# New Approach for Automatic Medical Image Annotation Using the Bag-of-words Model

Riadh BOUSLIMI, Jalel AKAICHI
Department of Computer Science
High Institute of Manage, Bouchoucha city
Le Bardo, Tunis, Tunisia
bouslimi.riadh@gmail.com, Jalel.akaichi@isg.rnu.tn

***ABSTRACT:*** *In this paper, we present a new approach for semantic automatic annotation of medical images. Indeed, the proposed approach uses the bag of words model to represent the visual content of the medical image combined with text descriptors based on term frequency–inverse document frequency technique and reduced by latent semantic to extract the co-occurrence between text and visual terms. In a first phase, we are interested in indexing texts and extracting all relevant terms using a thesaurus containing medical subject headings and concepts. In a second phase, medical images are indexed while recovering areas of interest which are invariant to change in scale such as light and tilt. To annotate a new medical image, we use the bag of words model to recover the feature vector. Indeed, we use the vector space model to retrieve similar medical images from the training database. The computation of the relevance value of an image according to a query image is based on the cosine function. To evaluate the performance of our proposed approach, we present an experiment carried out on five types of radiological imaging. The results showed that our approach works efficiently, especially with more images taken from the radiology of the skull.*

**Keywords:** Automatic Medical Image Annotation, Radiology, Information Retrieval, Latent Semantic, Bag of words, Feature Detection

**Received:** 18 January 2016, Revised 28 February 2016, Accepted 4 March 2016

© 2016 DLINE. All Rights Reserved

## 1. Introduction

In the last decade, a large number of medical reports containing textual information and digital medical images have been produced in various health care institutions. We distinguish several types of medical images including X-ray associated with medical reports such as *Magnetic Resonance Imaging* (MRI), *Computed Tomography* (CT), *Magnetic Source Imaging* (MSI), and *Magnetic Resonance Spectroscopy* (MRS). These medical images are usually stored in large databases and made available to various interested actors such as physicians, heath care professionals, researchers and students to diagnose patients and to provide valuable information for various purposes.

Observing the frequent use and large number of digital medical images, it becomes vital to develop new techniques and/or to enhance existing ones for information retrieval, which may improve the search efficiency in large databases of medical images. Among various advanced information retrieval practices, image annotation is considered as an essential task for the whole management of images' databases [1]. When medical images are annotated manually, for example by keywords, this becomes an

expensive and subjective task since it does not describe efficiently the content of the image and it may provide many errors leading to misleading interpretation. Therefore, manual annotation suffers from does many drawbacks and may lead to a loss of time and information, especially when it is performed on large images' databases [2] such as medical ones. To remedy to these limitations, automatic annotation of images is indispensable to perform efficient image retrieval, essential for many medical applications. It is essentially based on the combination of textual and visual information such as those composing medical reports. However, most systems which deal with multimedia medical reports exploit only the textual parts of these reports to extract the keywords and to attribute them to the considered medical images.

The standard approach is to represent the text as a bag of words [3] and to associate a weight to each word to characterize its frequency by the method called term frequency–inverse document frequency or *tf.idf*. The current challenge is to extend and to adapt this standard approach to applications involving medical images' usage. A natural orientation is to represent images using the bag of words model which has already shown its effectiveness in applications involving image annotation and/or retrieval [4] and [5].

The main intention of this work is to combine textual and visual information in order to build a feature vector that is reduced by the latent semantic indexing method. The method is based on the belief that words used in the equivalent contexts have a tendency to have analogous meanings; it is capable of extracting the conceptual content of a text by discovering associations between terms that take place in similar contexts. Accordingly, if a new unannotated medical image joins a medical report, our proposed system supports the automatic determination of the suitable keywords that will be associated to use the above method.

This paper is organized as follows:

• In Sect. 2, we present the state of the art covering different existing works related to the combination of both textual and visual information, and automatic annotation of medical images.

• In Sect. 3, we present a first contribution related to automatic indexation of medical reports which combines both textual and visual information using the bag of words model, and uses a MeSH thesaurus to control the indexed vocabulary. The combination of textual and visual information will be reduced using the latent semantic to build co-occurrences between words and regions of interest in medical image. The second contribution consists of an automatic annotation of a new medical image joining the database. It is performed through the comparison of two histograms using measure of intersection.

• In Sect. 4, we present the automatic medical image annotation approach.

• In Sect. 5, we present and discuss the experimentation.

• In Sect. 6, we conclude and present future work.

**2. Modeling Medical Images with Bag of Visual Words**

Bag of visual words also known as Bag of features or Bag of key points is a technique used for image representation in various research themes [6], [7], [8], [9] and [10]. This approach describes an image using a dictionary composed of different visual words which are considered as local image patterns focusing on significant semantic information about the image. Finding objects in images and matching images with others are generally carried out directly using local features computed on images [11]. In broader contexts, such as the categorization of images [12] or information retrieval [13],[9] and [10], these local features are grouped using an unsupervised classification algorithm to form a vocabulary of visual words as illustrated in Algorithm 1 inspired from research works described in [4] [14] and [15].

**Algorithm 1. Creation of a Vocabulary of Visual words;**

**Input :** Image collection;

**Output :** Vocabulary of visual words;
Begin
**Foreach** image in the set of collected images Do
Detect the points of interest of the considered image;
Describe the points of interest the considered image;
**EndFor**
Define visual words using unsupervised clustering;
Return the vocabulary of visual words;
**End.**

The most common approach is what is called dynamic cloud or k-means algorithm presented in [16] and [17]. The classification step patterns, using the above approach, can be seen as a step of adaptive sampling of the feature space. This makes it possible to reduce the size of this space and then to compute histograms associated with the occurrences of visual words. A visual word is interpreted as a class of patterns which exists frequently in the considered collected medical images. The vocabulary of visual words is then used to represent the image as illustrated in the following main steps establishing Algorithm 2.

**Algorithm 2. Bag of visual word representation;**
**Input :** Image collection;
**Output :** Weighted vectors of visual words;
Begin
**Foreach** image in the set of collected images do
Detect points of interest of a new image;
Description of points of interest;
Associate text descriptors with visual words;
Using the determined vocabulary by Algorithm 1;
Represent the image by a weighted vector of visual words;
**EndFor**
**End.**

The regions of the image are associated with visual words according to their provided descriptions. The image can then be represented using a weighted vector of visual words. This representation that uses the model of visual bag of words depends on many parameters such as the detection manner and the provided description of points of interest [18], the embraced classification algorithm [14], the size of the visual vocabulary, the number of visual words computed on image, the normalization approach [19], etc.

Many studies related to images categorization showed that the detection of regular points of interest gives the best results, especially when the vocabulary size is important [14]; [19]; we think that is the case of medical images which are characterized by a rich visual vocabulary. As for the representation of textual documents, visual words are weighted for each processed image. The occurrence of visual words, in an image, is used; however the discriminating power of these is also taken into account [15]. Some approaches use a binary weighting of visual words by considering the occurrence or non-occurrence of words in images [19]. A selection of visual words to consider representation is sometimes done by retaining only those that maximize the mutual information between a word and a class [19]. One of the current challenges concerning the description of visual information is to take into account the spatial information. Several approaches have been proposed in this area, based for example, on a regular cutting of the image according to a pyramidal decomposition or by constructing a visual phrase [20], [21] and [22].

The choice of the representation of images is quite difficult, mainly because of the existence of a semantic gap between the description and the interpretation of these images. Approaches describing images as bags of words are more and more used; however textual and visual information was considered separately. We think that to describe multimedia documents, such as medical reports, in an efficient manner, it is essential to combine textual and visual information in order to satisfy querying,

analysis and/or mining users' requirements.

### 3. Indexing Medical Reports

The medical report includes documents which describe partially or completely the patient state. Documents, which stand either on texts or medical images, are usually described, respectively with textual and visual terms. The two modalities are generally processed separately using each the bag of words model. They are then represented as a *tf.idf* weighted vectors characterizing the frequency of each word either visual or textual. The representation used for both modalities can be combined by a late fusion methodology and then queries, destined to retrieve multimedia information, can be performed more efficiently. This general methodology is presented by Algorithm 3.

**Algorithm 3. Indexing Medical Reports;**

**Input :** Medical reports;
**Output :** Dictionary of medical vocabulary;
**Begin**
**Foreach** report in the set of medical reports do
**Foreach** text in the considered report do
Perform a cleaning task of the text using a stop word database;
Construct tf.idf weight vector;
Generate the visual codebook;
Determine vectors of visual words;
Generate visual word's Matrix;
**EndFor**
**Foreach** image in the considered report do
Determine regions of interest;
Generate the visual codebook;
Determine vectors of visuals words;
Generate visual word's Matrix;
**EndFor**
**EndFor**
Combine word's Matrix and visual word's Matrix with LSA;
Return dictionary of medical vocabulary;
**End;**

#### 3.1 Textual Indexation of Medical Reports

To represent a text document as a vector of weights, it is necessary, as a first step, to define a text index terms or vocabulary. To ensure this, we apply a lemmatization which is the process of assembling together the different transformed forms of a word so they can be investigated as a solitary item, and then remove stop words for all man- aged medical text documents. Indexation is then performed using the software provided by Lemur project [23]. Each document is represented, following the model of Salton [3], as a weight vector (1):

$$d_i^T = (w_{i,1}, ..., w_{i,j}, ..., d_{i,|T|}, ...,) \quad (1)$$

where *wi,j* represents the weight of the term *tj* in a document di . This weight is computed as the product of two factors $tf_{i,j}$ and *idfj* . Note that $tf_{i,j}$ factor corresponds to the frequency of appearance of the term *tj* in the document di and *idfj* factor measures the inverse of the frequency of the word present in the related corpus. Consequently, the weight *wi, j* is even higher than the term *tj* is frequent in document di and uncommon throughout the corpus. Note that, for the computation of *tf* and *idf* factors, we use the formulation proposed by authors in [24]. The variable $tf_{i,j}$ is defined as follows (2) :

$$tf_{i,j} = \frac{k_1 n_{i,j}}{n_{i,j} + k_2 \ (1 - b + b) \ \frac{|d_i|}{d_{avg}}} \quad (2)$$

where $n_{i,j}$ represents the number of occurrences of a term $t_j$ in a document $d_i$, $|d_i|$ represents the document $d_i$ size and $d_{avg}$ describes the average size of all documents belonging to the corpus. The elements $k_1$, $k_2$ and $b$ are three constants that have the respective values 1 (one), 1 (one) and 0.5.

**3.2 Visual Indexing of Medical Reports**

The representation of the visual modality in medical reports is performed in two phases: The first consists of the creation of a visual vocabulary V and the second is concerned with the medical image representation using the fashioned vocabulary.

The vocabulary V associated with the visual modality is obtained using the bag of words model [25]. The process, of doing accomplishing this task, involves three main steps: The first consists on the selection of regions or points of interest, the second is concerned with the description determination which per by computing a regions' or points' descriptors, and the third consists of the combination of the descriptors' classes establishing the visual words.

We use two different approaches for the first two steps. The first step is based on the characterization of images by regions of interest which are detected using the affine invariant form descriptor for maximally stable extremal regions or MSER [26]. Categorized images are then represented by their bounding ellipses delimiting regions [27] which are then designated by the descriptor Scale Invariant Feature Transform or SIFT [11]. For image matching and recognition, SIFT structures are first taken out from a set of reference images and then gathered into a specific database. A new image is matched separately by comparing each of its features to those related to images stored in the database. Note that, discovering aspirant matching features is based on Euclidean distance of the associated feature vectors.

For the second step, the clustering is performed by applying the k-means algorithm the set of descriptors to obtain *k* cluster descriptors where each cluster center corresponds to a visual word. The *k*-means is an unsupervised clustering algorithm that classifies the input images into manifold classes based on their characteristic distance from each other. The algorithm undertakes that the data features form a vector space and attempts to find regular clustering in them.

The representation of an image using the vocabulary previously defined is used to compute a weight vector $d^V_i$ exactly as what is usually performed for a textual modality. For visual words associated with the image, we first compute the descriptors of the points or regions of interest of the image, and then we associate to each descriptor, the nearest word belonging to the vocabulary, in the sense of the Euclidean distance.

**3.3 Combination of Modalities with Latent Semantic**

In this section, we show how the two vocabularies are combined using the technique of latent semantic. Latent Semantic Indexing or LSI was used in the field of information retrieval [28] and [29]. It is an indexing and retrieval method that uses a technique called Singular Value Decomposition SVD to identify patterns related to the relationships between the terms and concepts contained in an unstructured collection of texts. LSI is based on the principle that words that are used in the same contexts tend to have similar meanings. A key feature of LSI is its ability to extract the conceptual content of a body of text by establishing associations between those terms that occur in similar contexts. This technique is intended to reduce the indexing matrix into a new space sensible dimensions which tend to express more ''*semantic*'' by correlating semantically related terms that are latent in a collection of texts.

This reduction, envisioned to show the hidden semantic relationships in the co-occurrence, allows for example to reduce the effects of synonymy expressing the state of being a synonym and polysemy conveying the capacity for a sign such as word to have multiple related meanings. It is also used to index without translation or dictionary use, parallel corpora; i.e. documents in different languages, but supposed to be translations of each other.

Technically, the LSI method is a business process of a matrix *M* representing the co-occurrence between terms and documents. Indeed, it is considered as the *SVD* of the matrix *M*, where $M_{i,j}$ describes the occurrences of term *i* in document *j* . The goal is to

compute the matrices $U$, $\Sigma$ and $V$ such that:

$$M = U \sum V^t \quad (3)$$

Where

- $U$ is the matrix of eigenvectors of $MM^t$
- $V^t$ is matrix of eigenvectors of $M^t M$
- $\Sigma$ is the diagonal matrix of singular values $r \times r$

This transformation allows the representation of the matrix $M$ as a product of two different sources of information: The first matrix $U$ related to documents and the second matrix $M$ associated to the terms.

Using the $k$ largest eigen values of $\Sigma$ and truncating the matrices $U$ and $V$ accordingly, we obtain an approximation of rank $k$ of $M$:

$$M_k = U_k \sum_k V_k^t \quad (4)$$

where $k < r$ is the dimension of the latent space. This reduction in size allows to capture information considered as important and eliminate the less significant information regarded as a noise produced by the redundant information such as synonymy and polysemy.

It should be noted that the choice of the parameter $k$ is difficult because it must be large enough not to lose information, and small enough to play its role of redundancy reducer.

### 3.4 Automatic Medical Image Annotation

The automatic medical image annotation is quite similar to the indexation process described in the previous section. As input, it has a medical image to be annotated and the purpose is to find a correlation between the textual and the visual information to be able to automatically assign a new textual annotation to the new image. To perform this task, we need to compute the joint probability between words and visual terms. We adopt for this an approach similar to the method described in [30] and which is enhanced by non-negative matrix factorization to engender multimodal image representations that incorporate visual features and text information. It is also able to determine a set of latent factors that correlate multimodal data in the same representation space. Algorithm 4 illustrates the process of medical image annotation.

**Algorithm 4. Bag of Visual Word Representation;**

**Input :** Unannoted image;

**Output :** Annotation words;

**Begin**

Detect the points of interest;

Describe the points of interest;

Associate textual and visual words descriptors;

Generate a histogram $Q$;

**Foreach** image $I$ in the set of annotated images do

Match $Q$ with histogram $Pi$ of image $I$;

Determine scorei and store it in the score table;

**EndFor**

Choose the best score $I$;

Determine annotation words associated to the best score;

Return annotation words and associate them to the new image;

**End.**

For an unannotated medical image a vector $Q$ is constructed; then Algorithm 4 attempts to determine, according $Q$, the most similar image to the unannotated image using the vector model with a similarity measure. In fact, the more the two representations contain the same elements, the more likely they provide a high value of the representation of the same information. The computation of the relevance value of an image to the query Retrieval Status Value or RSV is determined, based on the similarity of the query vector, using the cosine function [30] as follows.

$$RSV(Q,D) = \frac{\vec{Q}.\vec{D}}{|\vec{Q}| \, X \, |\vec{D}|} = \frac{\Sigma w_{iQ} \, X \, w_{iD}}{\sqrt{\Sigma w_{iQ}^2} \; X \sqrt{w_{iD}^2}}$$

The results are then sorted in a descending order; the image that has the highest score is retained. We attribute terms associated with the image that has the best score to the considered unannotated image.

## 4. Experimental Evaluation

### 4.1 Data Test and Evaluation Criteria

The relevance of the proposed approach is evaluated on a collection extracted from more than a thousand medical reports including radiology images classified in five categories: Thorax, abdomen, lumbar spine, pelvis and skull. Each report is obviously composed not only of radiological images, but also textual documents. We processed five types of reports for each type of radiological images; Figure. 1 presents an example of radiological images handled in our carried experiment.

We started by the reports' pre-indexing using the *tf.idf* combined with visual bag of visual words' techniques. To determine the co-occurrence between the textual and visual descriptors and also to reduce the semantic gap between these two descriptors, we use the latent semantics matrix. At this level, each region of radiological image is labeled by textual words representing each block of them. To evaluate the automatic annotation of a radiological images' process, we use the criterion of Mean Average Precision (MAP), which is a standard criterion in information retrieval used to evaluate the relevance of the obtained results.

### 4.2 Results

A preliminary step of our experiment is to index the medical reports using the model that combines the textual and visual descriptions reduced by latent semantic. Table 1 summarizes the values obtained for each MAP testing. We note that our approach works well on images of radiology of the skull rather than on the other types of images such as Thorax, Abdomen, Lumbar spine and Basin. These global observations are confirmed by the curves of precision/ recall shown in Figure. 1

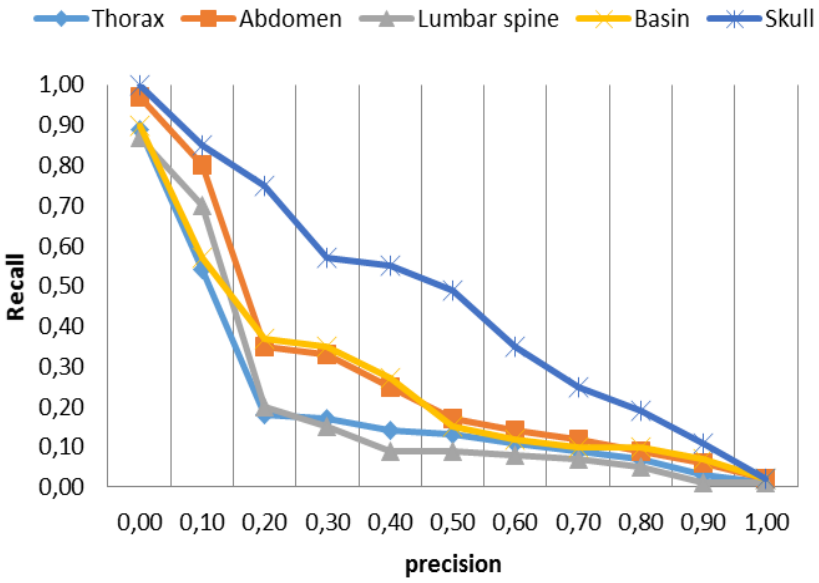


Figure 1. Curve of precision/ recall for different types of radiological images

| Type of radiographic image | MAP |
|---|---|
| Thorax | 0.1057 |
| Abdomen | 0.1326 |
| Lumbar spine | 0.2554 |
| Basin | 0.2826 |
| Skull | 0.3359 |

Table 1. Average Precision Results Obtained For Different Types of Radiological Images

## 5. Conclusion

In this paper, we proposed a new approach for semantic automatic annotation of medical images. It used the bag of words model to characterize the visual content of the medical image combined with text descriptors based on term frequency–inverse document frequency technique and reduced by latent semantic to extract the co-occurrence between text and visual terms. A medical report contains many elements such as medical images accompanied by text descriptions. In a first phase, we are interested to index texts, while minimizing the maximum indexing field, and to extract all relevant terms using a medical thesaurus. In a second phase, medical images are indexed while recovering areas of interest which are invariant to change in scale such as light and tilt. To annotate a new medical image, we use the bag of words model to recover the feature vector. The LSA technique significantly improves the quality of the annotation by combining different visual feature (i.e. local descriptor and global feature), and reduce the complexity computations on large matrices while maintaining good performance. Indeed, we use the vector space model to retrieve similar medical images from the database training. The computation of the relevance value of an image according to a query image is based on the cosine function. To evaluate the performance of our proposed approach, we present an experiment carried out on five types of radiological imaging. The results showed that our approach works efficiently, especially with more images taken from the radiology of the skull.

In future work, we strive to improve understanding the user expect label and to ameliorate the accuracy of the unlabelled images in the training database by testing further image descriptors. Furthermore, we plan to implement semantic web services to communicate with other information resources to acquire relevant biomedical information from distributed heterogeneous resources and then to test our framework on larger sets of annotated data.

## References

[1] Wang, X., Feng, D. (2009). Image registration for biomedical information integration, In Data mining and medical knowledge management: cases and applications, Information Science Reference, *Hershey*, p. 122–136.

[2] Kim, J., Kumar, A., Cai, T., Feng, D. (2011). Multi-modal Content Based Image Retrieval in Healthcare: Current Applications and Future Challenges, In Joseph Tan (Eds.), New Technologies for Advancing Healthcare and Clinical Practices, Pennsylvania, USA, 44–59.

[3] Salton, G.., Wong, A., Yang , C. (1975).A vector space model for automatic indexing, *Commun ACM* 18 (11) 613–620.

[4] Sivic, J., Zisserman, A. (2003). Video Google: a text retrieval approach to object matching in videos, *In*: Proceedings of the international conference on computer vision, 2, p. 1470–1477.

[5] Selvi, S., Kavitha , C. (2014). Radiographic medical image retrieval system for both organ and pathology level using bag of visual words, *In J Eng Sci Emerg Technol*, 6 (4) 410–416.

[6] Caicedo, J., Cruz-Roa, A., Gonzalez Fabio, A. (2009) .Histopathology image classification using bag of features and kernel functions, *In*: Conference on Artificial Intelligence in Medicine, *Ser. Lecture Notes in Computer Science*, 5651, 126–135.

[7] Jégou, H., Douze, H., Schmid , C. (2010). Improving bag-of-features for large scale image search, *In J Comput Vision*, 87 (3) 316–336.

[8] Rahman, M., Antani, S., G. (2011). Thoma Biomedical cbir using bag of keypoints in a modified inverted index. In International symposium on computer-based medical systems, ser. CBMS'11, 1–6.

[9] Wang, J Y., Li, Y P., Zhang, Y., Wang, C., Xie, H. L., Chen, G. L., Xiao, X. (2011). Bag-of-features based medical image retrieval via multiple assignment and visual words weighting, *IEEE Trans Med Image*, p. 3–30

[10] Wang, J., Li, Y., Zhang, Y., Xie,H., Wang, C. (2011). Boosted learning of visual word weighting factors for bag-of-features based medical image retrieval, *In*: Image and graphics (ICIG), 6$^{th}$ International Conference on, p. 1035–1040.

[11] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints, *In J Comput Vision*, 60 (2) 91–110.

[12] Avni, U., Konen, E., Sharon, M., Goldberger, J. (2011) .X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words, *IEEE Trans Med Image*, p. 3–30.

[13] Wu, M., Sun, Q., Wang , J. (2012).Medical image retrieval based on combination of visual semantic and local features, *Int J Signal Process Image Process Pattern Recognition*, 5 (4) 43–56.

[14] Jurie, F., Triggs, B. (2005). Creating efficient codebooks for visual recognition, *In*: ICCV'05 10$^{th}$ IEEE international conference on computer vision, p. 604–610.

[15] Yang, J., Jiang, Y., Hauptmann, A. , Ngo , C. (2007). Evaluating bag-of- visual-words representations in scene classification, *In*: MIR'07: *International Workshop on Multimedia Information Retrieval*, p. 197–206.

[16] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, *In*: 5$^{th}$ Berkeley symposium on mathematical statistics and probability, p 281–297.

[17] Diday, E. (1971).Une Nouvelle Méthode en Classification Automatique et Reconnaissance de Formes: La Méthode des Nuées Dynamique, Revue de Statistique Appliquée 19 (2), p. 19–33.

[18] Mikolajczyk, K., Schmid, C. (2004). Scale & affine invariant interest point detectors, *Int J Comput Vision*, 60 (1) 63–86.

[19] Nowak, E., Jurie, F., Triggs, B. (2006).Sampling strategies for bag-of- features image classification, *In:* ECCV'06: 9$^{th}$ European conference on computer vision: workshop on statistical learning in computer vision, p. 490–503.

[20] Lazebnik, S., Schmid, C., Ponce, J. (2006) . Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, *In*: CVPR'06: *IEEE computer society conference on computer vision and pattern recognition*, p. 2169–2178.

[21] Cao, Y., Wang, C., Li, C, Z., Zhang, L. (2010). Spatial-bag-of-features, *In*: CVPR'10: 23$^{rd}$ *IEEE conference on computer vision and pattern recognition.*

[22] Albatal, R., Mulhem, P., Chiaramella, Y. (2010). Phrases Visuelles pour l'Annotation Automatiques d'Images, CORIA'10. *In*: 7e Conférence en Recherche d'Information et Applications, 3–18.

[23] Lemur. (2013). http://www.lemurproject.org/

[24] Robertson, S., Walker, S., Hancock-Beaulieu, M., Gull, Lau, M. (1994) .Okapi at Trec-3, *In:*Text retrieval conference, p .21–30.

[25] Csurka, G. ., Dance, C., Fan, L., Willamowski , J., Brayn, C. (2004). Visual categorization with bags of keypoints, *In*: ECCV'04 Workshop on statistical learning in computer vision, p. 59–74.

[26] Matas, J. , Chum, O., Martin, U., Pajdla, T. (2002). Robust Wide baseline stereo from maximally stable external regions, *In*: Proceedings of the British machine vision conference (BMVA) p. 384–393.

[27] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir , T., Van Gool , L. (2005).A comparison of affine region detectors, *Int J Comput Vision* 65, p. 43–72.

[28] Deerwester, S., Dumais, S., Furnas, G.., Landauer, T., Harshman, R. (1990). Indexing by latent semantic analysis, *J Am Soc Inf Sci*, 41 (6) 391–407.

[29] Landauer, T., Foltz, P., Laham, T. (1998). Introduction to latent semantic indexing, *Discourse Process*, 25 (5) 259–284.

[30] Caicedo, J., Ben-Abdallah, J., Gonzalez, F., Nasraoui, O. (2012). Multimodal representation, indexing, automated annotation and retrieval of image collections via non-negative matrix factorization, *Neurocomputing*, 76 (1) 50–60.