DIMENSIONALITY REDUCTION OF VISUAL FEATURES USING SPARSE PROJECTORS FOR CONTENT-BASED IMAGE RETRIEVAL

Romain Negrel[⋆] *David Picard*[⋆] *Philippe-Henri Gosselin*[⋆ †]

{romain.negrel,picard,gosselin}@ensea.fr
[⋆] ETIS/ENSEA - University of Cergy-Pontoise, CNRS, UMR 8051, France
[†]Inria Rennes Bretagne Atlantique, France

ABSTRACT

In web-scale image retrieval, the most effective strategy is to aggregate local descriptors into a high dimensionality signature and then reduce it to a small dimensionality. Thanks to this strategy, web-scale image databases can be represented with small index and explored using fast visual similarities. However, the computation of this index has a very high complexity, because of the high dimensionality of signature projectors. In this work, we propose a new efficient method to greatly reduce the signature dimensionality with low computational and storage costs. Our method is based on the linear projection of the signature onto a small subspace using a sparse projection matrix. We report several experimental results on two standard datasets (Inria Holidays and Oxford) and with 100k image distractors. We show that our method reduces both the projectors storage cost and the computational cost of projection step while incurring a very slight loss in mAP (mean Average Precision) performance of these computed signatures.

***Index Terms*—** Image retrieval, Image databases, Indexes, Sparse matrices

1. INTRODUCTION

In this paper, we focus on dimensionality reduction methods for content-based image retrieval (CBIR) on web scale datasets. CBIR is a very dynamic research field that rises many challenges, and which has proposed a variety of efficient methods to solve the problem of similarity search. Many of these methods are based on the use of highly discriminative local descriptors [1] (*e.g.*, HoG [2], SIFT [3]). Initially, the similarity between two images [4] was computed directly on the sets of local descriptors extracted from the images. However, the computing cost of the pairwise similarity between two sets of local descriptors is prohibitive due to the large number of extracted local descriptors. To solve this problem, methods aggregating local descriptors in a unique signature [5, 6, 7, 8, 9] were proposed. It has been shown that these signatures retain the discriminatory power of local descriptors with similarity measures of low computational cost (*e.g.*, dot product). However signatures that perform well are very large [10], and the storage cost becomes prohibitive for web scale datasets. As we detail later, several methods to reduce signatures dimensionality have been proposed. These methods provide low dimensional signatures with low storage cost and good discriminatory power. However, they incur a high projection cost: the memory required to store the projectors and the computational cost to perform the projection are often prohibitive, and this is precisely what we investigate in this paper.

More specifically, we propose a new method to dramatically reduce the memory footprint and the computational cost of such projections. This method is based on the low-rank approximation of Gram matrix with a sparse projection matrix. Our main novel contributions are:

- The introduction of a sparsity constraint in the Gram matrix low-rank approximation problem;
- The introduction of a correction matrix to correct the errors induced by the coarse solution of the sparse low-rank approximation problem.

The rest of the paper is organized as follows: First, we give an overview of the related work in dimensionality reduction. Then we explain our proposed method in section 3. In section 4, we evaluate our method on Inria Holidays [11] and Oxford datasets and we discuss and compare the performance with the state of the art methods, before we conclude.

2. RELATED WORKS

In this section, we present current methods for the reduction of visual signatures, as well as their main advantages and drawbacks. More specifically, we focus on methods that compute linear projectors in Hilbert spaces:

$$\mathbf{y} = \mathbf{P}^\top \mathbf{x}, \tag{1}$$

with $\mathbf{y}$ the signature reduced at N dimensions, $\mathbf{P}$ the projectors matrix and $\mathbf{x}$ the original signature of W dimensions. The choice of linear projectors can be explained by their simplicity and their ability to deal with large datasets. The choice of unsupervised training can be explained by the difficulty of obtaining a ground truth for image similarity search.

The most popular approach to learn the projection matrix is the Principal Component Analysis (PCA) [12] which selects components of largest variance. PCA is well known in data analysis as it provides a compact representation of the data while guaranteeing the lowest reconstruction error of the data. This approach has shown to retain the discriminating power of signatures while greatly reducing the dimensionality in many state of the art papers [8, 13, 10]. However, the criterion of data reconstruction does not guarantee the preservation of the discrimination power of signatures.

To solve this problem, the authors of [14] propose a similar approach based on the low-rank approximation of the Gram matrix. The authors propose to compute the projectors such that the original similarity between two signatures is retained. To this purpose, they propose to solve the following problem:

$$\begin{aligned} \mathbf{P}_N^* = &\arg\min_{\mathbf{P}_N} ||\mathbf{X}^\top\mathbf{X} - \mathbf{X}^\top\mathbf{P}_N\mathbf{P}_N^\top\mathbf{X}||_F^2 \\ &\text{s.t. } \mathbf{P}_N \in \mathcal{M}_{W,N} \text{ with } N < L \ll W, \end{aligned} \tag{2}$$

with $\mathbf{X}$ the L signatures of the training set and $\mathbf{P}_N$ the projectors matrix. The closed form solution of this problem is:

$$\mathbf{P}_N^* = \mathbf{X}\mathbf{T}_N\mathbf{L}_N^{-1/2}, \tag{3}$$

with $\{\mathbf{T}_N, \mathbf{L}_N\}$ the N principal eigenvectors and eingenvalues of Gram matrix $\mathbf{X}\mathbf{X}^\top$. Furthermore, they propose to use the dot product associated with Mahalanobis distance as a new similarity measure. This similarity measure gives a better discrimination power and it can be integrated in the projectors matrix as follows:

$$\mathbf{P}_N^{'*} = \mathbf{P}_N^{'*}\mathbf{L}_N^{-1/2} = \mathbf{X}\mathbf{T}_N\mathbf{L}_N^{-1}. \tag{4}$$

This method provides projectors which drastically reduce the size of original signature while retaining the similarity between two signatures. For example, visual signatures of hundreds of thousands of dimensions can be reduced to a few hundreds and with similar retrieval performance.

However, all these current methods suffer from a major drawback: the size of projectors is as large as the size of visual features. This implies that the memory cost and computational cost of the projection have an order of $O(W \times N)$ with N the dimension of subspace. Thus, given the high values of W (*e.g.* at least hundreds of thousands), the corresponding projection matrix quickly becomes very large, the projection itself is not scalable. For instance such matrix is thus difficult to spread on a computational grid, or simply too large to fit in the available memory.

3. PROPOSED METHOD

In this section, we present our main contribution: a new projection matrix for significantly reducing the size of large signatures with a low storage cost and computational cost. Our projection matrix is based on optimizing the reconstruction of the Gram matrix of a training set with a sparsity constraint.

In order to obtain the sparse projectors that provide a better approximation of the Gram matrix, we must solve problem (2) constrained with ℓ_0 norm:

$$\begin{aligned} \mathbf{U}_N^* = &\arg\min_{\mathbf{U}_N} ||\mathbf{X}^\top\mathbf{X} - \mathbf{X}^\top\mathbf{U}_N\mathbf{U}_N^\top\mathbf{X}||_F^2 \\ &\text{s.t.} \begin{array}{l} \mathbf{U}_N \in \mathcal{M}_{W,N} \text{ with } N < L \ll W \\ ||\mathbf{u}_i||_0 = M, \forall i, \end{array} \end{aligned} \tag{5}$$

with M the number of non-zero entries by columns of matrix $\mathbf{U}_N$. However, since this problem is NP-hard, it is very complex to obtain an exact solution.

We propose to reformulate this problem by decomposing the projection matrix $\mathbf{U}_N$ into two matrices:

$$\mathbf{U}_N = \hat{\mathbf{P}}_N\mathbf{R}_N, \tag{6}$$

with $\hat{\mathbf{P}}_N$ a sparse matrix of W by N, $\mathbf{R}_N$ a full square matrix. Dimensionality reduction is then performed in two steps: (*i*) a first step of high dimensionality reduction with a sparse approximation of full projection matrix $\mathbf{P}_N^*$ and (*ii*) a second projection step with a low cost full matrix. The second projection step allows to correct the errors introduced by the sparse approximation $\hat{\mathbf{P}}_N^*$.

3.1. The sparse matrix $\mathbf{P}_N$

To compute the sparse projection matrix, we propose to compute the sparse approximation of the projectors $\mathbf{P}_N^*$ obtained by solving the problem without the constraint of sparsity. For this, we solve the following problem:

$$\begin{aligned} \hat{\mathbf{P}}_N^* = &\arg\min_{\hat{\mathbf{P}}_N} ||\mathbf{P}_N^{'*} - \hat{\mathbf{P}}_N||_F^2 \\ &\text{s.t. } ||\hat{\mathbf{p}}_i||_0 = M, \forall i. \end{aligned} \tag{7}$$

The problem (7) has a closed form solution obtained by thresholding the smallest values of the original projectors:

$$\hat{p}_{ki} = p_{ki}\,\mathrm{h}(|p_{ik}| - \nabla_i), \forall k \tag{8}$$

with h the Heaviside step function, $\nabla_i \in \mathbb{R}^+$ the threshold selected to satisfies the sparsity constraint.

The solution of this problem is a very simple but coarse approximation of the solution of the problem (5). Indeed, this solution does not take into consideration the correlations between projectors.

3.2. The correction matrix $\mathbf{R}_N$

To correct the errors introduced by the sparse projection matrix, we propose to compute a correction matrix in the low dimensional space. For this, we propose to compute the matrix $\mathbf{R}_N$ such that the Gram matrix of corrected signature, $\mathbf{G}_{\hat{\mathbf{P}}_N^*\mathbf{R}_N} = \mathbf{X}^\top\hat{\mathbf{P}}_N^*\mathbf{R}_N\mathbf{R}_N^\top\hat{\mathbf{P}}_N^{*\top}\mathbf{X}$, is as close as possible to the Gram matrix of signature obtained with full projectors $\mathbf{G}_{\mathbf{P}_N^{'*}} = \mathbf{X}^\top\mathbf{P}_N^{'*}\mathbf{P}_N^{'*\top}\mathbf{X}$. To obtain this correction matrix, we solve the following problem:

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} ||\mathbf{G}_{\mathbf{P}_N^{'*}} - \mathbf{Y}_{\hat{\mathbf{P}}_N^*}^\top\mathbf{W}\mathbf{Y}_{\hat{\mathbf{P}}_N^*}||_F^2, \tag{9}$$

with $\mathbf{Y}_{\hat{\mathbf{P}}_N^*} = \hat{\mathbf{P}}_N^{*\top}\mathbf{X}$ and $\mathbf{W} = \mathbf{R}_N\mathbf{R}_N^\top$. This problem is convex and, for $S < L$, it has a closed form solution:

$$\mathbf{W}^* = \mathbf{Y}_{\hat{\mathbf{P}}_N^*}^+\mathbf{G}_{\mathbf{P}_N^{'*}}\mathbf{Y}_{\hat{\mathbf{P}}_N^*}^{+\top}, \tag{10}$$

with $\mathbf{Y}_{\hat{\mathbf{P}}_N^*}^+ = (\mathbf{Y}_{\hat{\mathbf{P}}_N^*}\mathbf{Y}_{\hat{\mathbf{P}}_N^*}^\top)^{-1}\mathbf{Y}_{\hat{\mathbf{P}}_N^*}$ the pseudoinverse of $\mathbf{Y}_{\hat{\mathbf{P}}_N^*}$ matrix. We obtain the correction matrix $\mathbf{R}_N^*$ by factorization of $\mathbf{W}^*$ to this propose we use the eigen decomposition:

$$\mathbf{R}_N^* = \mathbf{V}\mathbf{D}^{1/2}, \tag{11}$$

with $\{\mathbf{V}, \mathbf{D}\}$ the eigenvectors and eigenvalues of $\mathbf{W}^*$.

The full projection for the dimensionality reduction is then decomposed into two projections:

$$\begin{aligned} \hat{\mathbf{y}}_i &= \hat{\mathbf{P}}_N^{*\top}\mathbf{x}_i \\ \mathbf{y}_i &= \mathbf{R}_N^{*\top}\hat{\mathbf{y}}_i \end{aligned}. \tag{12}$$

For convenience, we define the sparsity constraint independently from the dimension of the input vectors. To this end, we note by τ the rate of zero values in a sparse matrix:

$$\tau(\hat{\mathbf{P}}_N^*) = \frac{\text{Number of zero values in } \hat{\mathbf{P}}_N^*}{\text{Number of values in } \hat{\mathbf{P}}_N^*}. \tag{13}$$

In the case of our matrix $\hat{\mathbf{P}}_N^*$ constrained by ℓ_0 norm, we have the following relation between M and τ:

$$\tau(\hat{\mathbf{P}}_N^*) = \frac{W - M}{W}. \tag{14}$$

The cumulative computational cost of these two projections is $\mathcal{O}\left((N + (1-\tau) \times W) \times N\right)$. For high sparsity rate (*e.g.* $\tau =$

τ \ W	82k		415k		1.7M	
	$\mathcal{O}_c$	$\mathcal{O}_m$	$\mathcal{O}_c$	$\mathcal{O}_m$	$\mathcal{O}_c$	$\mathcal{O}_m$
0.0%	21,1M	160MB	106M	811MB	435M	3,2GB
90.0%	2,2M	16.3MB	10,7M	81.3MB	43,5M	332MB
99.0%	275k	1.9MB	1,1M	8.4MB	4,4M	33.5MB
99.9%	87k	420kB	172k	1.1MB	501k	3.6MB

Table 1. Costs of dimensionality reduction step as function of input signature size and sparsity rate with $N = 256$ (with $\mathcal{O}_c$ the computational cost in operations and $\mathcal{O}_m$ the storage cost).

99%), this cost is very low in comparison with the computational cost with full projectors. The storage cost is also drastically reduced: $\mathcal{O}\left((N + 2 \times (1 - \tau) \times W) \times N\right)$, considering that the values of sparse projectors are stored in couples (index, value).

Table 1 shows the computational and storage cost of the proposed method. We observe that high sparsity rate allows to strongly reduces the costs of dimensionality reduction step.

4. EXPERIMENTS

In this section, we present the reference datasets and the signatures we use to evaluate our proposed method. We discuss the performance of our method and compare it with the state of the art.

4.1. Datasets and Signatures

We use two well known benchmarks: Inria Holidays dataset and Oxford dataset. The Holidays dataset contains 1,491 images (typically personal holiday photographs) gathered in 500 groups. The Oxford dataset is a set of images (from Flickr) representing various Oxford landmarks; it contains 5,062 images of 11 landmarks. To evaluate the robustness of our method, we use 100k images distractor randomly extracted from ImageNet dataset. The ImageNet dataset is a set of high-quality images extracted from Flickr. For all training, we use Holidays Flickr60k dataset, which is a set of high-quality images randomly extracted from Flickr; it contains 60,000 images without ground truth.

For all the above mentioned images, we perform a two step preprocessing: (a) image resizing (to a maximum width of 512 pixels); (b) histogram equalization. Furthermore, we use two types of local descriptors: a texture descriptor HOG (128-dimensional) [2] and a color descriptor (96-dimensional) proposed by Perronnin et al. [15]. We extract these descriptors on a regular dense grid of 3×3 pixels and at 4 scales. We use these descriptors to compute "Vectors of Locally Aggregated Tensors" (VLAT) [9] and "Fisher Vector" (FV) [16] signatures as follows. For each descriptor, we compute: a VLAT signature with a cluster-wise PCA that preserves 80 dimensions by cluster; and a FV signature with a PCA on local descriptors that preserves 80 dimensions. The HOG and color signatures are subsequently concatenated. Then, we perform a power-normalization (for all experiments set to 0.1) and ℓ_2-normalization of the concatenated signatures.

In the following, we denote the signatures used for the experiments by their abbreviations prefixed by the dimentionnality reduction method used if any, and as suffix the size of the visual codewords. More precisely, we use the prefix "C" to denote the dimentionnality reduction method proposed in [14]; "CS" to denote our method without the correction matrix; and "CSR" to denote our method with the correction matrix. For example, "CS-VLAT-256" is a signature VLAT computed with 256 visual codewords and dimensionality of reduced with our proposed method without the correction matrix.

Sign.	Dim.	Holidays	Oxford
FV-64[17]	4k	59.5	31.7
VLAD-64[17]	4k	55.6	30.4
VLAD-64[18]	8k	62.2	50.0
VLAD-64[14]	8k	-	36.6
VLAT-64[14]	528k	**66.4**	**54.2**
FV-64	20k	78.6	43.3
FV-256	82k	82.0	49.3
VLAT-64	415k	81.6	58.4
VLAT-256	1.7M	**84.0**	**59.8**

Table 2. Evaluation of computed signatures and comparison of state-of-the-art results on Holidays and Oxford dataset (mAP in %).

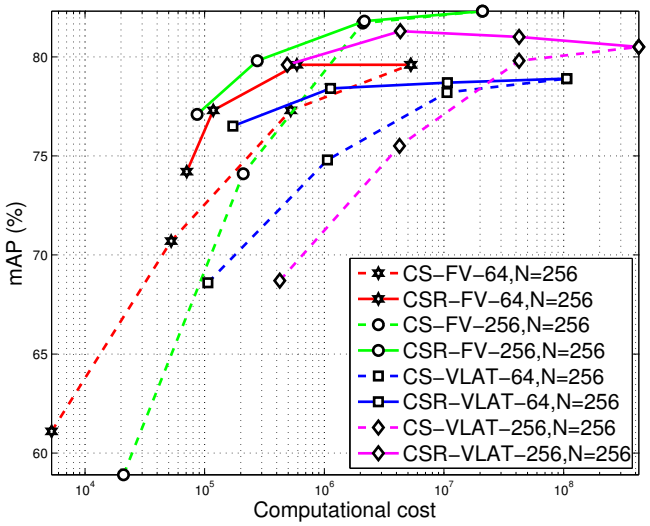

Fig. 2. mAP evolution on Holidays dataset as function of dimensionality reduction step computational cost for different sparsity rates (*i.e.*, $\tau \in \{99.9\%, 99\%, 90\%, 0\%\}$).

As a baseline, Table 2 shows the mAP on Holidays and Oxford datasets obtained by the unreduced signatures. This table is divided in two parts: in the top part, we report several results obtained in related state of the art; and, in the bottom part, we report the results of our implementation of the same methods. Comparing with the results of state of the art, we see that we obtain much better performance. This is mainly due to our use of two types of local descriptors as well as larger visual codebooks than those of the state of the art.

4.2. Parameter Analysis

Here, we study the behavior of the reduced signatures as function of the parameters of our method. All experiments are done using the Holidays dataset. We consider a web-scale context in which the final signature must be small ($N = 256$). We compute the sparse projectors $\hat{\mathbf{P}}_N^*$ and the correction matrix $\mathbf{R}_N^*$ for different sparsity rates $\tau \in \{99.9\%, 99\%, 90\%, 0\%\}$ (note that for $\tau = 0\%$: $\hat{\mathbf{P}}_N^* = \mathbf{P}_N^*$ and $\mathbf{R}_N^* = \mathbf{I}$). Figure 2 shows the mAP evolution on Holidays

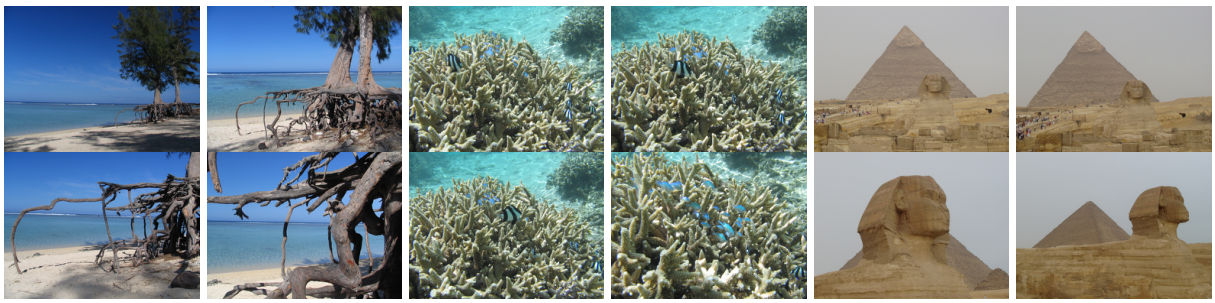

Fig. 1. Images from Holidays dataset [11].

dataset as function of the computational cost of our dimentionality reduction method. This cost is induced by the two projection steps: the projection of the signature on the sparse projectors; and the projection on the correction matrix. The continuous and dashed curves represent the performance of the reduced signatures with and without the correction step respectively. We note that at equivalent computation cost, the high dimensional projectors are less sensitive to high sparsity rate. We see that the correction step allows to correct well the errors introduced by sparse projectors. Moreover, we observe that the higher the sparsity rate, the more effective the correction step. We also observe that the mAP performance is similar for any type of signatures. Thus, our method is not sensitive to a particular signature choice. For instance, with "CSR-FV-256" signature and a sparsity rate of 90%, the computational cost is divided by a factor of about 9 without any performance loss. With the same "CSR-VLAT-256" signature, but with a sparsity rate of 99.9%, the computational cost is divided by a factor of about 1000 while incurring a mAP loss of only 0.9.

4.3. Comparison with the State of the Art

In the following, we compare the performances of our dimensionality reduction method with state of the art methods. We compute the signatures: FV-64, FV-256, VLAT-64 and VLAT-256 reduced at 128 dimensions. We compute the performance of these signatures on Holidays and Oxford datasets with and without the addition of 100k distractors.

The performance results are reported in Table 3 which is divided in three parts. In the top part, we report several results obtained in related state of the art. The middle part illustrates the results obtained by reproducing the dimensionality reduction method proposed in [14] for fairness of comparison. We note that this method preserves the performance of the original signature on Holidays dataset but on Oxford dataset the performance is degraded. This is probably caused by using the local color descriptor that is not relevant on this dataset. In the bottom part, we report the results of the proposed method. We observe, in the two datasets, that our method provides the same robustness as the state of the art when adding the distractors. However, our method has the great advantage of having a much lower dimensionality reduction cost. For example, in the case of C-VLAT-256, the loss in mAP performance is of 6.4 on Holidays and of 0.6 on Oxford. Using our method, in the case of CSR-VLAT-256, the loss in mAP performance is of 6.3 on Holidays and of 1.0 on Oxford while dividing the computational cost by a factor of around 1000.

Name	Dim.	Holidays		Oxford	
			+100k		+100k
FV-64-PCA[17]	128	56.5	**38.0**	24.3	-
VLAD-64-PCA[17]	64	44.7	32.0	-	-
VLAD-64-PCA[18]	128	-	-	32.5	26.6
C-VLAT-64[19]	256	72.3	58.0	-	-
VLAD-64-PCA[14]	128	-	-	32.7	25.6
C-VLAT-64[14]	128	57.3	-	**54.3**	**46.6**
C-FV-64	128	77.4	69.9	36.6	35.7
C-FV-256	128	79.9	**74.0**	38.0	37.2
C-VLAT-64	128	75.4	69.3	**42.1**	**41.3**
C-VLAT-256	128	78.2	71.8	39.0	38.4
CSR-FV-64, $\tau = 0.9$	128	77.5	69.9	35.9	34.7
CSR-FV-256, $\tau = 0.9$	128	**80.6**	73.9	37.8	36.9
CSR-VLAT-64, $\tau = 0.99$	128	75.6	68.8	40.9	39.9
CSR-VLAT-256, $\tau = 0.999$	128	76.9	70.6	35.6	34.6

Table 3. Evaluation of the reduced signatures robustness by adding of 100k distractors and comparison of state-of-the-art results on Holidays and Oxford dataset (mAP in %).

5. CONCLUSION

In this paper, we introduce a method to reduce the dimensionality of image signatures with very low storage and computational complexities. Our method consists in a linear projection of the image signature in a low dimensional subspace thanks to sparse projectors. The sparse projectors are initialized using a sparse approximation of dense projectors, and then corrected using a small matrix. This second step allows to better recover the discriminative power of reduced signatures. We have carried out experiments on Inria Holidays and Oxford datasets, which showed that our dimensionality reduction method provides close performance to the state of the art, while reducing storage and computational complexities of a factor in between 10 to 1000 times.

6. REFERENCES

[1] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005.

[2] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[3] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] Philippe Henri Gosselin, Matthieu Cord, and Sylvie Philipp-Foliguet, "Kernels on bags of fuzzy regions for fast object retrieval," in *image processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 1, pp. I–177.

[5] Josef Sivic and Andrew Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 1470–1477.

[6] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3360–3367.

[7] Florent Perronnin and Christopher Dance, "Fisher kernels on visual vocabularies for image categorization," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[8] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez, "Aggregating local descriptors into a compact image representation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3304–3311.

[9] David Picard and P-H Gosselin, "Improving image similarity with vectors of locally aggregated tensors," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 669–672.

[10] Jorge Sánchez and Florent Perronnin, "High-dimensional signature compression for large-scale image classification," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1665–1672.

[11] Herve Jegou, Matthijs Douze, and Cordelia Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Computer Vision–ECCV 2008*, pp. 304–317. Springer, 2008.

[12] Christopher M Bishop and Nasser M Nasrabadi, *Pattern recognition and machine learning*, vol. 1, springer New York, 2006.

[13] Matthijs Douze, Arnau Ramisa, and Cordelia Schmid, "Combining attributes and fisher vectors for efficient image retrieval," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 745–752.

[14] R. Negrel, D. Picard, and P.-H. Gosselin, "Web-scale image retrieval using compact tensor aggregation of visual descriptors," *MultiMedia, IEEE*, vol. 20, no. 3, pp. 24–33, 2013.

[15] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, "Improving the fisher kernel for large-scale image classification," in *Computer Vision–ECCV 2010*, pp. 143–156. Springer, 2010.

[16] Florent Perronnin, Yan Liu, Jorge Sánchez, and Hervé Poirier, "Large-scale image retrieval with compressed fisher vectors," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3384–3391.

[17] Hervé Jégou, Florent Perronnin, Matthijs Douze, Cordelia Schmid, et al., "Aggregating local image descriptors into compact codes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1704–1716, 2012.

[18] Jonathan Delhumeau, Philippe-Henri Gosselin, Hervé Jégou, and Patrick Pérez, "Revisiting the vlad image representation," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 653–656.

[19] Romain Negrel, David Picard, and Philippe-Henri Gosselin, "Compact tensor based image representation for similarity search," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2425–2428.