SETIT 2009
5th International Conference: **S**ciences of **E**lectronic,
Technologies of **I**nformation and **T**elecommunications
March 22-26, 2009 – TUNISIA

Scale and Rotation Invariant Local Feature Using Harris Laplace Detector in Color Textured Images

Younès RAOUI*,**, Michel DEVY** and El Houssine BOUYAKHF*

Faculté des Sciences, Université Mohamed V Rabat, Maroc
yraoui@laas.fr
bouyakhf@fsr.ac.ma

**Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS Toulouse, France*
michel@laas.fr

Abstract : In this article, 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. For that we define a new measurement from the component of the Taylor expansion of luminance. Then 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 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:. Results are presented on classical images of the COIL data base.

Keywords: color texture, Harris Laplace, Koendering model of color, KNN, interest points, object recognition, indexation

INTRODUCTION

The feature extraction from images allows to do object recognition. This is an important task for localization. Local features are the best concept in feature extraction. They are divided into two types, corner detectors like Harris or Hessian matrices, scale invariant detectors like Harris laplace where the Harris or Hessian matrices find the corners and the Laplacian select the scale.

The Harris Affine detector [7] uses a multiscale version of the Harris detector. It is well known that the Harris detector itself is an amelioration of the Moravec detector that computes the variations of intensities in a windowed region. In order to make this matrix independent to the image resolution, we use a matrix adapted to the scale changes with the automatic scale selection concept introduced by Lindberg. Precisely they are based on the eigen values of the second moment matrix.

The SIFT features are invariant to image translation, scaling, rotation and partially invariant to illumination changes and 3D projections [2]. Based on scale space theory, a difference of Gaussian function is used over multiple scales to identify potential interest

points invariant to scale and orientation. At these candidates, a detailed model is fit to determine locations and scales. Then, an orientation is assigned to each keypoint by computing the local image gradient directions. Finally, in each region around the keypoint, image gradients are measured.

The SURF detector is based on the computation of features and the descriptor. It uses the DoG operator as a Hessian matrix to improve time performance. SURF describes with a distribution of Haar wavelets responses within the interest point neighborhood with a use of integral images [8].

In our article, we propose to compute a new descriptor by extracting local features from color textured images. The descriptor contains entries for scale, orientation and local frequencies around each feature. We will use the Harris Laplace in order to compute coordinates of each feature and the characteristic scales. Then the descriptor is constructed by computing around each descriptor a distribution of the Gabor wavelets in order to integrate both color and texture.

We start by describing some theory of texture and color, then we present the principle of the Harris Affine detector and our method that integrate color. Then we show our approach for matching by KNN and hough transform

1. STATE OF THE ART

1.1. Color measurement :

Color is the perceptual result of light in the visible domain. The colorimetry allows, in the field of image processing, to analyse the spectrum by focusing on the spectral content of the trichromatic stimulis. The Gaussian color model developped by Koendering allows to measure color with derivatives of the the energy distribution.

This model is defined as the Taylor development of the energy distribution to the second order.

Let's have this development, where C is the color luminance :

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

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

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

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

E is the electromagnetic energy of the image.

$G_\lambda(.)$ and $G_{\lambda\lambda}(.)$ are the Gaussian derivatives.

$C(\lambda,\sigma_\lambda)$ represents the color function.

The Gaussian Color model provides a tool to measure colored object reflectance. It explores the spectral and spatial structure of color. It extends the spectral formulation where the well known N jet is used :

$$E_{\lambda^m x^{-n}}(\lambda,\vec{x}) = E(\lambda,\vec{x})*G_{\lambda^m x^{-m}}(\lambda,\vec{x},\sigma_\lambda,\sigma_{\vec{x}}) \quad (5)$$

and $G_{\lambda^m x^{-n}}(\lambda,\vec{x};\sigma_\lambda,\sigma_{\vec{x}})$ is the color receptive field.

λ_0 =520nm and σ_λ =55nm which corresponds to the Hering basis (human color model). At these values, the three components of the Gaussian color model very well approximate the CIE 1964 XYZ for colorimetry by the linear transform.

There are two important entities, the spectral differential quotient which represents a linear combination of given RGB sensitivities and differential quotient which is the gaussian derivation filter.

1.2 . Harris Laplace detector:

We explain in the following the principle of the Harris Laplace detector :

In fact, it is known that there are two classes of invariant features, local and global. The Harris Laplace detector gives local invariant features. These features are invariant to scale. For that, it extracts corners from the image.

The local invariant features can be traced back to the work of Moravec on sterio matching using corner detector. It was ameliorated by Harris by considering the following limitations :

1. Anisotropic response :

Only a discrete change of 45° was considered by Moravec. An analytic developpement is used instead to deal with small changes.

Denoting the image intensities by E, the change S produced by a shift (x,y) is given by :

$$S_{x,y} = \sum_{u,v} W_{u,v}[E_{x+u,y+v} - E_{u,v}]^2 = \sum_{u,v} W_{u,v}[xX + yY + o(x^2, y^2)]^2 \quad (6)$$

$$X = E \otimes (-1,0,1) = \frac{\partial E}{\partial x} \quad (7)$$

$$Y = E \otimes (-1,0,1)^T = \frac{\partial E}{\partial y} \quad (8)$$

With,

$$Y = E \otimes (-1,0,1)^T = \frac{\partial E}{\partial y} \quad (9)$$

2. The response is noisy, so a Gaussian window for smoothing is used :

$$W_{u,v} = \exp^{-\frac{u^2 + v^2}{2\sigma^2}} \quad (10)$$

3. The operator responds quickly to the variation of the edge, only the minimum of the energy is considered :

$$S(x, y) = (x, y)M(x, y)^T \quad (11)$$

Where :

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (12)$$

E is similar to an autocorrelation function and M is its form at origin.

I t was proven that α and β the eigen values of M are invariant to rotation.

And R is an important entity defined by :

$$R = Det - kTr^2 \quad (13)$$

So for edges, R<0

Besides, the scale is evaluated by representing the image as a pyramid such as each level corresponds to a scale. The scales are computed with the equation $S_n = \eta^n s_0$ □where $\eta = 1.4$. Also, the points are extracted by detecting the maximum in the 8 neighborhood of point x.

2 Our method :

2.1 Integration of Color in the Harris Detector :

We want to generalize Harris Laplace detector to use color images.
We have the second moment matrix for gray scale images given by :

$$M(x, y, \sigma_I, \sigma_D) = \sigma_D g(\sigma_I) * \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \quad (14)$$

We want to represent Lxx Lxy and Lyy, derivatives of color.
Where σ_D and σ_I are respectively the derivation and the integration scales.
La is the derivative of E in the a direction

Let's define the equations :

$$L_{\lambda xx}' = L_{xx}' + L_{\lambda xx}' + L_{\lambda\lambda xx}' \quad (15)$$
$$L_{\lambda xy}' = L_{xy}' + L_{\lambda xy}' + L_{\lambda\lambda xy}'$$
$$L_{\lambda yy}' = L_{yy}' + L_{\lambda yy}' + L_{\lambda\lambda yy}'$$

These functions are obtained by computing :

$$L_{xx}' = \int E(x, y, \lambda)G_{xx}(x, y)dxdy \quad (16)$$
$$L_{\lambda xx}' = \int E_\lambda(x, y, \lambda)G_{xx}(x, y)dxdy$$
$$L_{\lambda\lambda xx}' = \int E_{\lambda\lambda}(x, y, \lambda)G_{xx}(x, y)dxdy$$
$$L_{xy}' = \int E(x, y, \lambda)G_{xy}(x, y)dxdy$$
$$L_{\lambda xy}' = \int E_\lambda(x, y, \lambda)G_{xy}(x, y)dxdy$$
$$L_{\lambda\lambda xy}' = \int E_{\lambda\lambda}(x, y, \lambda)G_{xy}(x, y)dxdy$$
$$L_{yy}' = \int E(x, y, \lambda)G_{yy}(x, y)dxdy$$
$$L_{\lambda yy}' = \int E_\lambda(x, y, \lambda)G_{yy}(x, y)dxdy$$
$$L_{\lambda\lambda yy}' = \int E_{\lambda\lambda}(x, y, \lambda)G_{yy}(x, y)dxdy$$

The best linear transform from RGB to the Gaussian color model is given by:

$$G = MC^T \quad (17)$$

Where:

$$M_g = \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix} = \begin{pmatrix} \alpha & & \\ & \beta & \\ & & \gamma \end{pmatrix}$$

(18) [6]

Where α β γ are eigen values of Mg.

Let us note C=(r g b)   (26) the three color components

So, G= $\begin{pmatrix} \alpha r \\ \beta g \\ \gamma b \end{pmatrix}$

By using the equations (19), we obtain the convolutions :   Represents the

$$L'_{\lambda xx} = G_{xx} * (\alpha r + \beta g + \gamma b)$$

$$L'_{\lambda xy} = G_{xy} * (\alpha r + \beta g + \gamma b)$$

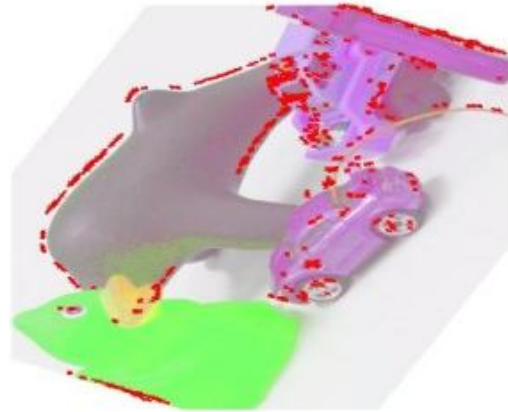$$L'_{\lambda yy} = G_{yy} * (\alpha r + \beta g + \gamma b)$$

(19)

Figure 1 : The second image was generated by changing the luminance and the contrast and rotating the first image.

Consequently, the second moment matrix for color image is given with :

$$M(x, y, \sigma_I, \sigma_D) = \sigma_D g(\sigma_I) * \begin{bmatrix} L'_{\lambda xx} L'_{\lambda xy} \\ L'_{\lambda xy} L'_{\lambda yy} \end{bmatrix}_{(20)}$$

At each scale, the image is processed to extract image orientations. The Rij is computed using the pixel differences.

$$R_{ij} = a \tan 2(A_{ij} - A_{i+1, j}, A_{i, j+1} - A_{ij})$$

(21)

2.2   Texture measurement :

Since Daugman [4,5] has generalized the Gabor function to model the receptive fields of the selective single cells, they have been widely used .

We use gabor wavelets in order to extract the significant frequencies in each region

Gabor wavelets :
Generally, Gabor filter can be viewed as a modulation of a Gaussian envelope and sinusoidal plane of a particular frequency and orientation [9].

We will use a class of Gabor wavelets instead of Gabor function to insure orientation and scale invariance.

In order to increase the efficiency of the Gabor filter, as suggested in [1] we define a class

of Gabor wavelets.

$$T_{pq}(x, y) = a^{-p} t(x', y'), a > 1, p, q \in Z \quad (22)$$

Where,

$$x' = a^{-p} (x \cos \theta + y \sin \theta), \quad (23)$$

$$y' = a^{-p} (-x \sin \theta + y \cos \theta), \quad (24)$$

Image transformations	Match %
Increase contrast by 2	95,5%
Descrease luminance by 2	97%
Rotate by 20 degrees	94,5%
Scale by 0.7	97,25%

Figure 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 applying KNN method

Where

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

By applying convolution of Gabor wavelets described before, we concatenate our descriptor with texture which is composed by 9 entries for

X	Y	σ	σ	σ	σ	r	r	r	r	t

space   scale   orientation   texture

Figure 3 : structure of our descriptor, 2 bin for space, 4 for scale, 4 for orientation, 9 for texture

each keypoint.

(a)

(b)

Figure 4 :Features extracted both with SIFT detector (a) and our detector (b)

The size of the descriptor is 19* the number of keypoints.

2.3 Recognition by indexation [2]

In order to evaluate how disdriminant and invariant are our interest points, they are used in order to learn object models from a set of images where these objects are presented to the system, on a uniform background, and 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})

Every point is associated with its attribute vector (with 19 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.

During the recognition step, points (or features) are extracted from the request image the system. The matching associates each feature to its nearest neighbor by using the closest euclidean distance in the attribute space. To ameliorate the performance, this distance is compared to the second closest from another object. Matches are retained if the distance ratio is lower than 0.8 which eliminates 90% and discards less than 5% of the correct matches. In the implementation, the kd tree is used to find the closest point of (O(logN)) in order to find a reliable matching, 3 features are sufficient. A hough transform entry is created for predicting the model location, orientation and scale from match hypothesis. It consists of finding all object poses that correspond to a feature. When clusters of features vote for the same object pose, the probability that the interpretation is correct is much higher. However this doesn't consider that a 3D object has 6 degrees of freedom pose space and maybe there are no rigid deformation. To take into account these details, a bread bin size of 30 degrees for orientation, a factor of 2 for scale and 0.25 times the maximum projected training image dimension(using prediction scale) for location are considered. Also the Lowe's method vote for the 2 closest bins in each dimension in order to avoid the problem of boundary effect in bin assignment. For improving the search of correct matches, we join a hash table to the hough transform voting where the key is the view number and the other elements are location,scale and image orientation.

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

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} tx \\ ty \end{bmatrix} \quad (25)$$

$$\text{Or} \begin{bmatrix} x & -y & 1 & 0 \\ y & x & 0 & 1 \\ & & & \end{bmatrix}.\begin{bmatrix} m \\ n \\ tx \\ ty \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \quad (26)$$
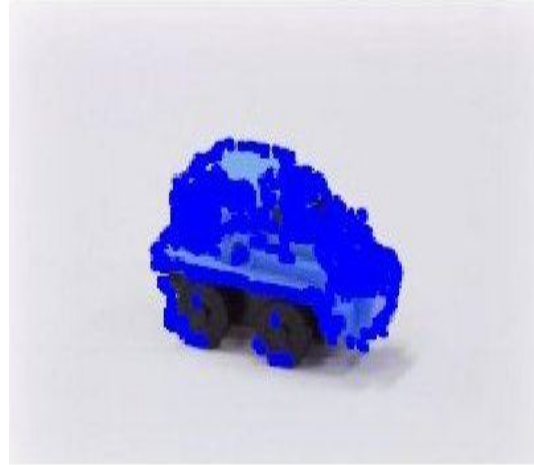
with $m = s\cos\theta$ and $n = s\sin\theta$

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 (26) using the pseudo-inverse

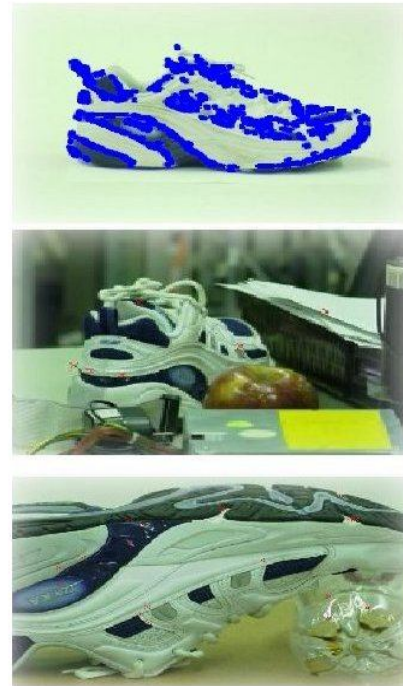solution:

$$x = [A^T A]^{-1} A^T b$$

Figure 5 : Image on the bear, with features. recognition results are shown with interest points learnt on image from the COIL database http://www-cvr.ai.uiuc.edu

$$e = \sqrt{\frac{2\|Ax - b\|^2}{r - 4}} \quad (27)$$

If it matches an exixting 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 exixting model view. In such case, the similarity transform solution is used to transform the new training image into the coordinates of the model view

We construct our model with the entries of position scale and   orientation . We don't add the texture features in the database because we shouldn't integrate them in the hough transform. We put 4 entries for scale, 2 for positionand 4 for orientation. Then we extract features from the test image and we apply the Lowe's method

We construct our model by extracting features from each view of the object. This allows to have a view invariant descriptor. We store the position, scale and orientation entries in a database. The descriptor of texture is only for more distinctiveness. We extract features from the test image. For matching, we apply first the Knn method to discard the features that aren't

in the test image, then we apply a hough transform to cluster reliable model hypotheses. A hash table is created to predict the model location, orientation and scale from the match hypotheses

The steps of the algorithm that implement the hough transform :

-Get the nearest neigbours from the database.
-Set up the orientation bins
-Set up the scale bins
-Loop over all of the models in the database.
-Set up the position bins
-Initialize the histogram to zero
-Histogram the translation into the four nearest bins
-Histogram the rotation into the 2 nearest bins
-Histogram the scaling into the 2 nearest bins
-Split the x and y indices of the tranlation bin
-Accumulate the histogram and record which points were placed in which bin
-Perform nonmax suppression on the histogram
-Return any candidate transformations

3. CONCLUSION :

In this article, we have suggested a new method for extracting local features from color textured images. For that we have improved the Harris Laplace detector by using the Koendering color model. We have proposed a combination with the hough transform for matching. In perspectives, we aim to apply our descriptor to view clustering.
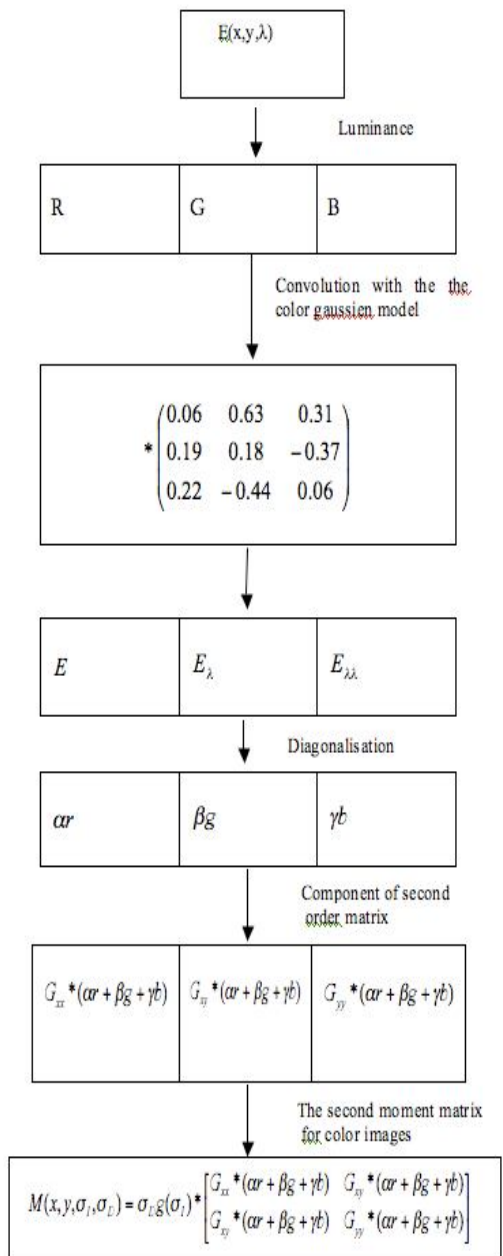
$E(x,y,\lambda)$

Luminance

| R | G | B |

Convolution with the the color gaussien model

$$* \begin{pmatrix} 0.06 & 0.63 & 0.31 \\ 0.19 & 0.18 & -0.37 \\ 0.22 & -0.44 & 0.06 \end{pmatrix}$$

| E | E_λ | $E_{\lambda\lambda}$ |

Diagonalisation

| αr | βg | γb |

Component of second order matrix

| $G_{xx}*(\alpha r+\beta g+\gamma b)$ | $G_{xy}*(\alpha r+\beta g+\gamma b)$ | $G_{yy}*(\alpha r+\beta g+\gamma b)$ |

The second moment matrix for color images

$$M(x,y,\sigma_I,\sigma_L)=\sigma_I^2 g(\sigma_I)*\begin{bmatrix} G_{xx}*(\alpha r+\beta g+\gamma b) & G_{xy}*(\alpha r+\beta g+\gamma b) \\ G_{xy}*(\alpha r+\beta g+\gamma b) & G_{yy}*(\alpha r+\beta g+\gamma b) \end{bmatrix}$$

Figure 6 : the integration of color in the Harris Laplace detector

REFERENCES

[1] Manjunath B.S. , Ma W.S. " Texture Features for Browsing and Retrieval of Image Data", IEEE Trans. on Patt. Anal. And Mach. Intell., 18, 837-842 (1996)

[2] David G. Lowe "Object Recognition from Local Scale Invariant Features"

[3] Gabor D "Theory of communication". Journal I.E.E.,

[4] Daugman J. G, 1980" Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile", Vision Research, vol. 20, pp. 847-856,(1980)

[5] Daugman, J.G. 1985"Uncertainty Relation for Resolution in space, Spatial Frequency, and OrientationOptimized by Two-Dimensional Visual Cortical Filters," J. Optical Soc. Amer. , vol.2, no.7,pp.1,160-1,169

[6]Geusebroek J M, Boomgaard R,Smeulders A W M, Geerts H, "Color Invariance"

[7] Mikolajzyk K , Schmid C "Scale & Affine Invariant Interest Point Detectors"

[8] Bay H Tuytelaars T, Gool L V "SURF: Speeded Up Robust Features"

[9] Lee T.S, " Image Representation Using 2D Gabor Wavelets", ", IEEE Trans. on Patt. Anal. And Mach. Intell., vol. 18, no. 10,(1996)

[10] Raoui Y, Bouyakhf E.H. and Devy M 2007 "Unified Color Texture Descriptor Using Gaussian Color Model and Rayleigh Density Function" SETIT Hammamet Tunisia