International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 1, No. 4, December 2010

Image Indexing for Object Recognition and Content Based Image Retrieval

Younès Raoui[1,2], El Houssine Bouyakhf[1] and Michel Devy[2]

[1]Laboratory of Computer Sciences, Artificial Intelligence and Form Recognition (LIMIARF) Faculty of Sciences, Mohamed V University, 4 Street Ibn Battouta B.P. 1014 RP, Rabat, Morocco[*]

[2]Laboratory of Analysis and Architecture of Systems(LAAS), Institut National Polytechnique, 7 Street Colonel Roche, Toulouse, France[**]

Abstract: In this work, we present two new color texture image indexing techniques based on a colortexture framework which constructs features with reasonable dimensionality. The first descriptor, using the Gaussian color model to extract the color features and the Gabor wavelets to extract texture features, is appropriate in many Content- Based Image Retrieval systems to improve their performances. The second descriptor proposed is a modified version of Harris descriptor which performs well in object recognition for color textured images while assuring repeatability and invariance to scale and rotation.

Keywords: Content based retrieval systems, Gaussian color model, Feature points, color texture framework, Rayleigh distribution, Hough transform, Object recognition.

1. Introduction

Nowadays, image databases know proliferation for different applications (data mining, appearance-based description of environments in robotics). It requires effective ways of retrieving visual information. The computer community is more involved in the development of automatic systems for such purposes. Mechanisms for browsing and researching are the keys for efficient usage of the digital resources. Such works aim to decrease the computational complexity of the retrieval algorithms without loosing the accuracy of search results.Traditionally, retrieval processes were based on textual features associated with images, which are easily automated(Liapis and Tzir- itas, 2004). However, the widespread synonymy and polysemy in natural languages affects the precision and the recall of such systems. Moreover, new forms cannot be handled. The apparition of a new concept named Content Based Image Retrieval (CBIR)â€ was the turning point. The goal of such system is to interact directly with the essence of images images (Yeh and Kuo,2003) . Swain and Ballard were among the pioneers of this approach (Yoo et al., 2005).. The hypothesis under the CBIR approach is that the statistical measures of similarity among representations of image attributes do correlate with subjective visual similarity .

In the first part of our paper, we present color texture attributes describing an image which used to feed our method of content based image retrieval. Our goal is to develop an effective mechanism to overcome several limitations which are related to existing systems. To do this, we propose a new image description method exploiting the results of the theory of color texture measurement which is developed by Hoang et al (Hoang et al., 2005) in order to probe an observed scene. However, we will modify this method so as to be invariant to rotation and scale by using a class of Gabor wavelets. In other hand, it is known that the dimensionality of Gabor filter's outputs is high. Bhagavathy et al demonstrated that such outputs have a Rayleigh tendency (. We will use this result to compute a more compact color texture feature. Besides, we present a work about the extraction of feature points from images.To extract these visual features, a significant number of detection and description methods have been developed in recent years : SIFT (Scale Invariant Feature Transform) (D.Lowe, 1999) which has been the source of inspirations for many derivations. SURF(Speeded Up Robust Features) (H. Bay, 2006) and CenSure (Center Surround Extremas) both used for their speed. SIFT is invariant to image scaling translation, to rotation and partially to illumination changes and affine or 3D projections. Features are detected with an approach that identifies stable points in scale space. They are created by representing the blurred image in multiple orientation planes and multiple scales. Whereas SURF detector computes the features and their descriptor. It uses the DoG operator as a Hessian matrix to improve time performance. It describes using a distribution of Haar wavelets responses in neighborhood of the interest points as well as the use of integral images. The extraction of invariant features can be performed using either conditions intrinsic to the objects which are included in the image or on the intrinsic properties of the objects. However, the second approach is shown to be less complex (Hoang et al., 2005).. Formally, two objects or two appearances of objects are equivalent under a group of transformations T if they belong to the same equivalence class . In section I, we present our local feature by developing the detection scheme and then we present our descriptor. In section II, we give an evaluation of our descriptor and we show results of objects recognition. In section III, we present our global descriptor where the color-texture measurement theory is explained. Then, we propose our image descriptor,as well as we

describe how this method has been implemented, and we give some preliminary results.In Section IV we give a summary of the paper and propose some future works.

2. Local descriptor

2.1 Detector computation

We use the Harris detector(K.Mikolajzyk and C.Schmid, 2004) because of its suitability in time performances. It is based on the second moment matrix also called the auto correlation matrix. It is convenient for the computation of the local structure into images. And finally, should be adapted to scale in order to make it independent on the resolution of the image. The second moment matrix adapted to the scale is defined as :

$$M(x, y, \sigma_I, \sigma_D) = \sigma_D g(\sigma_I) \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} \quad (1)$$

The gaussian kernel has to be discretized and cropped even if it is optimal for scale space analysis.

The local derivatives are computed by a gaussian kernel determined by the scale σ_D. The derivatives are then averaged in the neighborhood of the point by smoothing with a gaussian of size σ_I.

So the eigen values represent two principal changes in the neighborhood of the point. This allows the extraction of points for which both curvatures are significant so that the signal be important in the orthogonal directions.(junctions corners ..). These points are stable regarding to lightning conditions.

$$cornerness = det((M(x, \sigma_i, \sigma_D))) - \alpha.trace^2(M(x, \sigma_i, \sigma_D)) \quad (2)$$

The Harris detector applies only to gray scale images. Montesinos et al generalized it to color images(Montesinos et al., 1998). . In our case, we propose to use Gaussian color model to integrate color in the Harris detector. This is done by diagonalysing the explicit matrix of Koendering kernel(mentioned above) given in (Geusebroek et al., 2001a) with :

$$M_g = \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix} \quad (3)$$

And ponderate the three components of the Gaussian derivatives by $\alpha, \beta and \gamma$

Algorithm 2

```
for i← Feature Point-2 to Feature Point+2 do
    box ← Create a box around a point i
    result ← box * Gaussians(scales)
    Compute norm(result)
    V ← 4 highest values
    Add V to the Feature Point
end for
```

Figure 1: Features extracted both with SIFT detector(gray scale image) and by our detector

2.1.1 Scale computation

We want to select a characteristically scale of a local structure for which a function attains an extremum over scales. The selected scale is characterized because it measures the scale at which there is maximum similarity between the feature detection operator and the local image structure. We use the following algorithm for computing the characteristically scales :

Algorithm 1

1: in $\leftarrow$ *Readimage*
2: R $\leftarrow in(1)$
3: B $\leftarrow in(2)$
4: G $\leftarrow in(3)$
5: $M \leftarrow \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix}$
6: Diagonalize M $\rightarrow (\alpha,\beta,\gamma)$
7: S $\leftarrow \alpha.R + \beta.G + \gamma.B$
8: $L_{xx} \leftarrow G_{xx}.S$
9: $L_{xy} \leftarrow G_{xy}.S$
10: $L_{yy} \leftarrow G_{yy}.S$
11: $M(x,y,\sigma_I,\sigma_D) \leftarrow \sigma_D g(\sigma_I) \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix}$

2.1.2 Orientation assignment

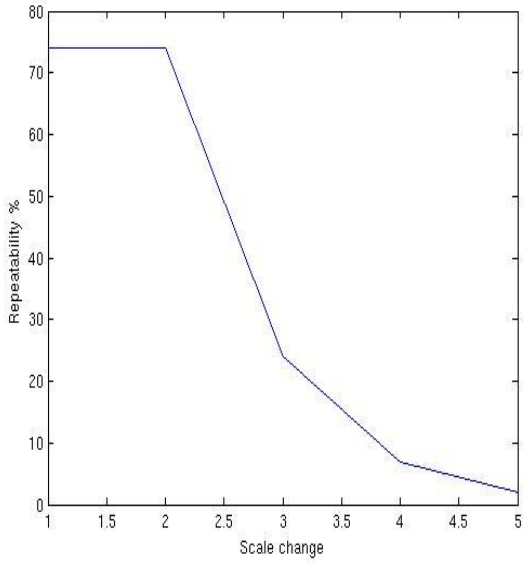


Figure 2: Repeatability of our detector

In order to be invariant to rotations, we compute orientation for interest points. That's why we use the same method as in SIFT. In fact it has good performances comparing to other descriptors because it mixes localized information and the distribution of gradient related features. So, for each scale the image is processed to extract image orientations. We compute pixel differences :

$$R_{ij} = a\tan(E_{ij} - E_{i+1,j}, E_{i,j+1} - E_{ij}) \quad (4)$$

Where E is the pixel value. We compute the orientation for the feature point and all points around it (4 points) and we concatenate these orientations in the same array.

Table 1: Structure of our descriptor, 2 bin for space, 21 for scale, 5 for orientation, 9 for texture.

x y	σ	r	t
space	scale	orientation	texture

2.2 Descriptor computation

We compute a texture invariant by using Gabor wavelets at different orientations and scales (B.S.Manjunath and W.S.Ma, 1996).. We use 9 kernels and we do convolution with small windows around each feature point(Gabor, 1946a)(J.G.Daugman, 1980)..

3. Recognition by indexing

We present a standard evaluation for our detection and description and we discuss results obtained in a real life object recognition.

• Standard evaluation :

Our descriptor is evaluated using the image sequences provided in Ponce Group Dataset and from the Amsterdam library of object images. We use images of textured and structured scenes.

3.1 Repeatability

This score shows the detected interest points which are found in both images relative to the lowest total number of interest points. However, only the part of the image that is visible in both images is taken into account. Our detector is compared to the DoG detector by Lowe and the Harris and Hessian Laplace detector proposed by Mikolajzyk. We compute the ratio between the number of point to point correspondances and the minimum number of points detected in the image. Then we take into account only points that are in the images. The algorithm that computes this ratio is presented in (K.Mikolajzyk and C.Schmid, 2004)..

1. The error in relative point location is less than 1.5 pixel: $\| x_a - Hx_b \| < 1.5$, where H is the matrix of homography.

2. The error in the image surface covered by point neighborhoods is S <0.4. In the case of scale invariant points the surface error is:

$$e_s = |1 - s^2 \frac{min(\sigma_a^2, \sigma_b^2)}{max(\sigma_a^2, \sigma_b^2)}| \quad (5)$$

where $\sigma_a and \sigma_b$ are the selected point scales and s is the actual scale factor recovered from the homography between the images (s > 1).

We estimate the stability of our descriptor by subjecting our images to affine projections, contrast and brightness changes. We can estimate the location of each key detected in the first image by predicting for knowledge of transform parameters in the transformed image. The following table shows the overall stability of the keys to image transformations. Each line of the table shows a particular image transformation. The column gives the percent of descriptors that have a matching descriptor in the transformed image.

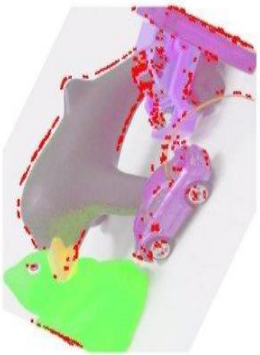


Figure 3: The second image was generated by changing the luminance and the contrast and rotating the first image.

Table 2: Image transformation are applied to a sample of four images. This table gives percent of keys that are found at matching location scale and orientations by apply- ing KNN method

Image transformations	Match%
Increase constrast by 2	95.5%
Decrease luminance by 2	97%
Rotate luminance by 2	94.5%
scale by 0.7	97.25%

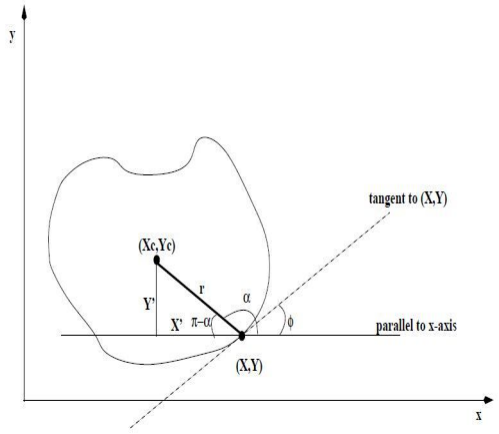


Figure 4: Generalized Hough Transform, special case: fixed orientation and size.

3.1.1 Object Recognition

In order to evaluate how discriminant and invariant are our interest points, we use them in order to learn object models from a set of images, in which these objects are presented to the system, on a uniform background. Then to recognize objects amongst the learnt ones from other scenes with possible occlusions. The learning step allows to build a data base, with associations between objects, views and set of interest points:

$$(object_i, view_{ij}, point_{ijk}) \quad (6)$$

We use 2 steps like Lowe's approach (D.Lowe,1999) in the matching of an image against a database of features. First, we associate each feature to its nearest neighbor by using the closest Euclidean distance in the attribute space.Then, we use the Generalized Hough Transform to predict model orientation and scale from match hypotheses. So we look for all object poses that correspond to a single feature. When cluster of features vote for the same object pose, the probability for having done a correct interpretation is much higher. In the following, we present the algorithm of the generalized hough transform.

• Hough transform

Every point is associated with its attribute vector (with 27 floating values). Some close points in the attribute

space, could belong to different views of different objects. Here in this preliminary version of our recognition system, we do not cluster these points. The generalized Hough Transform can be used to detect arbitrary shapes. It requires the complete specification of the exact shape of the target object.

• Preprocessing step
(a) Pick a reference point (e.g., (xc, yc))

(b) Draw a line from the reference point to the boundary.

(c) Compute ϕ (i.e., perpendicular to gradient's direction).

(d) Store the reference point (xc, yc) as a function of ϕ (i.e., build the R-table)

$$\phi_1 : (r_1^1, \alpha_1^1), (r_2^1, \alpha_2^1), .$$

$$\phi_2 : (r_1^2, \alpha_1^2), (r_2^2, \alpha_2^2), .$$

$$\phi_n : (r_1^n, \alpha_1^n), (r_2^n, \alpha_2^n), .$$

- The R-table allows us to use the contour edge points and gradient angle to recompute the location of the reference point.

Before integrating information in the new image, we should make a geometrical verification with the formula of the similarity transform.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} s\cos(\theta) & -s\sin(\theta) \\ s\sin(\theta) & s\cos(\theta) \end{pmatrix} + \begin{pmatrix} tx \\ ty \end{pmatrix} \quad (7)$$

We note that this equation Ax=b shows a single match between the model and the test image. In order to provide a solution, 3 matches are required. The least square solution can be determined by solving the equation (31) using the pseudo-inverse solution:

$$x = [A^T A]^{-1} A^T b \quad (8)$$

$$e = \sqrt{2 * \| Ax - b \|} \quad (9)$$

If it matches an existing model view, there are two cases : if e > T a new model view is formed from this training image. Else, the new training image is combined with the existing model view. In such case, the similarity transform solution is used to transform the new training image into the coordinates of the model view.

4. Experimental results

4.1 Setting

The recognition method presented above was tested on a dataset that contains 160 training and 51 test images against a uniform background.The test scenes contain between zero and five objects from the learning set, for a total of 79 occurrences. The viewpoint changed significantly between pictures containing a same object.

4.2 Results

The following table summarizes for each object, the obtained data for training and model (results are shown in figure 8).

Table 3: Results of model constructing from Ponce Group Dataset

	Apple	Teddy	Box	shoes	Car
	Learning				
Images	29	20	16	16	28
Features	100	100	100	100	100
	Recognition				
Images	5	7	5	4	6
Features	100	100	100	100	100

In order to recognize objects in a cluttered scene, we proceed as follows : We extract features from the training images in the database. These feature points are matched to the cluttered scene. They are shown on the figures in 11 as red particles on each detected object. We set a threshold at 100 features to improve the efficiency of the extraction.

5. Global Descriptor

5.1 Color and texture measurement

We study the physical aspect of color texture attributes. It is known in physical based approach that the observation process can be viewed as the integration of the energy density over spatio-spectral dimensions(Geusebroek et al., 1999).. We aim to develop a color texture frame work to be used in CBIR tasks. It is well known that the scale space theory developed by Taizo Iijmia (Iijima,
1959) is suggesting that the probes should have a Gaussian shape in order to avoid extra details when observed at a coarser scale (Koenderink and van Doorn, 1987).. The Gaussian scale space is in fact the prototype of linear scale spaces which has connection to the linear diffusion process.

5.1.1 Color measurement

Color is one perceptual result of light in the visible range 400-700 (nm) penetrating into the retina. In the last years, the analysis of color was based on the colometry where the spectral content of the tri-chromatic stimulus are matched

by humans, driving in the well known XYZ color matching function . The color can moreover be measured through the integration of the electromagnetic energy of the image over a spectral bandwidth with the use of Gaussian color model. The Gaussian color model is defined as the Taylor development of the energy distribution to the second order derivatives . One of the advantages of this method is that it reflects human vision process. Let's have the development to the second order of the filtered energy distribution : $C(\lambda,\sigma_\lambda)$

$$C(\lambda,\sigma_\lambda) = C(\lambda_0,\sigma_\lambda) + \lambda C_\lambda(\lambda_0,\sigma\lambda) + \frac{1}{2}\lambda^2 C_{\lambda\lambda}(\lambda_0,\sigma_\lambda) + o(\lambda) \quad (10)$$

$$C(\lambda_0,\sigma_\lambda) = \int E(\lambda)G(\lambda,\lambda_0,\sigma_\lambda)d\lambda \quad (2) \quad (11)$$

$$C_\lambda(\lambda_0,\sigma_\lambda) = \int E(\lambda)G_\lambda(\lambda,\lambda_0,\sigma_\lambda)d\lambda \quad (12)$$

$$C_{\lambda\lambda}(\lambda_0,\sigma_\lambda) = \int E(\lambda)G_{\lambda\lambda}(\lambda,\lambda_0,\sigma_\lambda)d\lambda \quad (13)$$

Where : $G_\lambda(.)$ and $G_{\lambda\lambda}(.)$ denote derivatives of the Gaussian with respect to λ . It is shown that $C(\lambda,\sigma_\lambda)$ represents a theoretical measurement of the color in an image.

5.1.2 Texture measurement

Since Daugman(Koenderink, 1984) (Daugman, 1985) has generalized the Gabor function proposed by(Gabor, 1946b) to model the receptive fields of the selective single cells, they have been widely used within the image processing community. Generally, Gabor filter can be viewed as a modulation of a Gaussian envelope and sinusoidal plane of a particular frequency and orientation .

5.2 Color texture framework

The observation process is viewed as a convolution of the energy as it bounces from a scene with the Gaussian function in the spatio-spectral space. Let's have $CT(x,y,\lambda)$ the function describing our observation:

$$CT(x,y,\lambda) = \iiint E(x,y,\lambda)G(x,y,\lambda,\sigma_s,\sigma_\lambda,\lambda_0)d\lambda$$

Besides, since the texture is characterized with the local spatial frequency, we choose to work in the frequency spectral space with coordinates (u,v, λ):

$$CT(u,v) = \int E(u,v,\lambda)G(u-u_0,v-v_0,\lambda-\lambda_0,\sigma_s,\sigma_\lambda)d\lambda = G(u-u_0,v-v_0,\sigma_s)\int E(u,v,\lambda)G(\lambda-\lambda_0,\sigma_\lambda)d\lambda \quad (15)$$

By applying the equations above we obtain:

$$CT_n(u,v,\lambda) = G(u-u_0,v-v_0,\sigma_s) \quad (16)$$

$$* \int E(u,v,\lambda)G_n(\lambda-\lambda_0,\sigma_\lambda)d\lambda \quad (17)$$

The multiplication in the frequency domain is equivalent to the convolution in the spatial domain. Thus,

$$CT_n(x,y) = Gb(x,y) * \int E(x,y,\lambda)G_n(\lambda-\lambda_0,\sigma_\lambda)d_\lambda) \quad (10)$$

where

$$G_b(x,y) = \frac{1}{2\pi\sigma\beta} e^{-pi\frac{(x-x_0)^2+(y-y_0)^2}{\sigma_s}} e^{i[\Upsilon_0 x + v_0 y]} \quad (19)$$

$(x_0, y_0)$is the center of the receptive field in the spatial domain. (Υ_0, v_0)optimal spatial frequency of the filter in the frequency domain. σ_s is the standard deviation of the elliptical Gaussian along x and y

5.3 Color texture descriptor

In the current study, the color texture descriptor is based on the unified frame work $CT_n(u,v)$ developed in the last subsection. We will use a class of Gabor wavelets instead of Gabor function to insure rotation and scale invariance. The idea of our present contribution is to compute a descriptor based on spatio-spectral properties of the image electro-magnetic energy. In order to increase the efficiency of the Gabor filter, we consider a class of Gabor wavelets.

$$T_{p,q}(x,y) = a^{-p}t(x',y'), a>1, p,q \in Z \quad (20)$$

where

$$x' = a^{-p}(x\cos(\theta) + y\sin(\theta)) \quad (21)$$

$$y' = a^{-p}(-x\sin(\theta) + y\cos(\theta)) \quad (22)$$

where

$$\theta = \frac{q\pi}{K} \quad (14)$$

K is the total number of orientations

$$CT_{npq}(x,y) = T_{pq}(x,y) * \int E(x,y,\lambda)G_n(\lambda-\lambda_0,\sigma_\lambda)d_\lambda) \quad (23)$$

$$CT_{npq}(x,y) = | A_{npq}(x,y) + jB_{npq}(x,y) | \quad (24)$$

However, the efficiency of this color-texture framework is adversely affected by the high dimensionality and computational complexity of the Gabor filter. This problem was treated in(Bhaga vathy et al., 2003).We propose then the following approach to overcome it : In fact, Dunn and Higgin demonstrated that the 1-dimensional Gabor filter outputs have a Rice distribution distribution (D.Dunn and Higgins, 1995). This property was extended by Bhagavathy, et al to the 2 dimensional case (Lazebnik et al., 2003). Thus, we have

$$A_{npq}(x,y) = R_{npq}(x,y)cos(\theta_{npq}(x,y)) \quad (25)$$

where $R_{pqn}(x,y)$ has the following PDF,

$$f_R(r) = \frac{r}{\sigma^2} exp(-\frac{r^2 + A_0^2}{2*\sigma^2}) I_0(\frac{A_0 r}{\sigma^2}) \quad (26)$$

and $I_0(\frac{A_0 r}{\sigma^2})$ represents the zero-order modified Bessel function of the first order.

$R_{pqn}(x,y)$ is a function containing spatio-spectral informations of the image. Therefore, we consider $R_{pqn}(x,y)$ as a local feature of color texture attributes within an image. p,q are respectively the scale, the orientation relative to texture. And n is the color channel. Besides, it is shown that the Rice PDF can vary from the Rayleigh PDF for small A_0 to approximate Gaussian PDF to large A_0. The first case is the most convenient since it concerns a wide range of textures. Bhagavathy, et al demonstrate that it is possible to compute the Rayleigh descriptor from the Gabor one by using the following equation [?]:

$$\Gamma^2_{mnq} = \frac{1}{2}(\Xi^2_{pqn} + \sigma^2_{pqn}) \quad (27)$$

Ξ_{pqn}, σ_{pqn} are respectively the mean and standard deviation corresponding to the channel n. The dimensionality of Rayleigh filter is almost 50% of Gabor filter with only a trade off of less than 3% on the error rate.

Principle:

(a) Compute descriptors for the color textured image based on the mean and the standard deviation for s scales, k orientations and 3 channels.

(b) Compute the Rayleigh feature using the Gabor one for each channel. The dimentionality of the resulting vector is 3*(s*k+2)

(c) Concatenating the 3 descriptors for each channel

Thus for the each spectral channel k = 0,1,2, we obtain three descriptor vectors:

$$D_{\Xi\Sigma 0} = [\Xi_{000}, \sigma_{000}, ..., \Xi_{s-1,k-1,0}, \sigma_{s-1,k-1,0}, \Xi_{I0}, \sigma_{I0}] \quad (28)$$

$$D_{\Xi\Sigma 1} = [\Xi_{001}, \sigma_{001}, ..., \Xi_{s-1,k-1,1}, \sigma_{s-1,k-1,1}, \Xi_{I1}, \sigma_{I1}] \quad (29)$$

$$D_{\Xi\Sigma 2} = [\Xi_{002}, \sigma_{002}, ..., \Xi_{s-1,k-1,2}, \sigma_{s-1,k-1,2}, \Xi_{I2}, \sigma_{I2}] \quad (30)$$

We construct the final descriptor by concatenating the three channel descriptors:

$$D_{\Xi\sigma} = (D_{\xi\sigma 0}, D_{\xi\sigma 1}, D_{\xi\sigma 2}) \quad (31)$$

From the $D_{\Xi\sigma}$, the Rayleigh parameter of the output distribution is given by :

$$\Gamma^2_{mnq} = \frac{1}{2}(\Xi^2_{pqn} + \sigma^2_{pqn}) \quad (32)$$

5.4 Implementation and experiments

5.4.1 Implementation

In order to evaluate the proposed descriptor of image retrieval , we use an RGB camera. As mentioned by Geusbroek and al., the best linear transform from RGB to the Gaussian color model is given by:

$$G = MC^T \quad (16)$$

Where

$$M_g = \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix} \quad (33)$$

C=(R G B)

We use also Gabor wavelets which configuration is proposed in [Geusebroek et al].

Table 4 : filter configuration

scales	4, 3.5, 2.95, 2.35, 1.75
central frequencies	0.05, 0.08, 0.14, 0.22, 0.33
orientations	$-\pi/4$,- $\pi/2$, $\pi/4$, $\pi/2$

5.4.2 Experiments

Experiments were performed on a Pentium 4 with 4 GB of main memory and 40 GB of storage. The programs have been implemented in MATLAB. We evaluate our method using 144 images. The images are captured from different view points and under different scales and orientations. They represent moreover different themes such as landscapes, animals, monuments and people. Each image is

represented by a vector of 66 elements (5 scales, 4 orientations and 3 channels). The following figure shows an example of simulations.

Figure 5: The query image

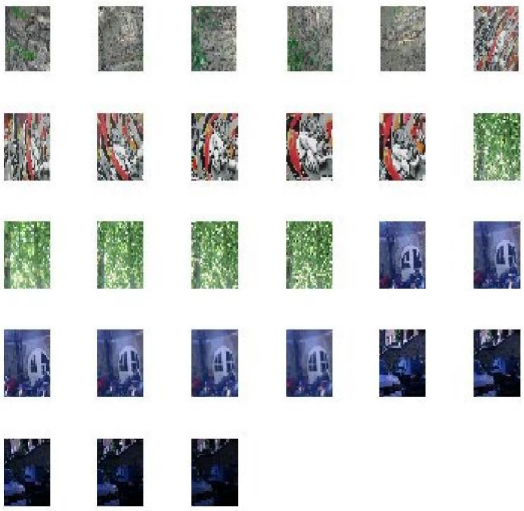


Figure 6: Retrieved images, from the most similar to the least similar

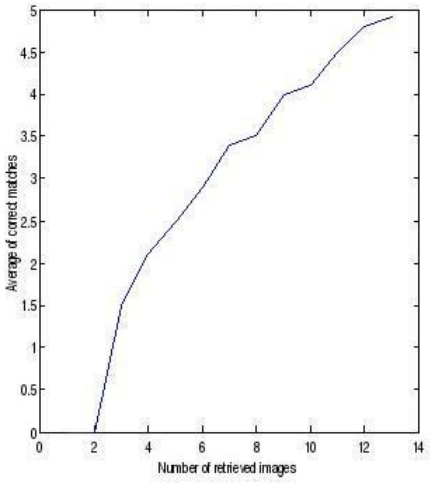
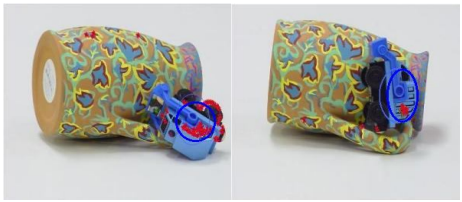


Figure 7: Results for retrieving images using our global feature

We take a sample of 10 images of a particular scene. Then we apply our global descriptor on each of these images. We do matching by comparing these descriptors and the query image descriptor. We compute for that the coefficient of pairwise linear correlation between each pair of the query vector and the vectors of images. Then best match is obtained by looking for the maximum correlation. The figure 3 shows the average of correct matches according to the number of retrieved images. We observe that the average is growing with the number of images.

Object (A)



Object (C)

Object (D)

Object (E)

Object (F)

Object (G)

Object (H)

Figure 8: Recognition of objects from A to H learned by our method Objects are captured by our camera in LAAS. Others are from Ponce Group Dataset and from the Amsterdam library of object images. Recognized objects are identified with red points on the objects and with a blue circle on it .

6. Conclusion

In this paper, we have presented a new color texture descriptor. We have also developed color-texture framework which constructs feature with reasonable dimensionality. This descriptor might be used in many Content-Based Image Retrieval systems to improve their performances. Then, we have presented a performing interest point detection-description scheme which allows to do object recognition for color textured images. Future work will aim at optimizing the code and applying it to robotic navigation.

Références

[1] Bhagavathy, S., Tesic, J., and Manjunath, B. (2003). On the rayleigh nature of gabor filter outputs. In ICIP.

[2] B.S.Manjunath and W.S.Ma (1996). Texture features for browsing and retrieval of image data. In IEEE Trans. on Patt. Anal. And Mach. Intell., 18, 837-842.

[3] Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. In J. Optical Soc. Amer.

[4] D.Dunn and Higgins, W. E. (1995). Optimal gabor filters for texture segmentation. In IEEE Trans. Image Proc.

[5] D.Lowe (1999). Object recognition from local scale invariant features.
Gabor, D. (1946a). Theory of communication. In Journal I.E.E.

[6] Gabor, D. (1946b). Theory of communication. In Journal I.E.E.

[7] Geusebroek, J., den Boomgaard, R. V., Smeulders, A., and Dev, A. (2000). Color and scale: The spatial structure of color images. In ECCV.

[8] Geusebroek, J., R.Boomgaard, Smeulders, A. W. M., and Geerts, H. (2001a). Color invariance.

[9] Geusebroek, J. M., den Boomgaard, R., A.Smeulders, and Geerts, H. (2001b). Color invariance. In IEEE Trans. on Patt. Anal. And Mach. Intell.

[10] Geusebroek, J.-M., den Boomgaard, R. V., Smeulders, A. M., and Dev, A. (1999). Color invariant edge detection. In Scale-Space Theories in Computer Vision.
[11] H. Bay, T.Tuytelaars, L. G. (2006). Surf: Speeded up robust features.
[12] Hoang, M., Geusebroek, J., and A.W.M.Smeulders (2005). Color texture measurement and segmentation. In ELSEVIER, Signal Processing.
[13] Iijima, T. (1959). Basis theory of pattern observation. In Papers of Technical Group on Automata and Automatic Control.
[14] J.G.Daugman (1980). Two-dimensional spectral analysis of cortical receptive field profile. In Vision Research, vol. 20, pp. 847-856.
[15] K.Mikolajzyk and C.Schmid (2004). Scale affine invariant interest point detectors.
[16] Koenderink, J. (1984). The structure of images. In Bio.Cybern.
[17] Koenderink, J. and van Doorn, A. (1987). Representation of local geometry in the visual system. In Biological Cybernetics.
[18] Lazebnik, S., Schmid, C., and Ponce., J. (2003). Sparse texture representation using invariant neighbourhoods. In Proc. of the Conf. on Computer Vision and Pattern Recognition, Madison.
[19] Lee, T. (1996). Image representation using 2d gabor wavelets. In IEEE Trans. on Patt. Anal. And Mach. Intell.
[20] Liapis, S. and Tziritas, G. (2004). Color and texture image retrieval using chromacity histograms and wavelet frames. In IEEE Trans. Multimedia.
[21] Manjunath, B. and Ma, W. (1996). Texture features for browsing and retrieval of image data. In IEEE Trans. on Patt. Anal. And Mach. Intell.
[22] Montesinos, P., Gouet, V., and Deriche, R. (1998). Differential invariants for color images. In International Conference on Pattern Recognition.
[23] Ravela, S. S. (2003). On multi-scale differential features and their representations for image retrieval and recognition. In Ph.d University of Massachusetts Amherst.
[24] Yeh, C. and Kuo, C. J. (2003). Iteration-free clustering algorithm for nonstationary image database. In IEEE Trans. Multimedia, vol.5, No. 2.
[25] Yoo, H., Park, H., and Jang, D. (2005). Expert system for color image retrieval. In ELSEVIER, Expert Systems with Applications.