Survey of the Adequate Descriptor for Content-Based Image Retrieval on the Web: Global versus Local Features

Hichem Bannour* — **Lobna Hlaoua**** — **Bechir Ayeb*****

Départements des sciences d'informatiques,
** Institut Supérieur des Sciences Appliquées et de Technologies de Sousse (ISSATS),*
*** Ecole Supérieure des Sciences et de Technologie de Hammam Sousse (ESSTHS),*
**** Faculté des Sciences de Monastir (FSM), Tunisia.*

*Hichem.Bannour@issatso.rnu.tn**
*Bannour.Hichem@yahoo.com – lobna1511@yahoo.fr** – Ayeb_b@yahoo.com****

ABSTRACT. *The need for efficient content-based image retrieval has increased hugely. Two methods are recognized for describing the content of images: using global features and using local features. In this paper, we propose two methods for image retrieving based on visual similarity. The first one characterizes images by global features, when the second is based on local features. In the global descriptor attributes are computed on the whole image, whereas in the local descriptor attributes are computed on regions of the image. The aim of this paper is to compare global features versus local features for Web images retrieval.*

RÉSUMÉ. *On reconnait actuellement, dans les systèmes de recherche d'image par contenu, deux méthodes pour la description du contenu des images : à travers des attributs locaux ou à travers des attributs globaux. Dans ce papier, nous proposons deux méthodes pour la recherche d'image qui sont basées sur la similitude visuelle. La première caractérise les images par des attributs globaux, alors que la seconde est basée sur les attributs locaux. Concernant le descripteur global, les attributs sont calculés sur l'ensemble de l'image, alors que pour le descripteur local, les attributs sont définis sur les régions de l'image. L'objectif de ce papier est d'évaluer les performances des attributs locaux contre les attributs globaux pour la recherche des images Web par contenu.*

KEYWORDS: *Content-based image retrieval, image segmentation, image features, local descriptor, global descriptor.*

MOTS-CLÉS : *Recherche d'image par contenu, segmentation d'image, attributs d'image, descripteur local, descripteur global.*

1. Introduction

The digit contents are being generated with an exponential speed. As the amount of collections of digital images increases, the problem of finding a desired image in the web becomes a hard task. Therefore, an efficient method to retrieve digital images on the Web is required.

There are two approaches to image retrieval: Text-Based approach and Content-Based approach.

– The former solution is a more traditional approach, which indexes images by using keywords. The keyword indexing of digital images is useful, but it requires a considerable level of effort and is often limited to describe image content.

– The alternative approach, the content-based image retrieval, also called CBIR, indexes images by using the low-level features of the digital images, and the searching task depends on features being automatically extracted from the image.

Most current CBIR systems retrieve images from a collection on the basis of the low-level features of images, such as color, texture, and shape. Almost all these systems are founded on the premise that images can be characterized by global descriptors to retrieve purposes in a database (Flickher *et al.,* 1995, Wu *et al.,* 2004, Quack *et al.,* 2004, Pi *et al.,* 2005). The global descriptor consists of features computed on the whole image. For example, in (Rubner *et al.,* 1997) authors proposed a Histogram search algorithms to characterize an image by its color distribution or histogram, they proposed the earth mover's distance (EMD) using linear programming for matching histograms.

However, in most cases the images represent a scene consisting of different objects (or regions), and therefore, a description of these regions should allow a better representation of the image content. The solution consists in separating the different regions of the image using a segmentation algorithm, then to use the appropriate features calculated on each region of the image to describe (Liu *et al.,* 2000, Jing *et al.,* 2004, Chen *et al.,* 2002). These features constitute the local descriptor. For example, The Stanford SIMPLIcity system (Wang *et al.,* 2001) uses statistical classification methods to group images into rough semantic classes, such as textured-non textured, graph-photograph.

In this paper we propose two methods for content based image retrieval. Our methods describe a given image on the basis of color and textures features, and are based on statistical moments for color characterization and the Tamura features (Tamura *et al.,* 1978) for texture description. Our methods, namely GDIR and LDIR, use respectively global features and local features for image description. Compared to other works, GDIR and LDIR proved to achieve higher accuracy.

Another issue in this work is to evaluate the accuracy of global descriptors versus local descriptors for image characterization and retrieval in the Web domain. In (Shyu *et al.,* 1998), authors compared local and global descriptor for medical image retrieval. They concluded that the empirical evaluation of their current implementa-

tion illustrates that local features significantly improve retrieval performance in the domain of HRCT of the lung. But, their experiments still miss details to compare efficiently the two descriptors.

Motivated by the above considerations and the lack of an accurate comparison in the literature between the two descriptors, we propose in this work to evaluate the accuracy of local versus global descriptor for web image retrieval.

The rest of the paper is structured as follows. In Section 2, we present the features used for image description. In Section 3, we introduce the global image description method. Section 4 illustrates the local image description method. Simulation and retrieval results will be reported in Section 5. The paper is concluded in Section 6.

2. Image Features

In this section, we introduce the image features used by our two methods for images description.

2.1. *Color Features*

The statistical moments is considered to be invariant to image shift, rotation and scale. Actually, moments also represent fundamental geometric properties of a distribution of random variables. In this proposal we used the statistical moments for color description. The used color descriptor is composed by the following attributes:

– Colors expectancy:

$$E_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij} \quad [1]$$

– Colors variance:

$$\delta_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - E_i\right)^2\right)^{\frac{1}{2}} \quad [2]$$

– Skewness:

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - E_i\right)^3\right)^{\frac{1}{3}} \quad [3]$$

Where P_{ij} is the (i,j) $pixelcolor$, N is the total number of pixels in the image.

These values allow to estimate the average color, the dispersion of color values from the average and the symmetry of their distribution in a region of the image (or respectively on the whole image).

2.2. *Textural Features*

In this work, we used Tamura texture features as texture descriptor of an image in the database. Tamura et al. (Tamura *et al.,* 1978) took a different approach based on psychological studies on human visual perception. They developed computational approximations for six different visually meaningful texture properties, namely, *coarseness*, *contrast*, *directionality*, *line-likeness*, *regularity*, *and roughness*. However, only three of the six proposed features correspond strongly to human perception and are widely used. These features are *coarseness*, *contrast* and *directionality* which describe respectively the "coarse vs. fine", "high vs. low" and "directional vs. non-directional" of a textured regions. In this proposal, we use these three described features in both descriptors.

3. Global Image Descriptor

The global image descriptor is composed by color and texture features being computed on the entire image.

The texture features are not always an accurate description of the image because they are computed on the whole image. Therefore, in the retrieval process we provide two alternatives to user, the first one is based on color features, the second is based on combined features (color and texture). When the retrieval based on color is fruitless, the user can use the other alternative. By integrating these two options, retrieval accuracy may be improved significantly.

4. Local Image Descriptor

The local image description is founded on the premise that images can be characterized by attributes computed on regions of the image. To separate the different regions of a given image we used an image segmentation method. So to compute our local image descriptor, we use the SOM algorithm to separate the homogeneous regions, than for each region we calculate the color and texture features described in section 2 - An example of local color descriptor is shown in Figure 3.

4.1. *Image Segmentation by Color Clustering*

Different types of neural networks have been proposed for the segmentation of color image (Dong *et al.,* 2005) (Wang *et al.,* 2003) (Ong *et al.,* 2002). However, SOM has the advantages of nonlinear projection, topology preserving, prominent visualization and rapid convergence, which makes it particularly useful for the color clustering (Kohonen, 1995).

For the broad domain images, such as the images in World Wide Web or in images library, precise object segmentation is nearly as difficult as image understanding. However, semantically precise segmentation is not needed to our system because our approach is insensitive to segmentation. In this work we chose to use the Self Organizing Map "SOM" for image segmentation.
The SOM is structured as a two-layer neural network with a rectangular topology as shown in Figure 1. Three inputs (R,G and B) are fully connected to the neurons on a 2-D plane. Each neuron is a cell containing a template against which inputs are matched. The template is the weight values to the neuron i, which is represented by $w_i = [w_{i1}, w_{i2}, w_{i3}]^T$. The SOM training has the following procedure:

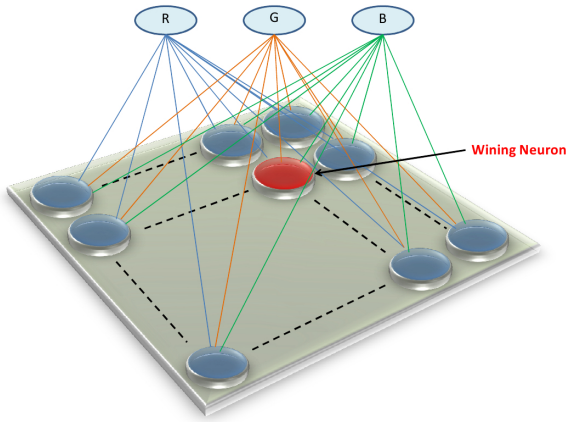

Figure 1. *SOM Topology.*

1) Initialization: Define the SOM map, and set the size of SOM to $n * n$. Set the neighborhood radius to n and the learning rate to 1. Randomly initialize the weight vector. The neighborhood type is Gaussian. The SOM training is successively performed by two phases. The weight vectors of the map are ordered in the first phase, and fine-tuned in the second phase.

2) Input: The input colors are randomly initialized. The image colors are reiteratively used to train the network for few times. During the training, each color point is cyclically chosen from the data set, and presented to all neurons on the map simultaneously.

3) Competitive Process: At time t, color point x(t)=[r(t),g(t),b(t)], T is presented to the network. The winning neuron c is computed with the shortest distance between the color point and weight vectors by the formula:

$$\|x(t) - w_c(t)\| = \min_i \|x(t) - w_i(t)\| \qquad [4]$$

$$where: \quad c = \min_i \|x(t) - w_i(t)\| \qquad [5]$$

4) Cooperative Process: The topological neighbors of winning neuron c are determined by the Gaussian function centered at neuron with the effective scope of $\Re_c(t) = [c_{k-1}, c_k, c_{k+1}]$.

5) Adaptive Process: The weights of winning neuron c and its neighbor neurons are updated within the neighborhood using formula 6 when $k \in \Re_c(t)$

$$w_i(t+1) = w_i(t+1) + \alpha h_{c_i}(t)[x(t) - w_i(t)] \qquad [6]$$

where, $\alpha(t)$ is the learning factor, and $h_{ci}(t)$ is the neighborhood function centered around the winning neuron .

6) Iteration: The next color point is presented to the network at time $t+1$. the learning rate α is decreased to $\alpha(t+1) = \alpha(0)(1 - t/T)$. The neighborhood radius is decreased to $\Re_c(t+1) = \Re_c(t)(2 - t/T)$. The new winning neuron is chosen by repeating the procedure from step 2 until all iterations have been made $t = T$.
T is the number of color points for training.

5. Experiments

Our methods has been implemented with a general-purpose image database including about 100 000 pictures, which are stored in JPEG format with size 384*256 or 256*384. To perform our proposal results, we evaluate retrieved images on the basis of local descriptor and global descriptor. The remaining experimental results are evaluated in terms of precision and recall. We used also the accuracy measurement to compare our results, which is the mean of recall and precision. The assessments are giving according to 10 classes, each containing 100 pictures, defined in the COREL database (COR 1999).

5.1. *Simulation*

Figure 2 demonstrates the results of image segmentation. Figure 2(a) represents the input image. 2(b) shows the obtained image after a segmentation by SOM with a map sized to 16*16 and 2(c) illustrates the obtained color classes by the same network. In 2(d) image segmentation by a SOM map of to 2*2 and in 2(e) the obtained color classes by the same network.

For the rest of our experiments, we used the following parameters for SOM:

– SOM size is 2*2.

– The neighborhood radius $\Re_c(t)$ is 1.

– the learning rate α is 0.8 and decreasing with time following this formula: $\alpha(t+1) = \alpha(0)(1 - t/T)$.

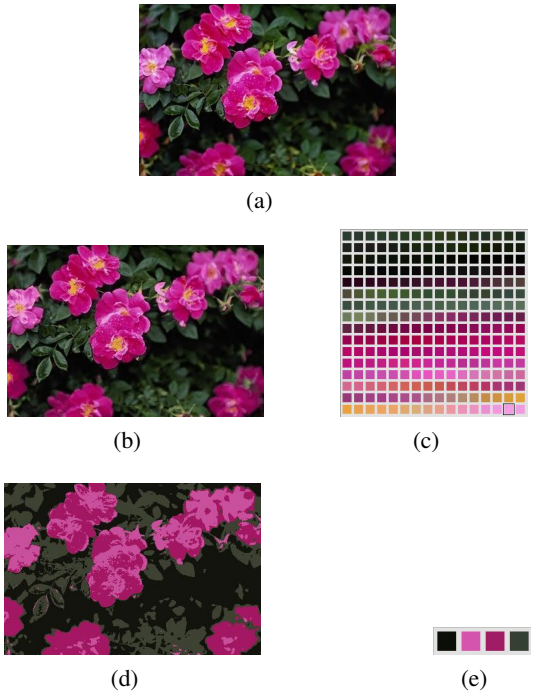

Figure 2. *Image segmentation using SOM*

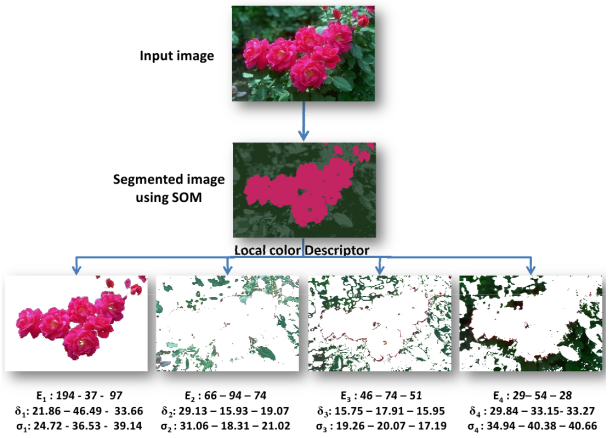

Figure 3. *An example of local color descriptor.*

In figure.3 we illustrate the process to obtain the local color descriptor. After image segmentation, we compute for each region the colors expectancy E_i, the colors variance δ_i and the Skewness σ_i, and put these values in the image vector descriptor.

5.2. *Similarity Measures*

In this section, we describe the similarity measures that we used for image retrieval. Each image in the database is represented by a vector descriptor containing both color and texture features which have been described above. In the retrieval process, for a given query we evaluate the relevance of each image according to a distance measurement defined as follows:

$$d = \sqrt{\sum_{X,X' \in F} \sum_{i=1}^{n} \sum_{j=1}^{3} (\min_{k=1}^{n}(X_{ij} - X'_{kj}))^2} \quad [7]$$

where :

– F is the set of image features $F = \{E_i, \delta_i, \sigma_i, C_i, Co_i, D_i\}$
with E_i is the Expectancy (Equation.1), δ_i is the Variance (Equation.2), σ_i is the Skewness (Equation.3), C_i is the Contrast, Co_i is the Coarseness and D_i is the Directionality.

– X and $X^{'}$ are features of respectively the query image and the target image.

– n is the number of regions in the image.

– 3 is the size of color components (R,G,B).

This equation corresponds to the Euclidean distance, which allows to measure the similarity of two images according to the used features F, on each color components (R,G,B). For both images, we compute the distance between features computed on a region of the query image, and the most close region on the target image. This distance is applicable to the local descriptor and the global descriptor. In global descriptor the number of regions in the image is 1, while in local descriptor the number of regions is n.
The retrieval result is a set of images ranked according to the scores given by the above equation.

5.3. *Connection to other works*

Because we have access to the SIMPLIcity system (Wang *et al.,* 2001), we compare the accuracy of our methods to it using the same COREL database. SIMPLIcity had been compared with the original IBM QBIC system and found to perform better. Also, we compare our methods to the EMD-based color histogram system (Rubner

et al., 1997) to prove that statistical moments work faster and give better results than histograms. To qualitatively evaluate the accuracy of our methods over the image database, we randomly pick 10 query images, namely, Africa people, beach, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and food.

To perform a fair comparison to the other works, we used the same experimental protocol than the one of SIMPLIcity. Precision within the first 100 retrieved images was computed for our methods, SIMPLIcity and EMD-based color histogram. Recall was not used in the SIMPLIcity experiments, so in this experiment too.

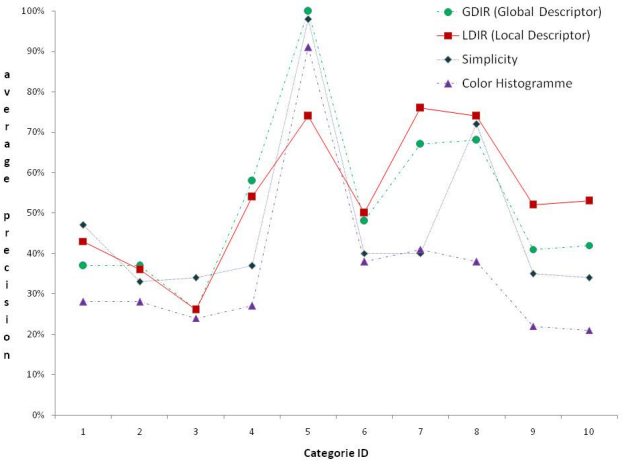

Figure 4. *Comparing global and local descriptor with SIMPLIcity and color histogram methods on average precision.*

Figure 4 shows the performance of our methods when compared to the SIMPLIcity and the EMD-based color histogram systems. Clearly, the color histogram-based matching systems perform much worse than the GDIR and LDIR systems in almost all image categories. To compute the feature vectors over 100 000 color images of size 384*256 requires approximately 120 minutes for the GDIR, and 652 minutes for LDIR, making a computation time per image of 0.072 sec for GDIR and 0.391 for LDIR. So, it is clear that our methods based on statistical moments work faster and give better results than the histogram based method.

Except for the Africa people, buildings and dinosaurs category, our methods has achieved better results than SIMPLIcity. For the other categories the difference between our methods and the other systems is quite significant. On average, the precision of GDIR and LDIR are higher than those of SIMPLIcity and EMD-based color histogram, and respectively equal to 52%, 54%, 47%, and 36%.

5.4. *Local descriptor vs. global descriptor*

Figure 5 shows that the accuracy of the local descriptor follows a linear curve, while the global descriptor curve varies according to the sought image. Thus, the accuracy of the global descriptor depends on images in the database and the query used for retrieving propose, while the local descriptor is more robust to these criteria. Also, the plot shows that the local descriptor accuracy is higher than the global one, except to images representing a bus, this is because of the different backgrounds of this category of images.

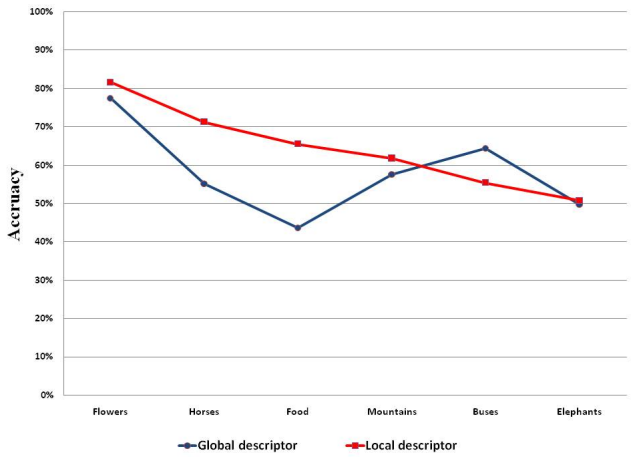

Figure 5. *Evaluation of descriptors accruacy on the COREL image base.*

Table 1 shows the results obtained for each of the seven category we defined into the COREL Database. The "size" row is the number of correct images in the database, the "Res" rows are the number of images returned by our methods when performing a query, the "Pre" rows are the precision, the "Rec" rows are the recall and the "Acc" rows represent the accuracy. Precision rate allows to estimate the relevant images ratio and the recall rate allows to estimate the ratio of relevant images omission.

A comparison between the global descriptor results and those of the local descriptor shows that, in the most of cases, the local descriptors can improve significantly the precision of the retrieval result. However, the recall is almost the same for both descriptors. Note that the accuracy of the local descriptor is also better than the global one. The average values confirm these findings more clearly.

However, the experiment with synthetic images (buses) shows that the global descriptor allows a better retrieval result. The system achieves an accuracy of 99% with the global descriptor, when it achieves an accuracy of 76.37% with the local descriptor.

From these results, we can see that the local descriptor achieves a higher accuracy when the desired image possesses several meaningful regions. However, when the

Table 1. *Obtained results on a subset of the COREL database using global descriptor and local descriptor.*

Category	size	Global Descriptor results				Local Descriptor results			
		Res	Pre (%)	Rec (%)	Acc (%)	Res	Pre (%)	Rec (%)	Acc (%)
Flowers	100	125	68,80	86,00	77,40	119	88,24	75,00	81,62
Horses	100	149	44,30	66,00	55,15	108	68,52	74,00	71,26
Food	100	86	37,21	50,00	43,60	122	59,02	72,00	65,51
Mountains	100	180	41,11	74,00	57,56	135	52,59	71,00	61,80
Buses	100	99	64,65	64,00	64,32	106	53,77	57,00	55,39
Elephants	100	86	53,49	46,00	49,74	97	51,55	50,00	50,77
Average values			**51,59**	**64,33**	**57,96**		**62,28**	**66,50**	**64,39**
Dinosaurs	100	98	100,0	98,00	99,00	119	69,75	83,00	76,37

image possess insignificant backgrounds, like in synthetic images where backgrounds do not represent any relevant information, the global descriptor is more useful.

Finally, we notice an important property during our experiments, is that the global descriptor allows a better Recall for the first 20 retrieved images; however the local descriptor allows a better recall on the total retrieved image.

6. Conclusion

In this paper, we proposed two methods of content based image retrieval according to visual similarity. The first method consists in indexing the images automatically through global features calculated on the whole image, while the second consists in indexing the image using features calculated on the regions of the image. An empirical assessment of the two methods shows that the local descriptor significantly improves the performance of research in the Web domain as it can retrieve more relevant images.

However, these methods are still limited to visual similarity retrieving and the used descriptors are often describing a statistical relationship on images features. This implies that searching task is semantically very poor and usually presents a very low individual meaning. So it is clear that there are a large semantic gap between the extracted features and the semantic level of the users' expectation expressed throw their queries. Hence, we plan to perform our image retrieval system using the semantic features.

7. References

Chen Y., Wang J. Z., « A Region-Based Fuzzy Feature Matching Approach to Content Based Image Retrieval », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, n° 9, p. 1252-1267, September, 2002.

COR, « Corel image database », 1999.

Dong G., Xie M., « Color clustering and learning for image segmentation based on neural networks », *IEEE Transactions on Neural Networks*, vol. 16, n° 4, p. 925-936, July, 2005.

Flickher M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., D.Steele, Yanker P., « Query by Image and Video Content: The QBIC System », *IEEE Computer*, vol. 28, n° 9, p. 23-32, September, 1995.

Jing F., Li M., Zhang H.-J., Zhang B., « An efficient and effective region-based image retrieval framework », *Image Processing, IEEE Transactions on*, vol. 13, n° 5, p. 699-709, May, 2004.

Kohonen T., « Self-Organizing Maps », Springer-Verlag, Berlin, Germany, 1995.

Liu F., Xiong X., Chan K. L., « Natural Image Retrieval based on Features of Homogeneous Color Regions », *SSIAI '00: Proceedings of the 4th IEEE Southwest Symposium on Image Analysis and Interpretation*, IEEE Computer Society, Washington, DC, USA, p. 73, 2000.

Ong S. H., Yeo N. C., Lee K. H., Venkatesh Y. V., Cao D. M., « Segmentation of color images using a two-stage self-organizing network », *Image and Vision Computing*, vol. 20, n° 4, p. 261-271, April 1, 2002.

Pi M., Mandal M., Basu A., « Image retrieval based on histogram of fractal parameters », *Multimedia, IEEE Transactions on*, vol. 7, n° 4, p. 597-605, Aug., 2005.

Quack T., Mönich U., Thiele L., Manjunath B. S., « Cortina: a system for large-scale, content-based web image retrieval », *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, ACM Press, New York, NY, USA, p. 508-511, 2004.

Rubner Y., Guibas L. J., Tomasi C., « The earth movers distance, multi-dimensional scaling, and color-based image retrieval. », *APRA Image Understanding Workshop*, p. 661-668, May, 1997.

Shyu C. R., Brodley C. E., Kak A. C., Kosaka A., Aisen A., Broderick L., « Local versus Global Features for Content-Based Image Retrieval », *CBAIVL '98: Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries*, IEEE Computer Society, Washington, DC, USA, p. 30, 1998.

Tamura H., Mori S., Yamawaki T., « Texture Features Corresponding to Visual Perception », , vol. 8, n° 6, p. 460-473, 1978.

Wang J. H., Rau J., Liu W. J., « Two-stage clustering via neural networks », *IEEE Transactions on Neural Networks*, vol. 14, n° 3, p. 606-315, May, 2003.

Wang J. Z., Li J., Wiederhold G., « SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, n° 9, p. 947-963, 2001.

Wu H., Lu H., Ma S., « WillHunter: interactive image retrieval with multilevel relevance », vol. 2, p. 1009-1012 Vol.2, Aug., 2004.