A diffusion approach for interactive image retrieval

Hassan Tabout, Youssef Chahir, Abderrahman Sbihi

A diffusion approach for interactive image retrieval

H. Tabout *, Y. Chahir **, A. Sbihi **

* *LASTID, Ibntofail University, Morocco*

h.tabout@ieee.org

** *GREYC - CNRS UMR 6072*

Université de Caen, France

youssef.chahir@info.unicaen.fr

*** *Abdelemalek Essaadi University, Morocco*

sbihi@cari-info.org

RÉSUMÉ.

ABSTRACT. *We study in this paper the problem of using multiple-instance semi-supervised learning to solve image Relevance feedback problem. Many multiple-instance learning algorithms have been proposed to tackle this problem; most of them only have a global representation of images. In this paper, we present a semi-supervised version of multiple instance learning. By taking into account both the multiple-instance and the semi-supervised properties simultaneously. A novel graph-based diffusion algorithm is developed, in which global and local information are used. Experimental results show promising results of the proposed method for a test database containing more than 2000 color seaweed images.*

MOTS-CLÉS : *Bouclage de pertinence, diffusion, Apprentissage multi-instance, images d'algues*

KEYWORDS: *Relevance Feedback, Diffusion, Multi-Instance Learning, Seaweed images*

1. Introduction

The present work is interested in indexing and retrieving algae images. Seaweeds or macroscopic marine algae, one of the important

marine living resources could be termed as the futuristically promising plants. These plants have been a source of food, feed and medicine. Agar, Carrageenan and Alginate are popular examples of products extracted from seaweeds. These have been used as food for human beings, feed for animals, fertilizers for plants and source of various chemicals. Seaweed products are used in our daily lives in one or the other way. For example, some seaweed polysaccharides are employed in the manufacture of toothpastes, soaps, shampoos, cosmetics, milk, ice creams, meat, processed food, air fresheners and a host of other items [1, 2]. However, many species are deemed to be harmful [3]. The aim of this study is to manage and to use multimedia metadata to facilitate access to our biodiversity heritage. This is done by considering, on the one hand, the tools of analysis and image processing for the description of these images' contents and on the other hand, the installation of a search and navigation system in such a base.

Recently, with the rapid development in various multimedia technologies, large collections of digital pictures have sprung up easily and users would like to retrieve and browse these collections. Consequently, Content-Based Image Retrieval (CBIR) has gained significant interest in computer vision field [4, 5]. CBIR systems rely on low-level features automatically extracted from image, such as color, texture and shape, to retrieve relevant images most similar to a query. However, Such approaches suffer from the fact that the comparison is usually performed using features extracted automatically from each image. These features should comply with the human perception. This requirement is a difficult challenge ; the difficulty comes from the semantic gap between low level image representation and higher level concepts by which human understand images. Since an image contains several regions of interest ; each region may have different contents and represent different semantic meaning. It is basic then, to segment an image into blobs and extract visual features from each blob. Hence, many local image indexing methods have been proposed [5, 6, 7] to alleviate the problem of the semantic gap. However, such methods may not be sufficient, since they do not take into account human judgement. An ideal image retrieval system from a user perspective would involve what is referred as semantic retrieval. The straightforward solution is to annotate each image manually with keywords and then search on those keywords using a text search

engine. The underlying principle of this approach is that keywords can capture the semantic content of images more precisely, and thus provide better means to organize and search an image database. However, manual annotation is not scalable and very expensive when the volume of data becomes important.

Relevance feedback algorithm tries to learn the relationship between the content of an image and its semantic meaning. Therefore, it seeks to retrieve images that best describe the visual content of the query image. After retrieval sessions, a mapping between semantic meaning and visual content is learned and the system dynamically learns the user's intention and gradually boosts its retrieval performance. In this paper, we formulate image relevance feedback as a supervised learning problem and present a novel solution using Multiple-Instance Learning (MIL). In our framework, images are viewed as bags, each of which contains a few instances corresponding to the segmented image regions.

The remainder of the paper is organized as follows. The next section summarizes related works. Section 3 describes the used segmentation approach. MIL-based bag feature representation is presented in section 4. Section 5 provides a brief review of the overall algorithm. In section 6, simulation results and evaluation are provided, and finally, concluding remarks are offered in section 7.

2. Related works

Image segmentation has attracted much attention in the computer vision applications, as a prior step for the recognition of different image elements or objects. Several algorithms have been introduced to tackle this problem. In this paper we provide a brief review of the related work that is most relevant to our approach. Targhi et al. have proposed in [8] a novel segmentation approach based on the Eigentransform. The transform provides a measure of roughness by considering the eigenvalues of a matrix which is formed by inserting the grey values of a square patch around a pixel directly into a matrix of the same size.

Recently, graph-based approaches have been gained significant interest to image segmentation [9, 10]. The basic idea of the underlying

approaches is the construction of a weighted graph G = (V, E) whose each node represents a pixel of the image and the weight of an edge is some measure of the dissimilarity between the two pixels connected by that edge (e.g., intensity, color, or some other local feature). This graph is partitioned into components in a way that minimizes some specified cost function of the vertices in the components and/or the boundary between those components. The procedure used here, is based on a diffusion model. We define a random walk through a local window by assigning a transition probability to each link. Chahir et al. have used diffusion on graph to recognize facial expressions [11] . In [12], the authors have used a diffusion kernel based nonlinear approach to identify the modality relationship between visual and text modalities. The authors in [13, 14] have used random walks approach to interactive image retrieval frameworks.

Much works on multiple-instance learning have been developed to object-based image retrieval problem [15]. Multiple-instance learning was first introduced by Dietterich et al. [16] in the drug activity prediction problems, and then broadly used in content-based image retrieval and image annotation [15, 17, 18]. In multi-instance learning, the training set is composed of many bags each contains many instances. The goal of a MIL algorithm is to generate a classifier that will classify unseen bags correctly. A bag is positive if at least one instance in it is positive and negative if all the instances in it are negative. In region based image retrieval, Images are segmented into small regions, and each region represents an instance. Under MIL setting, each query image which contains the target object is considered as a positive bag, while the other negative labeled images are considered as negative ones. Objects containing or within the target object are considered as positive instances, while the others are negative instances.

There have been much works on applying semi-supervised learning (SSL) to solve practical problems by propagating label information. Zhu has presented in [19] a detailed semi-supervised learning survey. However, most of them are only interested in single-instance setting, whereas image retrieval is often presented as a multiple instance learning problem. In this paper we have considered both multiple-instance and semi-supervised properties.

Related to multiple-instance semi supervised learning, Rahmani and Goldman [20] have presented a Graph-Based Semi-Supervised Learning Method to address object-based image retrieval task, that transforms any MI problem into an input for a graph-based single-instance semi supervised learning method that encodes the MI aspects of the problem simultaneously working at both the bag and point. In [21] Zhou et al. have proposed a Multiple Instance learning by Semi-Supervised Support Vector Machine (MissSVM) algorithm, which tackles multi-instance problems using semi-supervised learning techniques, in particular, a special semi-supervised SVM.

3. Texture Segmentation

Region based representation of images is an effective way to improve accuracy in image retrieval. Global features do not always represent salient objects seen in an image. Such features are computationally effective but provide rough representation of the image content. So, higher retrieval performance will be more effective with taking into account more precise information. Our approach to segmentation uses a spectral approach based on a random walks process.

Given a graph and a starting node, we move to a neighbor of it at random ; then we select randomly a neighbor of this node and we move to it etc. the random sequence of selected nodes is a random walk on a graph. Let $G = (V, E)$ be a connected graph with n nodes and m edges. A random walker on graph G moves from a node u to a neighbor node v with the probability :

$$p(u, v) = \frac{w(u, v)}{d(u)} \tag{1}$$

Where : w (u, v) is the edge weight between vertex u and v and d (u) is total weight of edges incident to vertex u. The sequence of random vertices (v_t : t=0, 1, 2...) is a Markov chain. We denote by P the matrix of transition probabilities of this Markov chain. So :

$$P = D^{-1}W \tag{2}$$

Where : D denote the diagonal matrix with D (u, u) = d (u) and W is the similarity matrix.

A classical construction of weight matrix is based on Gaussian kernel of 0-mean and variance σ. Let d (u, v) be some general distance measure between nodes u and v, then the weight w(u,v) will be computed as follows :

$$W(u, v) = \exp(-\frac{d(u, v)}{\sigma^2}) \qquad (3)$$

The goal here is to extract local and non local information. The graph is then constructed from the following local and non-local definition :

$$W(u, v) = \begin{cases} \exp\left(-\frac{\|u-v\|_2}{\sigma_1^2} - \frac{\|F(u)-F(v)\|_2}{\sigma_2^2}\right) & \text{if } u, v \in V_{pXp}(u); \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

$V_{pXp}(u)$ is a square window of size p around u. The feature vector F is a texture descriptor based on the mean of anisotropy, contrast and polarity throughout the window.

We proceed by computing the eigenvalues of the transition matrix P and sorting them in decreasing order. Finally we compute texture descriptor at each pixel by using Eigen transform.

$$\Gamma(G) = \sum_{i=1}^{w} \lambda_i \qquad (5)$$

Where w is the number of Eigen values.

In the context of image segmentation, we have proposed a novel method based on the trace of the Markov matrix. A brief description of the proposed algorithm is summarized as follows : Given an image X

1) We consider a square neighborhood of size p x p pixels around each pixel.

2) We compute a texture descriptor F throughout the window and we assign this descriptor to the pixel. This descriptor will be used in matrix similarity (equation 4).

3) We construct the Markov Matrix P

4) We measure texture in each pixel by computing the trace of the matrix P.

To test the performance of each algorithm several criteria were used. Respectively, tp, tn, fp, and fn, stand for the number of pixels being labelled as true positive, true negative, false positive, and false negative.

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \tag{6}$$

$$Precision = \frac{tp}{tp + fp} \tag{7}$$

$$Recall = \frac{tp}{tp + fn} \tag{8}$$

Experimental results are presented in Table 1. The algorithms performed comparably, in terms of pixel accuracy, precision and recall. However, the Eigen Transform method require more time than the proposed algorithm.

Tableau 1 – *Average performance on tested images.*

Method	Accuracy	Precision	Recall
Sum of Eigen values	0.70	0.74	0.75
Trace of P	0.69	0.74	0.74

In order to compare our method with an existing one, we have chosen the technique of Shi and Malik (Ncut). We have processed a group of images with our segmentation method and compared the results to Ncut algorithm. The Ncut algorithm uses the optimal parameters given by authors [10]. Figure 1 shows a comparison between the proposed approach and Ncut algorithms. According to the segmentation results on these images, we note that our algorithm best localize regions in the processed image compared to the Ncut method.
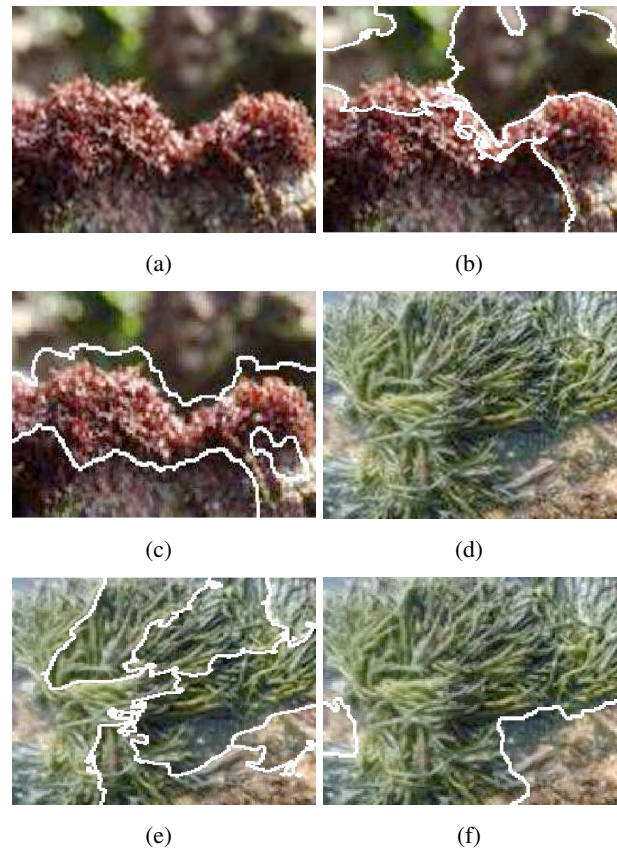
Figure 1 – *A comparison between the proposed approach and Ncut algorithms.(a) and (d). Original images. (b) and (e) the output of Ncut algorithm. (c) and (f) the output of our algorithm.*

4. Graph-based multi instance semi supervised learning

The goal of this paper is to predict the labels of the unlabeled images. Semi supervised learning is very useful in such problem. In this paper, we presented a graph-based learning algorithm. The key to semi-supervised learning problems is the prior assumption of consistency, which means : (1) nearby points are likely to have the same label ; and (2) points on the same structure (typically referred to as a cluster or a manifold) are likely to have the same label [22]. The first assumption is

local, while the second one is global. We have integrated both local and global information during learning.

In this section we will introduce the relevance feedback procedure in detail. We will describe a Multi-Instance and Semi-Supervised Learning method by combining local and global information.

4.1. *Local Representations*

In multi-instance learning, the training set is composed of many bags each includes many instances. A bag is positively labeled if it contains at least one positive instance ; otherwise it is labeled as a negative bag. The task is to learn some concept from the training set for correctly labeling unseen bags.

In region-based image retrieval, each region is an instance, and the set of regions that comes from the same image can be treated as a bag. We annotate an image as positive if at least one region in the image has the semantic meaning of the requested species. Given an image containing several regions, we can expect that at least one region will correspond to the user interest need even if segmentation may be imperfect. Hence, the image retrieval problem is in essence identical to the MIL setting. Each image in the image database is segmented. Color and texture features of each region are computed. The image is then seen as a bag containing many instances (feature vectors).

Based on the segmentation results, the representative color feature for each region is calculated by the color moments. The representative texture feature for each region is computed by the average anisotropy, polarity and contrast.

4.2. *Global representations*

Inaccurate image segmentation may make the MIL-based bag feature representation imprecise and therefore decrease the retrieval accuracy. We add global-feature to address this problem. In order to compensate the limitations associated with the specific color space and the specific texture representation, we construct the global features in a

different manner as used in creating the regional features. Color histograms represent color features; whereas the representative texture feature is computed by the average energy in each high frequency band after 2-level wavelet decompositions. After the global features of all the training images are obtained, they are fed into another set of a global bag.

5. Algorithm

The interactive system consists in asking the user questions such that his/her responses make it possible to reduce the semantic gap according to the following steps :

Step1. The system compares the query image with each image database using feature vectors. Similarity measurement, in the first retrieval, is carried out by Euclidean distance and the most similar images are returned and displayed by achieving the first retrieval stage.

Step2. The user annotates displayed images as positive or negative items according to his/her interest need.

Step3. The system then predicts the label of the unlabeled images :

Let X denotes the image set $X = \{ x_1, ..., x_l, x_{l+1}, ..., x_n \} \subset \Re^m$ and a label set $L = \{ +1, -1 \}$, the fist l points $x_i (i \leq l)$ are labeled as $y_i \in L$ and the remaining points $x_u = (l + 1 \leq u \leq n)$ are unlabeled. Here y_i is equal +1 if the image x_i belongs to the user's class of interest and $y_i = -1$ otherwise. We define a function F; which assigns a value y_i to image x_i [23] by :

$$F = (I - \alpha \, S)^{-1} y \qquad (9)$$

Where I denotes the identity matrix and α is a parameter in $(0; 1)$;

In graph-based learning, feature vectors are arranged in a weighted undirected graph. The graph is characterized by a weight matrix W, whose elements $Wij >= 0$ are similarity measures between vertices i and j, and by its initial label vector. The commonly used weight is defined by the equation 3.

The proposed algorithm is summarized as follows :

1) We use the segmentation approach described in 3.1, to segment the images in database and obtain the regions of each image. We compute the visual features of every image in the database and build the local similarity matrix W_L by using K-nearest neighbor in order to connect each vertex only to its k-nearest neighbors. We use average Hausdorff distance in equation 3 to compute distance between two bags of images.

2) Extract global features and Construct the global similarity matrix Wg. We use here Euclidean distance in Equation 3. The final similarity is defined as :

$$W = W_L * W_g \qquad (10)$$

3) Construct the matrix $S = D^{-1/2} * W * D^{1/2}$ in which D is a diagonal matrix and $D_{ii} = \sum_{j=1}^{n} wij$

4) labelling the unlabelled images by Computing F (equation 9).

6. Discussion and results

The algae image database contains algae images of various classes : Gelidium, Codium, Blidingia...There are more than 2000 images. Images are segmented using the algorithm described above, only regions larger than a threshold are selected. A 12 dimensional low-level feature vector is extracted from each region, which includes 9 color features and 3 texture features. For global bag, each image is indexed by 54 dimensional feature vector. This vector includes 48 components for color histogram generated in HSV color space and 6 for texture.

To evaluate the proposed algorithm, a total of 40 images, one from each species, were selected as the query images. For each query, the top 16 images were retrieved to provide necessary relevance feedback. Using this method, in the ideal case all the top 16 retrievals are from the same species. The performance was measured in terms of average retrieval rate of the 40 query images, which was defined by [23] :

$$\text{Retrieval rate} = \frac{\text{relevant images}}{\text{class size}} \times 100\%.$$

Table 2 exhibits the average retrieval rate, where *iter* denotes the number of iterations. The following observations were made from the results.

Firstly, the performance with the relevance feedback method after three iterations was substantially better than the noninteractive scheme (iter=0) and the improvements are important. Secondly, after three sessions of interactive learning, our method gave a promising performance : on average 93.54% of the correct images are in the top 16 retrieved.

Tableau 2 – *Average Retrieval Rate (%) For 40 Query image.*

	Average Rate Retrieval (%)
iter=0	70.26
iter=1	90.06
iter=2	92.96
iter=3	93.54

In the second experience we measured retrieval performance for each species. Retrieval effectiveness can be defined in terms of precision and recall rates. A precision rate can be defined as the percent of retrieved images similar to the query among the total number of retrieved images. A recall rate is defined as the percent of retrieved images, which are similar to the query among the total number of images similar to the query in the database.

Figure 2 shows the recall precision graphs correspond to a 4 species : gelidium sesquipedale, codium fragile, Fucus spiralis and blidinga minima after third round of relevance feedback. It is clear from the graphs that the use of proposed approach improves the performance. Moreover, adopting the relevance feedback mechanism presents a slight increase in terms of consuming time.

7. Conclusion

In this paper, we have introduced a novel framework for region based algae image retrieval. We have proposed a method of segmenting images based on a spectral decomposition with a local random walks model. We formulate interactive image retrieval as a semi supervised learning problem and present a novel solution using Multiple-Instances. This allows bridging the semantic gap from the low level description. The user can access directly to the objects of interest, specifically algae species and find images which contain the requested species. The evaluation showed that the proposed approach gives good results according to our test image database.

In the future, we will work on a larger algae database by considering more species ; we plan to present an image annotation system by mining and refining more relevant semantic information and building more suitable connection between image content features and available semantic information.
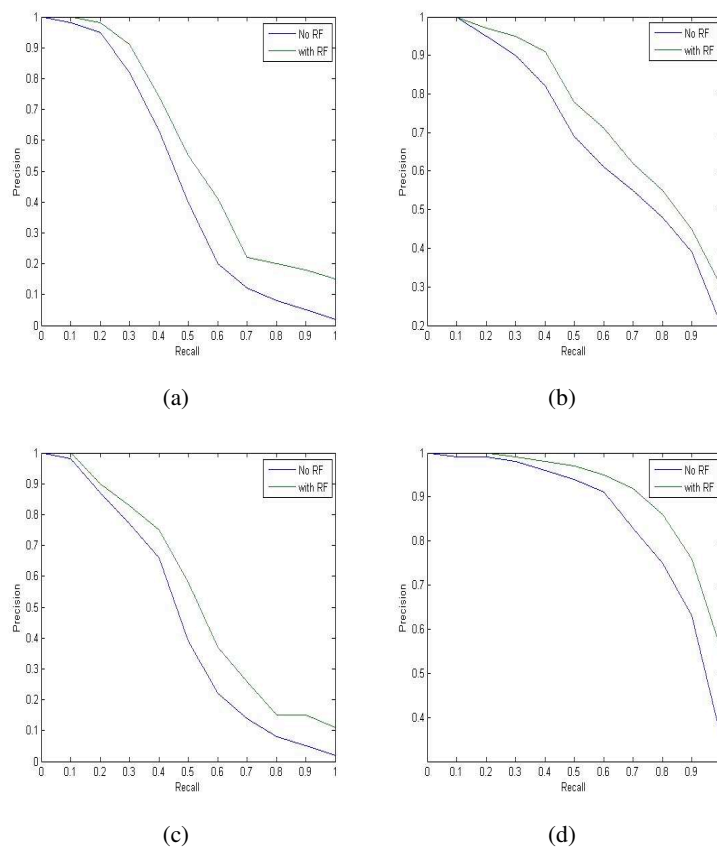
Figure 2 – *Precision and recall graph comparing interactive and noninteractive methods of 4 species : (a) Gelidium sesquipedale, (b) Codium fragile, (c) Fucus spiralis and (d) Blidinga minima.*

Références

[1] V.K. Dhargalkar and N. Pereira,"Seaweed : Promising plant of the millennium" Science and Culture, vol.71 (3-4), pp 60-66, 2005.

[2] S. Ayyappan, "Seaweed Cultivation and Utilization" National Academy of Agricultural Sciences, pp 22, 2003.

[3] R. Akallal, S.Alaoui, T.Givernaud and A.Mouradi " Intoxications alimentaires associes aux efflorescences d'algues marines nocives " Revue Marocaine, 2001.

[4] A.W. M. Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, .Content based image retrieval at the end of the early years, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349.1380, 2000

[5] C.Carson, S. Belongie, H. Greenspan, and J.Malik "Blobworld : Image segmentation using Expectation-Maximization and its application to image querying " IEEE Transaction on pattern analysis and machine intelligence, vol 24, N 8, pp 1026-1038,2002.

[6] J. Ashley et al. Automatic and semiautomatic methods for image annotation and retrieval in QBIC. In SPIE Proc. Storage and Retrieval for Image and Video Databases, pages 24-35, 1995.

[7] Khanh Vu et al." Image Retrieval Based on Regions of Interest" IEEE Transactions on knowledge and data engineering, vol. 15, no. 4, July/august 2003, pp. 1045-1049.

[8] Targhi Tavakoli, A.Hayman, E.Eklundh, J.Shahshahani "The eigen-transform and applications". ACCV (1), pp70-79. 2006.

[9] P. Soundararajan and S. Sarkar, "An in-depth study of graph partitioning measures for perceptual organization." IEEE Trans. Pattern Anal.Mach. Intell. Vol. 25, no. 6, pp. 642-660, 2003.

[10] J. Shi and J. Malik, "Normalized cuts and image segmentation" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, AUGUST 2000.

[11] Y. Chahir, Y. Zinbi et K. Aziz, "Catgorisation des expressions faciales par marches alatoires sur graphe", CORESA 2007.

[12] Rajeev Agrawal et al. "Application of Diffusion Kernel in Multimodal Image Retrieval", MIPR 2007.

[13] H.Sahbi et al." Graph Laplacian for Interactive Image Retrieval", ICASSP, 817-820. April 4 2008.

[14] Jana Urban and Joemon M. Jose. "Adaptive Image Retrieval using a Graph Model for semantic Feature Integration". MIR'06, October 26-27, Santa Barbara, California, USA. 2006

[15] Changhu Wang et al. "Graph-based Multiple-Instance Learning for Object-based Image Retrieval". International Multimedia Conference, pp 156-163. 2008.

[16] Dietterich, T. G., Lathrop, R. H., and Lozano-Perez, T. "Solving the multiple instance problem with axis parallel rectangles". Artif. Intell. Vol 98, pp 31-71. 1997

[17] Changbo Yang and Ming Dong, "region-based image annotation using asymmetrical support vector machine-based multiple-instance learning". Computer Vision and Pattern Recognition (CVPR'06. pp 2057 - 2063. 2006

[18] C. Yang, M. Dong, F. Fotouhi, "Region based image annotation through multiple instance learning", ACM Conference on Multimedia (MM'05), Singapore, pp. : 435-438, Nov., 2005

[19] Zhu, X. "Semi-Supervised Learning Literature Survey". Computer Sciences Technical Report 1530, University of Wisconsin-Madison, 2005.

[20] Rahmani, R. and Goldman, S. A. (2006). "MISSL : multiple instance semi-supervised learning". In Proc. of ICML, 2006.

[21] Zhi-Hua Zhou and Jun-Ming Xu, "On the Relation Between Multi-Instance Learning and Semi-Supervised Learning". IMC, pp. 1167-1174. 2007.

[22] Dengyong Zhou et al. "Learning with Local and Global Consistency". Advances in Neural Information Processing Systems 16, pp 321-328. 2004.

[23] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Machine Intell. vol.18, no. 8, pp. 837-842, 1996.