Adaptive and optimal combination of local features for image retrieval

Neelanjan Bhowmik[1,2], Valérie Gouet-Brunet[1], Lijun Wei[1], and Gabriel Bloch[2]

[1]Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mande, France
[2]Nicéphore Cité, 34 quai Saint-Cosme, 71100 Chalon-sur-Saône, France

Abstract. With the large number of local feature detectors and descriptors in the literature of Content-Based Image Retrieval (CBIR), in this work we propose a solution to predict the optimal combination of features, for improving image retrieval performances, based on the spatial complementarity of interest point detectors. We review several complementarity criteria of detectors and employ them in a regression based prediction model, designed to select the suitable detectors combination for a dataset. The proposal can improve retrieval performance even more by selecting optimal combination for each image (and not only globally for the dataset), as well as being profitable in the optimal fitting of some parameters. The proposal is appraised on three state-of-the-art datasets to validate its effectiveness and stability. The experimental results highlight the importance of spatial complementarity of the features to improve retrieval, and prove the advantage of using this model to optimally adapt detectors combination and some parameters.

Keywords: CBIR, interest points, feature combination, spatial complementarity, regression model.

1 Introduction

We are interested in content-based image similarity search, with the aim of better organizing and mining in the voluminous, complex image datasets. This work focuses on local image descriptors, where the extraction of feature points plays an imperative part in the process. The substantial number of available local feature descriptors in the present literature of Computer Vision and content-based image retrieval (CBIR), with respective advantages and drawbacks, makes it arduous to determine the most relevant descriptors for a given task and a given dataset. Thus it requires a framework making possible to evaluate the effectiveness of given descriptors on a specific dataset. It is also possible to combine descriptors to improve the content representation, such as in [5, 17], or to learn the best combination of given descriptors from a representative dataset [6]. All the different descriptors involved may not have the same relevance, and in addition, their distinctiveness may be different from one image to another. We think that it is important to appraise the complementarity between such local features, and in this work we focus on the complementarity of the detected points in the image, by exploiting statistical criteria of spatial analysis, in order to give the possibility to combine several descriptors, optimally for each image of the dataset and not only globally for the dataset. We propose a regression model with multiple complementarity criteria

of the feature detectors (which measure different properties of the detectors), such as distribution [7], contribution [9] and cluster-based measure [15]. Here, these criteria are evaluated for the combination of couples of detectors, but can easily be generalized to sets of detectors. Mean average precision (mAP) is incorporated to train the model and then anticipates the proper detector combination for a new dataset and new query images. Additionally, we demonstrate that this proposal allows to optimally fit some other parameters, here the best k during the k-nearest neighbor search.

The rest of the paper is organized as follows: Section 2 revisits the related work existing on the combination of descriptors, Section 3 explains the proposed methodology, Section 4 is dedicated to the experiments performed to evaluate our proposal, followed by conclusions in Section 5.

2 Related work

Several approaches of descriptors combination are available in the literature of CBIR. They are usually categorized as *early* and *late* fusion approaches [2, 23], based on the combination step position in the entire process according to the retrieval/learning step. The most common approach of early fusion is to combine multiple features into a single representation before exploiting it for retrieval/learning [23]. For instance, different shape properties are combined for image retrieval in [21], where genetic programing is used to find the suitable combination function for image descriptors, globally for the dataset. In the weighted early fusion approach proposed by [26], the weight values for different features such as color and texture are varied over the range of (0,1) to find the best appropriate weight values and then these features are combined depending on the assigned weight.

In the late fusion category, multiple features are learned or retrieved first separately, and then the responses or decisions are merged at the later stage [23]. In general for image retrieval, late fusion strategies are carried out in two primary ways such as, consolidate the rank responses and combining the different similarity scores for a query. The final output is obtained by cumulative ranked responses of the feature descriptors. In this context, a retrieval framework, based on genetic programing and relevance feedback, was proposed [8], where multiple sets of retrieved images are consolidated and then the rank list of the most relevant images to the query is returned. Other approaches, such as in short-term and long-term based learning [25], positive and negative feedbacks of the users are considered to construct semantic space and the final output. In the late fusion based image classification proposed by [27], each descriptor classifier output is combined by a weighted voting strategy where the voting weight is decided by the accuracy of the individual classifier.

Sometimes, fusion takes place during the retrieval/learning step, and is often called *intermediate* fusion. For example, in work [5], features fusion is performed during the retrieval step with multiple inverted index. Classification based on multiple kernels involves a learning step based on individual classifier and on the combination of weighted classifier [19].

In some of the previous approaches, the fusion step may not only rely on the fusion strategy but also on the selection of the features. For example, a hybrid method [18] is

proposed for simultaneous feature adaptation and feature selection, for a given dataset; here the parameter optimization during feature extraction and feature selection are carried out on a subset of dataset images by employing mixed gravitational search algorithm. In the work of [28], a rank based graph fusion technique is proposed by combining deep learning features, global and local features and the best feature combination is selected globally for a dataset based on the retrieval performance. [24] proposed a method for local selection of image features for similarity search and similarity graph construction, by computing local laplacian score and feature sparsification and considering the importance of the local neighborhood of each image point with respect to the image. Note that in all these approaches, the optimal combination of features is carried out globally for a whole dataset and not locally for each image.

3 Prediction of the complementarity between local detectors

This section is dedicated to the presentation of our proposal. In Section 3.1, we revisit the criteria used to evaluate the complementarity of several point detectors, and Section 3.2 describes how they are integrated in the whole prediction framework with a regression model.

3.1 Evaluation criteria of complementarity between keypoints

Our hypothesis is that better the detections are spread in the image, better the content is described, first because the detections would have more chance to describe the many areas of the image, and second because distant detections should statistically increase the variety of the associated descriptions, making the whole content description more distinctive. Therefore, we exploit several detectors of various natures and measure the spatial complementarity of two detectors, which will be exploited in our prediction model. The presentation is restricted to pairs of detectors, but can easily be generalized to the complementarity of sets of detectors. Let us consider the sets of keypoints extracted from an image by two detectors, $D_a = \{d_a^1(x_a^1, y_a^1), \ldots, d_a^n(x_a^n, y_a^n)\}$, $|D_a| = n$ and $D_b = \{d_b^1(x_b^1, y_b^1), \ldots, d_b^m(x_b^m, y_b^m)\}$, $|D_b| = m$.

Analysis of the spatial coverage. One of the key measurement criteria is coverage [7], describing how well the sets of points are distributed over an image. It is expected to gain a larger distribution if the points from two detectors occupy different locations in the image, which can traduce a better complementarity between the detections, producing a better representation of the content. First, a keypoint, *e.g.* $d_a^i(x_a^i, y_a^i)$, is considered as a reference point and Euclidean distances (ED_j^i) are calculated with other $(n+m-1)$ points of $D_a \cup D_b$. If two points detected by the two detectors are the same, there is no effect on the overall distribution. In order to neutralize the effect of the extreme outliers on the overall spatial distribution of $D_a \cup D_b$, the coverage measure is based on harmonic mean. The mean of the distances is computed as:

$$EDMean_{nm}^i = \frac{n+m-1}{\sum_{j=1, j\neq i}^{n+m-1}(1/ED_j^i)} \quad (1)$$

This step is reiterated for each keypoint of D_a and D_b considering each keypoint as a reference. The distribution complementarity score (DCS) is computed as:

$$DCS = \frac{n+m}{\sum_{i=1}^{n+m}(1/EDMean_{nm}^{i})} \tag{2}$$

Higher distribution scores, which are normalized between 0 to 1, indicate the better distribution of the points in the image.

Contribution measure. The contribution criterion [9] is a measure of the amount of dissimilar points detected by two detectors. It is possible that two detectors extract a certain number of same keypoints (p) for an image. The same detected points reduce the contribution measure of D_b over D_a and vice versa. The contribution of D_b over D_a is computed as:

$$Contribution_{D_b|D_a} = \frac{n-p}{n} \tag{3}$$

The overall complementarity between D_a and D_b is measured by:

$$CnCS = min(Contribution_{D_b|D_a}, Contribution_{D_a|D_b}) \tag{4}$$

If the detected points between the detectors are different, the score is 1; increasing number of common points reduces this score.

Cluster-based measurement of complementarity. Based on spatial clustering, this measure [15] determines how the different detectors extract similar local structures in a cluster. The clusters are generated in the image space from extracted points of D_a and D_b, using a clustering algorithm (*e.g.* k-means). Each cluster $(c_j, j = 1 \dots k)$ may contain points from D_a and/or D_b. Points from D_a and D_b in cluster c_j, *i.e.* resp. F_{jD_a} and F_{jD_b}, contribute to the total number of points (F_j) present in c_j. The frequency of the points from D_a and D_b in c_j is computed as:

$$p_{jD_a} = \frac{|F_{D_a}|}{|F_j|} \quad \& \quad p_{jD_b} = \frac{|F_{D_b}|}{|F_j|} \tag{5}$$

The whole complementarity score can be computed as:

$$ClCS = 1 - 2.\frac{1}{k}\sum_{j=1}^{k} min(p_{jD_a}, p_{jD_b}) \tag{6}$$

When p_{jD_a} or p_{jD_b} is close to 1 and the other is close to 0, the score is close to 1; it implies a better complementarity of the detectors.

3.2 Image retrieval based on regression model and complementarity measures

To perform image retrieval, we propose to learn a regression model based on the three complementarity criteria revisited in Section 3.1, on the number of detected interest

keypoints per image (Kp) and on mean Average Precision (mAP) as retrieval output, which is a summarized measure of quality across all the queries by averaging average precision. The objective is to predict the best detector combinations for an image dataset, and also to fit some parameters. Other configurations of criteria were tested, and we obtained the best results with this configuration.

To illustrate, in the Fig. 1 we consider two combinations of two detectors (hesaff-mser and har-colsym where hesaff-mser being generally more efficient on the considered dataset Paris_DB), and for each complementarity score and each query image, we plot the difference of scores between the two combinations vs. the corresponding difference in the mAP. We observe that globally, complementarity scores values increase with mAP values (most of the points are in the area related to positive axes). Therefore, We assume that the relationship between the complementarity criteria and the mAP is general for all image datasets, and we employ a classical linear regression model.

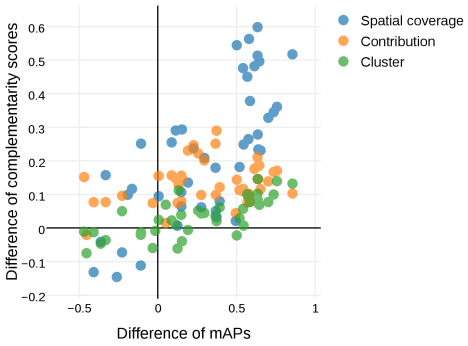

Fig. 1. Relationship between complementarity scores differences and mAP differences, for each query image and two combinations of two detectors.

Training of the regression model.
The training step is decomposed into three steps involving complementary criteria and mAP:

1. Several detectors, *e.g.* $D_1, \ldots, D_x$, are used to extract keypoints from images, leading to x sets of keypoints. Here, we consider the C_x^2 couples of detectors $(D_i, D_j)_{i \neq j}$ and compute for them the three complementarity scores, for each image of the dataset. We also keep the number of keypoints per image.
2. One mAP is then computed for the images dataset described with a couple of detectors $(D_i, D_j)_{i \neq j}$, using a classical approach of query-by-example retrieval able to use several descriptors jointly, such as in [5]. We obtain C_x^2 mAPs.
3. Finally, the relationship between the different combination of complementarity scores and the retrieval output (mAP) is learned by a linear regression model. Regression coefficients, such as adjusted $R2$, are calculated to analyze the model fitness and to determine the best fitted model for the prediction for the given model inputs and the output.

Prediction of the best detector combination. The prediction steps of the best detector pair for a new dataset are:

1. The detectors $D_1, \ldots, D_x$ extract keypoints on each image of the new dataset. The three complementarity scores of the detector pairs are computed, similarly to step 1 of Section 3.2.
2. For each detector combination $(D_i, D_j)_{i \neq j}$, we predict the mAP, called mAP', using previously trained regression model. The complementarity scores of each detector pair are the inputs for the regression model. The outputs mAP' are predicted using the model parameters and the inputs.
3. The detector pair with the highest mAP' is selected as the suitable detector pair for image retrieval on the new dataset.

These training and prediction steps are presented by considering pairs of detectors, but can be generalized to any sets of detectors, based on the generalization of the complementarity criteria. The approach of prediction presented above predicts the best detector combination *globally* for a given dataset. It can be directly employed to predict the best combination *for each query image*, which can be different from an image to another; in the experiment Section 4, we will see that the quality of image retrieval can be improved even more by considering such an image-by-image prediction. We will also see that the regression model can be employed to predict some other parameters, such as the k during k-NN retrieval.

4 Experiments and evaluation

This section presents and discusses the experiments conducted to evaluate our contributions.

4.1 Framework of evaluation

The experiments are conducted on three image datasets, illustrated in Fig. 2:

1. Paris_DB: this dataset is a public benchmark[1] consisting of 6412 images collected from Flickr by searching for Paris landmarks.
2. Oxford_DB: this public benchmark[2] consists of 5062 images collected from Flickr by searching for particular 11 Oxford landmarks.
3. Holiday_DB: this dataset is a public benchmark[3] consisting of 1491 images includes a large variety of scene types.

We have selected 7 detectors from characteristically diverse categories such as blob, corner, symmetry, etc.: Hessian affine (hesaff) [16], color symmetry (colsym) [10], MSER (mser) [14], Harris (har) [22], Star (star) [1], binary robust invariant scalable keypoints (brisk) [11] and oriented and rotated BRIEF (orb) [20]. Extracted keypoints are described by three complementarity local descriptors, *i.e.* SIFT [13], SURF [3] and SC [4], and used jointly in an image retrieval system designed for bag-of-word descriptors combination ('FII') [5]. Performances are presented with mean Average Precision

[1] http://www.robots.ox.ac.uk/ vgg/data/parisbuildings/
[2] http://www.robots.ox.ac.uk/ vgg/data/oxbuildings/
[3] https://lear.inrialpes.fr/ jegou/data.php

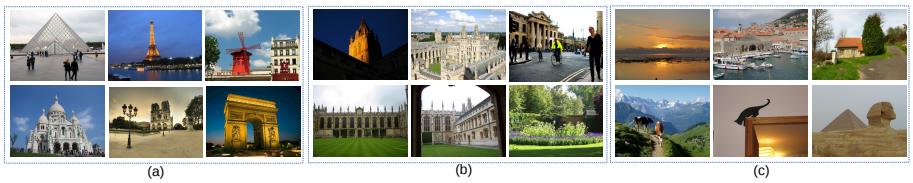

Fig. 2. Samples from the three benchmarks: (a) Paris_DB, (b) Oxford_DB and (c) Holiday_DB.

(mAP). Codebook size and value of k during nearest neighbor (k-NN) search are two important parameters of the 'FII' system. Optimal codebook size used for Paris_DB and Oxford_DB is 1500000 words. For the Holiday_DB, 30% of the total description points of each detector combination is selected as codebook size. Parameter k is varied in between 2 to 10 for optimal combination of nearest neighbors and detectors.

4.2 Global prediction of the detectors combination performance

In this section, we present the prediction results of detector combinations using the linear regression model. It is trained with Paris_DB as described in Section 3.2. Model

Dataset	Detector pair	mAP'	Detector pair	mAP'	Detector pair	mAP'
Paris_DB	hesaff-colsym	0.512	hesaff-mser	0.548	hesaff-har	0.481
	hesaff-star	0.547	hesaff-orb	0.501	hesaff-brisk	0.520
	colsym-mser	0.429	colsym-har	0.384	colsym-star	0.457
	colsym-orb	0.410	colsym-brisk	0.405	mser-har	0.481
	mser-star	0.526	mser-orb	0.510	mser-brisk	0.507
Oxford_DB	hesaff-colsym	0.501	hesaff-mser	0.537	hesaff-har	0.615
	hesaff-star	0.524	hesaff-orb	0.503	hesaff-brisk	0.523
	colsym-mser	0.482	colsym-har	0.554	colsym-star	0.459
	colsym-orb	0.360	colsym-brisk	0.504	mser-har	0.579
	mser-star	0.492	mser-orb	0.465	mser-brisk	0.527
Holiday_DB	hesaff-colsym	0.442	hesaff-mser	0.461	hesaff-har	0.427
	hesaff-star	0.450	hesaff-orb	0.392	hesaff-brisk	0.441
	colsym-mser	0.402	colsym-har	0.415	colsym-star	0.354
	colsym-orb	0.338	colsym-brisk	0.413	mser-har	0.400
	mser-star	0.395	mser-orb	0.376	mser-brisk	0.415

Table 1. Different detector combinations and predicted mAP (mAP') using the regression model.

inputs, the complementarity scores, *i.e.* distribution, contribution, cluster and number of keypoints (Kp), of detectors pairs are computed for Paris_DB. The mAP is calculated using the 'FII' approach on Paris_DB. The best fitted model 'Kp-Distribution-Contribution-Cluster - mAP' is selected, based on the highest adjusted $R2$ value, for further prediction experiments on test datasets, *i.e.* Oxford_DB and Holiday_DB. For predictions on the test datasets, procedure of Section 3.2 is applied by computing complementarity scores of the detector pairs. The predictions of the detector pairs using

previously trained 'Kp-Distribution-Contribution-Cluster - mAP' model are presented in the Table 1, with associated mAP'. Due to the space limitation, we only present the selected representative detector pairs prediction results. We observe that detector pairs, 'hesaff-har' and 'hesaff-mser' are associated with the best predicted mAP. Thus, we consider them as the best combinations for image retrieval on these datasets.

4.3 Effective performances for image retrieval

In this section, the effective image retrieval results ($mAP^\natural$), for Paris_DB and the test datasets Oxford_DB and Holiday_DB, are presented. Due to space limitation, in Table 2 we only present results associated with the two best predicted pairs and one worst predicted pair using 'FII' retrieval. For Oxford_DB, the best effective result should be obtained with 'hesaff-har' pair (see Table 1). Indeed, the highest effective mAP ($mAP^\natural$) is achieved with this combination (see Table 2). Also, the $mAP^\natural$ of 0.269 is achieved with 'colsym-orb' which is the worst performing pair. For Holiday_DB, although the $mAP^\natural$ are not in the same range as the predicted mAP, the sorted sequence of mAP' reflects the one of effective retrieval results. This first set of experiments confirms us that the complementarity scores, employed with the linear regression model, are able to correctly estimate the performance of a detectors pair for image retrieval, then to enable the use of the best detector pair for a dataset.

We also compare our results (see Table 2, related to 'LF' rows) with one of the state-of-the-art late fusion ('LF') image retrieval technique [17]. We selected the two best performing detector pairs for 'LF' retrieval for each dataset. The comparison results demonstrate that our proposed detector combination selection approach and then 'FII' image retrieval outperforms the 'LF' retrieval. In Table 3, the retrieval results with

	Paris_DB			Oxford_DB			Holiday_DB		
	Detector pair	k-NN	$mAP^\natural$	Detector pair	k-NN	$mAP^\natural$	Detector pair	k-NN	$mAP^\natural$
FII	hesaff-mser	2	0.589	hesaff-har	2	0.549	hesaff-mser	2	0.683
	hesaff-star	2	0.570	mser-har	2	0.456	hesaff-star	2	0.666
	har-colsym	2	0.371	colsym-orb	2	0.269	colsym-orb	2	0.499
LF	hesaff-mser	2	0.541	hesaff-har	2	0.450	hesaff-mser	2	0.630
	hesaff-star	2	0.535	mser-har	2	0.334	hesaff-star	2	0.599

Table 2. Effective mAP ($mAP^\natural$) of detector pair using 'FII' and 'LF' [17] technique.

selected single detector are presented in order to compare with detector pair results of Table 2. These results demonstrate the relevance of the use of several detectors in the representation of the content. In addition to 'FII' image retrieval system, the additional computation of our proposed framework includes computation of different features, complementarity between the features and regression model for prediction.

4.4 Effect of k-NN parameter on retrieval and its prediction

In this section, we present retrieval results in Table 4 by varying k during k-NN retrieval of the closest neighbors ($k = 2, 5, 10$) and observe the consequence on $mAP^\natural$. The best

Paris_DB			Oxford_DB			Holiday_DB		
Detector	k-NN	$mAP^\natural$	Detector	k-NN	$mAP^\natural$	Detector	k-NN	$mAP^\natural$
hesaff	2	0.546	hesaff	2	0.498	hesaff	2	0.646
mser	2	0.523	har	2	0.421	mser	2	0.505

Table 3. $mAP^\natural$ of single detector using 'FII' for the different datasets.

Dataset	Detector pair	$mAP^\natural$		
		$k=2$	$k=5$	$k=10$
Paris_DB	hesaff-mser	0.589	0.571	0.531
		0.591 (adaptive k 2,5 & 10)		
Oxford_DB	hesaff-har	0.549	0.547	0.533
		0.567 (adaptive k 2,5 & 10)		
Holiday_DB	hesaff-mser	0.683	0.677	0.670
		0.691 (adaptive k 2,5 & 10)		

Table 4. $mAP^\natural$ for all datasets by varying k-NN and adapting it with prediction model.

$mAP^\natural$ is obtained with $k=2$ for all datasets. The accuracy difference is 1.8% between $k=2$ and $k=5$ for 'hesaff-mser' in Paris_DB, while it is 1.6% between $k=2$ and $k=10$ for 'hesaff-har' with Oxford_DB. During the search for the nearest neighbors of the query point, higher values of k might include dissimilar neighbors in the k-NN lists. By using our model, it is possible to adapt the best value of k *for each query image* instead of finding it globally for the dataset. The procedure of Section 3.2 is applied for a detector combination, by varying k ($k=2,5,10$) and the prediction mAP obtained allows to adapt k to each query. In Table 4, underlined mAP correspond to $mAP^\natural$ obtained by adapting k to each query. The accuracy is increased by 0.2%, 1.8% and 0.8% for Paris_DB, Oxford_DB and Holiday_DB resp. compared to the previous best with $k=2$. Figure 3 shows the distribution of the k selected adaptively across the queries for all datasets. The majority of the best results are associated with $k=2$ followed by $k=5$ and $k=10$. For example with Oxford_DB, approximately 47% of the queries are executed with $k=2$, 44% with $k=5$, and only 9% with $k=10$.

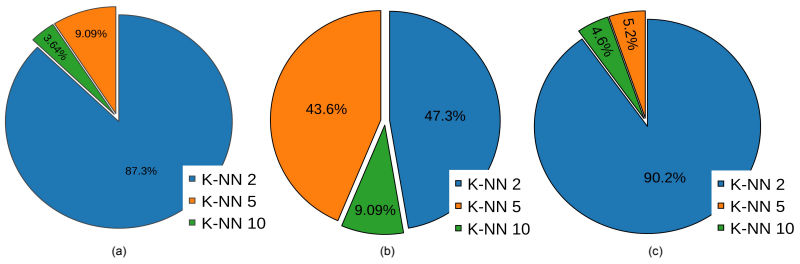

Fig. 3. Distribution of predicted k values across the queries: (a) 'hesaff-mser' for Paris_DB. (b) 'hesaff-har' for Oxford_DB (c) 'hesaff-mser' for Holiday_DB.

4.5 Image-by-image prediction of the best detector combination

In this section, we refine the results obtained in Sections 4.2, 4.3 and 4.4 by adapting the selection of the best detector combination to each image, by applying the prediction strategy of Section 3.2 to each query image. Six different combinations of $mAP^\natural$ obtained with two best detector pairs and three k-NN value ($k = 2, 5, 10$) are consolidated. In Table 5, we observe that $mAP^\natural$ is increased by 0.8%, 2.5% and 21.1% compared to previous best with $k = 2$ for Paris_DB, Oxford_DB and Holiday_DB. The achieved retrieval accuracy for Holiday_DB is 0.894, which is one of the best in the state-of-the-art to our knowledge, compared to Ref. [12] which is also based on bag of words. As depicted in Fig. 4, the majority of the selections are done with $k = 2$. For Paris_DB, 90.9% are selected for $k = 2$ of both pairs of detectors, while 5.49% are with $k = 5$. Most of the selections (85%) are done with 'hesaff-har' for Oxford_DB, while 15% are from 'mser-har'. Even if the statistical analysis (Fig. 3 and 4) has highlighted the dominance of some particular detectors pairs and values of k, we observe that using other ones adaptively allows to refine the results notably.

Dataset	Detector pair	$mAP^\natural$		
		$k = 2$	$k = 5$	$k = 10$
Paris_DB	hesaff-mser	0.589	0.571	0.531
	hesaff-star	0.570	0.544	0.512
	Adaptive detector combination	0.597 (Adaptive k 2,5 & 10)		
Oxford_DB	hesaff-har	0.549	0.547	0.533
	mser-har	0.456	0.430	0.420
	Adaptive detector combination	0.574 (Adaptive k 2,5 & 10)		
Holiday_DB	hesaff-mser	0.683	0.677	0.670
	hesaff-star	0.666	0.661	0.650
	Adaptive detector combination	0.894 (Adaptive k 2,5 & 10)		

Table 5. $mAP^\natural$ obtained for all the datasets, by selecting optimal detector pairs and optimal value k for each query image.

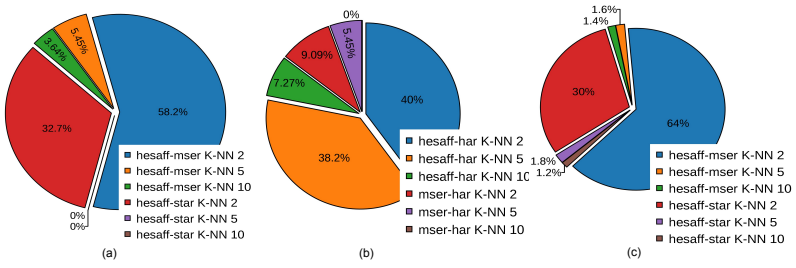

Fig. 4. Distribution of predicted values of k and detectors pairs across the queries: (a) 'hesaff-mser' & 'hesaff-star' for Paris_DB. (b) 'hesaff-har' & 'mser-har' for Oxford_DB (c) 'hesaff-mser' & 'hesaff-star' for Holiday_DB.

5 Conclusions

The main contribution of our approach is the possibility to select adaptively the best detector combination for each query in query-by-example image retrieval. The proposal rests on the use of spatial complementarity criteria between local features and on a linear regression model that models the relationship between complementarity and optimal performances during retrieval. Even if the statistical analysis highlights the dominance of some detectors pairs and values of k, we observe that using other ones adaptively allows to refine the results favorably. The conducted experiments clearly highlight the impact of the spatial complementarity of the selected features on the image retrieval performance: the higher complementarity scores imply a more distinctive representation of the content. The proposed framework can effectively reduce the overall experimental time by narrowing down the choice of detectors, and the adaptive selection of some parameters, such as k during the nearest neighbor retrieval, improves even more the retrieval accuracy. It is easily possible to extend this framework to the evaluation of the complementarity between multiple detectors.

Acknowledgments. The authors are grateful to Nicéphore Cité, Institut national de l'information géographique et forestière (IGN) and French project POEME ANR-12-CORD-0031 for the financial support.

References

1. Agrawal, M., Konolige, K., Blas, M.: Censure: Center surround extremas for realtime feature detection and matching. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 5305, pp. 102–115. Springer Berlin Heidelberg (2008)
2. Atrey, P.K., Hossain, M.A., Saddik, A.E., Kankanhalli, M.S.: Multimodal Fusion for Multimedia Analysis: A Survey. Multimedia Systems 16, 345–379 (2010)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). Computer Vision and Image Understanding 11(3), 346–359 (Jun 2008)
4. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition using Shape Contexts. Pattern Analysis and Machine Intelligence 24(4), 509–522 (Apr 2002)
5. Bhowmik, N., Gonzalez V, R., Gouet-Brunet, V., Pedrini, H., Bloch, G.: Efficient fusion of multidimensional descriptors for image retrieval. In: International Conference on Image Processing. pp. 5766–5770 (Oct 2014)
6. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: an experimental comparison. Information Retrieval 11(2), 77–107 (2008)
7. Ehsan, S., Clark, A.F., McDonald-Maier, K.D.: Rapid online analysis of local feature detectors and their complementarity. Sensors 13(8), 10876 (2013)
8. Ferreira, C., Santos, J., da S. Torres, R., Goncalves, M., Rezende, R., Fan, W.: Relevance feedback based on genetic programming for image retrieval. Pattern Recognition Letters 32(1), 27 – 37 (2011), image Processing, Computer Vision and Pattern Recognition in Latin America
9. Gales, G., Crouzil, A., Chambon, S.: Complementarity of feature point detectors. In: Richard, P., Braz, J. (eds.) VISAPP (1). pp. 334–339. INSTICC Press (2010)

10. Heidemann, G.: Focus-of-attention from local color symmetries. Pattern Analysis and Machine Intelligence 26(7), 817–830 (July 2004)
11. Leutenegger, S., Chli, M., Siegwart, R.: Brisk: Binary robust invariant scalable keypoints. In: International Conference on Computer Vision. pp. 2548–2555 (Nov 2011)
12. Li, X., Larson, M., Hanjalic, A.: Pairwise geometric matching for large-scale object retrieval. In: Computer Vision and Pattern Recognition. pp. 5153–5161 (June 2015)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
14. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of the British Machine Vision Conference. pp. 36.1–36.10 (2002)
15. Mikolajczyk, K., Leibe, B., Schiele, B.: Local features for object class recognition. In: International Conference on Computer Vision. vol. 2, pp. 1792–1799 Vol. 2 (Oct 2005)
16. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision 60(1), 63–86 (2004)
17. Neshov, N.: Comparison on Late Fusion Methods of Low Level Features for Content Based Image Retrieval. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E., Appollini, B., Kasabov, N. (eds.) Artificial Neural Networks and Machine Learning, Lecture Notes in Computer Science, vol. 8131, pp. 619–627. Springer Berlin Heidelberg (2013)
18. Rashedi, E., Nezamabadi-pour, H., Saryazdi, S.: A simultaneous feature adaptation and feature selection method for content-based image retrieval systems. Knowledge-Based Systems 39, 85 – 94 (2013)
19. Risojevic, V., Babic, Z.: Fusion of Global and Local Descriptors for Remote Sensing Image Classification. IEEE Geoscience and Remote Sensing Letters 10(4), 836–840 (2013)
20. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: International Conference on Computer Vision. pp. 2564–2571 (Nov 2011)
21. da S. Torres, R., Falcao, A.X., Goncalves, M.A., Papa, J.P., Zhang, B., Fan, W., Fox, E.A.: A genetic programming framework for content-based image retrieval. Pattern Recognition 42(2), 283 – 292 (2009), learning Semantics from Multimedia Content
22. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(5), 530–534 (May 1997)
23. Snoek, C.G.M., Worring, M., Smeulders, A.W.M.: Early Versus Late Fusion in Semantic Video Analysis. In: Proceedings of the 13th Annual ACM International Conference on Multimedia. pp. 399–402. ACM, New York, NY, USA (2005)
24. Sun, J.: Local selection of features for image search and annotation. In: Proceedings of the 22Nd ACM International Conference on Multimedia. pp. 655–658. ACM, New York, NY, USA (2014)
25. Wacht, M., Shan, J., Qi, X.: A short-term and long-term learning approach for content-based image retrieval. In: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. vol. 2, pp. II–II (May 2006)
26. Yue, J., Li, Z., Liu, L., Fu, Z.: Content-based Image Retrieval using Color and Texture Fused Features. Mathematical and Computer Modelling 54(3-4), 1121–1127 (2011), mathematical and Computer Modeling in Agriculture
27. Zhang, W., Qin, Z., Wan, T.: Image Scene Categorization using Multi-Bag-of-Features. In: Proceedings of International Conference on Machine Learning and Cybernetics. vol. 4, pp. 1804–1808 (2011)
28. Zhou, Y., Zeng, D., Zhang, S., Tian, Q.: Augmented feature fusion for image retrieval system. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. pp. 447–450. ACM, New York, NY, USA (2015)