

Applied Mathematical Sciences, Vol. 5, 2011, no. 42, 2109 - 2136

Global and Local Image Descriptors for Content Based Image Retrieval and Object Recognition

Younes RAOUI

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

El Houssine BOUYAKHF

Faculty of Sciences, Mohamed V University, 4 Street Ibn Battouta B.P. 1014 RP, Rabat, Morocco

Michel DEVY

Laboratory of Analysis and Architecture of Systems(LAAS), 7 Street Colonel Roche, Toulouse, France

Fakhita REGRAGUI

Faculty of Sciences, Mohamed V University, 4 Street Ibn Battouta B.P. 1014 RP, Rabat, Morocco

Abstract

This article develops two descriptors global and local for colored images. The first allows to do image retrieval while the second is applied to object recognition. First we combine color and texture attributes in a unique framework. Our approach is based on the physical properties of light bouncing from a scene. Furthermore, it is claimed that the distribution of Gabor filter outputs can also be provided by a Rayleigh. Filter. Thus, we deduce the Rayleigh-Gaussian framework from the Gabor-Gaussian one. Then we compute a descriptor based on statistical measures performed on Rayleigh-Gaussian features. Secondly, we propose a new detection function for interest points and a new characterization of such points. The detection function extends the Harris Laplace method. Our function extracts features invariant to rotation and scale from color images. We use the second moment as a basis of corner decision. Around each feature point we compute a texture descriptor using the Gabor filter. As an evaluation of this new class of interest points, we have implemented a recognition method of 3D objects by indexing a data base on object views; this method is similar to the Lowe method, except that that SIFT features and descriptors are replaced by the ones proposed in our approach.

This recognition method exploits a KNN classifier to match interest points from their descriptors, and then a Hough transform to cluster reliable point matches, voting for a consistent similarity transform.

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

1 INTRODUCTION

This paper presents two works on image description and retrieval. Firstly we present a new technique that computes a global descriptor of color texture images. Secondly we describe a method which computes a detector and a descriptors from colored images invariant to scale and rotation. In fact, The acquisition and use of images know an explosif growth. The retrieval systems called Picture archiving and communication systems(PACS) used limited text keywords and saved them as image headers. However, these keywords do not reflect the richness of features that contain the image itself. The Content Based Image Retrieval has received significant attention in the litterature. Their apparition is considered as a turning point because they allows an intelligent management of images [1]. Using this approach, users can provide an example image or image features (color, texture, shape ...) to find similar images.

Swain and Ballard were among the pioneers of this approach [2]. The hypothesis under the CBIR approach is that the statistical measures of similarity among representations of image attributes do correlate with subjective visual similarity [3].

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 that extracts global attributes which describes color and texture attributes. 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 [4] in order to probe an observed scene. 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 [5]. We will use this result to compute a more compact color texture feature.

In our second work, we present a novel approach 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) [6] which has been the source of inspirations for many derivations. SURF(Speeded Up Robust Features)[7] 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 [4]. Formally, two objects or two appearances of objects are equivalent under a group of transformations T if they belong to the same equivalence class [4].

Event though the Harris multiscale technique represents one of the state-of-the-art approaches to object detection/recognition, it has some deficiencies when applied to color textured images. Thus, we compute a detector that extracts corners from images with the well known Harris approach. We make this detector suitable to colored images by representing the image in the Gaussian color model developed by Koendering. To insure invariance to scale, we compute a characteristic vector of a local structure around each feature point.

Object recognition has been developed quickly in recent years because of advances in machine learning, modern approaches feature extraction techniques and greater availability of image databases. We apply our new features to recognize small learned objects in cluttered scenes. Up to now, the methods used are interested to objects that takes a large areas in a scene. Many recent approaches which make use of an object in real world scale such as a prior for the scale in the image. Hoeien et al [8] infers with an image jointly 3D object locations and scene information (3D surface, orientation). Moreover Gould et al [9] et Quiglez et al used object scale in recognition. They acquire high accuracy depth maps with laser scanners placed on a robot. Then object detector utilizes both surface variations, 3D shape and appearance to find objects. The paper is organised as follows : In section I, 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 II, we present our local feature by developing the detection scheme and then we present our descriptor. In section III, we give an evaluation of our descriptor and we show results of objects recognition. In Section IV we give a summary of the paper and propose some future works.

2 GLOBAL DESCRIPTOR

2.1 Color and texture measurement

We study the physical [10] 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[11]. 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 [12] is suggesting that the probes should have a Gaussian shape in order to avoid extra details when observed at a coarser scale [13]. The Gaussian scale space is in fact the prototype of linear scale spaces which has connection to the linear diffusion process.

2.1.1 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 [14]. 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 [14]. 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 : $C(\lambda, \sigma_\lambda)$

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

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

$$C_\lambda(\lambda_0, \sigma_\lambda) = \int E(\lambda) G_\lambda(\lambda, \lambda_0, \sigma_\lambda) d\lambda \quad (3)$$

$$C_{\lambda\lambda}(\lambda_0, \sigma_\lambda) = \int E(\lambda) G_{\lambda\lambda}(\lambda, \lambda_0, \sigma_\lambda) d\lambda \quad (4)$$

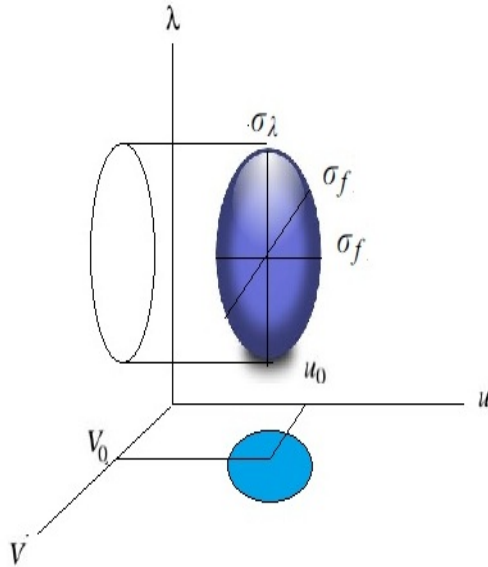


Figure 1: The 3D probe function

Where : $G_\lambda(\cdot)$ and $G_{\lambda\lambda}(\cdot)$ 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 [15].

2.1.2 Texture measurement

Texture is a key component of visual perception. Like color or shape, texture is an essential feature to consider querying. Many systems based on color or texture can be found. Among the characteristics of texture, some of the most commonly used is second order statistic that computes certain matrix values on which we can do several specific calculations. Since Daugman [16] [17] has generalized the Gabor function proposed by [18] 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 [19].

2.2 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:

$$CT(x, y, \lambda) = \int \int \int E(x, y, \lambda) G(x, y, \lambda, \sigma_s, \sigma_\lambda, \lambda_0) d\lambda \quad (5)$$

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) :

$$\begin{aligned} CT(u, v) &= \int E(u, v, \lambda) G(u - u_0, v - v_0, \lambda - \lambda_0, \sigma_s, \sigma_\lambda) d\lambda \quad (6) \\ &= G(u - u_0, v - v_0, \sigma_s) \int E(u, v, \lambda) G(\lambda - \lambda_0, \sigma_\lambda) d\lambda \end{aligned}$$

By applying the equations above we obtain:

$$\begin{aligned} CT_n(u, v, \lambda) &= G(u - u_0, v - v_0, \sigma_s) \quad (7) \\ &* \int E(u, v, \lambda) G_n(\lambda - \lambda_0, \sigma_\lambda) d\lambda \end{aligned}$$

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

$$CT_n(x, y) = Gb(x, y) * \int E(x, y, \lambda) G_n(\lambda - \lambda_0, \sigma_\lambda) d\lambda \quad (8)$$

where

$$G_b(x, y) = \frac{1}{2\pi\sigma\beta} e^{-pi \frac{(x-x_0)^2 + (y-y_0)^2}{\sigma_s}} e^{i[\Upsilon_0 x + v_0 y]} \quad (9)$$

(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

2.3 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, as suggested in [1] we consider a class of Gabor wavelets.

$$T_{p,q}(x, y) = a^{-p} t(x', y'), a > 1, p, q \in Z \quad (10)$$

where

$$x' = a^{-p}(x \cos(\theta) + y \sin(\theta)) \quad (11)$$

$$y' = a^{-p}(-x \sin(\theta) + y \cos(\theta)) \quad (12)$$

where

$$\theta = \frac{q\pi}{K}$$

K is the total number of orientations

$$CT_{npq}(x, y) = T_{pq}(x, y) * \int E(x, y, \lambda) G_n(\lambda - \lambda_0, \sigma_\lambda) d\lambda \quad (13)$$

$$CT_{npq}(x, y) = |A_{npq}(x, y) + jB_{npq}(x, y)| \quad (14)$$

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 [5]. 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 [20]. This property was extended by Bhagavathy, et al to the 2 dimensional case [21]. Thus, we have

$$A_{npq}(x, y) = R_{npq}(x, y) \cos(\theta_{npq}(x, y)) \quad (15)$$

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

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + A_0^2}{2 * \sigma^2}\right) I_0\left(\frac{A_0 r}{\sigma^2}\right) \quad (16)$$

and $I_0\left(\frac{A_0 r}{\sigma^2}\right)$ 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 [5]:

$$\Gamma_{mnq}^2 = \frac{1}{2}(\mu_{pqn}^2 + \sigma_{pqn}^2) \quad (17)$$

μ_{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:

1. Compute descriptors for the color textured image based on the mean and the standard deviation for s scales, k orientations and 3 channels.
2. Compute the Rayleigh feature using the Gabor one for each channel. The dimensionality of the resulting vector is $3*(s*k+2)$
3. Concatenating the 3 descriptors for each channel

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

$$D_{\mu\sigma 0} = [\mu_{000}, \sigma_{000}, \dots, \mu_{s-1,k-1,0}, \sigma_{s-1,k-1,0}, \mu_{I0}, \sigma_{I0}] \quad (18)$$

$$D_{\mu\sigma 1} = [\mu_{001}, \sigma_{001}, \dots, \mu_{s-1,k-1,1}, \sigma_{s-1,k-1,1}, \mu_{I1}, \sigma_{I1}] \quad (19)$$

$$D_{\mu\sigma 2} = [\mu_{002}, \sigma_{002}, \dots, \mu_{s-1,k-1,2}, \sigma_{s-1,k-1,2}, \mu_{I2}, \sigma_{I2}] \quad (20)$$

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

$$D_{\mu\sigma} = (D_{\xi\sigma 0}, D_{\xi\sigma 1}, D_{\xi\sigma 2}) \quad (21)$$

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

$$\Gamma_{mnq}^2 = \frac{1}{2}(\mu_{pqn}^2 + \sigma_{pqn}^2) \quad (22)$$

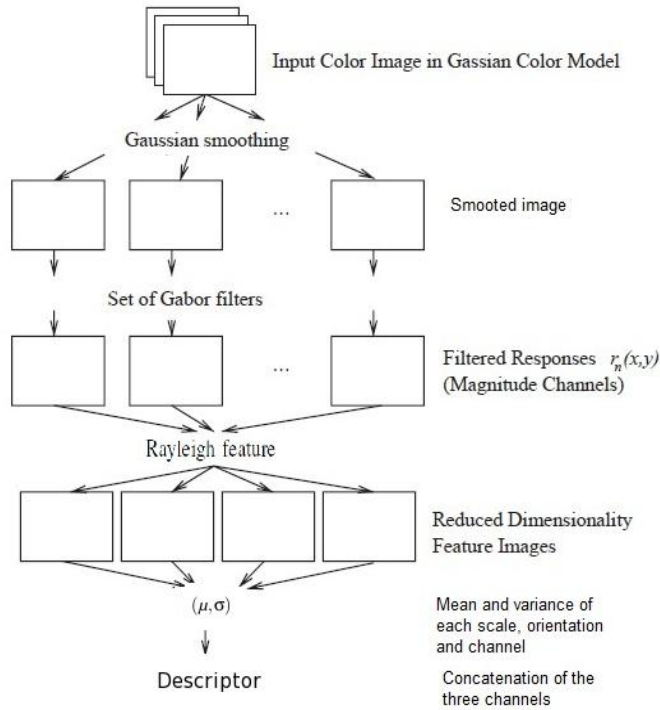


Figure 2: Color Texture description scheme

2.4 Implementation and experiments

2.4.1 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

$$M_g = \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix} \quad (23)$$

C=(R G B)

We use also Gabor wavelets which configuration is proposed in [22].

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

2.4.2 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.

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.

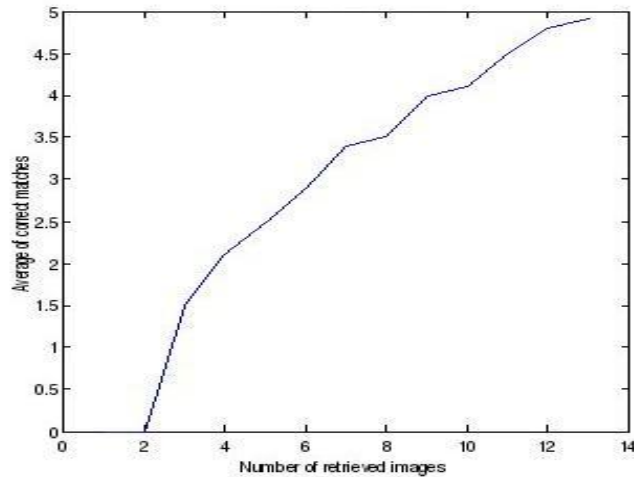


Figure 4: Results for retrieving images using our global feature

3 LOCAL DESCRIPTOR

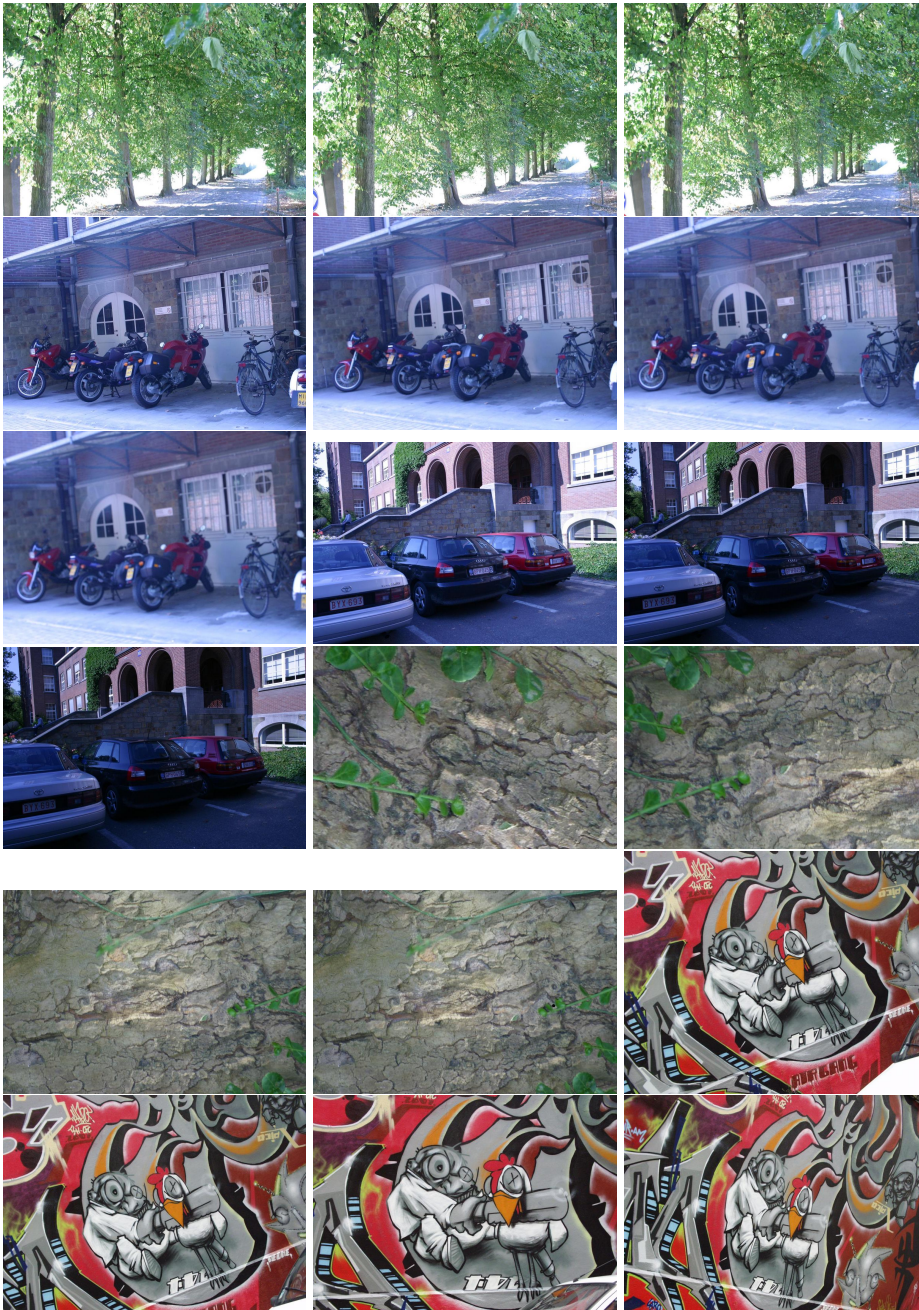
Our goal has been to develop both detector and descriptor that has a low dimensionality without scarifying its performance. We aimed to insure the distinctiveness of our descriptor while reducing its dimension. Many descriptors has been developed in literature in order to resolve problems like robot navigation and grasping. We get feel for what are the aspects contributing to the performances when doing experiments on object recognition. We show that our detector descriptor is more distinctive and equally repeatable. Moreover, the invariance of features relied on



Query image



Retrieved images



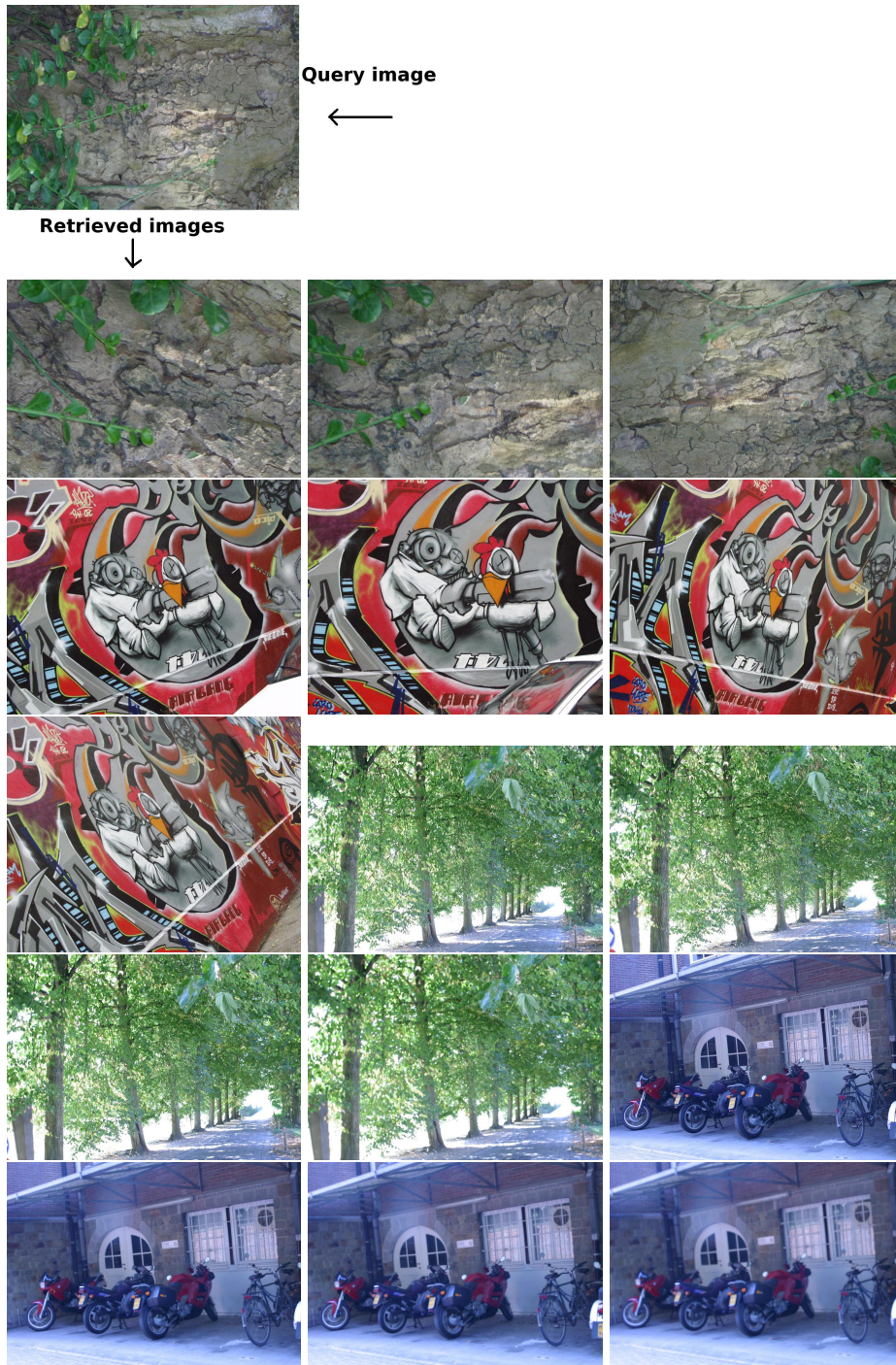


Figure 3: Retrieved images, from the most similar to the least similar

geometric and photometric deformations is of high importance for our application. We have been interested to the scale and image rotation invariance. In fact, these elements offer a compromise between feature complexity and robustness regarding to deformations. That's why we have considered second order effects such as skew, anisotropic scaling and perspective effects. Concerning the photometric deformation, we have assumed a Gaussian color model which represents the 3 components of the luminance in the Gaussian color model.

3.1 Detector computation

We use the Harris detector[23] 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 :

$$M(x, y, \sigma_I, \sigma_D) = \sigma_D g(\sigma_I) \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} \quad (24)$$

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.

$$cornerness = det((M(x, \sigma_i, \sigma_D))) - \alpha.trace^2(M(x, \sigma_i, \sigma_D)) \quad (25)$$

The Harris detector applies only to gray scale images. Montesinos et al generalized it to color images[24]. 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 [25] with :

$$M_g = \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix} \quad (26)$$

And ponderate the three components of the Gaussian derivatives by α, β, γ

Algorithm 1

```

1: in ← Readimage
2: R ← in(1)
3: B ← in(2)
4: G ← in(3)
5: M ←  $\begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix}$ 
6: Diagonalize M → (α,β,γ)
7: S ← α.R + β.G + γ.B
8: Lxx ← Gxx.S
9: Lxy ← Gxy.S
10: Lyy ← Gyy.S
11: M(x,y,σI,σD) ← σDg(σI)  $\begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix}$ 

```

Algorithm 2

```

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

```

3.1.1 Scale computation

We want to select a characteristic 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 characteristic scales :

3.1.2 Orientation assignment

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 :

$$R_{ij} = a \tan(E_{ij} - E_{i+1,j}, E_{i,j+1} - E_{ij}) \quad (27)$$

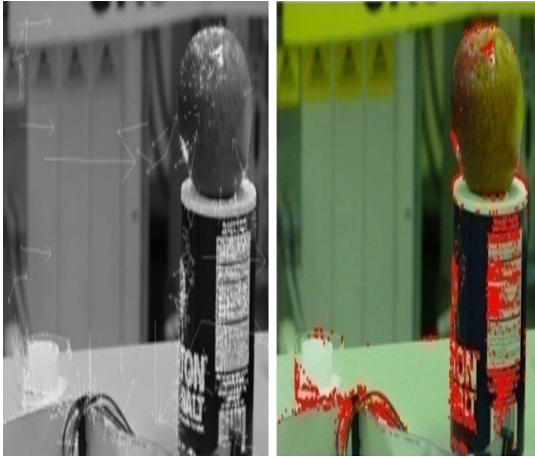


Figure 5: Features extracted both with SIFT detector(gray scale image) and by our detector

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.

x y	σ	r	t
space	scale	orientation	texture

Figure 6: Structure of our descriptor, 2 bin for space, 21 for scale, 5 for orientation, 9 for texture.

3.2 Descriptor computation

We compute a texture invariant by using Gabor wavelets at different orientations and scales [26]. We use 9 kernels and we do convolution with small windows around each feature point[27][28].

4 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.

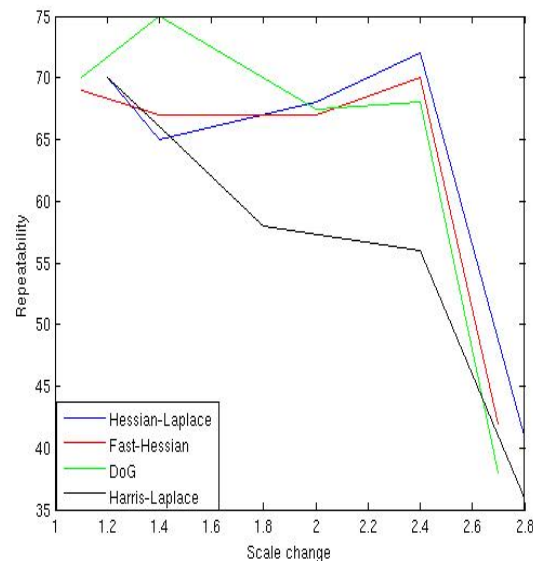
4.1 Repeatability

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 [23].

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 $S \leq 0.4$. In the case of scale invariant points the surface error is:

$$e_s = \left| 1 - s^2 \frac{\min(\sigma_a^2, \sigma_b^2)}{\max(\sigma_a^2, \sigma_b^2)} \right| \quad (28)$$

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



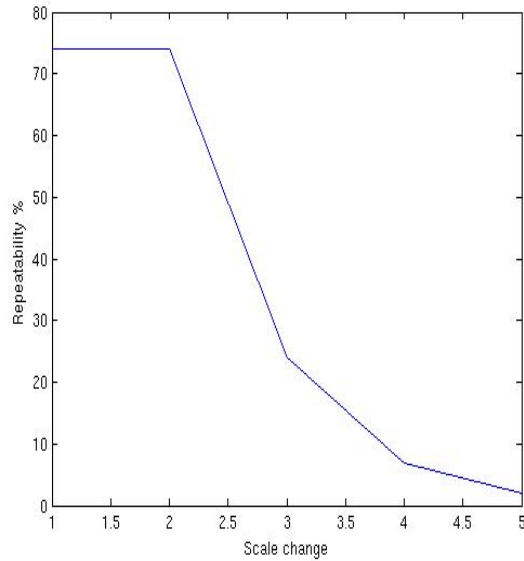


Figure 7: First Figure : Repeatability of Fast Hessian, DoG, Harris-Laplace, Hessian-Laplace interest points detectors with respect to scale, Second Figure : Repeatability of our detector descriptor with respect to scale

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.

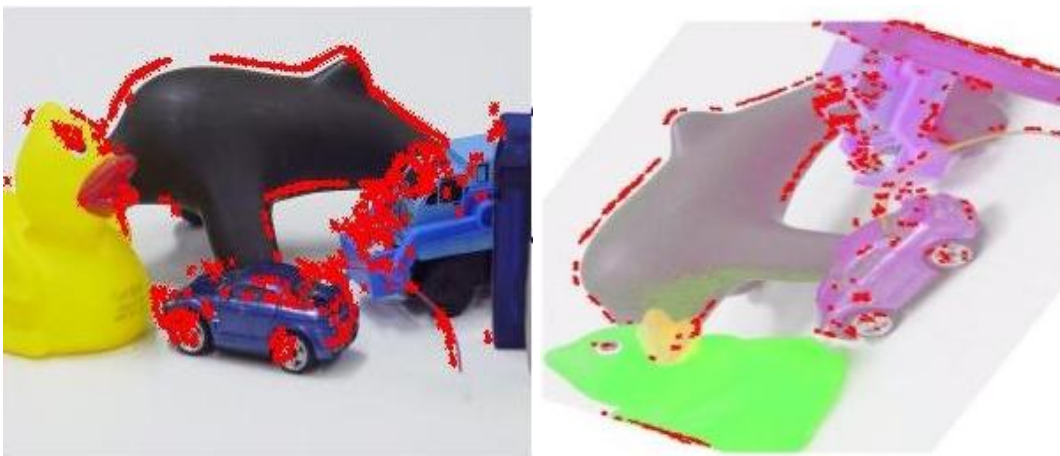


Figure 8: The second image was generated by changing the luminance and the contrast and rotating the first image.

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 9: 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 applying KNN method

4.1.1 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:

$$(object_i, view_{ij}, point_{ijk}) \quad (29)$$

We use 2 steps like Lowe's approach [6] 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.

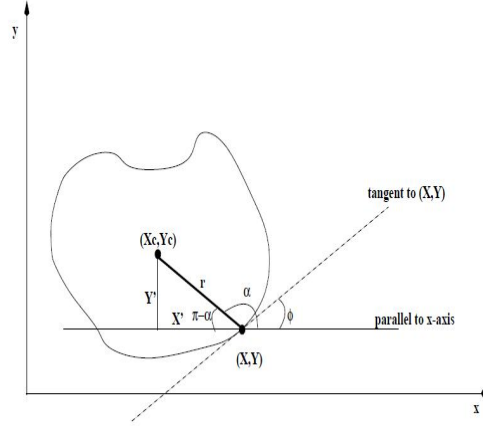


Figure 10: Generalized Hough Transform, special case: fixed orientation and size.

- Preprocessing step

1. Pick a reference point (e.g., (x_c, y_c))
2. Draw a line from the reference point to the boundary.
3. Compute ϕ (i.e., perpendicular to gradient's direction).
4. Store the reference point (x_c, y_c) as a function of ϕ (i.e., build the R-table)

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

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

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

- 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.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} s \cos(\theta) & -s \sin(\theta) \\ s \sin(\theta) & s \cos(\theta) \end{pmatrix} + \begin{pmatrix} tx \\ ty \end{pmatrix} \quad (30)$$

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:

$$x = [A^T A]^{-1} A^T b \quad (31)$$

$$e = \sqrt{2 * ||Ax - b||} \quad (32)$$

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.

5 EXPERIMENTAL RESULTS

5.1 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.

5.2 Results

The following table summarizes for each object, the obtained data for training and model.

	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

Figure 11: Results of model constructing from Ponce Group Dataset

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.

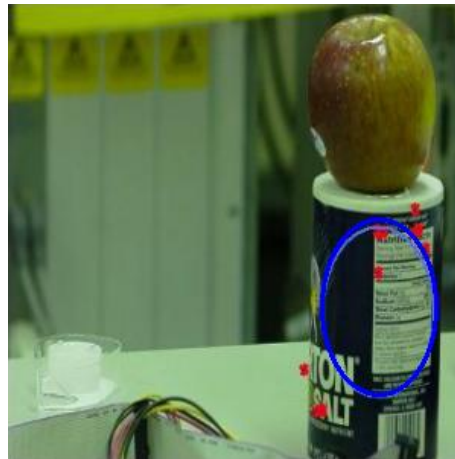


Object (A)





Object (B)





Object (C)



Object (D)

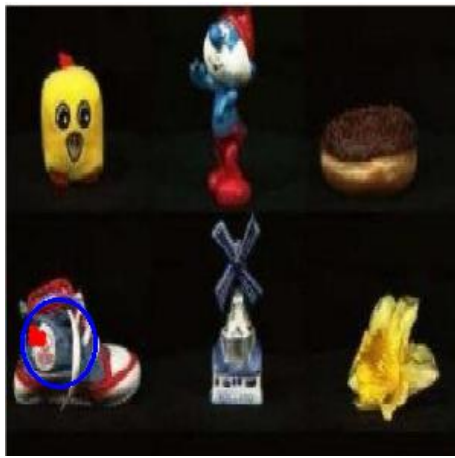


!!

Object (E)



Object (F)



Object (G)



Object (H)

Figure 12: 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.

References

- [1] C. Yeh and C. J. Kuo, "Iteration-free clustering algorithm for nonstationary image database," in *IEEE Trans. Multimedia*, vol.5, No. 2, 2003.
- [2] H. Yoo, H. Park, and D. Jang, "Expert system for color image retrieval," in *ELSEVIER, Expert Systems with Applications*, 2005.
- [3] S. S. Ravela, "On multi-scale differential features and their representations for image retrieval and recognition," in *Ph.d University of Massachusetts Amherst*, 2003.
- [4] M. Hoang, J. Geusebroek, and A.W.M.Smeulders, "Color texture measurement and segmentation," in *ELSEVIER, Signal Processing*, 2005.
- [5] S. Bhagavathy, J. Tesic, and B. Manjunath, "On the rayleigh nature of gabor filter outputs," in *ICIP*, 2003.
- [6] D.Lowe, "Object recognition from local scale invariant features," 1999.
- [7] L. G. H. Bay, T.Tuytelaars, "Surf: Speeded up robust features," 2006.
- [8] D. Hoiem, A. Efros, and M. Heber, "Putting objects in perspectives," in *CVPR*, 2006.
- [9] S. Gould, P. Baumstarck, M. Quigley, A. Y. Ng, and D. Koller, "Integrating visual and range data for robotic object detection," in *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2)*, 2008.

- [10] M. Quigley, S. Batra, S. Gould, E. Klingbeil, Q. Le, A. Wellman, and A. Y. Ng, "High-accuracy 3d sensing for mobile manipulation : Improving object detection and door opening," in *ICRA*, 2009.
- [11] J.-M. Geusebroek, R. V. den Boomgaard, A. M. Smeulders, and A. Dev, "Color invariant edge detection," in *Scale-Space Theories in Computer Vision*, 1999.
- [12] T. Iijima, "Basis theory of pattern observation," in *Papers of Technical Group on Automata and Automatic Control*, 1959.
- [13] J. Koenderink and A. van Doorn, "Representation of local geometry in the visual system," in *Biological Cybernetics*, 1987.
- [14] J. Geusebroek, R. V. den Boomgaard, A. Smeulders, and A. Dev, "Color and scale: The spatial structure of color images," in *ECCV*, 2000.
- [15] J. M. Geusebroek, R. den Boomgaard, A. Smeulders, and H. Geerts, "Color invariance," in *IEEE Trans. on Patt. Anal. And Mach. Intell.*, 2001.
- [16] J. Koenderink, "The structure of images," in *Bio.Cybern*, 1984.
- [17] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," in *J. Optical Soc. Amer.*, 1985.
- [18] D. Gabor, "Theory of communication," in *Journal I.E.E.*, 1946.
- [19] T. Lee, "Image representation using 2d gabor wavelets," in *IEEE Trans. on Patt. Anal. And Mach. Intell.*, 1996.
- [20] D. Dunn and W. E. Higgins, "Optimal gabor filters for texture segmentation," in *IEEE Trans. Image Proc.*, 1995.
- [21] S. Lazebnik, C. Schmid, and J. Ponce., "Sparse texture representation using invariant neighbourhoods," in *Proc. of the Conf. on Computer Vision and Pattern Recognition, Madison*, 2003.
- [22] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," in *IEEE Trans. on Patt. Anal. And Mach. Intell.*, 1996.
- [23] K. Mikolajczyk and C. Schmid, "Scale affine invariant interest point detectors," 2004.
- [24] P. Montesinos, V. Gouet, and R. Deriche, "Differential invariants for color images," in *International Conference on Pattern Recognition*, 1998.
- [25] J. Geusebroek, R. Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," 2001.

- [26] B.S.Manjunath and W.S.Ma, "Texture features for browsing and retrieval of image data," in *IEEE Trans. on Patt. Anal. And Mach. Intell.*, 18, 837-842, 1996.
- [27] D. Gabor, "Theory of communication," in *Journal I.E.E.*, 1946.
- [28] J.G.Daugman, "Two-dimensional spectral analysis of cortical receptive field profile," in *Vision Research*, vol. 20, pp. 847-856, 1980.

Received: December, 2010