Combination of image descriptors for the exploration of cultural photographic collections

Neelanjan Bhowmik,[a,b,*] Valérie Gouet-Brunet,[a] Gabriel Bloch,[b] and Sylvain Besson[c]
[a]University Paris-Est, LASTIG MATIS, IGN, ENSG, 73 Avenue de Paris, F-94160 Saint-Mande, France
[b]Nicéphore Cité, 34 Quai Saint-Cosme, 71100 Chalon-sur-Saône, France
[c]Nicéphore Niépce Museum, 28 Quai des Messageries, 71100 Chalon-sur-Saône, France

Abstract. The rapid growth of image digitization and collections in recent years makes it challenging and burdensome to organize, categorize, and retrieve similar images from voluminous collections. Content-based image retrieval (CBIR) is immensely convenient in this context. A considerable number of local feature detectors and descriptors are present in the literature of CBIR. We propose a model to anticipate the best feature combinations for image retrieval-related applications. Several spatial complementarity criteria of local feature detectors are analyzed and then engaged in a regression framework to find the optimal combination of detectors for a given dataset and are better adapted for each given image; the proposed model is also useful to optimally fix some other parameters, such as the k in k-nearest neighbor retrieval. Three public datasets of various contents and sizes are employed to evaluate the proposal, which is legitimized by improving the quality of retrieval notably facing classical approaches. Finally, the proposed image search engine is applied to the cultural photographic collections of a French museum, where it demonstrates its added value for the exploration and promotion of these contents at different levels from their archiving up to their exhibition *in* or *ex situ*. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.26.1.XXXXXX]

Keywords: content-based image retrieval; local descriptors; features complementarity; data fusion; cultural heritage; linked data.
Paper 16571SS received Jul. 1, 2016; accepted for publication Nov. 17, 2016.

1 Introduction

In the last decades, several museums and archive companies have started image digitization on a large scale. At the same time, the rapid development of the production of digital images confronts professional users and individuals with a huge number of images which are difficult to exploit due to the volume. In addition, image datasets are becoming more and more complex due to the diverse contents of the images, making their organization an essential stage for any dataset. When we consider the example of photographic museums (illustrated in Fig. 1 with photographs from Musée Nicéphore Niépce), image indexing, structuring, and retrieval steps are mostly done manually. These steps are strongly influenced by several factors, such as the use of the database, cultural background of the archivist, and his professional references. The cultural and professional references of the archivists instigate the standard of indexing and the thesaurus used in the museums, which are unknown to the general public.

When preparing an exhibition, the selection of photographs is usually driven by the registrar or the curator of the exhibition; a first filter takes place. Archivists extract keywords from photographs that might be of interest to the commissioner, thereby constituting a second filter. In the case of an exhibition project involving several institutions, these factors are increased. Additionally, managing different databases and the virtual sharing of iconographic collections are made difficult by the disparity of standards, the language barriers, the control of the software used, and the management of documentary parasites inherent in any search by keywords. Finally, based on lexical fields of language backgrounds, which are different from country to country, indexing and searching for images in a database may often be exclusive to one country.

Once established, the collection is then exhibited to the general public according to a given spatial organization when *in situ*. Virtual exhibitions usually consist of interacting via a website with the collection by selecting categories or keywords and looking at photographs through lists, thumbnails, or slideshows.

In the context of artwork and cultural heritage retrieval, a few recent content-based image retrieval (CBIR)-based approaches are available in the literature.[1,2] For cultural heritage retrieval, Vrochidis et al.[3] proposed a hybrid multimedia retrieval model, which combines low-level visual features-based retrieval and semantic annotation retrieval to find similar images. Yen et al.[4] developed an image retrieval system based on AdaBoost[5] and relevance feedback for painting image retrieval. SCULPTEUR[6] and MIRS[7] are two tools for museum multimedia retrieval. SCULPTEUR is designed for three-dimensional (3-D) and two-dimensional (2-D) content retrieval, while MIRS uses several low-levels visual features individually, such as color and texture features to retrieve similar images. In cultural multimedia retrieval, cross-domain retrieval is one of the applications where the goal is to find similar matches to a query of paintings, old postcards, and historical architecture photos in the characteristically different databases (see, e.g., contributions by Shrivastava et al.[8] and Russell et al.[9]). With similar ideas on the matching of cross-domain features, Aubry et al.[10]

*Address all correspondence to: Neelanjan Bhowmik, E-mail: neelanjan.bhowmik@ign.fr

Fig. 1 Examples of digitized photographs from the archives of the Musée Nicéphore Niépce, France.

proposed a technique to align 2-D descriptions of architectural site images with 3-D models. However, most of the approaches discussed usually consider a single feature rather than combining multiple features.

In this context, we are interested in CBIR tools to organize the voluminous and varied image datasets involved in museum collections. Indexing, comparing, and retrieving images by measuring the similarity of their content open very interesting perspectives when applied to museums. In addition to helping the archivist in automating the indexing step in large collections (e.g., automatic propagation of keywords, automatic linking of similar contents, and so on) while minimizing the subjective and cultural factors cited above, this paradigm may provide new standards and personalized tools to museum professionals and users, thus improving management of the collections by their experts as well as contributing to its enhancement for the general public.

In this work, we are interested in using local image descriptors, popular in the CBIR domain; they are like a fingerprint of interesting parts in the image content, producing a precise measure of comparison between images, and efficient indices within a dataset of images. A considerable number of local feature detectors[11–13] and descriptors,[14,15] with respective advantages and disadvantages, are present in the literature of CBIR.[16] It is always arduous to opt for the most pertinent features in image retrieval for a given dataset, and to determine the effectiveness of the features on a given dataset, an evaluation framework is usually required. The combination of multiple features, as already studied in several works,[17–19] improves the content representation. The different features may not have similar levels of aptness during the image retrieval process but it may be different from one image to another. One key emphasis is on the modeling of a suitable combination of image features for a given dataset, or better, for a given image. Here, we have chosen to focus on the modeling of the spatial complementarity between local image detectors because of the variety of digitized photographic contents of museums where a given detector may not have the same relevance from one image to another.

Our main contribution consists of the proposal of a regression model that involves several complementary statistical criteria of spatial analysis of feature detectors. Mean average precision (mAP), as evaluation measure of the quality of the content description, is incorporated to train the model and assists users in anticipating the proper detector combinations for each image of the dataset and each query image. With this model, other parameters, such as the k value during k-nearest neighbors (k-NN) search, can also be adaptively selected. This proposal, dedicated to query-by-example image retrieval in large datasets of images, is being employed to help the professionals of a French museum at different levels from the archiving up to the exhibition.

The rest of the article is organized as follows: Sec. 2 revisits the related work on descriptors combination. Section 3 presents our proposal, while Secs. 4 and 5 are dedicated to its evaluation and to its application for exploration in a digitized photographic museum collection, respectively, before concluding in Sec. 6.

2 Related Work on Descriptors Combination

Several approaches are available in the literature of CBIR for the combination of descriptors. They are usually categorized as "early" and "late" fusion approaches,[20,21] based on their 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.[21] In the early fusion approach proposed by Yu et al.,[22] several features, such as scale-invariant feature transform (SIFT),[14] histogram of oriented gradients,[23] and local binary patterns,[24] are concatenated into a single vector and then the concatenated vector is used for image retrieval. Genetic programming (GP) is used in one of the early fusion strategies proposed by da Torres et al.,[25] where several shape features are fused together for image retrieval. GP is exploited to find the suitable combination function for the image descriptors combination. A weight-based early fusion strategy is proposed by Yue et al.,[26] where different weights are assigned for different features, such as color and texture; the weight values are varied in between (0, 1) in order to find the most suitable values for each feature.

In the late fusion category, multiple features are first learned or retrieved separately, then the responses or decisions are merged at the later stage.[21] In the image retrieval context, late fusion strategies are carried out in two primary ways:

1. consolidating the rank responses and
2. combining the different similarity scores for a query.

The final output is obtained by cumulative ranked responses of the feature descriptors. In the late fusion strategy proposed by Ferreira et al.[27] based on relevance feedback

and GP, multiple sets of retrieved images were 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,[28] and positive and negative feedbacks of the users, are considered to construct semantic space and the final output. Several classical late fusion strategies for image retrieval and their comparisons are discussed by Neshov[18] and Chatzichristofis et al.[29] Late fusion strategy is also employed for image classification problems. For example, in Ref. 30, the output of the support vector machine classifiers, which are associated with the low-level descriptor, are combined by a weighted voting; the voting weight is decided by the accuracy of the individual classifier.

Sometimes, fusion takes place within the retrieval/learning step itself and is often called "intermediate" fusion. For example, multiple features are combined for CBIR by Bhowmik et al.[17] during the retrieval step by building a multiple inverted index. Classification based on multiple kernels involves a learning step based on an individual classifier and on the combination of weighted classifiers.[31,32] Other strategies of fusion exist, such as "sequential" fusion as in Ref. 33, where one descriptor is used as a filter for the other descriptors on the subset of image.

The fusion step may not only rely on the fusion strategy, but also on the selection of features. A hybrid method[34] is proposed for simultaneous feature adaptation and feature selection for a given dataset. Zhou et al.[35] proposed a rank-based graph fusion technique for image retrieval by combining deep learning features and global and local features. A method for local selection of image features for a similarity search and similarity graph construction was proposed by Sun.[36] All these approaches involving the optimal combination of features are carried out globally for the whole dataset.

3 Image Retrieval Based on a Prediction Model of the Complementarity between Local Detectors

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

3.1 *Evaluation Criteria of Complementarity between Keypoints*

In this work, our hypothesis states that the better the detections are spread throughout the image, the better the content is described, first because the detections would have a greater 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. To achieve this goal, we exploit several detectors of various natures, such as detectors of corners, of salient interest points, of local symmetries, of blobs, and so on, with the intention of maximizing the spatial distribution of these detections in the image. To measure this information, we have chosen to explore criteria that evaluate the spatial complementarity of two detectors, which will be exploited in our prediction model. We revisit three criteria in Secs. 3.1.1, 3.1.2, and 3.1.3. For the sake of clarity, 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 $D_b = \{d_b^1(x_b^1, y_b^1), \ldots, d_b^m(x_b^m, y_b^m)\}$, $|D_b| = m$.

3.1.1 *Analysis of the spatial coverage*

One of the key measurement criteria is coverage,[37] which describes how well the sets of points are distributed over an image. If the points from the two detectors occupy different locations in an image, then a high distribution is expected. This implies detectors are highly complementarity to each other and they produce detailed representations of the image 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), which are in image space, 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. 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 the harmonic mean. The mean of the distances is computed as

$$\mathrm{EDMean}_{nm}^i = \frac{n + m - 1}{\sum_{j=1, j\neq i}^{n+m-1} (1/\mathrm{ED}_j^i)}. \tag{1}$$

This step is reiterated for each keypoint of D_a and D_b considering each keypoint as a reference. The distribution complementarity score $(\mathrm{Di}_{\mathrm{cs}})$ is computed as

$$\mathrm{Di}_{\mathrm{cs}} = \frac{n + m}{\sum_{i=1}^{n+m} (1/\mathrm{EDMean}_{nm}^i)} \tag{2}$$

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

3.1.2 *Contribution measure*

The contribution criterion[38] is a measure of the number of dissimilar points detected by two detectors. It is possible that two detectors extract a certain number of the 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 $(\mathrm{Cn}_{D_b|D_a})$ is computed as

$$\mathrm{Cn}_{D_b|D_a} = \frac{n - p}{n}. \tag{3}$$

The overall complementarity between D_a and D_b is measured as

$$\mathrm{Cn}_{\mathrm{cs}} = \min(\mathrm{Cn}_{D_b|D_a}, \mathrm{Cn}_{D_a|D_b}). \tag{4}$$

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

3.1.3 *Cluster-based measurement of complementarity*

Based on spatial clustering, this measure[39] determines how the different detectors extract similar local structures in a cluster. The clusters are generated in the image space from the extracted points of D_a and D_b, using a clustering algorithm (e.g., k-means). Each cluster $(c_j, j = 1 \ldots k)$ may contain points from D_a and/or D_b. Points from D_a and D_b in cluster c_j, i.e., F_{jD_a} and F_{jD_b}, respectively, contribute to

the total number of points (F_j) present in c_j. The proportional number of points from D_a and D_b in c_j is computed as

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

The cluster complementarity score ($\mathrm{Cl_{cs}}$) can be computed as

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

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; this implies a better complementarity of the detectors.

Before extracting the keypoints of interest, all the images are normalized/scaled to the same size (1024× resolution) to make the measures computed from one image to another comparable.

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 complementarity criteria revisited in Sec. 3.1, the number of detected interest keypoints per image (Kp), and the mean average precision (mAP) as the image retrieval system output, which is a summarized measure of quality across all the queries by averaging precision as presented in Eq. 8 in Sec. 4.1. The objective is to predict the best detector combinations for an image dataset and also to fit some parameters. We assume that the relationship between the complementarity criteria and the mAP is general for all image datasets, and we employ the following linear regression model:

$$\mathrm{mAP} = \beta_1 \mathrm{Kp} + \beta_2 \mathrm{Di_{cs}} + \beta_3 \mathrm{Cn_{cs}} + \beta_4 \mathrm{Cl_{cs}}, \tag{7}$$

where β_i are the model coefficients. The block diagram of the proposed framework is depicted in Fig. 2 to give a general overview of the entire framework.

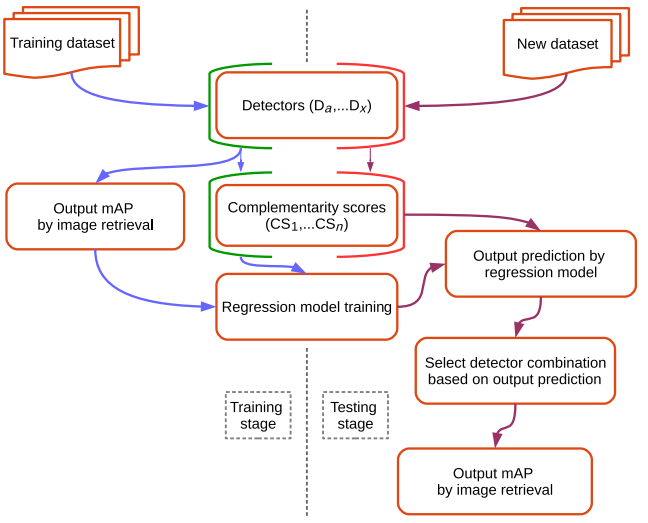

Fig. 2 Block diagram of the proposed image retrieval framework based on complementarity and regression model.

3.2.1 *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 (Kp) 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 which is able to use several descriptors jointly, such as in Ref. 17. We obtain C_x^2 mAPs.
3. Finally, the relationship between the complementarity scores and the retrieval output (mAP) is learned by a linear regression model according to Eq. 7. Coefficients of determination, i.e., adjusted R^2, which measures the explanatory power of regression model with multiple predictors, is calculated. Adjusted R^2 is effective in overcoming the overfitting issue of a model with multiple inputs, and it analyzes the model fitness and determines the best model for prediction for the given inputs and output.

3.2.2 *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 from each image of the new dataset. The three complementarity scores of the detector pairs are computed, similar to step 1 in Sec. 3.2.1.
2. For each detector combination $(D_i, D_j)_{i \neq j}$, we predict the mAP, called mAP^p, using a previously trained regression model. The complementarity scores of each detector pair are the inputs for the regression model. The outputs, mAP^p, are predicted using the model parameters and the inputs.
3. The detector pair with the highest mAP^p is selected as the suitable detector pair for image retrieval on the new dataset.

These training and predictions 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, which can be easily adapted. 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 one image to another; the three complementarity scores are simple criteria that can be computed very quickly online. In the experiment in Sec. 4, we will see that the quality of image retrieval can be improved even more by considering such an image-by-image prediction. We 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 public image datasets:

1. Paris_DB: this dataset (see first row of Fig. 3) is a public benchmark[40] consisting of 6412 images collected from Flickr by searching for 12 particular Paris landmarks.
2. Oxford_DB: this public benchmark[41] consists of 5062 images collected from Flickr by searching for 11 particular Oxford landmarks (see second row of Fig. 3).
3. Holiday_DB: this dataset is a public benchmark[42] consisting of 1491 images including a large variety of scene types. Examples are shown in the third row of Fig. 3.

We have selected seven detectors from characteristically diverse categories, such as blob, corner, symmetry, and so on, Hessian affine (hesaff),[11] color symmetry (colsym),[43] MSER (mser),[44] Harris (har),[45] Star (star),[12] binary robust invariant scalable keypoints (brisk),[46] and oriented and rotated BRIEF (orb).[13] Extracted keypoints are described by three complementary local descriptors (i.e., SIFT,[14] speeded-up robust features,[15] and shape contexts[47]) and are jointly used in a query-by-example image search engine ("FII")[17] designed for descriptor combination and based on a bag of words/codebook. The extracted visual features are clustered into a vocabulary of visual words which is more robust to small versatile content. The FII search engine, which performs well compared to state of the art methods,[17] works by integrating the responses to a query into a finer subdivision of multi-index structures, which are constructed from multiple codebooks. Performances are presented with mean average precision, i.e., mAP (in the range of 0 to 1), which is computed according to the equation below:

$$\mathrm{mAP} = \frac{1}{Q}\sum_{q_i \in Q} \mathrm{AP}(q_i) = \frac{1}{Q}\sum_{q_i \in Q}\left[\frac{1}{n_r}\sum_{n=1}^{n_r} \mathrm{Pr}_{q_i}(\mathrm{Re}_n)\right], \tag{8}$$

where $\mathrm{AP}(q_i)$, i.e., average precision, measures the mean over the precision (Pr_{q_i}) after each relevant retrieval at the n'th recall (Re_n), and n_r is the number of relevant retrievals for the i'th query. Codebook size and the value of k during nearest neighbors (k-NN) retrieval are two important parameters of the FII search engine. The optimal codebook size used for Paris_DB and Oxford_DB is 1,500,000 words. For the Holiday_DB, 30% of the total description points of each detector combination are selected as the codebook size. Parameter k is varied in between 2 and 10 for an optimal combination of the nearest neighbors and detectors. The values, which are highlighted in bold font in the tables, represent the best achieved values.

4.2 *Study of the Optimal Regression Model*

In this section, we discuss the training of the regression model with different combinations of model inputs, i.e., spatial complementarity scores, and then the selection of the best suitable regression model for prediction on the test datasets.

Paris_DB is used to train the regression model. Model inputs, the complementarity scores, i.e., distribution, contribution, cluster, and number of keypoints (Kp), of detector pairs are computed for the images of Paris_DB. The mAP is calculated using the FII search engine on Paris_DB. We trained our model with different combinations of the inputs and calculated adjusted R^2 to determine the best-fitted model. In Table 1, we present three different configurations and corresponding adjusted R^2. The highest value of adjusted R^2 is achieved with "Kp-distribution-contribution-cluster—mAP" model, and therefore this configuration will be used for further prediction on the test datasets.

However, to justify the selection of a particular configuration, we use Paris_DB for prediction with both models, i.e., Kp-distribution-contribution-cluster—mAP and "distribution-contribution—mAP." The prediction steps are explained in Sec. 3.2.2. The predicted mAP (mAP^p) is shown in Tables 2 and 3 correspondingly. The best performing detector pair is "mser-star" according to distribution-contribution—mAP, and "hesaff-mser" with Kp-distribution-contribution-cluster—mAP. When considering effective retrieval with the FII search engine, the best performing detector pair is hesaff-mser (with mAP = 0.593) compared to mser-star with mAP of 0.564. Hence, the best-fitted model, Kp-distribution-contribution-cluster—mAP, is selected for prediction on the test datasets and for the following experiments.

4.3 *Global Prediction of the Detectors Combination Performance*

In this section, we present the prediction results of detector combinations using the linear regression model.

Fig. 3 Samples from the three benchmarks used in our experiments: (a) row for Paris_DB, (b) row for Oxford_DB, and (c) row for Holiday_DB.

Table 1 Adjusted R^2 value calculation for the regression model with different combinations of the complementarity scores.

Dataset	Model inputs	Adjusted R^2 value
Paris_DB	Distribution-contribution	0.123
	Distribution-cluster	0.126
	Kp-distribution-contribution-cluster	**0.166**

The model is trained with Paris_DB as described in Sec. 3.2.1. Kp-distribution-contribution-cluster—mAP model is used for the prediction experiments on test datasets, i.e., Oxford_DB and Holiday_DB based on the adjusted R^2 score. For prediction on the test datasets, the procedure in Sec. 3.2.2 is applied by computing complementarity scores of the detector pairs. The predictions of the detector pairs are presented in Table 4, with the associated mAP^p. The regression errors obtained for Oxford_DB and Holiday_DB are 3.9% and 5.6%, respectively. We also generated a set of random data of a size similar to the one of Paris_DB to qualify the regression error with the test datasets. The regression error for random data is 19.3%, which confirms that the errors obtained on the test datasets are acceptable. We observe that detector pairs "hesaff-har" for Oxford_DB and hesaff-mser for Holiday_DB, are associated with the best predicted mAP, which was trained with Paris_DB. Thus, we consider them as the best combinations for the image retrieval on these datasets.

4.4 *Effective Performance for Image Retrieval*

In this section, the image retrieval results, i.e., mAP, using FII for the Paris_DB and the test datasets Oxford_DB and Holiday_DB, are presented. Here, mAP retrieval results with FII are denoted by effective mAP (mAP^e). In Table 5, we present results associated with the two best predicted pairs and the one worst predicted pair. For Oxford_DB, which was trained with Paris_DB, the best effective result should be obtained with the hesaff-har pair (see Table 4). Indeed, the highest mAP^e is achieved with this combination (see Table 5). Also, a mAP^e of 0.269 is achieved with "colsym-orb," which is the worst performing pair. For Holiday_DB, although the mAP^e is not in the same range as the predicted mAP, the sorted sequence of the predicted mAP is confirmed as a relevant prediction, and the highest mAP^e is obtained with the hesaff-mser combination. This first set of experiments confirms that the spatial complementarity scores employed with the linear regression model are able to correctly estimate the performance of a detector pair for image

Table 2 Detector combinations and mAP^p using Kp-distribution-contribution-cluster—mAP model.

	Detector pair	mAP^p	Detector pair	mAP^p	Detector pair	mAP^p
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
	har-star	0.481	har-orb	0.458	har-brisk	0.4795
	star-orb	0.456	star-brisk	0.481	orb-brisk	0.467

Table 3 Detector combinations and mAP^p using distribution-contribution—mAP model.

	Detector pair	mAP^p	Detector pair	mAP^p	Detector pair	mAP^p
Paris_DB	hesaff-colsym	0.500	hesaff-mser	0.517	hesaff-har	0.490
	hesaff-star	0.520	hesaff-orb	0.461	hesaff-brisk	0.486
	colsym-mser	0.521	colsym-har	0.496	colsym-star	0.538
	colsym-orb	0.418	colsym-brisk	0.453	mser-har	0.513
	mser-star	**0.542**	mser-orb	0.453	mser-brisk	0.519
	har-star	0.524	har-orb	0.468	har-brisk	0.513
	star-orb	0.482	star-brisk	0.533	orb-brisk	0.442

Table 4 Detector combinations and mAP^p using Kp-distribution-contribution-cluster—mAP model for test datasets.

	Detector pair	mAP^p	Detector pair	mAP^p	Detector pair	mAP^p
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
	har-star	0.575	har-orb	0.561	har-brisk	0.459
	star-orb	0.441	star-brisk	0.504	orb-brisk	0.492
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
	har-star	0.409	har-orb	0.420	har-brisk	0.3815
	star-orb	0.389	star-brisk	0.400	orb-brisk	0.424

Table 5 Effective mAP (mAP^e) of detector pair using the FII search engine for the different datasets.

	Detector pair	k-NN	mAP^e		Detector pair	k-NN	mAP^e
Paris_DB	hesaff-mser	2	**0.589**	Oxford_DB	hesaff-har	2	**0.549**
	hesaff-star	2	0.570		mser-har	2	0.456
	har-colsym	2	0.371		colsym-orb	2	0.269
Holiday_DB	hesaff-mser	2	**0.683**				
	hesaff-star	2	0.666				
	colsym-orb	2	0.499				

retrieval and then to enable the use of the best detector pair for a dataset.

In Table 6, the retrieval results with single detectors are presented to compare with detector pair results of Table 5. These results demonstrate again the relevance of the use of several complementary detectors in the representation of the content.

4.5 *Effect of k-Nearest Neighbors Parameter on Retrieval and Its Prediction*

The k-NN retrieval of the nearest neighbors is a central step in any image search engine. Here, the nearest neighbors search concerns the retrieval of the k points similar to the query point. The optimal value of k is not easy to determine because it concerns the retrieval of similar points, and this value cannot be predetermined intuitively or learned easily. In general, the problem is addressed by fixing the k value for the whole dataset after having tested the retrieval performances with different values. In this section, we present retrieval results in Table 7 by varying k ($k = 2, 5, 10$) and observe the consequence on mAP^e.

The best mAP^e is globally obtained with $k = 2$ for all datasets. The accuracy difference is 1.8% between $k = 2$ and $k = 5$ and 5.8% between $k = 2$ and $k = 10$ for hesaff-mser in Paris_DB, while it is 1.6% between $k = 2$ and $k = 10$ for hesaff-har with Oxford_DB. In the Holiday_DB, the accuracy difference is 1.3% between $k = 2$ and $k = 10$ for hesaff-mser. During the search for the nearest neighbors of the query point, higher values of k might include dissimilar

Table 6 Effective mAP of single detector using the FII search engine for the different datasets.

Paris_DB			Oxford_DB			Holiday_DB		
Detector	k-NN	mAP^e	Detector	k-NN	mAP^e	Detector	k-NN	mAP^e
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 7 Effective mAP for Paris_DB, Oxford_DB, and Holiday_DB by varying k-NN ($k = 2, 5, 10$) and adapting it by exploiting the prediction model.

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

neighbors in the k-NN lists. Nevertheless, we observed through dedicated experiments that the optimal value of k may be different from one query to another. By using our model, it is possible to adapt the best value of k for each query instead of estimating it globally for the dataset. The procedure in Sec. 3.2 is applied by varying k ($k = 2, 5, 10$) and the prediction mAP obtained allows to adapt k to each query. In Table 7, the mAP in bold font corresponds to the effective mAP^e 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, respectively, compared to the previous best ones with $k = 2$. It is interesting to observe in Fig. 4 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, ~47% of the queries are executed with $k = 2$, 44% with $k = 5$, and only 9% with $k = 10$; this result particularly highlights the importance of the automatic prediction of k for each query. Similarly, for Holiday_DB, 90.2% is selected with $k = 2$, 5.2% with $k = 5$, and 4.6% with $k = 10$.

4.6 *Image-by-Image Prediction of the Best Detector Combination*

In this section, we refine the results obtained in Secs. 4.3, 4.4, and 4.5 by adapting the selection of the best detector combination to each image and by applying the prediction strategy of Sec. 3.2.2 to each query image. Six different combinations of mAP^e obtained with two best detector pairs and three k-NN values ($k = 2, 5, 10$) are consolidated. In Table 8, we observe that mAP^e is increased by 0.8%, 2.5%, and 21.1% compared to the previous best with $k = 2$ for Paris_DB, Oxford_DB, and Holiday_DB, respectively. For Holiday_DB, the achieved retrieval accuracy is 0.894, which is one of the best in the state of the art to our knowledge, compared to Ref. 48 which is also based on bag of words. As depicted in Fig. 5, the majority of the selections are done with $k = 2$. For Paris_DB, 90.9% is selected for $k = 2$ for both pairs of detectors, while 5.45% is with $k = 5$. Most of the selections (85%) are done with hesaff-har for Oxford_DB, while 15% are from "mser-har." A similar trend is observed with Holiday_DB, where 67% is selected from hesaff-mser. Also for Holiday_DB, 94% is

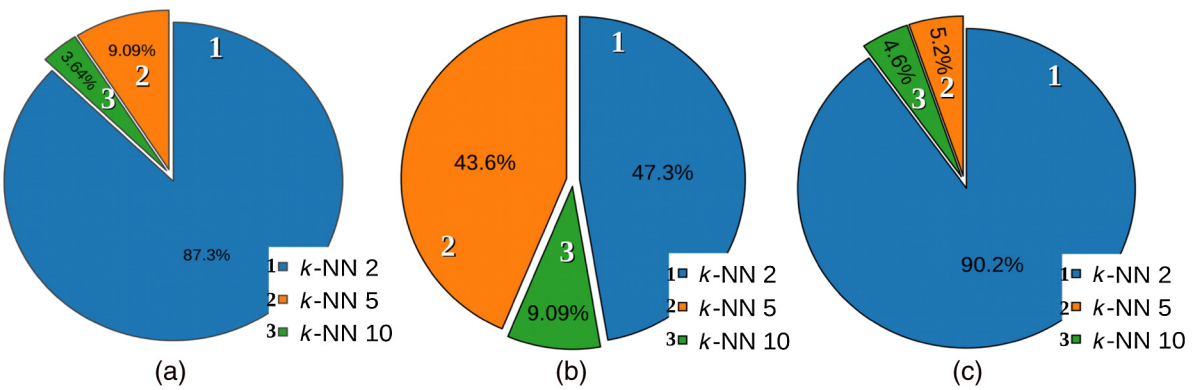

Fig. 4 Distribution of predicted k values across the queries for all datasets, (a) hesaff-mser for Paris_DB, (b) hesaff-har for Oxford_DB, and (c) hesaff-mser for Holiday_DB.

Table 8 Effective mAP obtained for all the datasets, by selecting optimal detector pairs and optimal value k for each query image.

Dataset	Detector pair	mAP^e $k=2$	mAP^e $k=5$	mAP^e $k=10$
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, and 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, and 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, and 10)	

selected from $k=2$ of both the detector pairs, 3.4% with $k=5$, and the remaining 2.6% from $k=10$.

Even if the statistical analysis (Figs. 4 and 5) has highlighted the dominance of some particular detector pairs and values of k, we observe that using others adaptively allows one to notably refine the results. More generally, these experiments also clearly highlight the impact of the spatial complementarity of the selected features on the retrieval performance.

5 Exploration of a Digitized Photographic Museum Collection

The proposed approach of image retrieval (i.e., the FII search engine enriched with the adaptive selection model) is currently applied to the exploration of the collection of a French museum, the Nicéphore Niépce museum[49], within the scope of the French project POEME[50]. Since its creation and thanks to the influx of donations and an active acquisition policy, the Niépce museum has had the aim of telling the whole photography history in its technical and artistic aspects as seen in its popular and commercial uses. Dedicated to photography, it proposes to explain all the aspects of the practice from its emergence in the 19th century to its current developments, and then covers all fields of photography. The digitization of its collections, which began in 1999 and goes on with several hundreds of thousands of digitized items, has enabled the museum to develop interactive multimedia devices for the general public which are presented in its showrooms and to make available some of those contents via virtual exhibitions on Internet. Some examples of the digitized contents of the museum are shown in Fig. 1.

Section 5.1 is dedicated to the evaluation of the proposal for the Niépce collection by fitting the main parameters similarly as in Sec. 4. Section 5.2 illustrates its application to the needs of the professionals of the museum and the promotion of their collections outside the museum.

5.1 *Evaluation of the Proposal for the Niépce Collection*

The same detectors, descriptors, and other parameters (such as k-NN values) as those used in Sec. 4.1 are used for these experiments. The results for the prediction of the best suitable detector combinations are presented in Table 9; combination "hesaff-orb" is associated with the highest predicted mAP, followed by hesaff-har. Therefore, these combinations are used in the following experiments.

Based on the results of Table 9, the effective retrieval results, i.e., mAP^e for those detector combinations and varying k-NN ($k=2$, 5, 10) are consolidated and presented in Table 10. The achieved mAP^e with adaptive selection is 0.651 (bold in the table), which is an accuracy increment of 3.9% compared to the previous best obtained with hesaff-orb and $k=2$. To understand this more precisely, Fig. 6 shows how the selections are distributed across the queries. Most of the selections, ~50%, are with $k=2$ from

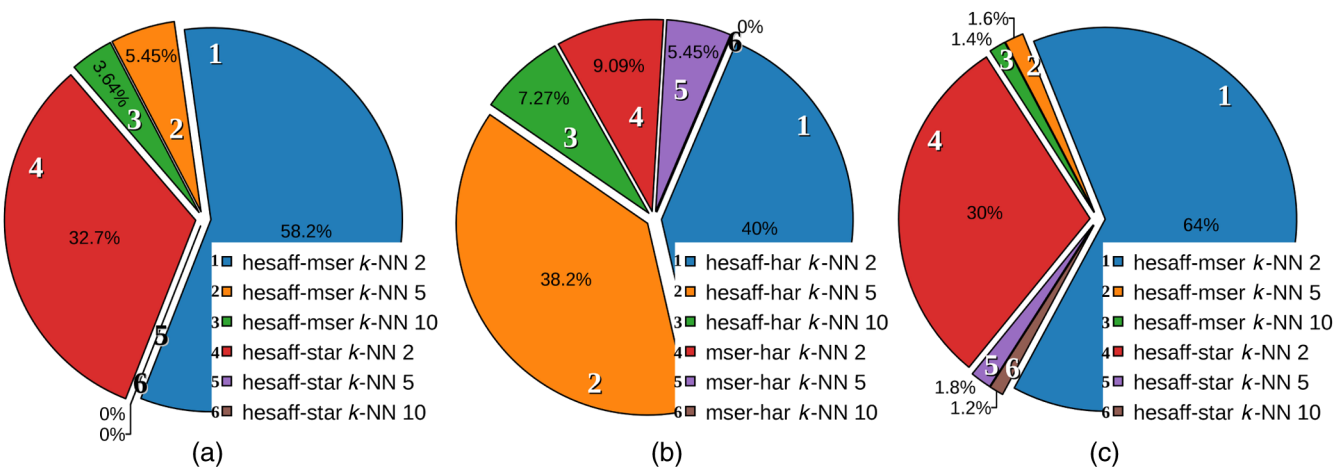

Fig. 5 Distribution of predicted values of k and detectors pairs across the queries: (a) hesaff-mser and "hesaff-star" for Paris_DB, (b) hesaff-har and mser-har for Oxford_DB, (c) hesaff-mser and hesaff-star for Holiday_DB.

Table 9 Different detector combinations and predicted mAP using the regression model, for the Niépce collection.

Detector pair	mAP^p	Detector pair	mAP^p	Detector pair	mAP^p
hesaff-colsym	0.461	hesaff-mser	0.431	hesaff-har	0.478
hesaff-star	0.468	hesaff-orb	**0.482**	hesaff-brisk	0.476
colsym-mser	0.419	colsym-har	0.439	colsym-star	0.435
colsym-orb	0.440	colsym-brisk	0.441	mser-har	0.420
mser-star	0.421	mser-orb	0.430	mser-brisk	0.429
har-star	0.452	har-orb	0.460	har-brisk	0.442
star-orb	0.450	star-brisk	0.433	orb-brisk	0.431

Table 10 Effective mAP obtained by selecting optimal detector pairs and optimal value k for each query image, for the Niépce collection.

	mAP^e		
Detector pair	$k = 2$	$k = 5$	$k = 10$
hesaff-orb	0.612	0.602	0.591
hesaff-har	0.609	0.584	0.550
Adaptive detector combination		**0.651**	
		(Adaptive k 2,5, and 10)	

both combinations. In addition, 60.5% is selected with the best performing pair, hesaff-orb, and the remaining 39.5% with hesaff-har. Again, the varied distribution of the optimal configurations selected justifies the relevance of the proposal. In the Appendix, we present and discuss additional and secondary experimental results on the Niépce collection.

5.2 *Examples of Retrieval and Applications*

In collaboration with the Nicéphore Niépce museum, the work presented in this paper is currently evaluated for several scenarios at different levels of museum needs, from archiving up to the exhibition, with the objectives of providing tools for the experts and to better highlight their photographic collections for the general public; Secs. 5.2.1 and 5.2.2 illustrate these objectives.

5.2.1 *Exploration of the Niépce collection*

The first image retrieval results, searching for a particular content in the Niépce collection, are depicted in Fig. 7. Here, the dominant configuration exhibited by the selection model is hesaff-orb as the detector combination and k-NN = 2.

The example (query image-1) concerns a large collection of scans with similar layouts. Here, the retrieved images are related to bridge and port constructions, and they exhibit the different construction stages of the same barrage and the other similar bridges or ports, and then allow to focus on and isolate a thematic subset of the collection. Being able to automatically link these contents helps the archivist in indexing

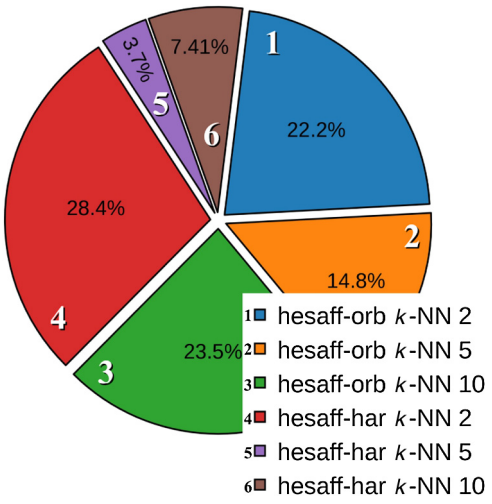

Fig. 6 Distribution of k and detector pairs across the queries for the Niépce collection.

the collections as they are digitized, and may help the curator and commissioner in the selection of interesting contents for an exhibition as a complement to traditional approaches of selection usually based on the experience, memory, and archivist indexing. It can also provide a new way of browsing the collection for a visitor, by querying it with particular photographs or parts of photographs (the local descriptors employed enable partial queries). The Niépce museum is currently investigating the design and creation of an immersive and interactive environment *in situ* which integrates the present CBIR search engine, with the ambition of proposing virtual exhibitions centered on the visitor who will have the possibility to organize the navigation by querying the collections through several modalities himself.

Another scenario is related to the variety of the sources of the Nicéphore Niépce museum, which keeps complete archives of photographers, such as negatives, contact sheets, and prints. Thus, different formats of the same image could exist on several physical media. The classification of these archives is often lost before arriving at the museum. The first task of a museum when it received a photographer's collection is to reorganize the collection and match each part of the archives to each other. For example, Jean Moral[51] was mainly a fashion photographer in the 1930s. His collection includes thousands of photo prints, images of photo negatives, and contact sheets of small print of 6×6 cm^2. Jean Moral published hundreds of images in fashion magazine Harper's Bazaar[52] (HB). Two-thirds of Jean Moral's collection is related to fashion and the images were never dated and captioned. To insert the date and caption the models, archivists are compelled to compare each image with each page of HB magazine where an image of Moral was published. Therefore, for an archivist, the magazines' exploration takes several hours for matching between the publication on the magazines and the photographer's collections. However, this tedious and rigorous manual work can be simplified and automated by introducing our image retrieval framework.

Two examples of the matching between photographer's collections, which were exploited in the HB magazines, are depicted in Figs. 8 and 9. The several hundreds of scanned pages of HB magazines are used as a dataset and Jean Moral's photograph's collections, i.e., photo print, and contact sheets

Fig. 7 Image retrieval example by querying the Niépce collection: for a query, the six best results are presented in decreasing order of similarity, from left to right.

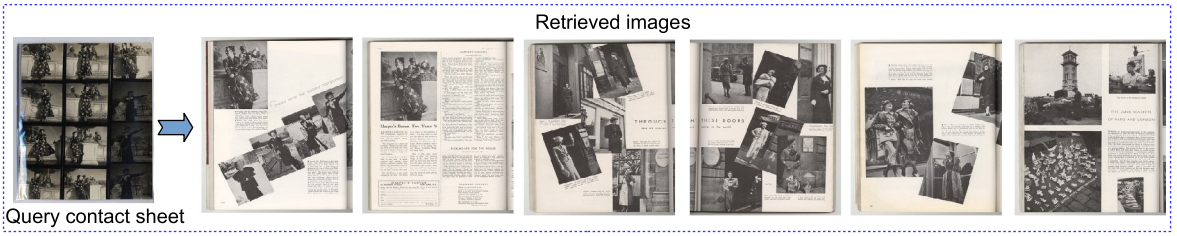

Fig. 8 Image retrieval example by querying in the Niépce HB collection: photo contact sheet of Jean Moral's collection is used as query. The six best results are presented in decreasing order of similarity, from left to right.

Fig. 9 Image retrieval example by querying in the Niépce HB collection: photo print of Jean Moral's collection is used as query. The six best results are presented in decreasing order of similarity, from left to right.

are queried in this dataset. In Fig. 8, the same photo from the contact sheet was identified in two different issues of HB, i.e., March 1935 and May 1935 issues, retrieved at the two first positions, respectively. Jean Moral's photo prints were queried in the HB collection, as depicted with the example of Fig. 9. Interestingly, here we discovered that the first two retrieved images do not correspond to the same item, but to very similar shooting contexts (they were published in the September 1937 HB issue), thus providing an additional insight on the photograph's collection to the archivist.

5.2.2 *Cross-domain image retrieval*

With the growing acquisition of contents in various professional and general public domains, cross-domain image

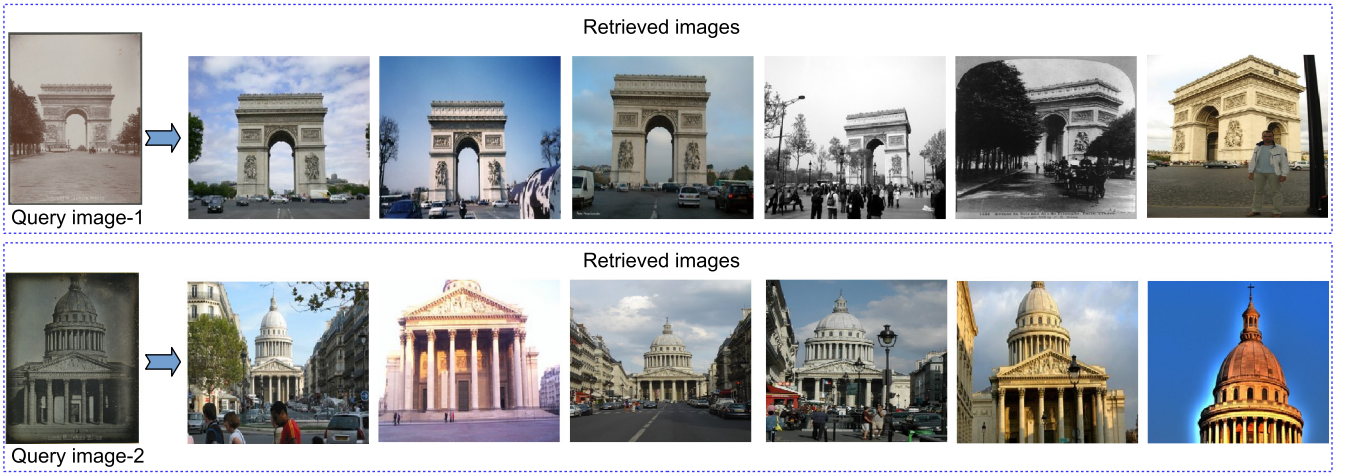

Fig. 10 Illustration of cross-domain image retrieval: the query images are taken from the Niépce collection while the dataset is Paris_DB (Flickr). For each query, the six best results are presented in decreasing order of similarity, from left to right.

retrieval is a topical subject that questions the problem of comparing, indexing, and searching for contents potentially acquired by different sources or modalities (different cameras, paintings, sketches, and son) and at different times. Here, we illustrate this problem with photographs of the Niépce collection and street-view imagery (e.g., Flickr, Google street-view). Figure 10 shows two examples of retrieval within the public Paris_DB dataset, with digitized old pictures of monuments from the Niépce collection as queries.

Although the query images were taken many years earlier with different technologies and different surroundings, we are able to retrieve images that represent the same monument or geographical area. Such cross-domain application is of great interest. By considering a georeferenced dataset (acquired with a mobile mapping system, for instance, the one of the French Mapping Agency[53]) such historical or cultural contents can be precisely relocalized such as in Ref. 54. Moreover, cross-domain linking opens the door to their promotion outside the museum by connecting them with official mapping databases and services, such as the French Géoportail[55] or a 3-D web mapping engine (e.g., itowns[56]), along the same lines as the Linked Data initiative, which intends to connect distributed data across the web.

6 Conclusions

The main contribution of the present work is the proposal of a regression model based on several spatial complementarity criteria among local image features, in order to estimate the optimal combination of detectors for the description of a given image within the scope of query-by-example image retrieval. The proposed model also allows to optimally fix some other parameters, such as the k in k-NN retrieval. Experiments and evaluations were performed on three public datasets.

Even if the statistical analyses presented highlight the dominance of some particular detector combinations (and values of k), we observed that adaptively using other ones —for some images—allows to refine the results favorably and notably. 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 the retrieval accuracy even more. The experiments were performed with couples of detectors, but this framework can easily be extended to the evaluation of the complementarity between multiple detectors by generalizing the complementarity criteria.

Finally, the proposed image search engine was experimented with the cultural photographic collections of the French museum Nicéphore Niépce, where it has demonstrated its potential for the exploration and promotion of these contents at different levels from their archiving up to their exhibition in the museum and outside, and their linking with other categories of contents, such as geographical mapping contents.

Appendix: Additional Experimental Results of the Niépce Collection

In this section, we present the Niépce collection experimental results, which are not included in Sec. 5.1. According to Table 9, the best two performing detector combinations are hesaff-orb and hesaff-har. Indeed, the best effective mAP (mAP^e) is achieved with the hesaff-orb combination (see Table 11). In accordance with the prediction result, the second best mAP^e is achieved with hesaff-har; similarly, the worst performing combination is obtained with "colsym-mser."

To compare, Table 12 provides results by using these detectors alone. Detector combinations produce superior results compared to a single detector. Again, these experiments on the museum image collections demonstrate the advantage of using detector combinations and the regression framework proposed.

Table 13 presents the retrieval results by varying k value ($k = 2$, 5, 10) and observing the change on mAP^e. We observe that the accuracy is significantly increased by exploiting the adaptive selection of k for each query image; the mAP^e (bold in the table) is increased by 1.1% and 2.1% compared to $k = 2$ and $k = 5$.

Table 11 mAP^e of detector combinations, for the Niépce collection.

Detector pair	k-NN	mAP^e
hesaff-orb	2	**0.612**
hesaff-har	2	0.609
colsym-mser	2	0.401

Table 12 mAP^e of single detectors, for the Niépce collection.

Detector	k-NN	mAP^e
hesaff	2	0.593
orb	2	0.467
colsym	2	0.399
mser	2	0.384

Table 13 mAP^e for the Niépce collection, by varying k-NN ($k = 2$, 5, 10) and adapting it with the prediction model.

	mAP^e		
Detector pair	$k = 2$	$k = 5$	$k = 10$
hesaff-orb	0.612	0.602	0.591
		0.623	
		(adaptive k 2,5, and 10)	

Acknowledgments

The authors are grateful to Nicéphore Cité, Institut National de l'Information Géographique et Forestière and French Project POEME ANR-12-CORD-0031 for the financial support.

References

1. C. Tsai, "A review of image retrieval methods for digital cultural heritage resources," *Online Inf. Rev.* **31**(2), 185–198 (2007).
2. S.-Q. Jiang et al., "Visual ontology construction for digitized art image retrieval," *J. Comput. Sci. Technol.* **20**(6), 855–860 (2005).
3. S. Vrochidis et al., *A Hybrid Ontology and Visual-based Retrieval Model for Cultural Heritage Multimedia Collections*, pp. 1–10, Springer US, Boston, Massachusetts (2009).
4. S. H. Yen et al., "A content-based painting image retrieval system based on adaboost algorithm," in *2006 IEEE Int. Conf. on Systems, Man and Cybernetics*, Vol. 3, pp. 2407–2412 (2006).
5. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997).
6. S. Goodall et al., *SCULPTEUR: Multimedia Retrieval for Museums*, pp. 638–646, Springer, Berlin, Heidelberg (2004).
7. F. A. Mahdi and A. Ibadi, "MIRS: museum image retrieval system using most appropriate low-level feature descriptors," *Int. J. Comput. Sci. Issues* **11**(5), 1 (2014).
8. A. Shrivastava et al., "Data-driven visual similarity for cross-domain image matching," *ACM Trans. Graphics* **30**, 1 (2011).
9. B. C. Russell et al., "Automatic alignment of paintings and photographs depicting a 3D scene," in *2011 IEEE Int. Conf. on Computer Vision Workshops (ICCV)*, pp. 545–552 (2011).
10. M. Aubry, B. C. Russell, and J. Sivic, "Painting-to-3D model alignment via discriminative visual elements," *ACM Trans. Graphics* **33**(2), 14 (2014).
11. K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vision* **60**(1), 63–86 (2004).
12. M. Agrawal, K. Konolige, and M. Blas, "Censure: center surround extremas for realtime feature detection and matching," *Lect. Notes Comput. Sci.* **5305**, 102–115 (2008).
13. E. Rublee et al., "ORB: an efficient alternative to sift or surf," in *2011 IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 2564–2571 (2011).
14. D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
15. H. Bay et al., "Speeded-up robust features (SURF)," *Comput. Vision Image Understand.* **110**, 346–359 (2008).
16. T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Inf. Retrieval* **11**(2), 77–107 (2008).
17. N. Bhowmik et al., "Efficient fusion of multidimensional descriptors for image retrieval," in *IEEE Int. Conf. on Image Processing (ICIP 2014)*, pp. 5766–5770 (2014).
18. N. Neshov, "Comparison on late fusion methods of low level features for content based image retrieval," *Lect. Notes Comput. Sci.* **8131**, 619–627 (2013).
19. T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Inf. Retrieval* **11**(2), 77–107 (2008).
20. P. K. Atrey et al., "Multimodal fusion for multimedia analysis: a survey," *Multimedia Syst.* **16**, 345–379 (2010).
21. C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early versus late fusion in semantic video analysis," in *Proc. of the 13th Annual ACM Int. Conf. on Multimedia*, pp. 399–402, ACM, New York (2005).
22. J. Yu et al., "Feature integration analysis of bag-of-features model for image retrieval," *Neurocomputing* **120**, 355–364 (2013).
23. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, Vol. 1, pp. 886–893 (2005).
24. T. Ojala, M. Pietikäinen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002).
25. R. da Torres et al., "A genetic programming framework for content-based image retrieval," *Pattern Recognit.* **42**(2), 283–292 (2009).
26. J. Yue et al., "Content-based image retrieval using color and texture fused features," *Math. Comput. Modell.* **54**(3–4), 1121–1127 (2011).
27. C. Ferreira et al., "Relevance feedback based on genetic programming for image retrieval," *Pattern Recognit. Lett.* **32**(1), 27–37 (2011).
28. M. Wacht, J. Shan, and X. Qi, "A short-term and long-term learning approach for content-based image retrieval," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2006) Proceedings*, Vol. 2 (2006).
29. S. A. Chatzichristofis, A. Arampatzis, and Y. S. Boutalis, "Investigating the behavior of compact composite descriptors in early fusion, late fusion, and distributed image retrieval," *Radioengineering* **19**(4), 726–734 (2010).
30. W. Zhang, Z. Qin, and T. Wan, "Image scene categorization using multi-bag-of-features," in *Proc. Int. Conf. on Machine Learning and Cybernetics*, Vol. 4, pp. 1804–1808 (2011).
31. V. Risojevic and Z. Babic, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.* **10**(4), 836–840 (2013).
32. T. Durand et al., "Semantic pooling for image categorization using multiple kernel learning," in *IEEE Int. Conf. on Image Processing (ICIP 2014)*, pp. 170–174 (2014).
33. Y. Cao et al., "Matching image with multiple local features," in *Proc. of 20th Int. Conf. on Pattern Recognition*, pp. 519–522 (2010).
34. E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems," *Knowl. Based Syst.* **39**, 85–94 (2013).
35. Y. Zhou et al., "Augmented feature fusion for image retrieval system," in *The Indian Council of Medical Research (ICMR 2015)*, pp. 447–450, ACM, New York, New York (2015).
36. J. Sun, "Local selection of features for image search and annotation," in *Association for Computing Machinery (ACM, MM 2014)*, pp. 655–658, ACM, New York, New York (2014).
37. S. Ehsan, A. F. Clark, and K. D. McDonald-Maier, "Rapid online analysis of local feature detectors and their complementarity," *Sensors* **13**(8), 10876–10907 (2013).
38. G. Gales, A. Crouzil, and S. Chambon, "Complementarity of feature point detectors," in *12th Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, pp. 334–339, INSTICC Press (2010).
39. K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Tenth IEEE Int. Conf. on Computer Vision (ICCV 2005)*, Vol. 2, pp. 1792–1799 (2005).
40. Paris_DB, http://www.robots.ox.ac.uk/~vgg/data/parisbuildings/.
41. Oxford_DB, http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/.
42. Holiday_DB, http://lear.inrialpes.fr/people/jegou/data.php.
43. G. Heidemann, "Focus-of-attention from local color symmetries," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 817–830 (2004).
44. J. Matas et al., "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Machine Vision Conf.*, pp. 1–10 (2002).
45. C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 530–535 (1997).
46. S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: binary robust invariant scalable keypoints," in *IEEE Int. Conf. on Computer Vision (ICCV 2011)*, pp. 2548–2555 (2011).
47. S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 509–522 (2002).
48. X. Li, M. Larson, and A. Hanjalic, "Pairwise geometric matching for large-scale object retrieval," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2015)*, pp. 5153–5161 (2015).
49. Nicéphore Niépce museum, http://en.museeniepce.com.
50. POEME Projet website, www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2%5BCODE %5D=anr-12-cord-0031.
51. Jean Moral, https://fr.wikipedia.org/wiki/Jean_Moral.
52. Harper's Bazaar, http://www.harpersbazaar.com/.
53. N. Paparoditis et al., "Stereopolis II: a multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology," *Revue Franaise de Photogrammtrie et de Tldtection* **200**(1), 69–79 (2012).
54. Y. Song et al., "6-DOF image localization from massive geo-tagged reference images," *IEEE Trans. Multimedia* **18**, 1542–1554 (2016).
55. Géoportail, https://en.wikipedia.org/wiki/Geoportail.
56. Q. D. Nguyen et al., "A 3D heterogeneous interactive web mapping application," in *IEEE Virtual Reality Conf.*, Arles, France, pp. 838–840 (2015).

Neelanjan Bhowmik is a PhD student at Université Paris-Est, MATIS laboratory of the French Mapping Agency (IGN), France. He received BE degree in electronics engineering from Rashtrasant Tukadoji Maharaj Nagpur University, India in 2008, and MSc degree in computer vision engineering from the University of Sheffield, UK in 2012. From 2008 to 2011, he worked as a technical associate at Tech Mahindra Limited, India. His current research interests include computer vision, image retrieval, and embedded systems.

Valérie Gouet-Brunet has been a senior researcher in the MATIS group of LaSTIG Lab. at the French Mapping Agency (IGN) since 2012. Since 2014, she has been the head of the MATIS. Formerly, she was an assistant professor at the Conservatoire National des Arts et Metiers, Paris, France. She obtained her PhD in computer science in 2000 and habilitation to direct researches in 2008. Her area of research concerns content-based image retrieval at large scale.

Gabriel Bloch has been a deputy director at Nicéphore Cité, France, since 2009. He was in-charge of technology consultant and R&D at Nicéphore Cité. During 2004 to 2007, he created a company specializing in 3-D computer graphics. After obtaining his master's degree in 2003 at the Engineering School Ecole Nationale Supérieure d'Arts et Métiers, France, he obtained a diploma specialized in virtual reality at the Ecole Nationale Suprieure d'Arts et Métiers Research Image Institute, France.

Sylvain Besson has been in charge of Musée Nicéphore Niépce, France and of its Inventory Documentation Department since 2007. He graduated in museology from École du Louvre, France, in 2001, and has been in charge of the computerization of the museum collections since 2003. He is also the coordinator of all museum projects related to new technologies and of their use in the field of iconographic research, of virtual exhibition and of putting online.