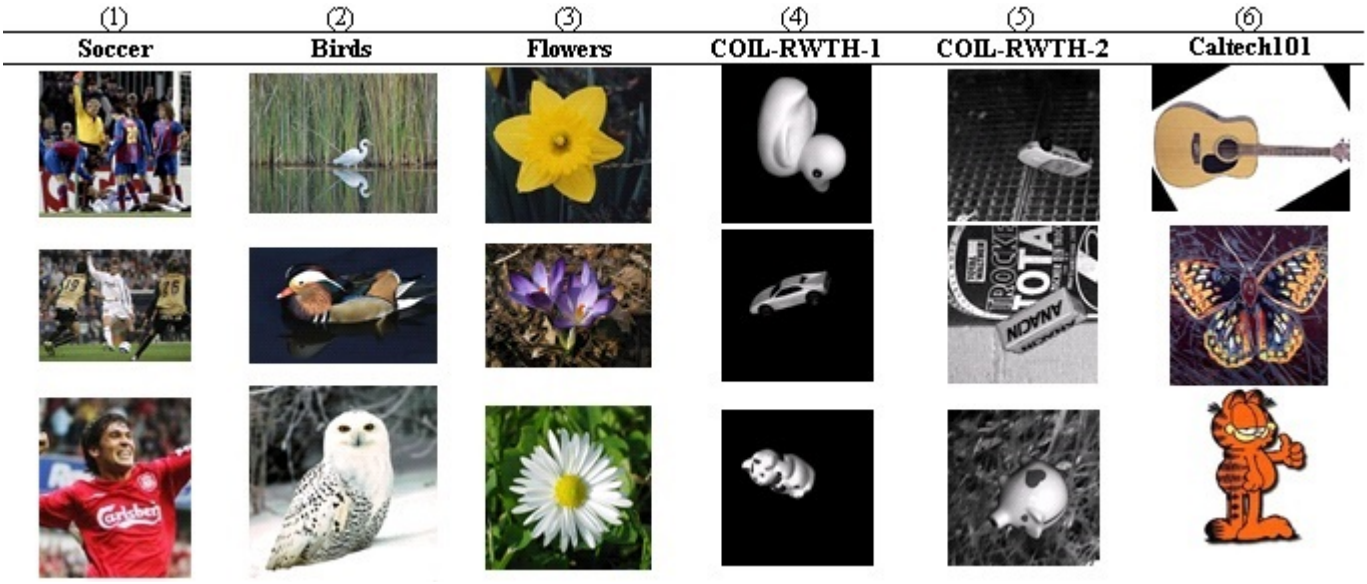


Fig. 1. *Col (1):* ***Soccer Dataset*** *samples. (up) Barcelona players with the referee, (middle) Real Madrid player, white, with two rivals of different team and (down) Liverpool player standalone. Col (2):* ***Birds Dataset*** *samples: (up) Egret, (middle) Wood duck and (down) Snowy owl. Col (3):* ***Flowers Dataset*** *samples: (up) Daffodil, (middle) Crocus and (down) Daisy. Col (4):* ***COIL-RWTH_1 Dataset*** *samples: (up) duck toy, (middle) car toy and (down) cat toy. Col (5):* ***COIL-RWTH_2 Dataset*** *samples: (up) car toy, (middle) medicine box and (down) piggy bank. Col (6):* ***Caltech101 Dataset*** *samples: (up) Spanish guitar, (middle) butterfly and (down) Garfield.*

In addition to the former six datasets we also validated our proposal using two subsets of ImageNet [38], the image database based on WordNet hierarchy nouns [39]. In ImageNet, each node of the mentioned hierarchy is depicted by a non-fixed number of images. WordNet [39] is an English lexical database where verbs, adverbs, nouns and adjectives are grouped into "synsets". Each one is a set of

cognitive synonyms with a distinct concept and it is related with other synsets by means of lexical and conceptual-semantic relations.

At the moment of the elaboration of this paper, the ImageNet database counts with 14,197,122 images with 21,841 synsets indexed, and it tries to provide an average of 1,000 images to illustrate each synset.

Based on the previous description and considering the already used datasets, we have selected two subsets of ImageNet that fulfil the following restrictions:

- Synsets are selected for fine-grained classification, i.e. classification among categories that are both visually and semantically very similar, like Birds [33] and Flowers [34], since it is one of the errors that a human is more susceptible to make than a computer [38].
- The selected synsets should contain images with only one object per image.
- At least 6 different objects, or categories, should be selected, as it is the minimum different number of classes in Birds dataset.

Fig. 2 presents samples from the two selected subsets of ImageNet. "ImageNet_Birds" contains 1400 images that show 7 species of birds (200 images per class). "ImageNet_Dogs" is composed by 1600 images of 8 different dogs species (200 images per class).

Fig. 2. *Row (1): ImageNet Sub-dataset "ImageNet_Birds". From left to right, the following synsets have been used: Chaja, Eagle, Frigate Bird, Peacock, Pelican, Sage Grouse and Whydah. Row (2): ImageNet Sub-dataset "ImageNet_Dogs". From left to right: Irish Setter, Irish Washer, Miniature Schnauzer, Sealyham Terrier,Vizsla, Welsh Springer, Afghan Hound and Yorkshire Terrier.*

	Images	Classes	Images per class	Training set Images	Training set Images per class	Test set Images	Test set Images per class
Soccer	280	7	40	62.5% - 175	25	37.5% - 105	15
Birds	600	6	100	75% - 450	75	25% - 150	25
Flowers	1360	17	80	75% - 1020	60	25% - 340	20
COIL-RTWH-1	5760	20	N/A	75% - 4320	N/A	25% - 1440	N/A
COIL-RTWH-2	5760	20	N/A	75% - 4320	N/A	25% - 1440	N/A
ImageNet_Birds	1400	7	200	75% - 1050	150	25% - 350	50
ImageNet_Dogs	1600	8	200	75% - 1200	150	25% - 400	50

Caltech101	9146	101	40-800	75% - 6860	N/A	25% - 2286	N/A

Table 3. *Summary of each dataset. N/A means that there is no uniform number of images per class. In Soccer we have used 62.5% instead 75% as the Soccer references [4][6] indicates, 25 images for training and 15 for test on each class = 62.5%. To avoid confusion with the Birds dataset [33] we will refer to the ImageNet subset of Birds as "ImageNet_Birds" from now on.*

3. Impact of the radius selection in the compass operator for image classification

3.1 Obtaining edge images using the compass operator

The compass operator [18] compares colour distributions and detects edges in RGB images. For each pixel, it compares the colour distributions on either side of an oriented circle for a number of orientations using the earth-movers distance, a robust histogram-matching method. The resulting edge image is computed using equation (6).

$$\boldsymbol{I_E = c_o\left(I_R, r_j\right)} \quad , \tag{6}$$

where $\boldsymbol{I_E}$ is the edge image obtained after applying the compass operator $\boldsymbol{c_o}$ with a radius $\boldsymbol{r_j}$ to the original image $\boldsymbol{I_R}$.

In [18] $\boldsymbol{r = 3\sigma}$ is defined, where $\boldsymbol{\sigma}$ is the standard deviation of the Gaussian used to weight the pixels in the compass operator, and $\boldsymbol{r}$ is the parameter radius.

3.2 Radius evaluation.

We evaluate how the number and size of the edges extracted using the compass operator affect the information contained on Edge-SIFT descriptors. Our objective is to assess the impact of the chosen radius in the image classification process. To do that, 14 different values for the radius parameter $\boldsymbol{r}$ for the compass operator have been evaluated: 0.75, 1.50, 2.25, 3.00, 4.00, 4.50, 5.25, 6.00, 6.75, 7.50, 8.25, 9.00, 15.00 and 30.00. These values have been selected following a homogeneous sampling process. As it is shown in **Fig. 3**, for the first radius chosen, $r = 0.75$, object's edges are barely visible and with the last value, $r = 30$, edges are blurred and ill-defined. The value used by Xie et al. [17] when they proposed the use of Edge-SIFT for image classification is $r = 4$, the same as in the literature.

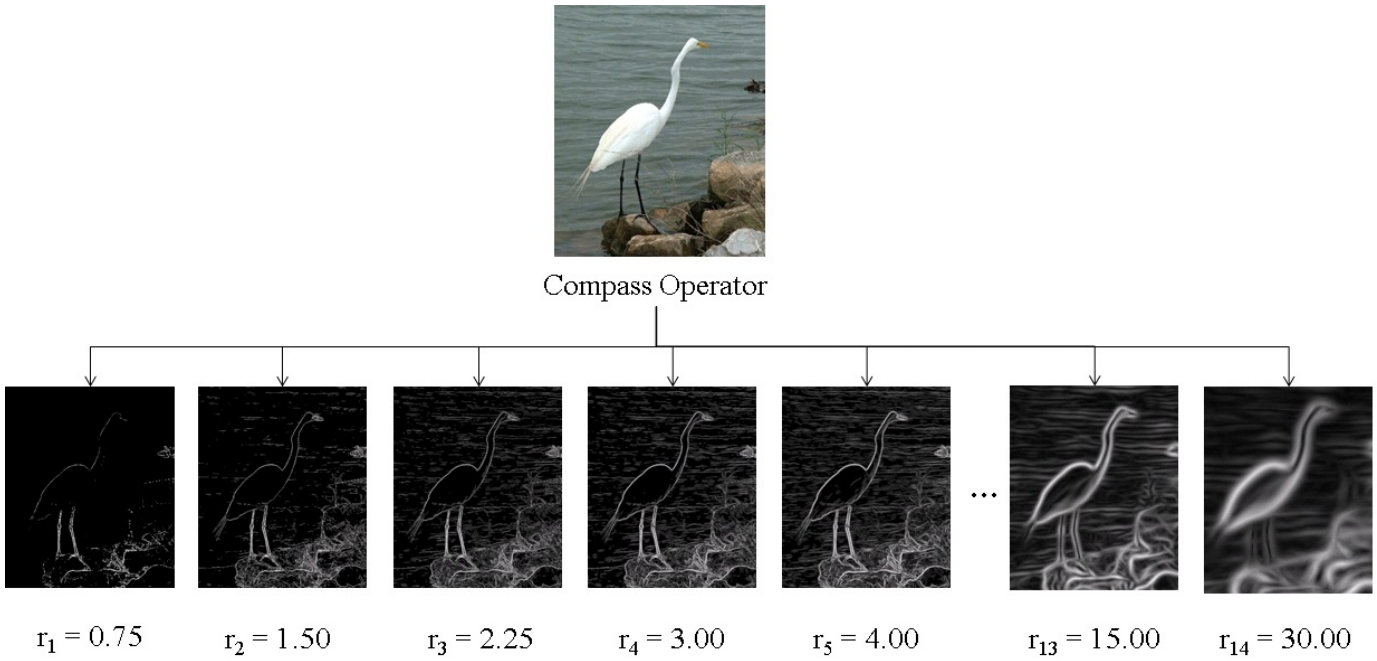


Fig. 3. *In the first row an image of an egret from Birds dataset. In the second row, edge images obtained after applying compass operator with different radius values to image on top.*

Fig. 4 shows the different BoWs that were computed using visual words from three different situations. In the first row (i) the dense SIFT descriptors are computed only from the original images. In the dotted rectangle (ii), Edge-SIFT descriptors are obtained from the edge images with the compass operator using different values of radius r_j. And in the last three rows of the diagram (iii), the BoW classification was carried out after concatenating dense SIFT (i) and Edge-SIFT descriptors (ii) as an early fusion stage to provide a more discriminative visual vocabulary. In this way, it is possible to compare the influence of the different values of ***r***, and how they affect the classification. There are other cutting-edge fusion techniques [40][41], but we use concatenation to fairly compare the approach of Xie et al. [17].

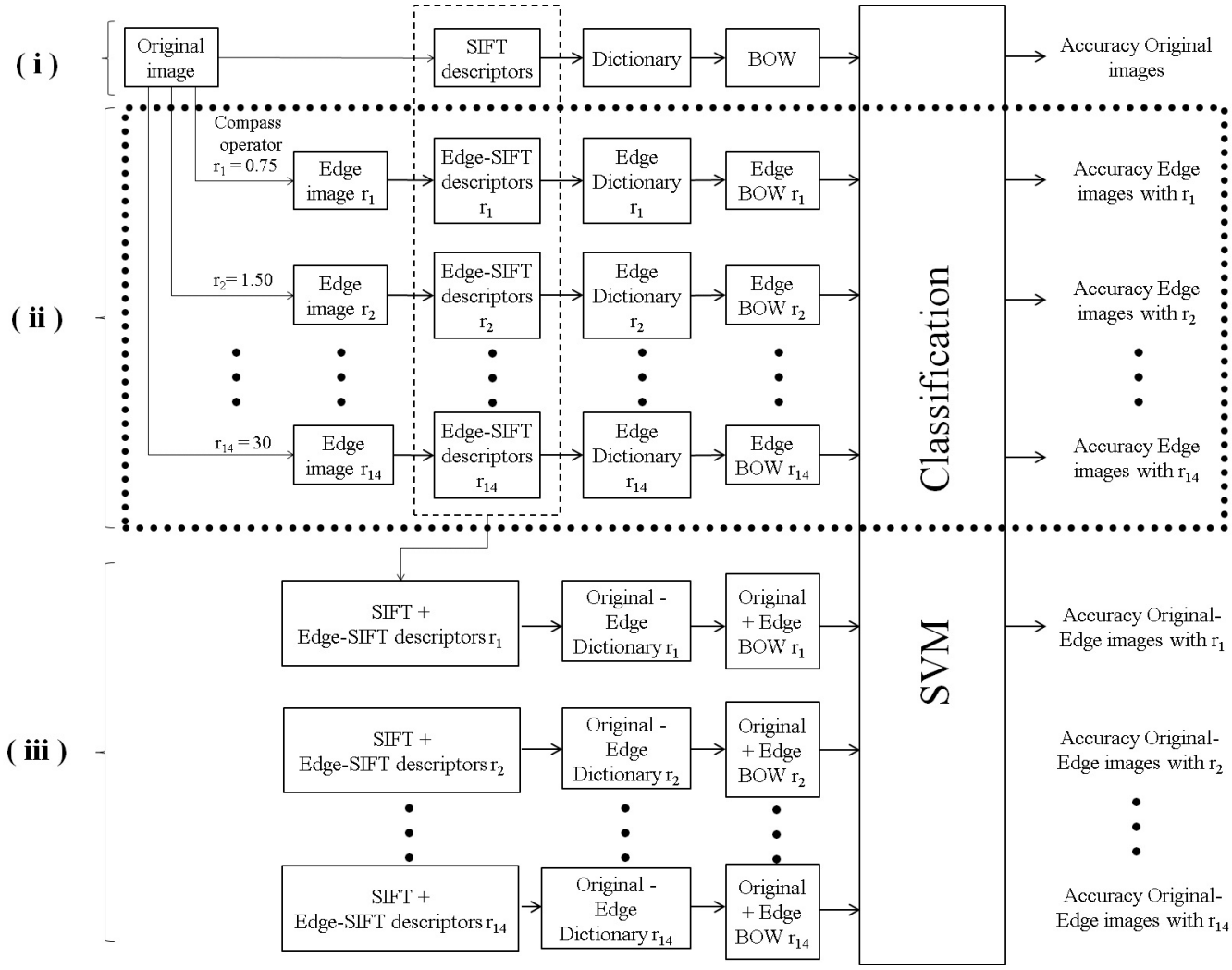


Fig. 4. *Diagram of the image classification process when using (i) dense SIFT descriptors, (ii) Edge-SIFT descriptors with different radius and (iii) SIFT+Edge-SIFT descriptor with the previous radius used.*

Edge-SIFT descriptors, (ii) in

Fig. 4, are used to evaluate the impact of the radius in classification performance. It represents how the standard BoW process is carried out for all the edge images by using the 14 different radii obtaining a classification result for each of them.

First for each of the 14 radii analysed within the compass operator, the edge images (6) of the dataset are calculated and then, dense SIFT descriptors are extracted from them. Hence, an edge image $\boldsymbol{E_n}$ can be expressed as:

$$\boldsymbol{d_{En}} = \{(\boldsymbol{d_1}, \boldsymbol{k_1})_{\boldsymbol{E}}, (\boldsymbol{d_2}, \boldsymbol{k_2})_{\boldsymbol{E}}, \ldots (\boldsymbol{d_k}, \boldsymbol{k_k})_{\boldsymbol{E}}\} \quad , \tag{7}$$

where $\boldsymbol{d_i}$ is the ***i*-th** descriptor extracted from an edge image $\boldsymbol{I_E}$ at the coordinates of the keypoint $\boldsymbol{k_i}$ and $\boldsymbol{k}$ stands for the number of descriptors extracted.

Once the descriptors of all the images are extracted for a given radius, the dataset can be expressed as a set of descriptors $\boldsymbol{D_{rj}}$:

$$\boldsymbol{D_{rj}} = \{\boldsymbol{d_{E1}}, \boldsymbol{d_{E2}}, \dots, \boldsymbol{d_{Em}}\} \quad , \tag{8}$$

where $\boldsymbol{d_{Em}}$ are the descriptors of the edge image $\boldsymbol{I_{Em}}$ computed with the radius $\boldsymbol{r_j}$.

Let $\boldsymbol{c}$ be the number of visual words (4) that represent the dataset $\boldsymbol{D_{rj}}$. The visual dictionary (4) for the dataset created using the descriptors extracted from the edge images with the radius $\boldsymbol{r_j}$ is represented as:

$$\boldsymbol{V_{rj}} = [\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_{sz}}] \quad , \tag{9}$$

where $\boldsymbol{w_i}$ represents the *i*-th visual word for the $\boldsymbol{r_j}$ radius. The size of this dictionary is 128 x $\boldsymbol{c}$.

Finally, the BoW, which will be used in the classification, is computed:

$$\boldsymbol{B_{rj}} = [\boldsymbol{H_{E1}}, \boldsymbol{H_{E2}}, \dots, \boldsymbol{H_{Em}}] \quad , \tag{10}$$

where $\boldsymbol{B_{rj}}$ is the Bag of Words matrix coded for the $\boldsymbol{r_j}$ value whose size is $\boldsymbol{m}$ x $\boldsymbol{c}$ and $\boldsymbol{H_{Em}}$ is the histogram of visual words that represents each image $\boldsymbol{I_{Em}}$.

The next table shows accuracy obtained using dense-SIFT descriptors, calculated as explained in the diagram of

Fig. 4(i).

	Soccer	**Birds**	**Flowers**	**COIL-RWTH-1**	**COIL-RWTH-2**	**ImageNet Birds**	**ImageNet Dogs**	**Caltech101**
Accuracy	48.00%	67.47%	62.24%	96.48%	50.60%	68.91%	43.70%	58,54%

Table 4. *Accuracy obtained in the seven datasets using only dense SIFT descriptors.*

Fig. 5 shows the results of the classification using the BoW built upon the Edge-SIFT and upon the fusion of dense and edge-SIFT descriptors, which allow us to visualise the radius influence for each dataset. It can also be observed how the fusion of dense SIFT plus Edge-SIFT improves the accuracy, as Xie indicated on his work, but here we want to remark that we are evaluating different sigma values.

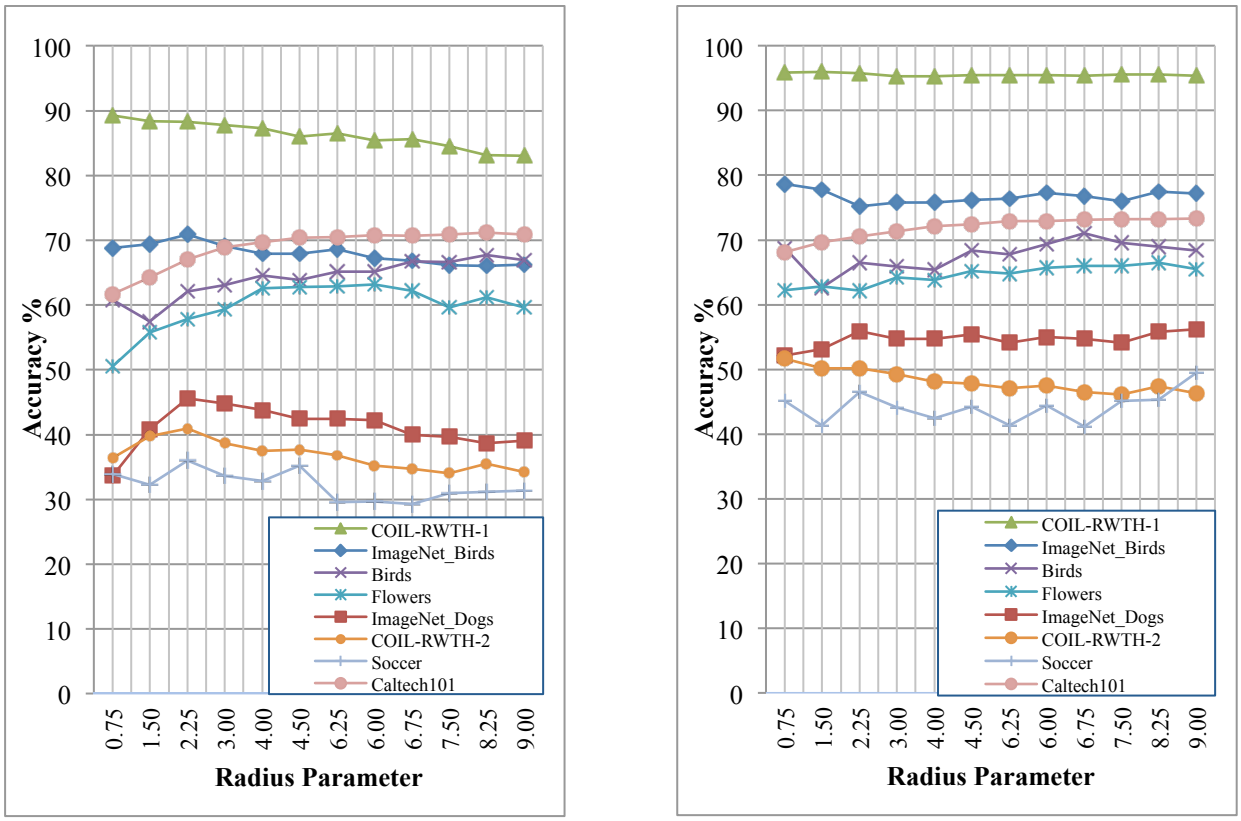


Fig. 5. *(i): Image Classification results using (left) Edge-SIFT and (right) dense SIFT+Edge-SIFT descriptors for different radius values.*

Fig. 5 (i) shows how much the classification accuracy depends on the radius value used in the compass operator. In **Fig. 5** (ii), this fact is validated also when using a fusion of dense SIFT and Edge-SIFT. Although Xie et al. [17] set $\boldsymbol{r = 4}$ in their experiments, it is clear that this value does not guarantee the best accuracy due to the different information that the edges present for different radii. It is especially remarkable how the accuracy in the COIL-RWTH-1 dataset drops as the radius increases when describing only with Edge-SIFT but it drops much less when fusion is used.

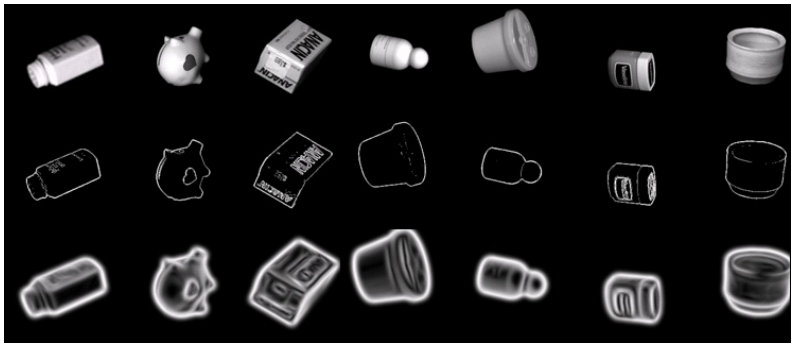


Fig. 6. *Some samples from COIL-RWTH-1 (1st row). Since there is no background all the features extracted belongs to the object. Edge Images after applying r=0.75 (2nd row) and r=30 (3rd row) show the influence in the details depending on radius.*

Fig. 6 depicts an example of how the higher the value of the radius of the compass operator, the more blurred the edges will be. As a result, more noise will be added to the edge descriptors extracted and, thus, the classification accuracy will be lower. On the other hand, when values of ***r*** are small, the information contained in the edges is clearer and, therefore, the information extracted by the descriptors is more suitable for the classification process. The accuracy for the COIL-RWTH-1 dataset in **Fig. 5** clearly decreases from $\boldsymbol{r = 0.75}$ to $\boldsymbol{r = 9.00}$.

This tendency is similar with Soccer and COIL-RWTH-2 datasets due the same reason. In these cases, however, the images have background, so the accuracy is lower and the evolution is not as clear as in COIL-RWTH-1 dataset. In contrast, regarding Flowers and Birds datasets, the worst results are achieved when ***r*** is lower than 3.00.

Different situations can lead to a poor edge definition. The background in the image can generate confusing information in the edge image like it is shown in the first column of **Fig. 7**. In some cases, the edge images do not contain enough information to describe the object as depicted in the second column of **Fig. 7**. The values that are higher than 15.00 will cause blurred objects, as it is shown in the 3rd to 6th columns of **Fig. 7**.

Fig. 7. *Birds and Flowers Original (1st row) and edge images (2nd row). 1st column shows for a Birds image the Edge Image after applying the compass operator with r=0.75. It can be observed that there are more edges from the background than from the bird. 2nd column shows for a Flower dataset image the resulting image after apply r=0.75 too. There are not enough edges to describe the object. Columns 3 to 6 show samples from both datasets with r=15 applied. The edge objects are blurred.*

Note that in **Fig. 5** radii 15 and 30 are not analysed. They were discarded because in the first 5 datasets studied it was possible to observe that these radii do not contain discriminative information, as shown in **Fig. 3**, **Fig. 6** and **Fig. 7**, since they are very time consuming.

3.3 Ideal radius selection and theoretical classification

In the previous subsection, we have pointed out how much the classification accuracy is affected by the value chosen for the radius parameter. Using the results shown in **Fig. 5** (i), in this section we present the idea that the image classification accuracy is improved when the appropriate radius value is selected for each image of the dataset.

Table 5 depicts a small sample of the SVM classification results for the experiment 1 in Soccer dataset with the 14 different ***r*** values. The last row of **Table 5** presents the global accuracy for each radius calculated as the number of hits for the values in its column divided by the number of images.

	0.75	1.50	2.25	3.00	4.00	4.50	5.25	6.00	6.75	7.50	8.25	9.00
Image_1	1	0	1	0	0	0	0	0	1	0	0	0
Image_2	0	0	0	0	0	0	0	0	0	1	0	0
Image_3	0	0	0	0	0	0	0	0	0	0	0	0
Image_4	1	1	1	1	1	1	1	1	1	1	1	1
Image_5	0	0	0	0	0	0	0	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…	…	…
Image_105	0	1	1	0	1	0	0	0	0	0	0	1
Accuracy	33.9	32.29	36	33.62	32.86	35.24	29.62	29.71	29.24	30.95	31.14	31.33

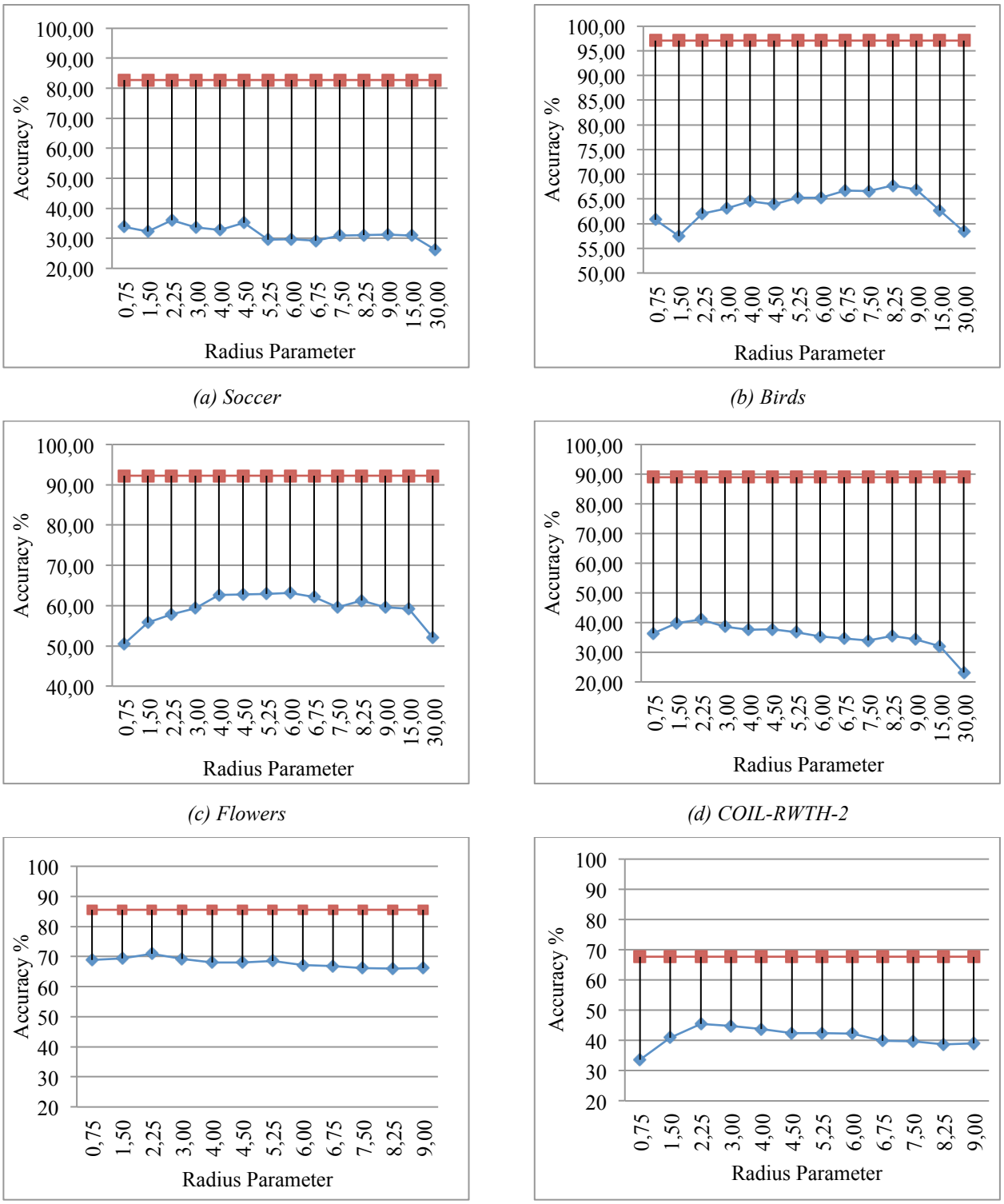
Table 5. *Soccer Test set 1 classification results per image for each radius parameter analysed. For each one of the 105 test images, values 1 or 0 stands for a right or wrong classification of the image when the specified radius was used.*

Let us now imagine that it was possible to choose, for each image, one of the radii that produced a hit, if any, e.g., for Image_1, radius 0.75, 2.25 or 6.75. The question that arises is: What would be the hit rate obtained for the whole dataset?

We have called this approach the "ideal radius selection". We have calculated this accuracy for the five different experiments defined in Section 2.3. The results obtained, presented as the average accuracy for those five runs, are shown in **Table 6.**

DS / Exp	1	2	3	4	5	**MEAN**
Soccer	76,2	82,9	83,8	85,7	84,8	**82,68%**
Flowers	87,1	95,9	94,1	92,6	91,2	**92,18%**
Birds	95,3	98	96,7	98	97,3	**97,06%**
COIL-RTWH-2	88,4	88,4	88,9	89,6	89,1	**88,90%**
ImageNet_Birds	86,86	83,71	87,43	84,00	85,43	**85,49%**
ImageNet_Dogs	66,50	68,50	68,25	69,75	65,50	**67,70%**
Caltech101	92,92	92,26	93,17	92,74	93,39	**92,90%**

Table 6. *Estimated accuracy using Edge-SIFT with "ideal radius selection". Each row indicates the best results obtained per experiment on each dataset. COIL-RWTH-1 is not used since it does not contain background information that penalizes the accuracy.*



(a) Soccer *(b) Birds*

(c) Flowers *(d) COIL-RWTH-2*

(e) ImageNet_Birds *(f) ImageNet_Dogs*

Fig. 8. *Theoretical maximum improvement in the accuracy for (a) Soccer, (b) Birds, (c) Flowers, (d) COIL-RWTH-2, (e) ImageNet_Birds (f) ImageNet_Dogs. For Edge-SIFT classification, the use of the same "r" parameter for all the dataset images (rhomboids in blue) versus the best "r" parameter selection for each image of the dataset (squares in red).*

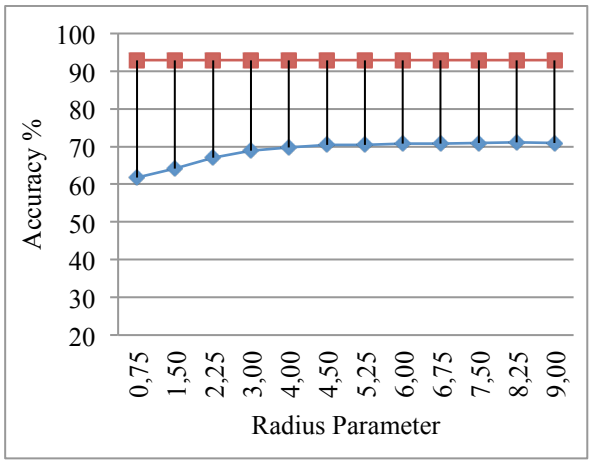


Fig. 9. *Theoretical maximum improvement in the accuracy for Caltech101 Dataset.*

Fig. 8 and **Fig. 9** show the maximum theoretical improvement that would be possible to achieve when applying, for each image, one of the radii that retain enough information to classify the mentioned image correctly.

It is important to highlight that the results shown in **Fig. 8** and **Fig. 9** might not be strictly representative of a real situation. **Table 6** was constructed based on a dictionary for BoWs created using the same radius for all the images. When $\boldsymbol{r}$ is selected for each image, a different set of descriptors is extracted and a different BoW process is carried out.

But, at the same time, it is also important to observe that, paying attention to the comparison between COIL-RWTH-1 and COIL-RWTH-2 performances, these results could be close to a plausible scenario if we were able to filter out all the descriptors that come from the background of the image. We would like to remind that all the background in COIL-RWTH-1 has been set to black and it contains exactly the same 20 objects than COIL-RWTH-2. Accuracy for COIL-RWTH-1 is 88.90% when $\boldsymbol{r = 0.75}$, where all the information from the image was stored in the main object edges. For COIL-RWTH-2, the average accuracy is 88.90%, almost the same value obtained in COIL-RWTH-1 classification from **Fig. 5** (i): 89.25% with $\boldsymbol{r = 0.75}$. Since the results in both datasets are very close, both share the same 20 objects and one of them has no background, it makes us think that this ideal selection of the radius makes certain sense.

Sadly, in the next section, we do not present a method to choose the best radius for each image. Instead, we introduce a method to estimate one of the best possible radii for the whole dataset, what it is also an improvement in relation with the current state-of-the-art situation.

4. Radius method estimation.

Based on the observations pointed out in the previous section, we propose a method to estimate, for each dataset, a radius that guarantees that the classification accuracy obtained using the fusion of Edge-SIFT and dense SIFT is better than the one achieved with the radius value proposed in the literature.

Firstly, the proposed method selects, using a *minimum dispersion criterion*, a small subset taken from a training set **(Table 1)**. Later, with this subset, the complete BoW pipeline for the 14 radii is followed and the accuracy for each of them is computed. We also present two strategies to obtain the best radius. One strategy uses one-random sub-sampling validation that is the fastest but exhibits Monte Carlo variation. The other strategy uses 3-random sub-sampling validation that is the most robust but slower than the previous one.

4.1 Minimum dispersion criterion of edges density for subset selection

Fig. 10 presents some examples of images from the Birds dataset after applying a compass operator with the smallest radius value $\boldsymbol{r = 0.75}$. We have chosen this radius because it is the one that yields the smallest number of contours in the edge image. In this way, when the background is uniform, as it is shown in 1st to 3rd columns of **Fig. 10**, most of the borders will come from the foreground objects. When several objects with high contrast are present in the background, as for example in the 4th to 6th columns of **Fig. 10**, a lot of short contours appear distributed along a great portion of the image. Our intention is to select those edge images with a high number of homogeneous (i.e., black or nearly black) areas, and a small number of areas with few borders. At the same time, we want to get rid of those images with a high number of edges distributed along most of the image. To do that, we study how the borders are scattered along an edge image and we keep the images that have the smallest dispersion using the following method.

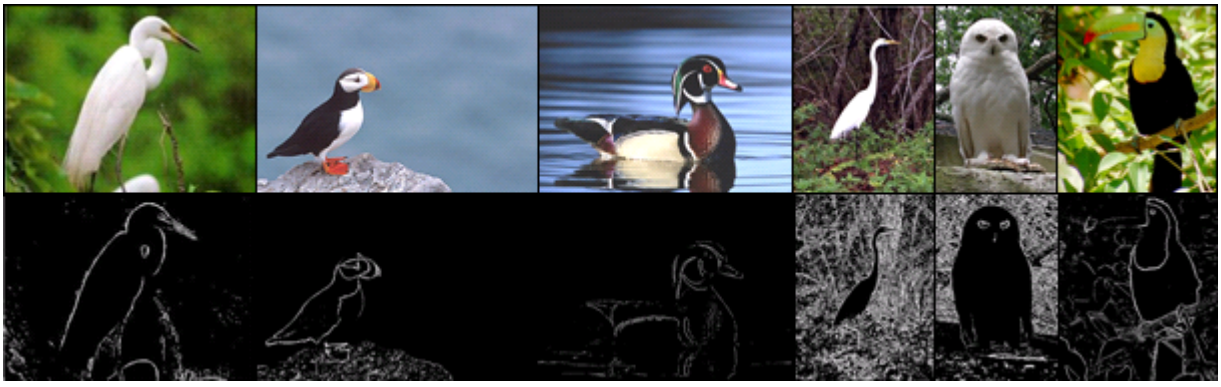


Fig. 10. *Birds dataset. Images to be accepted (Columns 1 to 3) and to be discarded (Columns 4 to 6) by the proposed method. The 2nd row depicts the edge images obtained after applying a compass operator with r_1=0.75 to the original images (1st row). The reason to accept or discard is based on the information provided by the edge images. Images selected (1st, 2nd and 3rd column) have a more homogeneous background what yields less scatter and short contours.*

First of all, given a dataset divided into training and test set, all the edge images of the training set are obtained using only the above-mentioned radius $\boldsymbol{r_1 = 0.75}$. Later, every edge image $\boldsymbol{I_E}$, is divided in a squared grid of $\boldsymbol{np\ x\ np}$ patches $\boldsymbol{p_{st}}$ where $\boldsymbol{s}$ and $\boldsymbol{t}$ denotes the row and column, respectively. This process results in the following matrix:

$$\boldsymbol{I_E} = \begin{bmatrix} \boldsymbol{p_{1,1}} & \cdots & \boldsymbol{p_{1,np}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p_{np,1}} & \cdots & \boldsymbol{p_{np,np}} \end{bmatrix} \tag{11}$$

Afterwards, a measure of the density of the edges in each path is obtained based on the number of contours that emerge after applying the compass operator. The average number of local peaks, $\overline{\boldsymbol{pk}}_{\boldsymbol{st}}$, of each patch $\boldsymbol{p_{st}}$ is calculated as the mean of the number of local maxima present in every row of that patch. Consequently, for each edge image $\boldsymbol{I_E}$, an Average Peaks Edge Image matrix, $\overline{\boldsymbol{PK}}\boldsymbol{I_E}$, of size $\boldsymbol{np \times np}$, is computed and represented in the following way:

$$\overline{\boldsymbol{PK}}\boldsymbol{I_E} = \begin{bmatrix} \overline{\boldsymbol{pk_{1,1}}} & \cdots & \overline{\boldsymbol{pk_{1,np}}} \\ \vdots & \ddots & \vdots \\ \overline{\boldsymbol{pk_{np,1}}} & \cdots & \overline{\boldsymbol{pk_{np,np}}} \end{bmatrix} \tag{12}$$

Later, the standard deviation ∇ of $\overline{\boldsymbol{PK}}\boldsymbol{I_E}$ is obtained by means of equation (13) and each image, $\boldsymbol{I_E}$, is represented by its $\boldsymbol{\nabla_{IE}}$ value:

$$\boldsymbol{\nabla_{IE}} = \left(\frac{1}{\boldsymbol{ts-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{t}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{s}} \left(\overline{\boldsymbol{pk_{ij}}} - \overline{\boldsymbol{PK}}\boldsymbol{I_E}\right)^{\boldsymbol{2}}\right)^{\frac{\boldsymbol{1}}{\boldsymbol{2}}}, \tag{13}$$

where $\boldsymbol{I_E}$ is any of the edge images, from the training set using the radius given by $\boldsymbol{r_1}$.

After that, the edge images in the dataset represented by its $\boldsymbol{\nabla_{IE}}$, are sorted in ascending order. As it was expected, it has been empirically observed that images with most of its representative information in the edges of the objects, instead on those of the background, have the lowest average peak standard deviation, $\boldsymbol{\nabla_{IE}}$ value.

$$\boldsymbol{D_{\nabla_1}} = \{(\boldsymbol{\nabla_{IE_1}}, \boldsymbol{\nabla_{IE_2}}, \ldots, \boldsymbol{\nabla_{IE_m}})\} \tag{14}$$

The subset in (14), that will be used to estimate the best radius for the whole dataset, is created selecting the 30% of the images with lower values of $\boldsymbol{\nabla_{IE}}$.

$$\boldsymbol{D_{E1\_Sel}} = \left\{\boldsymbol{I_{E_{r1}}} : \boldsymbol{\nabla_i} < \boldsymbol{P_{30}}\right\}, \tag{15}$$

where $\boldsymbol{P_{30}}$ is the 30th percentile of the $\boldsymbol{\nabla_{IE}}$ distribution and $\boldsymbol{D_{E1\_Sel}}$ stands for the selected subset. This percentage has been chosen empirically: it guarantees good results, which are presented in the next sections, and computational efficiency. The proposed algorithm is summarized in **Fig. 11**.

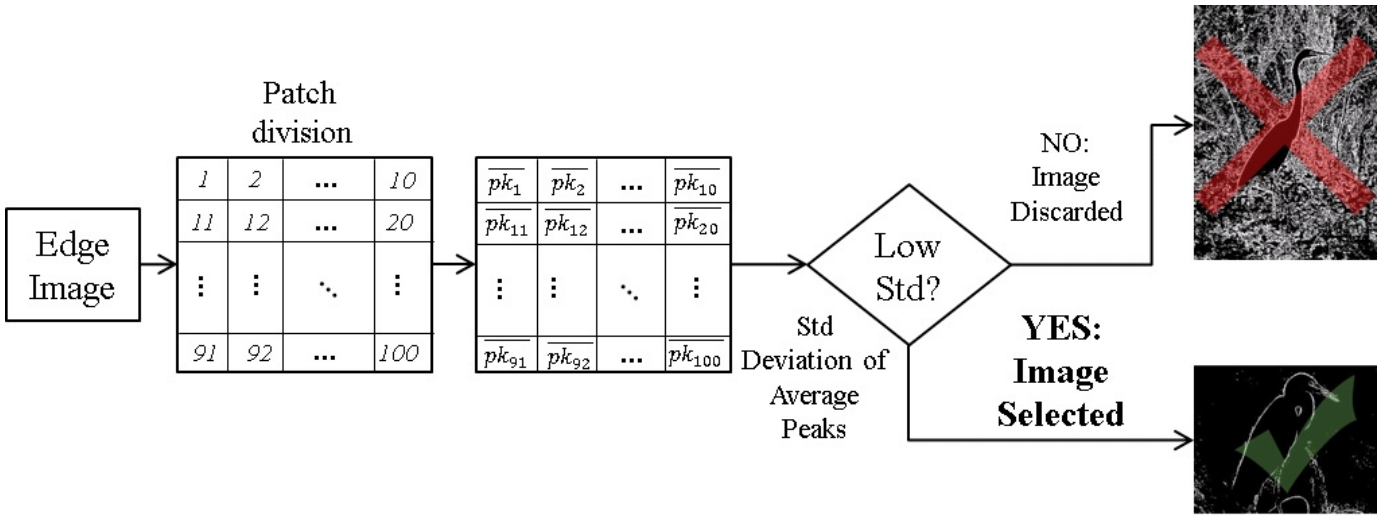


Fig. 11. *Proposed algorithm to select images for subset creation. In the example the grid is of size 10 x 10, with a total number of patches equals to 100. When the image has a low standard deviation of the average peak, it will be included in the selected subset.*

4.2 Radius estimation from the selected training subset using one-random sub-sampling validation

Once the training subset is obtained, the same process explained in

Fig. 4, but only for Edge-SIFT descriptors, is followed with the selected images. Therefore, a BoW matrix is computed for each one of the 14 radii previously mentioned. To do that, Edge-SIFT descriptors are computed and a dictionary is coded for each of the 14 radii. Thereafter, a classification using SVM is carried out. Finally, the ***r*** value that yields the best accuracy among the 14 computed is chosen and used to describe the whole dataset.

Therefore, we can say that the *predicted radius*, $\hat{\boldsymbol{r}}$, is obtained in the following way:

$$\hat{\boldsymbol{r}} = \boldsymbol{max}_{\boldsymbol{i}\in\{\mathbf{1},\ldots,\boldsymbol{n}\}}\, \boldsymbol{f}(\boldsymbol{r_i})\,, \tag{16}$$

where n= 14, and $f(r_i)$ is the *accuracy* for the selected training subset with the corresponding value of radius.

Fig. 12 illustrates the complete process of radius prediction for a single experiment, that is to say, for one random sub-sample using one of the training set configurations indicated in **Table 1**.

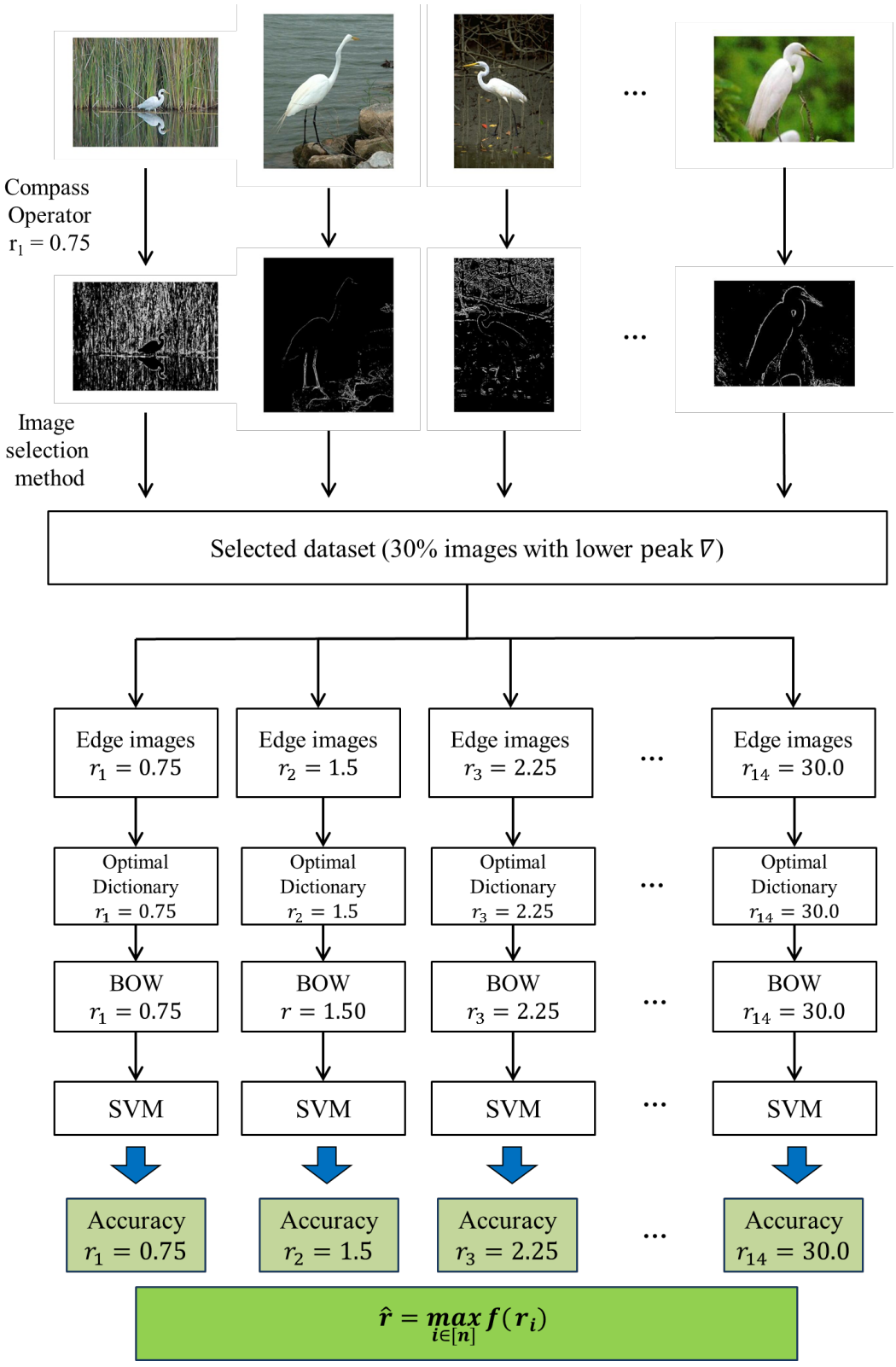


Fig. 12. *Diagram for one-test radius prediction method. If the above method is applied to three different training sets of images, the predicted radius will be the result of a voting between them.*

4.3 Radius estimation using 3-random sub-sampling validation

The estimation of the radius using one random sub-sample, as explained in previous section, is the fastest but not the most robust way of implementing the proposed method.

Number of Training set used	Soccer	Birds	Flowers	COIL-RTWH-2	ImageNet Birds	ImageNet Dogs	Caltech101
1	80,50%	77,40%	72,10%	81,80%	72,40%	73,20%	73,90%
3	41,40%	32,30%	16,40%	45,30%	17,20%	19,60%	21,70%

Table 7. *Time saved, expressed in percentage, when using the radius parameter prediction on the different datasets. First row present the savings when one-random sub-sampling validation is used and second row savings with 3-random sub-samplings with voting.*

Table 7 depicts that with a single sub-sample it is possible to estimate a good radius with time savings of around 70% or 80%.

The savings are the difference between calculating the accuracy of dense SIFT+Edge-SIFT with the radius estimated with our method, and for all the 14 indicated radii values using the complete training set as

Fig. 4. The drawback of using one-random sub-sample is that it exhibits Monte Carlo variation, and the results vary when the analysis is repeated with different random splits. Despite this variation, and based on the experiments carried out, our method always performs better than when using the default radius in Soccer and COIL-RWTH-2 datasets, using one-random sub-sampling validation. However, in the experiments carried out with Birds and Flowers datasets, our method does not guarantee that the estimated radius yields better accuracy than the default one using only one-random sub-sample.

Using a 3-random sub-sampling validation our proposal estimates a radius that outperforms the results obtained with radius 4.00 as proposed in the literature. In the next section, results for each dataset using both sub-sampling validation strategies are compared. The effect of the kernel used in the SVM is discussed too.

5. Results and discussion

To test our proposal we present the experiments carried out using the first five training sets from each dataset. Training sets from **Table 1** were used for radius prediction and test sets from **Table 2** were used to evaluate the accuracy with different radius parameters, including the predicted one. Using this setup, we noticed that only using 3-random sub-samples it is possible to always estimate a radius that yields better image classification accuracy than the one obtained when using $\boldsymbol{r = 4.00}$.

5.1 Soccer Dataset

As it is shown in

Fig. 1, most of the images in this challenging dataset include several objects and noisy background.

Table 8 shows an example about how the radius estimation method has been validated using one-random sub-sampling validation. **Table 9** shows the same results as **Table 8** but when the radius is estimated by voting with 3-random sub-sampling validation.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Training #	$r_1 = 0.75$	$r_2 = 1.50$	$r_3 = 2.25$	$r_4 = 3.00$	$r_5 = 4.00$
1	**35,14**	31,52	32,57	31,24	30,67
2	**35,52**	31,05	31,71	29,52	29,62
3	33,62	32,48	33,52	**34,29**	32,67
4	**34,57**	34,00	33,33	30,86	31,05
5	**34,19**	32,19	32,29	31,81	31,14

Max accuracy per test

Accuracy results for **SIFT + Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Test #	$r_1 = 0.75$	$r_4 = 3.00$	$r_5 = 4.00$	**Comparison**
1	**43,81**	N/C	**41,90**	43.81 > 41.90
2	**53,33**	N/C	**42,86**	53.33 > 42.86
3	N/C	**45,71**	**41,90**	45.71 > 41.90 => OK
4	**45,71**	N/C	**44,76**	45.71 > 44.76
5	**40,95**	N/C	**35,24**	40.95 > 35.84

Table 8. *Summary of the accuracy, in %, obtained with the 5 experiments carried out in Soccer Dataset using SVM with an intersection kernel. The radii have been estimated using one-random sub-sampling validation. Values have been highlighted as follows:* ***Bold*** *+ Shadowed****:*** *Accuracy obtained when the radius proposed in the literature, r = 4.00, was used for computing Edge-SIFT descriptors.* ***Bold****: Maximum accuracy obtained with Edge-SIFT descriptors on training sets with one-random sub-sampling validation. The r value that yields this result is chosen as the estimated one and used later for the whole dataset.* ***Bolded*** *+ underlined****:*** *Accuracy obtained using the previously estimated r in the whole dataset when dense SIFT and Edge-SIFT are fused.*

Despite it is bold, in **Table 8** the maximum accuracy for a single train set, and therefore its associated ***r*** value, is identified using a dotted - and blue - ellipsoid as an example. In the rest of the tables they will be only highlighted by a bold number as the caption of **Table 8** explains. For instance, in Test #3, the maximum accuracy obtained is 34.29%, with $r_4 = \mathbf{3.00}$. Using that r_4 in the whole dataset –second part of **Table 8**- the accuracy obtained was 45.71%. Comparing it with the accuracy (41.90%) yielded by r_5, the radius proposed in the literature, we confirm that there is an improvement. It is possible to check out that this is the same behaviour for the rest of the experiments. The increase in accuracy using the proposed method goes from 0.95 in Test#4 to 10.47 in Test#2 – both with $r_1 = 0.75$. In every single

experiment, the radius estimation method outperforms the baseline. It is worth highlighting that the improvement of 10.47 points obtained in Test#2, between 42.86% and 53.33%, represents an accuracy 24.4% higher.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Training #	$r_1 = 0.75$	$r_2 = 1.50$	$r_3 = 2.25$	$r_4 = 3.00$	$r_5 = 4.00$	**Radius estimated per test**
1	**35,14**	31,52	32,57	31,24	30,67	$r_1 = 0.75$
2	**35,52**	31,05	31,71	29,52	29,62	$r_1 = 0.75$
3	33,62	32,48	33,52	**34,29**	32,67	$r_3 = 2.25$
4	**34,57**	34,00	33,33	30,86	31,05	Voting => $r_1 = 0.75$
5	**34,19**	32,19	32,29	31,81	31,14	

Accuracy results for **SIFT + Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Test #	$r_1 = 0.75$	$r_5 = 4.00$	**Comparison**
1	**43,81**	**41,90**	
2	**53,33**	**42,86**	
3	53,33	**41,90**	
MEAN	50,16	**42,22**	50.16 > 42.22 => OK

Table 9. *Summary of the accuracy, in %, obtained with the 5 experiments carried out in Soccer Dataset using SVM with an intersection kernel. The radii have been estimated by voting with 3-random sub-sampling validation.*

Table 9 shows the radius estimation and the accuracy obtained in the whole dataset using a voting schema with 3-random sub-sampling validation. As it was done in the previous table, an example has been highlighted with dotted - and blue – ellipsoids. Among the 5 experiments carried out, 3 of them are chosen to predict the radius with the same process followed in the one-round case. For the **Table 9** example we selected the first three experiments, but it is possible to observe that any combination of 3 experiments would yield the same result. The maximum accuracy for tests 1, 2 and 3 points out three radius values, $r_1 = \mathbf{0.75}$, $r_1 = \mathbf{0.75}$ and $r_4 = \mathbf{3.00}$. Using a voting system, r_1 receives 2 out of 3 votes, so this is the predicted radius. In **Table 9** it can be observed that if the dense SIFT + Edge-SIFT descriptors were calculated with $r_1 = \mathbf{0.75}$ with three test-sets, an averaged accuracy of 50.16% would be obtained. This accuracy is 7.94 points higher than the average accuracy obtained if the literature radius $r_5 = \mathbf{4.00}$ was used for these three tests (42.22%). The maximum improvement is 9.20 points if we use the $r_1 = \mathbf{0.75}$

estimated by the voting of Tests#2, #3 and #5 predictions, being 40% to 49.20% the mean accuracies if $r_1 = \mathbf{0.75}$ and $r_5 = \mathbf{4.00}$ is used respectively.

As we will explain in more detail when we analyse the Flowers dataset, in case of a draw after the voting, our proposal is to choose the radius which is closer to the literature proposal.

To evaluate the radius estimation method both *linear* and *intersection* [31][32] kernels for SVM were used. Based on the experiments carried out, our proposal uses a linear Kernel in all datasets except for Soccer one. The presence of several objects of interest in the same image, i.e. more than one football player belonging to the same or different class, the presence of the referee and the noisy background make the intersection kernel more appropriate for the radius estimation method.

5.2 Birds Dataset

This is another complex dataset because the birds are sometimes surrounded by different types of environments, which may add noise to the descriptors, as it is shown in **Fig. 10**. There are other cases where the bird share some colours with the background and make the classification task difficult, as it is depicted in the 6th column of **Fig. 10**.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Training #	r_5 = 4.00	r_6 = 4.50	r_7 = 5.25	r_8 = 6.00	r_9 = 6.75	r_{10} = 7.50
1	63,60	62,47	59,40	61,40	62,47	**64,00**
2	62,13	61,33	61,27	63,67	62,67	**64,00**
3	60,13	62,20	62,00	62,93	62,13	**63,20**
4	61,07	59,20	60,80	62,20	63,40	**63,47**
5	60,20	60,00	61,93	62,60	61,93	**63,07**

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Test #	r_5 = 4.00	r_{10} = 7.50
1	**66,00**	**67,33**
2	**63,33**	**65,33**
3	**61,33**	**60,67**
4	**64,67**	**70,00**
5	**56,67**	**56,67**

Table 10. *Summary of the 5 experiments carried out in Birds Dataset.*

Based on the explanations of Soccer dataset, we perform the analysis of the 3-random sub-sampling validation through the voting system. The maximum increase in accuracy using the proposed method is 2.89 points if we select the radius $r_{10} = \mathbf{7.50}$ after the voting between Tests #1, #2 and #4 in **Table 10**. The improvement obtained will vary depending on the tests selected but the robustness of the method can be validated with different combinations of tests from **Table 10**. Those 2.89 points represents an improvement of 4.46%.

When the radius was estimated by using the one-random sub-sampling validation, it achieved an improvement of 5.33 points with $r_{10} = \mathbf{7.50}$ on Test#4. However, we observed that Test#3 predicted a radius that does not guarantee a better accuracy than when we use the literature one. Despite the rest of the sub-samples predicted a good radius parameter standalone, because Test#3 failed it is not 100% guaranteed that one-random sub-sampling validation provides a suitable radius to be used. Based on these results we suggest one-random prediction only for large time saving in the process and three-random sub-sampling validation for less time saving but a radius prediction that guarantees the accuracy improvement.

5.3 Flowers Dataset

This complex dataset, widely used in previous works [42][43][44], comprises 17 classes of flowers.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Training #	$r_5 = 4.00$	$r_6 = 4.50$	$r_7 = 5.25$	$r_8 = 6.00$	$r_9 = 6.75$	$r_{10} = 7.50$
1	59,47	60,32	**61,41**	60,91	60,00	60,29
2	59,85	**61,59**	61,26	61,09	60,06	59,50
3	59,65	**61,35**	61,00	61,21	54,12	54,12
4	60,12	**62,18**	61,18	61,15	60,76	60,09
5	59,82	60,74	60,68	**61,09**	60,71	60,38

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Test #	$r_5 = 4.00$	$r_6 = 4.50$	$r_7 = 5.25$	$r_8 = 6.00$
1	**61,18**	60,29	**60,00**	62,06
2	**63,24**	**64,71**	57,94	61,76
3	**65,00**	**65,59**	64,71	65,00
4	**61,76**	**63,53**	61,76	63,82
5	**61,76**	60,59	62,35	**61,47**

Table 11. *Summary of the 5 experiments carried out in Flowers Dataset.*

First, we perform the analysis of the 3-random sub-sampling validation through the voting system. The maximum increase in accuracy using the proposed method is 1.28 points if we select the radius $r_6 = \mathbf{4.50}$ after the voting between Tests #2, #3 and #4 in **Table 11**.

We will get different improvements according to the tests selected but the robustness of the method can be validated with different combinations of tests from **Table 11**. In this dataset, depending on which three tests are chosen for prediction it might be possible to have a draw in the voting system. If this is the case, the predicted radius will be the value closer to 4 (the radius used in the literature) among these three. In **Table 11**, the voting between Tests #1, #4 and #5 is a draw between r_6 , r_7 and r_8, but r_6 is chosen since it is the one closer to r_5. This conclusion is made under the consideration that most of the 5 tests performed predict r_6.

The 1.28 points early discussed represents a 2.02% of improvement if we use the estimated radius r_6 instead the r_5 proposed in literature.

Furthermore, if the radius is estimate with only one random sub-sample, it will achieve an improvement of 1.77 points. But, as it was the case in the Birds dataset, it can be observed how there is a test in **Table 11**, that predicts a radius which does not guarantee a better accuracy than the one with the literature radius - Tests#1 and #5 predict $r_7 = 5.25$ and $r_8 = 6.00$ respectively. Despite the rest of the tests predict a good radius parameter it is not guaranteed that one random sub-sampling will always provide a suitable value. Like Birds, this approach can be used only for large time saving, but to guarantee an accurate radius prediction we will use the 3-random sub-sampling validation.

5.4 COILRWTH-1 Dataset

The application of radius estimation method is not needed in this dataset since only contains objects with no background.

5.5 COIL-RWTH2 Dataset

This dataset comprises 20 classes of objects placed over non-homogeneous backgrounds. Some examples are shown in **Fig. 1**.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Training #	$r_1 = 0.75$	$r_2 = 1.50$	$r_3 = 2.25$	$r_4 = 3.00$
1	38,58	**39,76**	38,42	36,70
2	38,74	38,64	**39,49**	37,69
3	38,33	38,38	**39,07**	37,86
4	38,53	**39,59**	38,35	37,36
5	38,64	38,75	**39,11**	37,36

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Linear Kernel

Test #	$r_2 = 1.50$	$r_3 = 2.25$	$r_5 = 4.00$
1	**44,01**	42,91	**42,56**
2	44,50	**45,61**	**43,67**
3	45,61	**44,43**	**43,88**
4	**45,74**	45,19	**42,84**
5	44,22	**45,54**	**45,19**

Table 12. *Summary of the 5 experiments carried out in COIL-RWTH-2 Dataset.*

As in the case of Soccer dataset, COIL-RWTH-2 dataset can use both one and three-random sub-sampling validation methods.

The increase in accuracy using one-random sub-sample goes from 0.35 in Test#5 to 2.90 in Test#4– both with $r_3 = 2.25$ and $r_2 = 1.50$ respectively. In every single experiment, the radius estimation method outperforms the baseline.

The 2.90 points of improvement represents a 6.77% of improvement if we use the estimated $r_2 = 1.50$ instead the literature one, $r_5 = 4.00$.

The maximum increase in accuracy using 3-random sub-sampling is 2.03 points if we select the radius $r_2 = 1.50$ after the voting between Tests #1, #3 and #4 in **Table 12**.

The improvement obtained varies depending on the tests selected but the robustness of both one or three-random sub-sampling validations checked with different combinations of tests from **Table 12**.

5.6 ImageNet_Birds Dataset

The next table shows the predictions obtained in ImageNet_Birds ImageNet subset.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Training #	$r_2 = 1.50$	$r_3 = 2.25$	$r_5 = 4.00$	$r_6 = 4.50$
1	**71,14**	69,71	68,86	66,00
2	69,71	**71,71**	67,43	67,43
3	69,71	69,43	69,43	**71,43**
4	**71,14**	69,43	67,14	68,00
5	**69,71**	66,57	68,57	66,86

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Test #	$r_2 = 1.50$	$r_3 = 2.25$	$r_5 = 4.00$	$r_6 = 4.50$
1	**78,86**	75,43	**77,71**	76,57
2	77,14	**75,43**	**75,14**	75,14
3	77,14	74,86	**75,71**	**75,43**
4	**80,29**	74,86	**76,29**	76,57
5	**75,43**	75,43	**74,29**	77,43

Table 13. *Summary of the 5 experiments carried out in ImageNet_Birds Dataset.*

As it was explained in Section 5.1, we perform the analysis of the 3-random sub-sampling validation through the voting system. 2.10 points is the maximum accuracy obtained when radius $r_2 = 1.50$ after the voting between Tests #1, #4 and #5 in **Table 13**. The robustness of the method can be validated with different combinations of tests from **Table 13**. Those 2.10 points represents an improvement of 2.75%.

If we estimate the radius with the one-random sub-sampling validation, we will obtain an improvement of 4.00 points with $r_2 = 1.50$ on Test#4. Nevertheless, Test#3 predicted a radius that does not guarantee a better accuracy than when we use the literature one. Although the rest of the sub-samples predicted a good radius parameter standalone, as the Test#3 prediction it is not 100% precise, just one-random sub-sampling validation grants a suitable radius to be used. Therefore, we suggest one-random prediction only for large time saving in the process and three-random sub-sampling validation to get a radius prediction that guarantees the accuracy improvement.

5.7 ImageNet_Dogs Dataset

The next table shows the predictions obtained in ImageNet_Dogs ImageNet subset.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Training #	$r_3 = 2.25$	$r_4 = 3.00$	$r_5 = 4.00$	$r_6 = 4.50$
1	**47,00**	42,50	41,75	39,75
2	44,50	43,75	**45,25**	41,25
3	**45,75**	44,25	45,75	45,25
4	46,50	45,50	46,25	**47,00**
5	43,50	**45,25**	43,00	42,50

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Test #	$r_3 = 2.25$	$r_4 = 3.00$	$r_5 = 4.00$	$r_6 = 4.50$
1	**54,00**	54,75	**53,25**	54,50
2	56,25	56,00	**59,00**	56,00
3	**59,75**	55,75	**55,00**	55,00
4	55,75	55,25	**55,00**	**57,75**
5	54,00	**52,00**	**51,50**	53,75

Table 14. *Summary of the 5 experiments carried out in ImageNet_Dogs Dataset.*

We carry out our 3-random sub-sampling validation method in this ImageNet subset. A 2.08 points higher accuracy is obtained if we select the radius $r_3 = 2.25$ after the voting between Tests #1, #3 and #4 in **Table 14**. Those 2.08 points represents an improvement of 3.82%.

The accuracy is increased up to 4.75 points when the radius is estimated by using the one-random sub-sampling validation, with $r_3 = 2.25$ on Test#3. It is remarkable that after 7 datasets analysed and 5 predictions per dataset this is the unique time that the predicted radio match with the one indicated by Xie et al. [17], $r_5 = 4.00$ predicted by Test#2.

On one hand, in this dataset we suggest using one-random prediction for a large time saving and for achieving accuracy equal or greater than the one obtained by literature.

On the other hand, the three-random sub-sampling validation will get a radius prediction that guarantees the accuracy improvement.

5.8 Caltech101 Dataset

Once we have demonstrated the effectiveness of the CREIC method with small datasets, i.e, datasets with few classes, we tested it on Caltech101 obtaining the following results.

Radius estimation. Accuracy results for **Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Training #	$r_8 = 6.00$	$r_9 = 6.75$	$r_{10} = 7.50$	r_{11}=8.25	r_{12}=9.00
1	69,85	70,33	**70,37**	70,37	69,72
2	**70,59**	70,55	70,03	70,46	70,20
3	71,11	**71,42**	71,03	71,11	70,85
4	70,07	70,20	69,94	**70,37**	69,85
5	71,03	70,63	70,76	70,98	**71,37**

Accuracy results for dense **SIFT + Edge-SIFT** descriptors with the sampled subset. Intersection Kernel

Test #	$r_4 = 4.00$	$r_8 = 6.00$	$r_9 = 6.75$	$r_{10} = 7.50$	r_{11}=8.25	r_{12}=9.00
1	**71,68**	72,98	73,11	**72,94**	73,28	73,15
2	**72,20**	**73,15**	73,63	73,28	73,15	73,24
3	**72,42**	73,37	**73,07**	73,81	73,46	73,76
4	**72,11**	72,55	72,63	72,89	**73,24**	73,28
5	**72,11**	72,72	73,11	73,33	73,20	**73,28**

Table 15. *Summary of the 5 experiments carried out in Caltech101 Dataset.*

Finally, an analysis of the 3-random sub-sampling validation through the voting system is performed on Caltech101. We obtained an accuracy 1.26 points higher with $r_{10} = 7.50$ estimated after the voting between Tests #1, #4 and #5 in **Table 15**. Again, the effectiveness of the method is demonstrated through different combinations of tests from **Table 15**. The 1.26 points represent an improvement of 1.75%.

In this dataset, the use of one-random sub-sampling validation results in an improvement of 1.26 points too, with $r_{10} = 7.50$ on Test#1. This maximum improvement matches the best one obtained with the 3-random sub-sampling validation but with a 73.90% of time saving as it was described in **Table 7**.. To sum up, in Caltech101 dataset we suggest to use one-random prediction for large time saving and a radius prediction that guarantees the accuracy improvement.

6. Conclusions and future work

The Edge-SIFT descriptors are extracted from an edge image obtained after applying the compass operator with a certain radius. Their early fusion with dense SIFT was demonstrated by Xie et al. [17] to be capable of improving the image classification.

Edge images are strongly dependent on the radius of the compass operator. Working into the BoW framework, the influence of this radius parameter in the classification is shown using six different datasets.

Once this influence is confirmed, the performance obtained if the best radius could be selected for each image has been empirically assessed. In that case, the accuracy from all datasets improved from 26.19% to 83.50% in the best case – Soccer dataset.

The former demonstration emphasizes the importance of extracting suitable descriptors from edge images for dictionary construction. Based on these findings, an algorithm to estimate a suitable radius for any specific dataset is proposed. This estimation is carried out on the basis of a minimum dispersion criterion on the image edges. It is demonstrated that a low standard deviation on the local maxima patches are related to the presence of relevant edges and, thus, the chances of extracting more information from the object of interest. The estimated radius guarantees a better accuracy than taking the default one proposed in the bibliography when Edge-SIFT are fused with dense SIFT descriptors. All these findings are validated on four challenging datasets. The mentioned prediction method ensures a saving in time that can be quite significative if only one-random sub-sample is performed. The three-random sub-sampling validation will guarantee the radius prediction but the time saving would not be as good as when one sub-sample is used.

The results obtained show that, depending on the dataset, the improvement varies from 1.75% in Caltech101 to 24.4% in Soccer dataset.

The next step will be to improve the radius selection method using, for example, saliency maps. Another future research line could be the selection of the best radius per image, since we have demonstrated that it would yield even better results.

Acknowledgments

This work has been supported by the research project with reference DPI2012-36166 from the Spanish Government.

7. References

[1] D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, in :Int. J. Comput. Vision 60, 2, 2004, pp.91-110. doi:10.1023/B:VISI.0000029664.99615.94

[2] J. Xie, L. Zhang, J. You, S. Shiu, Effective Texture classification by texton encoding induced statical features, in: Pattern Recognition, 48(2), 2014, doi:10.1016/j.patcog.2014.08.014

[3] K. van de Sande, T. Gevers, C. Snoek, Evaluating Color Descriptors for Object and Scene Recognition, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, September, 2010, pp. 1582-1596, doi:10.1109/TPAMI.2009.154

[4] J. van de Weijer, C. Schmid, Coloring local feature extraction, in: ECCV, volume 3952, Springer, 2006, pp.334–348, doi:10.1007/11744047_26

[5] D. Vigo, F. Khan, J. van deWeijer, T. Gevers, The impact of color on bag-of-words based object recognition, in: ICPR, 2010, pp.1549 –1553. doi: 10.1109/ICPR.2010.383

[6] J. van de Weijer, C. Schmid, Applying Color Names to Image Description, in: Image Processing, 2007, in: ICIP 2007, in: IEEE International Conference on , vol.3, no., pp.III - 493,III - 496, Sept. 16 2007-Oct. 19 2007 doi:10.1109/ICIP.2007.4379354

[7] V. Gonzalez-Castro, J. Debayle, and J.C. Pinoli., Color adaptive neighborhood mathematical morphology and its application to pixel-level classification, in: Pattern Recognition Letters 47 (1), 2014, pp. 50-62, doi:10.1016/j.patrec.2014.01.007

[8] O. Penattia, E. Vallea, R. Torres, Comparative study of global color and texture descriptors for web image retrieval, in: Journal of Visual Communication and Image Representation, Vol. 23, Issue 2, February 2012, pp 359-380, doi:10.1016/j.jvcir.2011.11.002

[9] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, in: Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.24, no.7, Jul 2002, pp.971-987, doi:10.1109/TPAMI.2002.1017623

[10] G. Zhenhua, D. Zhang, A Completed Modeling of Local Binary Pattern Operator for Texture Classification, in: Image Processing, IEEE Transactions on , vol.19, no.6, June 2010, pp.1657-1663, doi:10.1109/TIP.2010.2044957

[11] J. Ylioinas, A. Hadid, Y. Guo, M. Pietikainen, Efficient image appearance description using dense sampling based local binary patterns, in: Computer Vision – ACCV, 2012. doi:10.1007/978-3-642-37431-9_29

[12] J. Ylioinas, X. Hong, M. Pietikainen, Constructing local binary pattern statistics by soft voting, in: Lecture Notes in Computer Science Volume 7944, 2013, pp 119-130, doi:10.1007/978-3-642-38886-6_12

[13] A. Borji, L. Itti, State-of-the-art in modeling visual attention, in: IEEE Transactions on. Pattern Analysis and Machine Intelligence, 35 (1) , 2013, pp. 185–207, doi:10.1.1.252.3616

[14] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, in: International Journal of Computer Vision, 1988, 1 (4): 321, doi:10.1007/BF00133570.

[15] A. Bosch, A. Zisserman, X. Muoz, Image Classification using Random Forests and Ferns, in: Computer Vision, 2007, in: ICCV 2007, in: IEEE 11th International Conference, Oct. 2007, pp.1,8,14-21, doi:10.1109/ICCV.2007.4409066.

[16] J. Feng, B. Ni, Q. Tian, and S. Yan, Geometric p-norm feature pooling for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 2609-2704, doi:10.1109/CVPR.2011.5995370

[17] L. Xie, Q. Tian, and B. Zhang, Spatial pooling of heterogeneous features for image classification, in: Image Processing, IEEE Transactions on, vol. 23, no. 5, 2014, pp. 1994–2008, doi:10.1109/TIP.2014.2310117

[18] M.A. Ruzon, C. Tomasi, Color edge detection with the compass operator, in: Computer Vision and Pattern Recognition, 1999, in: IEEE Computer Society Conference on. , vol.2, 1999, pp.166, doi:10.1109/CVPR.1999.784624

[19] J. Meltzer, S. Soatto, Edge descriptors for robust wide-baseline correspondence, in: Computer Vision and Pattern Recognition, 2008, in: IEEE Conference on , 23-28 June 2008, pp.1-8, doi:10.1109/CVPR.2008.4587684

[20] B. A. Maxwell and S. J. Brubaker, Texture edge detection using the compass operator, in: BMVC, volume II, September 2003, pp. 549– 558, doi:10.1.1.413.1690

[21] A.N. Evans, X.U. Liu, A morphological gradient approach to color edge detection, in: Image Processing, IEEE Transactions on , vol.15, no.6, June 2006, pp.1454-1463, doi:10.1109/TIP.2005.864164

[22] K. Khongkraphan, An Efficient Color Edge Detection Using the Mahalanobis Distance, in: Journal of Information Processing Systems. 10, 4, 2014, pp. 589-601, doi:10.3745/JIPS.02.0010

[23] A. Vedaldi and B. Fulkerson, Vlfeat: an open and portable library of computer vision, algorithms, in: Proceedings of the international conference on Multimedia (MM '10). ACM, New York, NY, USA, 2010, pp. 1469-1472, doi:10.1145/1873951.1874249

[24] http://www.vlfeat.org/ Last access January 23th, 2016.

[25] J. B. MacQueen. Some methods for classification and analysis of multivariate observations, in: Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, volume 1, 1967, pp. 281-297. doi:10.1.1.308.8619

[26] S. Lloyd, Least squares quantization, in: PCM, IEEE Trans. Inf. Theor. 28, 2, , 2006, pp. 129-137, doi:10.1109/TIT.1982.1056489

[27] G. Csurka, C. Bray, C. Dance, and L. Fan, Visual categorization with bags of keypoints, in: Workshop on Statistical Learning in Computer Vision, ECCV, 2004, pp. 1–22, doi:10.1.1.72.604

[28] V. N. Vapnik, The Nature of Statistical Learning Theory, Springer New York Inc., New York, NY, USA, 1995, doi:10.1.1.332.356

[29] J. A. Suykens and J. Vandewalle, Least squares support vector machine classifiers, in: Neural processing letters, 9(3), 1999, pp 293–300, doi:10.1023/A:1018628609742

[30] R.-E. Fan, K.-W. Chang, C.-J. Hsieh et al., LIBLINEAR: A library for large linear classification, in: J. Mach. Learn. Res., 2008, pp. 1817–1874, doi:10.1.1.140.9959

[31] S. Maji, A.C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, in: Computer Vision and Pattern Recognition, 2008, in : IEEE Conference on , 23-28 June 2008, pp.1-8, doi:10.1109/CVPR.2008.4587630

[32] C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, in: ACM Transactions on Intelligent Systems and Technology, 2011, 2:27:1--27:27, doi:10.1145/1961189.1961199

[33] 'The Ponce Group - Birds', http://www-cvr.ai.uiuc.edu/ponce_grp/data/ , Last access January 23th, 2016.

[34] M. E. Nilsback, A. Zisserman, A Visual Vocabulary for Flower Classification, in: Computer Vision and Pattern Recognition, 2006, in: IEEE Computer Society Conference on , vol.2, no., 2006, pp.1447-1454, doi:10.1109/CVPR.2006.42

[35] D. Keysers, M. Motter, T. Deselaers, and H. Ney, Training and Recognition of Complex Scenes using a Holistic Statistical Model, DAGM 2003, Pattern Recognition, 25th DAGM Symposium, Magdeburg, Germany, Volume LNCS 2781 of Lecture Notes in Computer Science, pp. 52-59, September 2003. doi:10.1.1.65.2980

[36] S. A. Nene, S. K. Nayar and H. Murase, Columbia Object Image Library (COIL-20), in: Technical Report CUCS-005-96, February 1996, doi:10.1.1.54.5914.

[37] L. Fei-Fei, R. Fergus and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004. doi: 10.1016/j.cviu.2005.09.012

[38] O. Russakovsky*, J. Deng* et al, (* = equal contribution), ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575v3, 2015.

[39] WordNet, A lexical database for English, http://wordnet.princeton.edu/, Last access January 23th, 2016.

[40] F. Basura et al., Discriminative feature fusion for image classification, in: Computer Vision and Pattern Recognition (CVPR), 2012, in: IEEE Conference on , vol., no., June 2012, pp.3434-3441, doi:10.1109/CVPR.2012.6248084

[41] X. Bai, C. Liu, P. Ren et al, Object Classification via Feature Fusion Based Marginalized Kernels, in: Geoscience and Remote Sensing Letters, IEEE , vol.12, no.1, Jan. 2015, pp.8-12, doi:10.1109/LGRS.2014.2322953

[42] J. Zhu, J. Yu, C. Wang et al., Object recognition via contextual color attention, in: Journal of Visual Communication and Image Representation, Volume 27, February 2015, pp. 44-56, doi:10.1016/j.jvcir.2015.01.003

[43] W. Hu; R. Hu; N. Xie et al., Image Classification Using Multiscale Information Fusion Based on Saliency Driven Nonlinear Diffusion Filtering, in: Image Processing, IEEE Transactions on , vol.23, no.4, April 2014, pp.1513-1526, doi:10.1109/TIP.2014.2303639

[44] T. Kobayashi, Low-Rank Bilinear Classification: Efficient Convex Optimization and Extensions, in: International Journal of Computer Vision, Volume 110, Issue 3 , 2014, pp 308-327, doi:10.1007/s11263-014-0709-5

Eduardo Fidalgo Fernández received the M. Sc. (2008) degree in Industrial Engineering and the Ph.D. degree in 2015, both from University of León. His current research interests are in the field of computer vision and pattern recognition, in particular, object recognition and invariant local features applied to image classification.

Enrique Alegre received the M.Sc. degree in Electrical Engineering in 1994 from the University of Cantabria and the Ph.D. degree in 2000 from the University of León. Currently, he is an Associate Professor in the Department of Electrical, Systems and Automation, University of Leon, Spain. His research interests include computer vision and pattern recognition applications to medical and industrial problems, and, more recently, machine learning and image processing applications for crime control and prevention.

Victor Gonzalez-Castro received the M.Sc. degree in Computer Science in 2006 and the Ph.D. degree in 2011, both from the University of León. During his Ph.D. he worked in image processing applied to semen quality control. He was a postdoctoral research assistant in the École des Mines de Saint-Étienne (France), working in adaptive mathematical morphology applied to classification of skin lesions. He is currently working as a lecturer in Medical Image Analysis at the Centre for Clinical Brain Sciences of the University of Edinburgh. His research interests include image processing, pattern recognition and machine learning applied to biomedical problems.

Laura Fernández-Robles received the BSc (2009) degree in Industrial Engineering from University of León where she graduated with honours (first class). In 2011 she received the MSc degree in Intelligent Systems in Engineering at University of León. At present, she is a PhD candidate at University of León, Spain and at the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, the Netherlands. Her current research interests are in the field of computer vision and pattern recognition, in particular, object recognition and local features description.