International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 10 (2018) pp. 018-027

Submit: 23 September, 2017; Accepted: 5 February, 2018; Published: 8 February, 2018

A New Content-Based Image Retrieval System Using Parzen Relevance Feedback and Kullback-Leibler Divergence

Mohammed Ennekhli[1], Rajae El Ouazzani[2] and Mohamed El Haziti[1]

[1]LRIT Associated Unit to the CNRST, URAC 29,
Faculty of Sciences, Mohammed V University, Rabat, Morocco
ennekhli@hotmail.com

[2]Information Technology and Multimedia Team,
Moulay Ismail University, EST, Meknes, Morocco
elouazzanirajae@gmail.com

***Abstract*: In content-based image retrieval (CBIR), many multimedia applications use visual distance to find a collection of images which share the same properties. However, visual distance between two images is often not suitable to semantic distance between the same images. In fact, the semantics term refers to the way how people interpret the image content. Currently, it is difficult to find good correspondences between high-level image semantics and low-level image features which create a "semantic gap". In this paper, we propose a new relevance feedback method which reduces the semantic gap between images. The key steps of our process are the following: At first, we compute the visual distance through the Kullback-Leibler Divergence (KLD). Then, we apply the Relevance Feedback to enhance the retrieval effectiveness by using three different machine learning algorithms: Gaussian Mixture Model (GMM), Support Vector Machines (SVM) and Parzen classifiers; thus, we learn relevant and irrelevant images according to user selection. Experimental results on 5000 images from the COREL database show that comparing to traditional approaches, Parzen classifier is effective and can significantly improve retrieval rates.**

Keywords: CBIR, active learning, Parzen, SVM, GMM, KLD, relevance feedback, semantic gap.

I. Introduction

The rapid growth of the computer network and multimedia technologies, provides a lot of image collections in our life. This huge mass of images requires therefore techniques that can retrieval the desired images easily and accurately. One of the most popular techniques in literature is content based image retrieval (CBIR). There are a lot of fields of application where CBIR system could be applied as the web searching, medical diagnosis, education, etc [1]. Conventional CBIR systems allow users to search for similar images to a query image by extracting low-level features such as color, texture, shape, or any other information that can be derived from the images itself. In recent studies, the researchers have found that using multiple found that using multiple visual features to represent an image could give better results than using a single feature. The merged low-level feature descriptors could be global features such as color histogram [2], [3], as they could be local feature like SIFT and SURF which usually refer to the bag-of-words (BoW) models [4], [5], [6]. However, searching for images by using visual content quickly reaches its limits, in fact that no direct link between high-level concepts and low-level features is available [7], as shown in Figure 1. Therefore, the best way to reduce the semantic gap in CBIR systems is to involve human element in the search process. Because, the computer could not know the way or how people interpret the image content, and could not know also what part of the image is important for the user.

(a) (b)

Figure. 1: Example of a semantic gap: images (a) and (b) have similar visual features but they are completely different in semantic contents.

Among the alternative solutions that have been proposed is the textual-based image retrieval methods that depend on the metadata associated with images, however, these methods require manual annotation of images. Barrios et al. [8] proposed to use the text present in title, description, and tags of the images to improve the results obtained with a standard content-based search. The shortcoming of using text information to retrieval images is required mass of manual annotation and time-consuming process. In addition, its performance is strongly dependent on the quality of tags, because people have different interpretations for the same image due to the disparity of their knowledge, intelligence, visual analysis, experiences and culture background. Moreover, the same image can have different meanings and contain several subjects (polysemy of the image). Take Figure 2 as an example, it has an image with various tags that we could not know which of them are more important to the user. To avoid this semantic ambiguity, relevance feedback systems have been introduced to allow the user to interact directly with the system interface without any external influence [9], [10]. Through these interactions, the system learns the user intent and renders results that can satisfy him.

Figure. 2: Tags of this image are: man, store, car, trees.

In this paper, we present a relevance feedback model as a tool to reduce semantic gap problem. The rest of the paper is organized as follows. In Section 2, we propose a computed approximation of KL-distance between two Gaussians mixtures. In Section 3, we explain the process of relevance feedback with GMM, SVM and Parzen algorithms. Section 4 presents the experimental results. Finally, in Section 5, we provide conclusions and future works.

II. Visual Similarity

Dealing with relevance feedback requires the computation of visual distance between the query image and each image in the database. In CBIR, images are indexed by their visual content, such as color, texture, shapes, etc. However, texture provides important information in image classification; it describes the content of many real-world images such as fruit, skin, clouds, trees, bricks, and fabric. Hence, texture is an important feature in defining high-level semantics for image retrieval. Among the various texture features, Gabor and wavelet features are widely used for image retrieval and have been reported to match well human vision study results [12], [13], [14]. In this paper, we characterize texture images via marginal distributions of their wavelet sub- band coefficients. The advantage of using the wavelet is that it provides a more flexible way of analyzing both space and frequency contents by allowing the use of variable sized windows. Therefore, wavelet transformation provides better representation of an image for feature extraction [15].

Before starting our processing, we must normalize the data because the range of pixel values varies widely. So, we need to put the pixel values on the same scale to ensure that each image has a similar data distribution. Data standardization is a way of normalizing the data and is widely used in many machine learning algorithms [16], where the mean is subtracted to the image and divided by its standard deviation as shown in equation (1). The distribution of such data would resemble a Gaussian curve centered at zero.

$$I_N = \frac{I-\mu}{\sigma} \quad (1)$$

where I is the original image, μ is the mean of the image, and σ is its standard deviation.

A. Generalized gaussian density modeling of wavelet coefficients

By accepting the hypothesis that the distribution of energy in frequency domain identifies the texture, Do and Vetterli [17] identified texture features through the conventional pyramid wavelet decomposition with three levels, they used the Daubechies maximally flat orthogonal filters with 8 in length (D4 filters) [18]. Wavelet coefficients are modeled by generalized Gaussian density (GGD) for each sub-band that allows us to reduce the model dimension.

A generalized Gaussian density G (α, β) is a function often used in multimedia signals [19], depending on two parameters $\alpha > 0$ and $\beta > 0$ as shown in equation (2):

$$G(\alpha,\beta) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp -\left(\frac{|x|}{\alpha}\right)^{\beta} \quad (2)$$

where $\Gamma(.)$ is the Gamma function as shown in equation (3):

$$\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1} dt, \ \ z > 0 \quad (3)$$

When $\beta = 2$, we find a Gaussian equation, when $\beta = 1$, it is a Laplacian distribution. β determines the distribution form, it is called the shape parameter. α determines the curve spreading, it is called the scale parameter.

For each scale s, we evaluate the scale parameter and the shape parameter of the generalized Gaussian density which describes well the behavior of wavelet coefficients at the scale s. We can use the maximum-likelihood (ML) estimator to estimate α and β parameters.

B. Image similarity measure

After extracting texture features, a similarity measure is conducted to compare the query image with each image in the database, then the most similar results will be sent to the user. Kullback-Leibler Divergence (KLD) is one of the most popular distance measures used in the CBIR [20], [17], [21]. KLD is a distance measure related to relative entropy between two probability distributions. For two discrete probability distributions P and Q, the KLD between them is defined by the equation (4):

$$D(P,Q)_{KLD} = \sum_i P(i) log \frac{P(i)}{Q(i)} \quad (4)$$

Then, we get the following analytic form for the KLD between two GGD as shown in equation (5):

$$D(P(.;\alpha_1,\beta_1),Q(.;\alpha_2,\beta_2))_{KLD} = \log\left(\frac{\beta_1\alpha_2\Gamma\left(\frac{1}{\beta_2}\right)}{\beta_2\alpha_1\Gamma\left(\frac{1}{\beta_1}\right)}\right) +$$

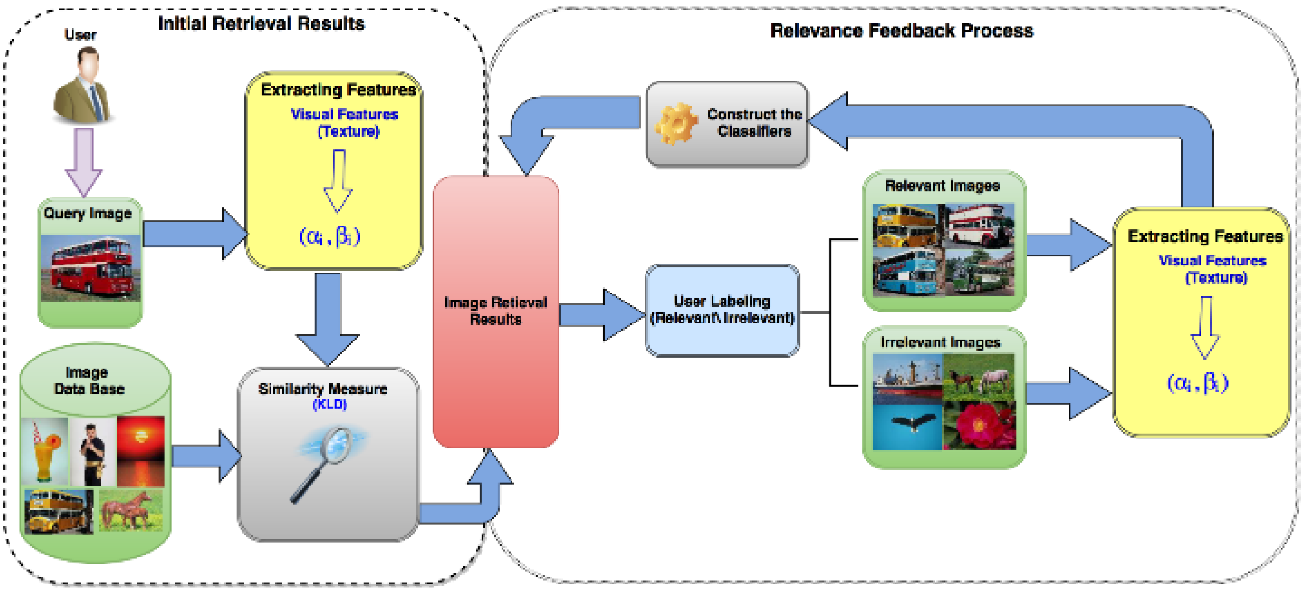


Figure. 4: Block diagram of our system

$$\left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma\left(\frac{\beta_2+1}{\beta_1}\right)}{\Gamma\left(\frac{1}{\beta_1}\right)} - \frac{1}{\beta_1} \qquad (5)$$

.

(α1, β1) et (α2, β2) are the GGD parameters of two sub-bands.

Therefore, the similarity measure between two subbands wavelet can be efficiently calculated by using the model parameters. We accept the reasonable assumption that wavelet coefficients in different subbands are independent. The global similarity distance between two images is precisely the sum of KLDs between corresponding pairs of subbands. We get a high similarity or stronger relevance when this similarity measure has a low similarity score. Once we compute the visual similarity across the KLD between the query image and each image in the database, we save the KLD values in a vector and we sort it in ascending order, then we get retrieval images located in the top (see Figure 3).

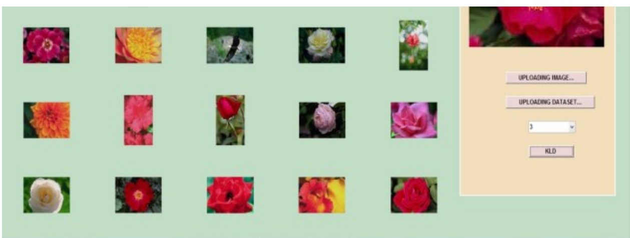


Figure. 3: Example of retrieval images related to the query image "pl flower" (from COREL) by using KLD measure

III. Semantic Similarity

A. Relevance feedback(RF)

When we get retrieval results, the user provides his feed-back towards responses: are images relevant or irrelevant to the query image? If the responses are irrelevant, the feedback loop is repeated many times until the user is satisfied. In our system, the user selects relevant images only, all other images are automatically considered as irrelevant. Figure 4 shows bloc diagram of our system. The involved steps are described in Algorithm 1.

The user interface of our system is divided into two blocks:

- **Results block (Rb)**: It is the main block that displays the retrieval results in which the user makes his final decision (indicate: the user selects the images of this block only in the first iteration) (see Figure 8).
- **Assist block (Ab)**: It displays images that are not yet labeled (see Figure 8). Hence, the user will label only unlabeled images to improve the performance of system with much more samples for training set.

Algorithm 1 The proposed scenario for RF in CBIR

1: Initially, we get the initial retrieval results in Rb by applying the KL-divergence as shown in Equation (5).
2: Select relevant images from the Rb.
3: Learn the system through a feedback algorithm by applying twice a machine learning:

$$\textbf{\textit{learning 1}}: x \in U,\ Ab \leftarrow Machinelearning(x)$$

$$\textbf{\textit{learning 2}}: x \in DB,\ Rb \leftarrow Machinelearning(x)$$

% U: unlabeled images
% DB: all database images
4: **if** the user is satisfied **or** results remain the same **then**
5: Feedback process will be stopped.
6: **else**
7: select new images from the Ab and return to step 3.
8: **end**

B. Active Learning

In general, high-level semantic features require the use of formal tools such as supervised or unsupervised machine learning techniques [22], [23]. In our system, we will take advantage of user/machine interactions in relevance feedback. RF uses active learning algorithms which converge quickly to the query concept after few iterations. The active learning can start its process with a small training set; later and after each iteration, the user will select samples to be added to the training set. The selection operation concerns relevant samples only. The success of active learning processes depends on a number of design decisions, including data preprocessing, choice of classification algorithm, and querying strategy. In this paper, we used three different machines learning algorithms to get comparative results: GMM, SVM (Support Vector Machines) and Parzen. We used the same training set structure in the three machine learning algorithms. Every image is represented by a vector with the 18 GGD model parameters. Therefore, the set of training data $\{x_1, \ldots, x_n\}$ is represented as vectors in space $X \subseteq R^{18}$ and their corresponding labels $\{y_1, \ldots, y_n\}$ belong to two separate classes $y_i \in \{-1, 1\}$.

1) GMM-Based Relevance Feedback

The Gaussian Mixture Model (GMM) is one of the semi-parametric techniques for estimating probability density functions [24]. In addition, GMM is widely used in classification field because it is robust and still relatively easy to use. The GMM density is parameterized by the mean μ, covariance matrices Σ and mixture weights π where $\Theta = (\pi_1, \ldots, \pi_k, \mu_1, \ldots, \mu_k, \Sigma_1, \ldots, \Sigma_k)$. Given a set of N independent and identically distributed samples $X = \{x_1, \ldots, x_N\}$ that were drawn from a GMM comprised of k Gaussian components, the probability density function (pdf) on point x is given by equation (6):

$$p(x|\Theta) = \sum_{i=1}^{k} \pi_i N(x|\Theta) \quad (6)$$

Where $\sum_{i=1}^{k} \pi_i = 1$ and $N(x|\Theta)$ is the multivariate Gaussian distribution:

$$N(x|\mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)) \quad (7)$$

The parameters Θ of the GMM are typically estimated by the expectation-maximization (EM) algorithm [25]. Its goal is to maximize the likelihood generated by each GMM. Since EM algorithm depends on initialization, we use k-means algorithm to initialize GMM parameters. Then, we iteratively apply expectation (E) step and maximization (M) step until the likelihood convergence.

Each class (relevant and irrelevant) of retrieval images is represented by a specific GMM. Then, we get a GMM1 for relevant images and a GMM2 for irrelevant ones. The learning process of GMMs parameters is presented in Figure 5.

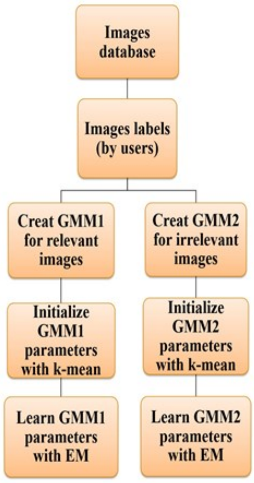


Figure. 5: The learning process of GMMs parameters

2) SVM-Based Relevance Feedback

SVM (Support Vector Machines) are often used to learn high-level concepts from low-level image features. More- over, it has been introduced by CBIR as a powerful relevance feedback tool; in fact, it performs fairly well in systems using global representations [26], [27], [28], [29]. As a core ma- chine learning technology, SVM has strong theoretical foundations, and excellent empirical successes [30]. Besides, the typical form of SVM can be used as a binary classifier which is very suitable to model our problem, since it has good sup- port ability for small sample training sets [27]. An SVM classifies data by finding the best hyperplane that leaves the maximum margin possible from both classes (Figure 6), such as, samples near to the separating hyperplane (support vectors) have low confidence to be classified correctly.

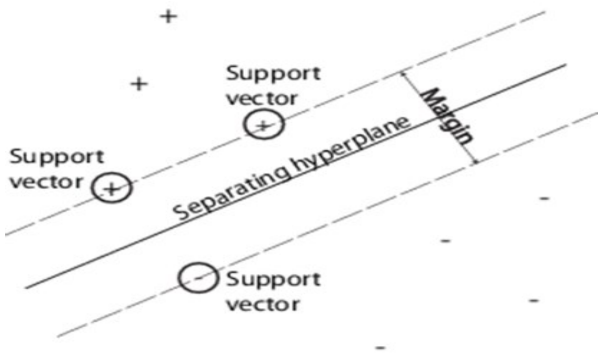


Figure. 6: A simple linear Support Vector Machine [31]

Consider the binary classification problem $\{(x_i, y_i)\}_{i=1}^{N}$, where x_i are the labeled samples and $y_i \in \{relevant, irrelevant\}$ the corresponding labels. Based on this training set, we train an SVM classifier. The decision function which classifies the example x into one of the two classes of retrieval images (relevant or irrelevant) is shown in equation (8):

$$f(x) = \sum_i y_i \alpha_i K(x_i, x) + b \qquad (8)$$

where b is a bias parameter, α_i is the positive Lagrange multipliers and $K(x_i, x_j)$ is the Gaussian Radial Basis Function (RBF) kernel [32].

3) Parzen-Based Relevance Feedback

Since we cannot make the assumption on distribution among samples in our study, so we will use a non-parametric probabilistic system. The principle of non-parametric probability density estimation is the delimitation of a region R_N around each point first, then the calculation of samples number in this volume, and finally the determination of the density as the ratio between the samples number in this region and the product of region volume with total number of samples [33]. Thus, we get an estimation of the probability density function with equation (9):

$$\hat{P}_N(x) = \frac{K_N}{NV_N} \qquad (9)$$

With:

N : the total number of samples,

K_N: the number of samples in the region R_N,

V_N : the volume of R_N .

The three required conditions to ensure the convergence of $\hat{P}_N(x)$ to $P_N(x)$ are:

$$\lim_{N\to\infty} V_N = 0, \qquad \lim_{N\to\infty} K_N = \infty \ \ and \ \ \lim_{N\to\infty} \frac{K_N}{N} = 0$$

In literature, Parzen window is defined as a learning method by neighborhood, it responds to the previous conditions. Furthermore, it is considered as a good candidate for learning in image retrieval systems. Taking a point with coordinates x in the description space (with P dimensions) and defining a volume (hypercube whose side is h_N with h_N the number of observations) around this point by: $V_N = h_N^p$. The influence function $\varphi(u)$ called **Parzen kernel** is defined by the expression (10):

$$\begin{cases} 1 & if \ \ |u_i| \le \frac{1}{2} \ \ for\, i = 1, \dots, P \\ 0 & \quad\quad else \end{cases} \qquad (10)$$

We get the number of examples in the hypercube and the density estimation by the equation (11):

$$\hat{P}_N(x) = \frac{1}{N}\frac{1}{V_N}\sum_{i=1}^{N}\varphi\left(\frac{x-x_i}{h_N}\right) \quad (11)$$

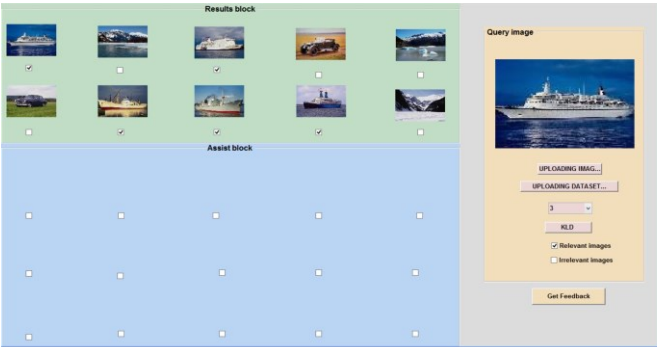
Therefore, we adjust the two following parameters to use Parzen kernel:

- **The volume**: to ensure the estimator convergence, the smoothing window dimension (the volume V_N) must satisfy these two conditions:

$$\lim_{N\to\infty} V_N = 0 \ \ and \ \lim_{N\to\infty} NV_N = \infty$$

For example, we can take: $V_N = \frac{V_1}{\sqrt{N}}$

Figure 7 shows that the choice of volume (V_1) plays a very important role in density estimation. If this volume is very large, then the estimator will tend to level density. Otherwise, if it is too small, the estimator locally follows the presence or absence of an example in the volume.



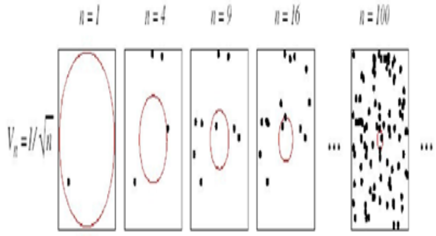


Figure. 7: The volume of the region R_n [34]

- **The kernel**: the choice of the Parzen kernel function is less sensitive. However, to smooth the density estimation which is discontinuous with the function φ(u), we often use other kernel functions such as Gaussian kernel, generalized to the multidimensional case whose expression is given in equation (12):

$$K(u) = \frac{1}{(2\pi)^{P/2}|\Sigma|^{1/2}}\exp\left(\frac{1}{2}u^T\Sigma^{-1}u\right) \quad (12)$$

where u ∈ Rp and Σ is the estimated covariance matrix on the examples description.

We get the final density estimation (13) by replacing equation (12) in equation (11):

$$\hat{P}_N(x) = \frac{\sum_{i=1}^{N}\exp\left(\frac{(x-x_i)^T\Sigma^{-1}(x-x_i)}{(2h_N)^2}\right)}{NV_N(2\pi)^{P/2}|\Sigma|^{1/2}} \quad (13)$$

Finally, to classify the example x into one of the two classes of retrieval images (relevant or irrelevant), we calculate its density estimation $\hat{P}(x)$ for each one of the two classes, then, we assign x to the class which has the highest density.

In practice, the power of Parzen kernels method comes from its generality (by the distribution hypothesis). However, this power is paid by the number of examples needed to have a good estimation, it grows exponentially with data dimension.

IV. Experimental Results

In this section, we present the experimental results to verify the effectiveness of proposed approaches. We perform experiments over 5000 images from Corel Photo Gallery [35]. Corel Photo Gallery is the most popular image database used in CBIR, based on semantic concepts, it comprises 10800 images. Images in the database are divided into 80 different concept groups (categories), e.g., dog, flower, tiger, train, etc.

In the reorganized database, each category includes over than 100 images, such as, we define images with the same semantic category as relevant.

Figures 8, 9, 10, 11,12 and 13 display retrieval results of the three first feedback iterations by using SVM and Parzen algorithms. Through the obtained results, it is clear that feedback performance is improved after each iteration and we notice that the Ab allows a quick convergence to desired images in a few feedback steps. Thus, satisfied results are gotten by three or five feedbacks only. We have presented experimental results related to SVM and Parzen only. In fact, the utilization of RF with GMM to reduce the semantic gap between images give very weak performances.

Figure. 8: (Iteration # 0) the retrieval results of relevance feedback based on SVM for the query image "obj ship

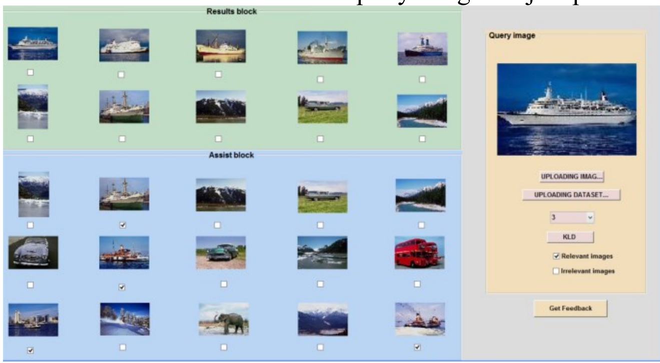


Figure. 9: (Iteration # 1) the retrieval results of relevance feedback based on SVM for the query image "obj ship"

Figure. 10: (Iteration # 2) the retrieval results of relevance feedback based on SVM for the query image "obj ship"

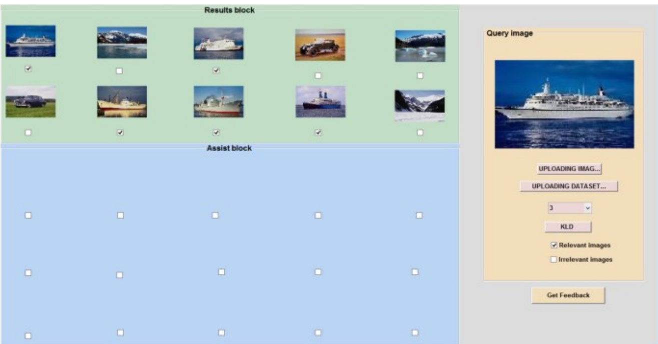
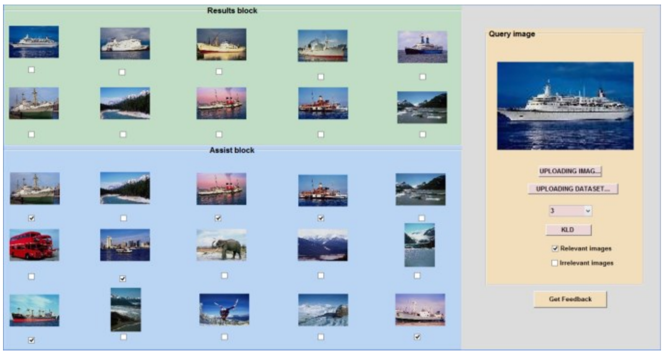


Figure. 11: (Iteration # 0) the retrieval results of relevance feedback based on Parzen for the query image "obj ship"

Figure. 12: (Iteration # 1) the retrieval results of relevance feedback based on Parzen for the query image "obj ship"

The effectiveness of our proposed approaches is realized by



two major criteria, namely precision and recall. They are used to measure related experimental evaluations and are defined in equations (14) and (15):

$$Precision = \frac{|Ret(q) \cap Rel(q)|}{|Ret(q)|} \quad (14)$$

$$Recall = \frac{|Ret(q) \cap Rel(q)|}{|Rel(q)|} \quad (15)$$

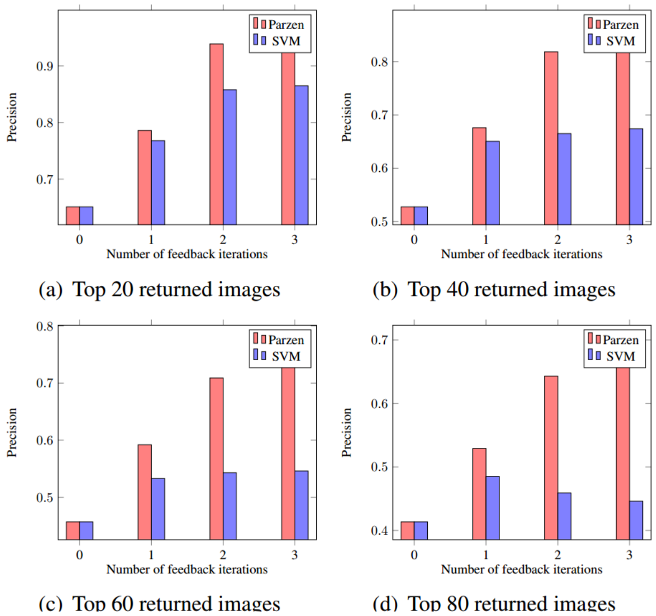
Where:

$Ret(q)$: are the retrieved images for the query image q.

$Rel(q)$: are the relevant images that are in the database for the query image q.

Figure. 13: (Iteration # 2) the retrieval results of relevance feedback based on Parzen for the query image "obj ship"

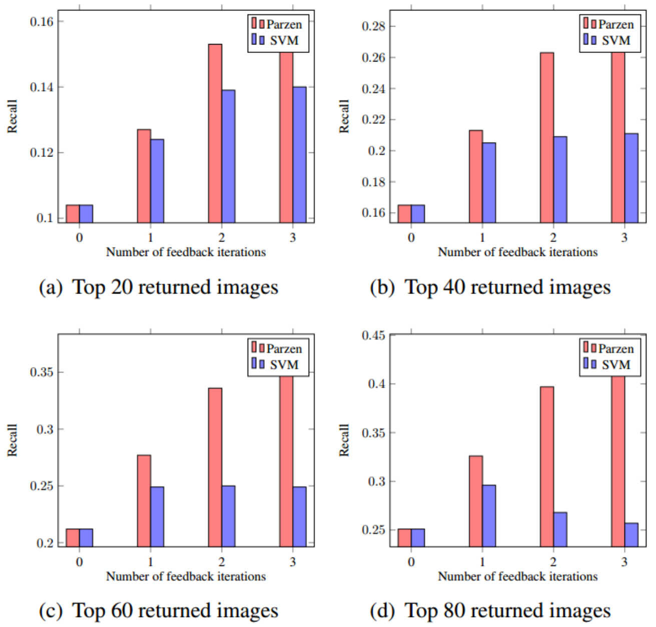
Our system performances are evaluated according to the number of images retrieved in Rb (20, 40, 60 and 80 images), several experiments have been realized. Figures 14 and 15 show the variations of the precision and recall rates respectively as the number of feedback iterations using SVM and Parzen. The retrieval results reach better performance starting from the second feedback iteration. Moreover, they clearly show that Parzen achieves substantially better retrieval precision and recall than SVM whatever the iteration number or the number of returned images.



(a) Top 20 returned images (b) Top 40 returned images

(c) Top 60 returned images (d) Top 80 returned images

Figure. 14: Relevance feedback learning precision

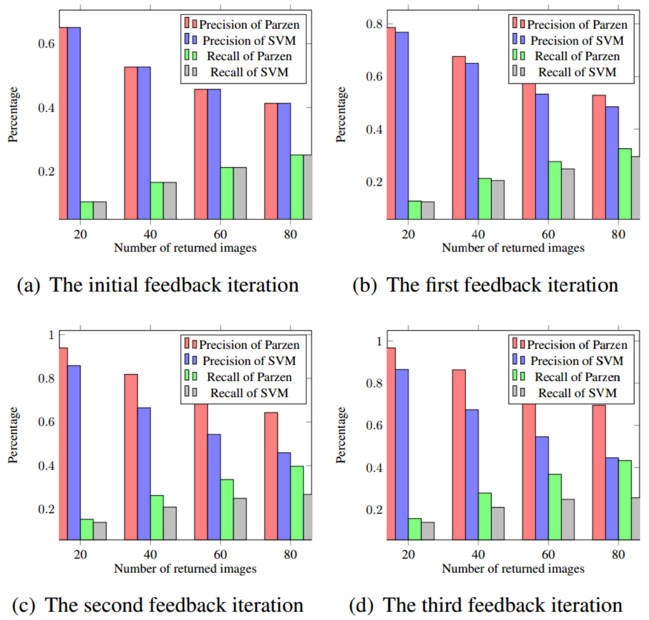


(a) Top 20 returned images (b) Top 40 returned images

(c) Top 60 returned images (d) Top 80 returned images

Figure. 15: Relevance feedback learning recall

The second set of analyses examines the impact of the number of returned images. Contrary to precision rate, Figure 16 shows that the recall rate is improved when the number of returned images increases. As shown in Figure 17, the performance of our Parzen RF retrieval system becomes better when the number of feedback iteration increases. Besides, Figure 18 illustrates that the performance of our SVM RF retrieval system becomes almost stable after the first feedback iteration.

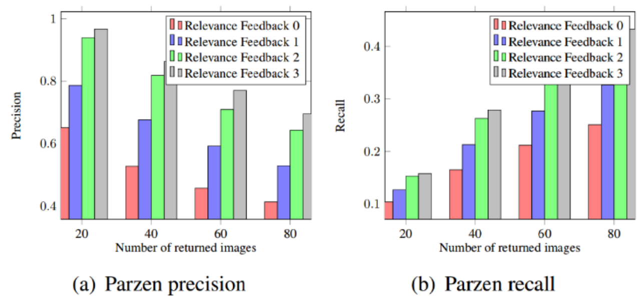


(a) The initial feedback iteration (b) The first feedback iteration

(c) The second feedback iteration (d) The third feedback iteration

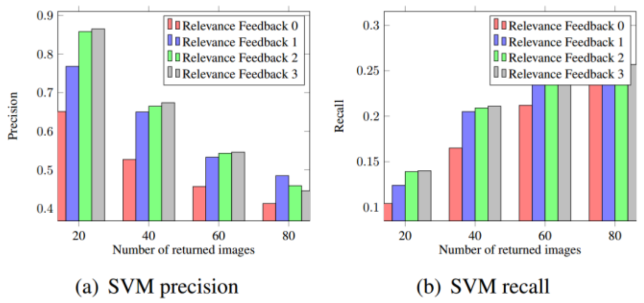
Figure. 16: Performance comparison between Parzen and SVM

Moreover, we also evaluate the dispersion measure of a dataset from its mean by calculating the standard deviation. Figure 19 presents the standard deviation comparison between SVM and Parzen in the first, second and third RF learning processes. We notice that Parzen is more stable than an SVM - based RF learning.



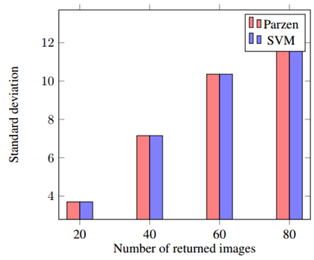
(a) Parzen precision (b) Parzen recall

Figure. 17: Parzen performance based on the number of returned images and the number of feedback iteration

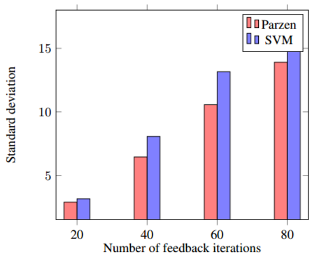


(a) SVM precision (b) SVM recall

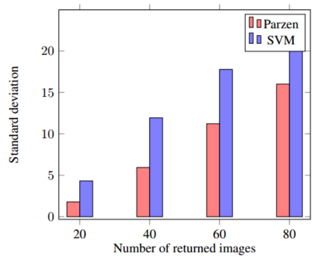
Figure. 18: SVM performance based on the number of returned images and the number of feedback iteration



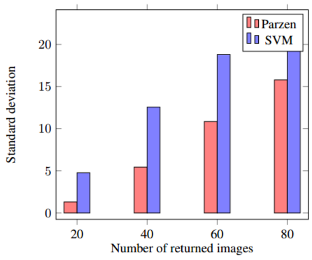
(a) The initial retrieval



(b) The first feedback iteration



(c) The second feedback iteration



(d) The third feedback iteration

Figure. 19: Retrieval standard deviation

To evaluate the performance of our method by comparing it to those that have been used in image retrieval systems ([36], [37], [29], [38], [39], [11]), we made a set CBIR experiments. Table 1 shows the comparison of the precision according to the number of images returned and the number of feedback iterations between the proposed method and the other tested methods. In all approaches, we can see that the precision increases when the number of iterations in- creases and decreases when the number of image returned increases. The experiment showed also a clear preference for our method.

Table 1. Precision according to the number of images returned and the number of feedback iterations.

	20				**40**				**60**				**80**			
	FI0	**FI1**	**FI2**	**FI3**	**FI0**	**FI1**	**FI2**	**FI3**	**FI0**	**FI1**	**FI2**	**FI3**	**FI0**	**FI1**	**FI2**	**FI3**
Wang et al. [36]	0.23	0.33	0.47	0.68	0.18	0.27	0.38	0.53	0.16	0.24	0.35	0.44	0.15	0.21	0.32	0.37
Liu and Wang [37]	0.23	0.33	0.51	0.61	0.18	0.28	0.43	0.50	0.16	0.24	0.39	0.44	0.15	0.22	0.35	0.40
Zhang et al. [29]	0.23	0.32	0.39	0.50	0.18	0.27	0.35	0.40	0.16	0.23	0.30	0.36	0.15	0.20	0.27	0.33
Huang et al. [38]	0.23	0.32	0.45	0.53	0.18	0.28	0.35	0.52	0.16	0.25	0.32	0.42	0.15	0.22	0.27	0.38
Wang et al. [39]	0.23	0.34	0.48	0.71	0.18	0.29	0.43	0.55	0.16	0.25	0.40	0.46	0.15	0.22	0.34	0.43
Proposed method (Parzen)	0.65	0.78	0.93	0.96	0.52	0.67	0.81	0.86	0.45	0.59	0.70	0.77	0.41	0.52	0.64	0.69

Another interesting parameter to evaluate the effectiveness of our proposed approach, is the necessary number of iterations to reach a high precision rate. Table 2 reveals the comparison of the efficiency among different approaches which presented in [11], [36] and [39]. The results show that our approach reaches the specific precision (80%) for the top 20 returned images in two iterations only, which means that the user doesn't need a lot of time to be satisfied.

Table 2. The minimum number of feedbacks for different approaches to reach the specific precision 80%.

Comparative approaches	**Minimum number of feedbacks**
Naïve QPM, QPM, Naïve QR, QR [11]	>6
QPM+QR [11]	5
QEX+QR [11]	6
X.-Y. Wang et al. [36]	4
X.-Y. Wang et al. [39]	4
NPRF [11]	2
Proposed method (Parzen)	2

V. Conclusions and Future Work

The main goal of the current study is to reduce the "semantic gap" between images through Parzen and SVM relevance feedback algorithms based on texture features. This is the first time that Parzen classifier is used in relevance feedback and the experimental results obtained reveal that Parzen performances are more effective than SVM and GMM within a very short term of feedback iterations. Hence, one of the issues that emerges from these findings is that the choice of an appropriate machine learning algorithm is a factor that clearly affects the performance of relevance feedback quality.

In the future, more information on image features to represent the image content would establish a greater degree of precision and recall. Also, the use of other machine learning algorithms can improve the performances of our RF retrieval system.

References

[1] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," in *Computer*, vol. 28, no 9, pp. 18-22, Sep 1995

[2] Talib, Ahmed, et al. "A weighted dominant color descriptor for content-based image retrieval.", Journal of Visual Communication and Image Representation, vol. 24, no 3, pp. 345-360,2013

[3] Wang, Ren-Jie, Ya-Ting Yang, and Pao-Chi Chang. "Content-based image retrieval using H. 264 intra coding features.", *Journal of Visual Communication and Image Representation,* vol. 25, no 5, pp. 963-969, 2014

[4] Murala, Subrahmanyam, and QM Jonathan Wu. "Expert content-based image retrieval system using robust local patterns.", *Journal of Visual Communication and Image Representation*, vol. 25, no 6, pp. 1324-1334, 2014

[5] Raveaux, Romain, Jean-Christophe Burie, and Jean-Marc Ogier, "Structured representations in a content-based image retrieval context." *Journal of Visual Communication and Image Representation,* vol. 24, no 8, pp. 1252-1268, 2013

[6] Lu, Zhiwu, et al. "Semantic sparse recoding of visual content for image applications.", *IEEE Transactions on Image Processing* vol. 24, no 1, pp. 176-188, 2015

[7] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying Ma, "A survey of content-based image retrieval with high-level semantics", *Pattern Recognition*, vol. 40, no 1, pp. 262-282, 2007

[8] Barrios, Juan Manuel, Diego Daz-Espinoza, an Benjamin Bustos. "Text-Based and Content-Base Image Retrieval on Flickr.", *Similarity Search and Applications, 2009. SISAP'09. Second International Workshop on. IEEE*, 2009

[9] Xu, Heng, Jun-yi Wang, and Lei Mao. "Relevance feedback for Content-based Image Retrieval using deep learning."; *Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. IEEE*, 2017

[10] Zhang, Yujiao. "Image retrieval algorithm based on Harris-Laplace corners and SVM relevance feedback.", Software *Engineering and Service Science (ICSESS)*, 7th IEEE International Conference on. IEEE, 2016

[11] Ja-Hwung Su, Wei-Jyun Huang, P.S. Yu, V.S. Tseng, "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns", *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no 3, pp. 360-372, 2011

[12] Ma, Wei-Ying, and Bangalore S. Manjunath. "Netra: A toolbox for navigating large image databases.", *Multimedia systems*, vol. 7, no 3, pp. 184-198, 1999

[13] James Ze Wang, Jia Li, Gio Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries", In *Advances in Visual Information Systems*, 4th International Conference, Lyon, France, November 2-4, 2000, Proceedings, pp. 360-371. 2000.

[14] Phillipe Salembier, Thomas Sikora, "Introduction to MPEG-7", *Multimedia Content Description Interface*, John Wiley & Sons, 2002.

[15] S. G. Mallat, "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no 7, pp. 674-693, 1980

[16] Grus, Joel. *Data Science from Scratch*, Sebastopol, CA: O'Reilly, pp 99-100, 2015

[17] M. N. Do, M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance", *Trans. Img. Proc*, vol. 11, no 2, pp. 146-158, 2002

[18] Ingrid Daubechies, *Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics*, 1992

[19] J.-R.Ohm, "Multimedia Communications Technology", *Ed. Springer*, pp. 53, 2004

[20] T.P. Minka, R.W. Picard, "Interactive Learning Using a Society of Models", *Pattern Recognition*, vol. 30, no 4, pp. 565- 581, 1997

[21] Don H. Johnson and Sinan Sinanovic,"Symmetrizing the Kullback-Leibler Distance", *IEEE Transactions on Information Theory*, 2000.

[22] A. Vailaya, M. A.T. Figueiredo, A. K. Jain, Hong-Jiang Zhang, "Image classification for content-based indexing", *Trans. Img. Proc*, vol. 10, no 1, pp. 117-130, 2001

[23] Jiebo Luo, Andreas E. Savakis, "Indoor vs outdoor classification of consumer photographs using low-level and semantic features", *In ICIP (2)*, pp. 745-748. 2001

[24] Dasgupta, S., "Learning mixtures of Gaussians. 1999", *Fortieth Annual IEEE Symposium on Foundations of Computer Science (FOCS)*

[25] S. Akaho, "Nonmonotonic generalization bias of Gaussian mixture models", *Neural Computat*., vol. 12, no 6, pp. 1411-1428, 2000

[26] Apostolos Marakakis, Nikolaos Galatsanos, Aristidis Likas, Andreas Stafylopatis, "Relevance feedback for content-based image retrieval using support vector machines and feature selection", *In Proceedings of the 19th International Conference on Artificial Neural Networks*, pp 942-951. Springer-Verlag, Berlin, Heidelberg, 2009

[27] Lingjun Li, Yihua Zhou, "Weight revision and svm-based relevance feedback algorithm for content-based image retrieval*", In Conference Anthology, IEEE*, pp 1-5. 2013

[28] Simon Tong, Edward Chang, "Support vector machine active learning for image retrieval", *In Proceedings of the Ninth ACM International Conference on Multimedia*, pp 107-118. ACM, New York, NY, USA, 2001

[29] Lei Zhang, Fuzong Lin, Bo Zhang, "Support vector machine learning for image retrieval". *In ICIP (2),* pp 721-724. 2001

[30] Nello Cristianini, John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, 2000

[31] MATLAB, "Support Vector Machines (SVM): User's Guide (R2015a)", *The MathWorks Inc*, Natick, Massachusetts, United States, 2015

[32] Bernhard Scholkopf, Kah-Kay Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, "Comparing support vector machines with gaussian kernels to radial basis function classifiers", *IEEE Transactions on Signal Processing*, vol. 45, no 11, pp. 2758-2765, 1997

[33] Emanuel Parzen, "On estimation of a probability density function and mode", *The Annals of Mathematical Statistics*, vol. 33, no 3, pp. 1065-1076, 1962

[34] Alan L. Yuille (2004),"Non-Parametric Learning" [PowerPoint slides], Retrieved from http://www.stat.ucla.edu/˜yuille/Courses/Stat231-Fall04/ Lecture7-Nonparametric.ppt

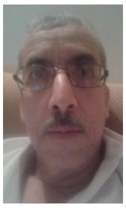
[35] J.Z. Wang, Jia Li, G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no 9, pp. 947-963, 2001

[36] Xiang-Yang Wang, Yong-Wei Li, Hong-Ying Yang, and Jing-Wei Chen, "An image retrieval scheme with relevance feedback using feature reconstruction and SVM reclassification", *Neurocomput*ing, vol.127, pp. 214-230, 2014

[37] Rujie Liu, Yuehong Wang, “SVM-based active feedback in image retrieval using clustering and unlabeled data”, *Pattern Recognition*, vol. 41, no 8, pp. 2645-2655, 2008

[38] S. Huang, Q. Wu, S. Lai, “Improved AdaBoost-based image retrieval with relevance feedback via paired feature learning”, *Multimedia Systems,* vol. 12, no 1, pp. 14-26, 2006

[39] Xiang-Yang Wang, Lin-Lin Liang, Wei-Yi Li, Dong-Ming Li, and Hong-Ying Yang, “A new SVM-based relevance feedback image retrieval using probabilistic feature and weighted kernel function”, *Journal of Visual Communication and Image Representation*, vol. 38, pp. 256-275, 2016



Author Biographies

Mohammed Ennekhli received his Master’s degree in Computer Science and Telecommunications from Faculty of Science, Mohammed V University, Rabat, Morocco in 2013. He is currently a PhD student since 2014 in computer Science in the Faculty of Sciences, member of the Research Laboratory in Computer Science and Telecommunications, Mohammed V University, Rabat, Morocco. His research interests include the treatment of multimedia information and image processing.

Rajae El Ouazzani received her Master’s degree in Computer Science and Telecommunication by the Mohammed V University of Rabat (Morocco) in 2006 and the PhD in Image and Video Processing by the High National School of Computer Science and Systems Analysis (Morocco) in 2010. From 2011, she is a Professor in the High School of Technology of Meknes, Moulay Ismail University in Morocco. Since 2007, she is an author of several papers in international journals and conferences. Her domains of interest include multimedia data processing and telecommunications.

Mohamed Elhaziti received his Master degree in computer science by Mohammed V University of Rabat (Morocco) in 1994 and he has the doctorate in image compression using fractal transformations by the same university in 2003. He has the PHD degree in 2013 by IBN Zohr University of Agadir (Morocco). He is an author of several papers in international journals and conferences. His domains of interest include image processing and complex networks. He
is currently a Professor in high school of technology (E.S.T) in SALE, Morocco.