EFFICIENT FUSION OF MULTIDIMENSIONAL DESCRIPTORS FOR IMAGE RETRIEVAL

Neelanjan Bhowmik[1,2]*, Ricardo González V.*[1,3]*, Valérie Gouet-Brunet*[1]*, Hélio Pedrini*[3]*, Gabriel Bloch*[2]

[1]University Paris-Est, IGN/SR, MATIS lab, 73 avenue de Paris, 94160 Saint-Mandé, France
[2]Nicéphore Cité, 34 quai Saint-Cosme, 71100 Chalon-sur-Saône, France
[3]Institute of Computing, University of Campinas, Campinas, Brazil, 13083-852

ABSTRACT

Due to the large diversity of existing feature descriptors in content-based image retrieval, the image contents can be better represented by the joint use of several descriptors in order to explore their potentially complementary characteristics. This paper presents and discusses a strategy for fusion of the different multidimensional features involved, based on inverted multi-indices and dedicated to similarity search. Image descriptors are quantized separately and efficiently through dimension reduction techniques, before being combined in the inverted multi-indices. To exhibit its effectiveness, the proposal is evaluated on two datasets having different contents and sizes, facing several state-of-the-art approaches of image descriptor fusion. The obtained results reconfirm that the joint use of several descriptions improves similarity search, and show that our fusion proposal outperforms other solutions, while manipulating lower or similar volumes of features.

Index Terms— CBIR, image descriptors, fusion, dimensionality reduction, inverted index.

1. INTRODUCTION

With the hasty growing of the media collection, it is imperative to develop effectual search processes, such as Content-Based Image Retrieval (CBIR) to access voluminous, complex and unstructured data efficiently. CBIR is used increasingly in order to organize, search and share the image and video collections [1]. The extraction of distinguished features from an image dataset and the measure of the resemblance between them is the core of CBIR. Hence, the key emphasis is on describing suitable image characteristics, which should coincide with the users vision and perception of similarity of the images [2] (*i.e.*, to account for the gap between low and high-level semantic concepts).

Nowadays, literature on image descriptors is very rich [1, 3, 4], providing several families to describe different image characteristics for different targets. It has already been demonstrated that combining different descriptions is propitious to better describe image contents. Several fusion approaches exist (see Section 2), therefore, it is pertinent to investigate the best strategy for combining image characteristics. Among all the existing solutions for describing image contents and organizing the extracted features in order to deal with large dataset, we focus on a recent approach called inverted multi-index [5]. This method suggests a data structure and a strategy to combine multidimensional features efficiently: it decomposes the image descriptor space into n desired subspaces. Then, the best responses to a query in each subspace are retrieved and combined into one response that ensures better result than the traditional approaches based on classical inverted indices.

In this work, we present a novel fusion method for efficiently combining multiple descriptors for image retrieval, based on the inverted multi-index approach, but amended it in several ways. This proposed approach allows to combine any number of multidimensional image descriptors by integrating their responses to a query in finer subdivisions.

This paper is organized as follows. Section 2 revisits the related works on fusion of descriptors for image retrieval, Section 3 describes the proposed methodology, Section 4 presents the experiments performed to evaluate our proposal, followed by conclusions in Section 5.

2. RELATED WORK

Recently, several techniques for fusion (combination) of image descriptors have been proposed in the literature. Fusion can be investigated differently according to the involved descriptors, the strategy of combination and the application targeted. In general, the existing fusion approaches can be categorized as *early* and *late* fusion approaches [6, 7], which refers to their relative position from the feature comparison or learning step in the whole processing chain.

Early fusion usually refers to the combination of the features into a single representation before comparison/learning [7]. The most widespread solution is to concatenate the feature vectors into a single vector, such as in [8] with SIFT [9], HOG [10] and LBP [11] features. Other approaches such as weight based color and shape early fusion in local color pixel classification descriptor [12] and weight based texture and color feature fusion [13] are proposed for image retrieval.

Late fusion refers to the combination, at the last stage, of the responses obtained after individual features comparison or learning [7]. When considering image retrieval, multiple ranked outputs of the multiple descriptors are aggregated to generate another concluding ranked output. This method of fusion can be implemented either score-based where it combines the different similarities or distances from the query, or ranked-based which considers the combination of the response ranks. The outputs to combine can be weighted to give more importance to particular descriptors, by fixing the weights *a priori* or, better, by learning them for a given content [14]. A comparison between the most classical late fusion approaches is discussed in [15, 16] for image retrieval. When considering classification, e.g. for image categorization, late fusion is performed slightly differently: it usually involves a weighted voting strategy from the outputs of the classifiers associated to the individual descriptors, such as in [17], which exploits a panel of BoF

The authors are grateful to Nicéphore Cité, French project POEME ANR-12-CORD-0031, FAPESP and Santander Program of International Exchange Mobility for the financial support.

representations of low-level descriptors, associated to several Support Vector Machine (SVM) classifiers. More sophisticated learning strategies (i.e. multiple kernels, boosting) simultaneously learn individual classifier and combination classifier weights [18, 19, 20]. Since they take place at different levels of learning, these approaches are sometimes categorized as *intermediate* fusion [19].

In the aforementioned approaches, all the different descriptors are exploited as the same level, however, some other methods, which could be gathered under the name *sequential* fusion, consider one descriptor as a filter before using another one on the remaining subset of images or regions in the images. For example, in [21] such a strategy was proposed for the fusion of Affine-SIFT, MSER and color moment features. Similarly, in [22], global image descriptors are first used for coarse similarity search, before exploiting more expensive local features in order to refine similarity search.

3. FUSION OF DESCRIPTORS WITH INVERTED MULTI-INDEXES

This section is dedicated to the presentation of our proposal. In Section 3.1, we revisit the approach proposed in [5] which our proposal rests on, whereas Sections 3.2 and 3.3 describe our contributions.

3.1. The inverted multi-index

The introduction of inverted multi-indices [5] opens higher sparse subdivision of the search space without affecting overall processing time compared to standard inverted indexing. Product quantization (PQ) [23] based method was proposed to improve the approximate nearest neighbor search [24]. Higher dimensional vectors are split into low dimensional subspaces of Cartesian product. These subspaces are quantized independently. The Euclidean distance between two vectors, which are epitomized by a subspace quantization indices, is computed through quantized codes. The overall process enhances the search quality by limiting the quantization noise. PQ is integrated with inverted indexing in order to avoid exhaustive search, hence it boots searching speed.

In inverted multi-indices [5], one high dimensional vector is decomposed into n smaller dimension sub-spaces; then n PQ codebooks are computed by clustering each of the n sub-space separately. It is constructed as a multi-dimensional table which contains n lists of ordered codewords from the n corresponding codebooks. A given query is split into the same sub-spaces and k nearest neighbors (k nearest codewords), from the corresponding codebook, are computed and stored in a list. These n lists of k-nn consist of codewords with associated inverted indices, are merged using multi-sequence algorithm [5] to generate final list comprises vector similar to the query. This approach is depicted in Fig. 1.

3.2. Combination of different descriptors

Our methodology is based on the method presented in Section 3.1. However, instead of decomposing one query vector into two equal sub-vectors, here the query is represented with several image descriptors, as illustrated in Fig. 1. The performed steps are:

1. A query image is represented by m descriptors (such as SIFT [9], SURF [25], SC [26]), leading to m descriptions of the content.
2. Then, m candidate lists of responses are built, where each list contains the k nearest codewords to the respective query, their respective distances and the set of associated images (we are not considering repeated image identifiers).
3. Since the distances from different lists are related to their descriptor space and characteristics, standard normalization is applied by using the maximum and minimum distances of the respective descriptor to them.
4. The candidate lists are combined through the multisequence algorithm, as proposed in [5], which returns final lists that consist of codewords and associated image ids sorted by their increasing distances from the query.
5. A voting algorithm is proposed to compute the frequency of the retrieved images. The voting algorithm generates a frequency list that consists of image ids and associated frequencies according to their occurrences. All the frequency lists are summed up to generate a final frequency list that consists of the most similar images ids retrieved for the query image.

This approach can be categorized as intermediate fusion because the candidate lists, related to closest words for each descriptor, are merged (and not the candidate lists of images, as with late fusion).

3.3. Feature dimensionality reduction

The joint use of different descriptors naturally leads to increase the volume of manipulated features, and then to slow down the computational processes. This drawback can be addressed by exploiting dimensionality reduction techniques, which decrease each multidimensional description dimension to its half or fourth part, while maintaining a good degree of accuracy, sometimes similar even higher to the one of the original description. In addition, each dimensionality reduction technique can bring some particular advantage: PCA [27] is able to remove noise from the descriptions, while PLS [28] can add distinctiveness as it takes into account class correlation. Consequently, we propose to use them to decompose the multidimensional description into smaller subspaces, instead of simply splitting it into several parts as in [5]. Indeed, we think that this alternative may conduce to establish finer subdivisions and then to determine nearest neighbors better, in addition to reduce the amount of features to manipulate.

4. EXPERIMENTS AND EVALUATION

This section presents and discusses, in Sections 4.2, 4.3 and 4.4, the experiments conducted to evaluate our contributions, after having presented the framework of evaluation used in Section 4.1.

4.1. Framework of evaluation

The experiments are conducted on two image datasets with different sizes and contents:

- COIL_DB: this dataset contains 600 synthesized images containing 100 objects with different orientations and viewpoints, from the well-known benchmark *COIL-100*[1], synthetically inserted on photographs as background (images with heterogeneous and complex contents downloaded from Internet). Examples are shown in first row of Fig. 2.
- Paris_DB: it is a public benchmark[2] consisting of 6412 images collected from Flickr by searching for 12 particular Paris landmarks; see examples in second row of Fig. 2.

The image descriptors employed in our experiments are local descriptors, suitable to retrieve objects in such datasets with cluttered

[1] http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php.
[2] http://www.robots.ox.ac.uk/∼vgg/data/parisbuildings/index.html.

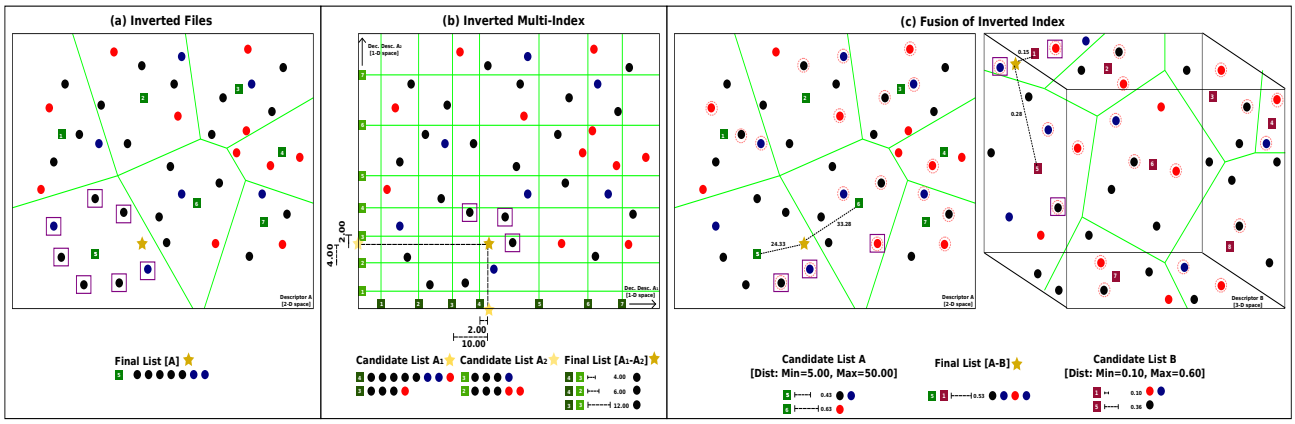

Fig. 1: Illustration of three different strategies for similarity search. For each one, three multidimensional words (colored circles) are distributed in the descriptor space. An image content, rich in interest points, can overpopulate some clusters, therefore, those clusters (represented by green numbered squares) have a strong impact on that image representation. In order to perform the task of finding the 3-NN for a given query point (yellow star), the three strategies proceed as follows. (a) Classical inverted file identifies the cluster to which the query belongs and retrieves all its associated descriptions inside. (b) Inverted multi-index subdivides the descriptor space into n subspaces ($n = 2$ here). Then the multi-sequence algorithm combines the nearest centroid to the query in each of the n subspaces, selecting the descriptions related to all the best combinations of subspace centroids. However, overpopulating descriptions from one image decreases the possibility of retrieving descriptions from other images with lower amounts of descriptions. (c) With our approach, fusion of inverted indices, n descriptors are used to find images that match the query with more similar characteristics (here a 2D descriptor A and a 3D descriptor B). Each cluster in a descriptor space represents only the nearest descriptions from each matched image (descriptions in hooped dotted circles). Furthermore, as n subspace responses are combined, we are able to obtain a direct rank of the images that better match the query image. This rank represents the images that are similar in all or most of all descriptor characteristics.

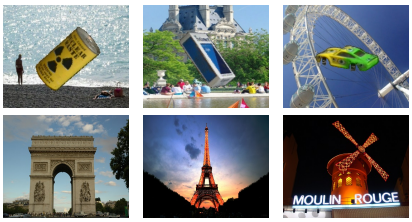

Fig. 2: Samples from two benchmarks used in our experiments.

contents. Among the descriptors tested, we concentrate on SIFT [9], SURF [25] and Shape Context ("SC") [26], which performed better individually for these datasets. All the descriptions used are quantized into words with the FLANN library hierarchical k-means[3].

The approaches and descriptors are evaluated in terms of quality of retrieval through precision and recall curves. Due to space limitations, we only present their mean Average Precision (mAP) which is a summarized measure of quality across the multiple queries by averaging average precision.

4.2. Parameter settings and baseline

The two main parameters of the proposed approach are the codebook size and parameter k of the individual k-nearest neighbor search (step #2 in Section 3.2). We varied the codebook size in range {25000,125000} for COIL_DB and in range {500000,2000000} for Paris_DB, and obtained the best mAP with 50000 words for COIL_DB and 1500000 words for Paris_DB, for the three used descriptors. Similarly, parameter k was varied from 1-NN to 5-NN, where 2-NN achieved the best results for all the descriptors and datasets. These parameters are used by default in the rest of the experiments.

To give a first insight into the proposal, we begin by evaluating the original version of the inverted multi-index [5] completed with our strategy of image voting (step #5 in Section 3.2) on a single descriptor, facing the classical approach based on Bag-of-Features (BoF) with tf-idf scores [29]. Table 1 shows the mAP achieved on the three descriptors used individually for both datasets.

Table 1: mAP results for individual descriptors.

	BoF with tf-idf			Inverted multi-index		
Benchmark	SIFT	SURF	SC	SIFT	SURF	SC
COIL_DB	0.10	0.09	0.10	0.55	0.44	0.53
Paris_DB	0.09	0.08	0.08	0.43	0.48	0.39

Approach based on BoF with tf-idf achieves a lower precision since this representation relies on a global description of the image associated to a global similarity measure (usually the Euclidean distance), less robust to the complex scenes present in the two datasets. Not surprisingly, the voting strategy of the inverted multi-index approach performs better since it searches for similar areas in different images by comparing groups of words.

4.3. Reduction strategies and their fusion

In this section, we evaluate the methodology presented in Section 3.3. The descriptors are used individually and each of them is decomposed with PCA and PLS dimensionality reduction techniques. On the two datasets, Table 2 shows the mAP obtained (i) before any reduction, with the simple splitting strategy of [5] (column "Descriptor"), (ii) after reduction, again with the simple splitting strategy (column "Reduction") and (iii) with the fusion of two reduced descriptions (column "Fusion"). The sub-index and super-index texts next to each descriptor indicate their original dimension or their reduced dimension preceded by the technique used (*e.g.* $\text{SURF}^{\text{PLS32}}$ indicates that description SURF was reduced with PLS down to 32 dimensions, and $\text{SC}^{\text{PCA12}}_{\text{PLS12}}$ that description SC was reduced both with PCA and PLS down to 12 dimensions, before fusing them with the inverted multi-index).

[3] http://www.cs.ubc.ca/research/flann/

Table 2: mAP results with reduced descriptions and their fusion.

	Descriptor	Reduction		Fusion
COIL_DB	$SIFT^{128}$	$SIFT^{PCA32}$	$SIFT^{PLS32}$	$SIFT^{PCA32}_{PLS32}$
	0.55	0.47	0.45	0.58
		$SIFT^{PCA64}$	$SIFT^{PLS64}$	$SIFT^{PCA64}_{PLS64}$
		0.47	0.46	**0.59**
	$SURF^{64}$	$SURF^{PCA20}$	$SURF^{PLS20}$	$SURF^{PCA20}_{PLS20}$
	0.44	0.36	0.33	0.43
		$SURF^{PCA32}$	$SURF^{PLS32}$	$SURF^{PCA32}_{PLS32}$
		0.37	0.36	**0.45**
	SC^{36}	SC^{PCA12}	SC^{PLS12}	SC^{PCA12}_{PLS12}
	0.53	0.40	0.38	0.50
		SC^{PCA18}	SC^{PLS18}	SC^{PCA18}_{PLS18}
		0.44	0.42	**0.54**
Paris_DB	$SIFT^{128}$	$SIFT^{PCA32}$	$SIFT^{PLS32}$	$SIFT^{PCA32}_{PLS32}$
	0.43	0.47	0.43	**0.49**
		$SIFT^{PCA64}$	$SIFT^{PLS64}$	$SIFT^{PCA64}_{PLS64}$
		0.48	0.42	**0.49**
	$SURF^{64}$	$SURF^{PCA20}$	$SURF^{PLS20}$	$SURF^{PCA20}_{PLS20}$
	0.48	0.45	0.44	0.49
		$SURF^{PCA32}$	$SURF^{PLS32}$	$SURF^{PCA32}_{PLS32}$
		0.49	0.47	**0.52**
	SC^{36}	SC^{PCA12}	SC^{PLS12}	SC^{PCA12}_{PLS12}
	0.39	0.31	0.28	0.33
		SC^{PCA18}	SC^{PLS18}	SC^{PCA18}_{PLS18}
		0.39	0.37	**0.42**

Irrespective of the dataset, the loss in precision is not proportional to the percentage of dimension reduction applied for single reduction; it achieves slightly lower or similar precision than with the original description. The column "Fusion" shows that the fused reduced descriptions of the original descriptor is able to achieve better results than the individual parent descriptions. This is due to the fact that the fused descriptors represent two complementarity subspaces which estimate better the approximation of nearest neighbors. In addition, the largest dimension of the manipulated features is the same as the one of its original description (*e.g.,* 32+32=64 for SURF).

4.4. Fusion of different descriptions

This section presents the results of the fusion obtained with the joint use of different descriptions, in reference to Section 3.2. We evaluate all the possible combinations of SIFT, SURF and SC descriptors, and compare our proposal ("FII" in the tables) to two state-of-the-art descriptor combination approaches: feature concatenation [8] based on BoF with tf-idf ("CBoF") and the best late fusion technique [16] ("LF"). Table 3 shows the performance obtained for the two datasets.

Table 3: mAP results with the fusion of different descriptors.

	Descriptors	CBoF [8]	LF [16]	FII
COIL_DB	SIFT-SURF	0.10	0.50	**0.57**
	SIFT-SC	0.10	0.52	**0.59**
	SIFT-SURF-SC	0.11	0.53	**0.61**
Paris_DB	SIFT-SURF	0.10	0.53	**0.55**
	SURF-SC	0.09	0.49	**0.53**
	SIFT-SURF-SC	0.11	0.53	**0.54**

Due to the poor scores obtained with the classical approach BoF with tf-idf on individual descriptors (see Table 1), the results obtained here with approach CBoF are low again, even if the fusion improves them slightly. The LF method performs better for the two datasets, but it is not able to outperform the FII method, whatever the combination. This is due to the fact that the former method only considers the associated neighbors to the nearest word to the query, while the FII considers several combinations of word neighbors. We also observe that the best configurations of descriptors are not the same for the two datasets: it is SIFT-SURF-SC for COIL_DB and SIFT-SURF for Paris_DB; descriptor SC penalizes the combined description for Paris_DB.

CBoF builds a high dimensional space by concatenating several descriptors. It requires more memory compared to LF approach. For FII, computational time is related to the number of fused descriptors, as the complexity of multi-sequence algorithm is $n \log n$, where n denotes the number of descriptors. However, the fusion of reduced descriptors shortens the computational time without compromising mAP, as it is experimented in the following Section 4.5.

4.5. Fusion of different reduced descriptions

Finally, we evaluate the combination of several descriptors with their reduced version. Table 4 shows the mAP obtained and associated averaged retrieval time for COIL_DB and Paris_DB. An Intel(R) Core(TM) i7-2670QM computer with CPU 2.20 GHz and 8 GB de RAM was used to measure the computational time. We observe that, for FII, the best mAP are similar to or even better than the ones obtained by fusing the original descriptions (Table 3), and always better than the ones of LF. Note that the total amount of dimensions for the fused reduced descriptions never exceed 128, which is the dimension of the largest description used alone (SIFT). However, computational time for LF is slightly faster since our algorithm performs several combinations to find the best nearest neighbors (the largest gap concerns Paris_DB with six combined descriptions).

Table 4: mAP and averaged retrieval time with the fusion of different reduced descriptors.

	Descriptors	LF [16]	Time (s)	FII	Time (s)
COIL_DB	$SIFT^{PCA64}$ $SURF^{PCA32}$	0.49	0.031	**0.57**	0.033
	$SIFT^{PCA32}$ $SURF^{PCA20}$ SC^{PCA12}	0.48	0.043	**0.58**	0.045
	$SIFT^{PCA32}_{PLS32}$ $SURF^{PCA20}_{PLS20}$ SC^{PCA12}_{PLS12}	0.50	0.079	**0.67**	0.083
Paris_DB	$SIFT^{PCA64}$ $SURF^{PCA32}$	0.53	0.524	**0.54**	0.609
	$SIFT^{PCA32}$ $SURF^{PCA20}$ SC^{PCA12}	0.49	0.738	**0.50**	0.928
	$SIFT^{PCA32}_{PLS32}$ $SURF^{PCA20}_{PLS20}$ SC^{PCA12}_{PLS12}	0.49	1.480	**0.51**	3.680

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed an approach to the fusion of multiple image descriptors based on an improved inverted multi-index structure. The experiments performed for similarity search on two datasets have demonstrated the relevance of their combination through this structure: the combination of different image characteristics clearly improves the content representation, and the strategy of fusion brings distinctiveness during nearest neighbor search. The proposal has demonstrated its superiority facing two state-of-the-art fusion approaches [8, 16]. In addition, we have shown that the use of complementary techniques of dimension reduction as description decomposition, PCA and PLS, contributes to improve distinctiveness during similarity search, while potentially reducing the volume of manipulated features, and then limiting the computational complexity despite the multiple descriptions involved.

The descriptors combined were chosen *a priori*, according to their presupposed complementarity. In the future, we plan to study measures of complementarity in order to combine optimal configurations of descriptors, before evaluating the whole similarity search engine at large scale.

6. REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 55:1–5:60, May 2008.

[2] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia Computing Communications and Applications*, vol. 2, no. 1, pp. 1–19, Feb. 2006.

[3] T. Tuytelaars and K. Mikolajczyk, *Local Invariant Feature Detectors: A Survey*. Hanover, MA, USA: Now Publishers Inc., 2008.

[4] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating Color Descriptors for Object and Scene Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

[5] A. Babenko and V. Lempitsky, "The inverted multi-index," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3069–3076.

[6] P. K. Atrey, M. A. Hossain, A. E. Saddik, and M. S. Kankanhalli, "Multimodal Fusion for Multimedia Analysis: A Survey," *Multimedia Systems*, vol. 16, pp. 345–379, 2010.

[7] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early Versus Late Fusion in Semantic Video Analysis," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2005, pp. 399–402.

[8] J. Yu, Z. Qin, T. Wan, and X. Zhang, "Feature Integration Analysis of Bag-of-Features Model for Image Retrieval," *Neurocomputing*, vol. 120, pp. 355 – 364, 2013.

[9] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, USA, 2005, pp. 886–893.

[11] T. Ojala, M. Pietikäinen, and T. Maenpää, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[12] P. A. S. Kimura, J. M. B. Cavalcanti, P. C. Saraiva, R. da Silva Torres, and M. A. Gonçalves, "Evaluating Retrieval Effectiveness of Descriptors for Searching in Large Image Databases," *Journal of Information and Data Management*, vol. 2, no. 3, pp. 305–320, 2011.

[13] J. Yue, Z. Li, L. Liu, and Z. Fu, "Content-based Image Retrieval using Color and Texture Fused Features," *Mathematical and Computer Modelling*, vol. 54, no. 34, pp. 1121–1127, 2011, mathematical and Computer Modeling in Agriculture.

[14] R. da S. Torres, A. X. Falcão, M. A. Gonçalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox, "A Genetic Programming Framework for Content-based Image Retrieval," *Pattern Recognition*, vol. 42, no. 2, pp. 283 – 292, 2009.

[15] D. Frank Hsu and I. Taksa, "Comparing rank and score combination methods for data fusion in information retrieval," *Information Retrieval*, vol. 8, no. 3, pp. 449–480, 2005.

[16] N. Neshov, "Comparison on Late Fusion Methods of Low Level Features for Content Based Image Retrieval," in *Artificial Neural Networks and Machine Learning*, ser. Lecture Notes in Computer Science, V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. Villa, B. Appollini, and N. Kasabov, Eds. Springer Berlin Heidelberg, 2013, vol. 8131, pp. 619–627.

[17] W. Zhang, Z. Qin, and T. Wan, "Image Scene Categorization using Multi-Bag-of-Features," in *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 4, 2011, pp. 1804–1808.

[18] P. Gehler and S. Nowozin, "On Feature Combination for Multiclass Object Classification," in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 221–228.

[19] D. Picard, N. Thome, and M. Cord, "An Efficient System for Combining Complementary Kernels in Complex Visual Categorization Tasks," in *Proceedings of 17th IEEE International Conference on Image Processing*, Sept 2010, pp. 3877–3880.

[20] V. Risojevic and Z. Babic, "Fusion of Global and Local Descriptors for Remote Sensing Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 836–840, 2013.

[21] Y. Cao, H. Zhang, Y. Gao, X. Xu, and J. Guo, "Matching Image with Multiple Local Features," in *Proceedings of 20th International Conference on Pattern Recognition*, 2010, pp. 519–522.

[22] D. Lisin, M. Mattar, M. Blaschko, E. Learned-Miller, and M. Benfield, "Combining Local and Global Image Features for Object Class Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, June 2005, pp. 47–47.

[23] R. Gray and D. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.

[24] H. Jegou, M. Douze, and C. Schmid, "Product Quantization for Nearest Neighbor Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 117–128, 2011.

[25] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 11, no. 3, pp. 346–359, Jun 2008.

[26] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, Apr 2002.

[27] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, "Dimensionality Reduction Through PCA over SIFT and SURF Descriptors," in *Proceedings of IEEE Conference on Cybernetics Intelligent Systems*, 2012, pp. 1–6.

[28] R. Rosipal and N. Krmer, "Overview and Recent Advances in Partial Least Squares," in *Subspace, Latent Structure and Feature Selection*, ser. Lecture Notes in Computer Science, C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, Eds. Springer Berlin Heidelberg, 2006, vol. 3940, pp. 34–51.

[29] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *Proceedings of Ninth IEEE International Conference on Computer Vision*, 2003, pp. 1470–1477 vol.2.