



# Spectral clustering and dimensionality reduction applied to Content based image retrieval with hybrid Descriptors

K Houari, M K Kholladi, Youssef Chahir

## ► To cite this version:

K Houari, M K Kholladi, Youssef Chahir. Spectral clustering and dimensionality reduction applied to Content based image retrieval with hybrid Descriptors. International Review on Computers and Software (IRECOS), 2010, International Review on Computers and Software, 5 (1), pp.14-21. hal-01883173

**HAL Id: hal-01883173**

**<https://hal.science/hal-01883173>**

Submitted on 27 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spectral clustering and dimensionality reduction applied to Content based image retrieval with hybrid Descriptors.

**K. Houari**

University Larbi Ben M'hidi,  
Oum El Bouaghi, Algérie  
hk\_houari@yahoo.fr

**M. K. Krolladi**

University Mentouri of Constantine, Algeria,  
Director of MISC Laboratory  
krolladi@yahoo.fr

**Y. Chahir**

University of Caen, France,  
Team of research on Image  
Processing, GREYC,  
URA CNRS 6072, campus II  
chahir@info.unicaen.fr

---

**Abstract** – The topic of research exposed in this paper concerns Content Based image retrieval in a heterogeneous high database. The increase of storage capacities and the evolution of compression techniques have generated an explosion of the digital information quantity. Their computerization opens a vast field of applications. In this setting we are interested more especially in the problem of the dimensionality reduction and spectral clustering of a heterogeneous database of images in order to image retrieval by the content. Our new gait described in this paper consists to:

- In first phase the description of the database images by a hybrid descriptor which are Interest SIFT points combined with texture descriptor given by the application of Wavelet transform. The descriptor is multi-dimensional, robust and invariant to changes and scales.
- In second phase the representation of the database images as a convex graph.
- In third phase the reduction of the space of representation by the application of an unsupervised spectral classification (The Spectral training uses information contained in the eigenvectors of the normalized matrix of transition to detect structures in data.) That will provide us classes of images that has shortcoming the Eigen-values calculated on the matrix of symmetry.

As last phase, we use the Nyström theory that will permit us, not to recalculate the all Eigen-values, but only the lasts one.

**Keywords:** image retrieval, Sift, clustering, texture, Nyström.

---

## I. Introduction

Algorithms of spectral classification for the unsupervised analysis of data offer a very effective tool for the exploration of the data structure.

Methods of Clustering [1, 2] have been used in various contexts and such disciplines 'dated mining', research of documents, image segmentation and classification of objects. The objective of clustering methods is the classification of objects on the basis of the criteria of similarity or chosen dissimilarity where groups (or classes) are a set of similar objects.

The Crucial aspect in the classification is the representation of models and the distance of similarity. Every model is usually represented by a set of descriptors of the studied system. It is important to note that a good choice of model representation can increase performances of the classification. The choice of the set of descriptors depends on the system. The system of representation being fixed, it is possible to choose the suitable similarity measure between models or objects. The most popular measure of dissimilarity for a metric representation is the distance, as Euclidian distance. [3].

Techniques of Clustering can be divided globally in two categories:

- Hierarchical,
- Partitioning.

Hierarchical Techniques of clustering [1,4,5] are capable to find the structure that is then divided in substructure and so forth recursively .the result is a hierarchical structure of groups that is called 'dendrogram'.

Methods of clustering 'Partitioning' try to get a unique partition of data without the other under-partitions as in hierarchical algorithm processes and are based always on the optimization of an appropriated objective function. The result is the creation of a hyper-surface separation between classes. The Standard methods of partitioning (ex., K-means, fuzzy c-means, SOM and neural gas) that use two centroides are unable to separate in the desired manner the two sets. The use of several centroides can solve the problem but by generating a complex description of a simple whole of data. For all these reasons modifications and a new approach were introduced.

Among the range of modifications we can mention the varieties of fuzzy c-means [6], but the major problem is that information on the forms of the clusters must be introduced as a preliminary. Recently, some methods of clustering which produce a nonlinear separation of hyper surfaces between the clusters were proposed. Theses algorithms can be divided into two great families:

Several methods of clustering were modified by incorporating the concept of kernels (ex: k-means, fuzzy c-means, SOM and neural gas). The use of the kernels makes it possible to implicitly represent the data in a space of great dimension called space of description; while carrying out a linear partitioning of this space of descriptors, the result of the partitioning of the space of input will be non-linear

Methods of Spectral clustering are based on concepts of the spectral theory of graphs. The basic idea is to construct a graph either weighted from data in entries. Each *vertices* or *nodes* represents a pattern (object) and every weighted *edges* or *lines* holds only in account the similarity between two patterns. In this case the problem of clustering amounts to a problem of graph portioning, what is treated by the spectral graphs theory. The heat of this theory is the calculation of the eigen values of the Laplace matrix of the weighted graph obtained from data. There is a narrow relation in fact between the second smaller value of Laplace and the partitioning of graph [7-8].

Lately, the interest for the reduction and the treatment of high dimension data turned toward the spectral clustering methods, because of numerous successes [34]. These methods use the spectral content of a similarity matrix (distance between each pair of data) to achieve the reduction and the partition of a data. More specifically, the eigenvectors are seen like a tool providing a representation of data in a space where it is well separated and can be grouped easily, and the eigen values

having the possibility to capture the conceptual features and perceptual of a set of data. This last mathematical property has not been taken practically ever in account for the assessment of performances of algorithms of the clustering spectral

In this article the first section describes the SIFT descriptor as well as the texture ones and in the second section we describe the spectral method of global diffusion applied for reducing the space representation. Conclusions and recommendations for the future works are presented in the last section.

## II. Extraction of SIFT points.

During these last years, the invariant descriptor to the Scale transformations "Invariant Features Transform (SIFT) feature" [16] received a lot of attention. It has been demonstrated in a recent survey [17] that he gives better results than other descriptors. Some minor modifications on the descriptor initial SIFT have been brought (PCA-SIFT [12], however Gloh-Sift [17], but the gain is not always obvious in all experimentations. Thus, in our present survey, we are interesting to the initial SIFT.

The SIFT method is used to extract the invariant interest points to transformations and scale from image in order to matching it with another image with different point of view of an object or a stage. The Descriptor is invariant to scale and rotation and he has been shown that he procures a robust matching through a certain degree of refine distortions, of changes of view in 3D, addition of noises, and of illumination changes. Descriptors are highly distinctive; in the sense that only one descriptor can be matched correctly with a big probability with a big database of several other image descriptors.

The algorithm consists in four main stages:

1. Scale-space extrema detection  
(Search all scales/locations for interest points);
2. Keypoint localization  
(Determine stable interest point location and scale)
3. Orientation assignment  
(Determine orientation of feature at key point)
4. Key point descriptor  
(Generate 128 dimension vector representing feature)

The first stage identifies the localizations and the scales which can be assigned under various points of sights of the same object.

The Detection of the localizations which are invariant with the scaling of the image can be achieved by the search of the points which are stable through all the scales, by using a continuous function of the space of scales. The only possible core of spaces of scales is the Gaussian function.

Of that, the space of the scales of images is defined like the function:  $L(x, y, \sigma)$ , which are produced from the convolution of the variable scale-Gaussian  $G(x, y, \sigma)$ , with the image of Input  $I(x, y)$ , i.e.:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

Where  $*$  is the operation of convolution and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The points of interests are detected by using the extreme in the space of scales in the function of Gaussian difference  $D$  convoluted with the image:  $I(x, y)$ :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

Where  $k$  is a constant factor multiplying which separates two close spaces. With an aim of detecting the local maxima and the minima of:  $D(x, y, \sigma)$ , each point is compared has its eight neighbors in the current image and has its new neighbors of the image of scale to the top and below. It is selected so only it is higher or lower than all its neighbors.

Once the point of interest candidate was found by comparing the pixel has its neighbors; the following stage consists to make a filtering detailed in the vicinity of the point to detect the localization, the scale, and the ratio of curve.

This information makes it possible to reject the points which have a weak contrast (and which is sensitive to the noise) or are located on edges. A quadratic function 3D is applied to the local points

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

Where  $D$  and its derivatives are evaluated at the simple point  $X = (x, y, \sigma)^T$ . The localization of the extrema is obtained by considering the derivatives compared to  $X$ , which gives us:

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}$$

Who is a linear system 3x3, easy to solve.

The value of the function at the extrema is:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X}$$

Who is very useful for the unstable rejection of the extreme with a weak contrast?

At this stage, the algorithm rejects also the interest points which have slightly defined nodes, i.e.: the points which have in the function of Gaussian Difference a broad curve around an edge with low curve in the perpendicular direction.

By assigning a consistent orientation, based on the properties of the local image, the descriptor of the interest point can be represented relative with this orientation by ensuring however invariance to the rotation of the image.

For each simple image,  $L(x, y)$  on this scale, the magnitude of the gradient,  $m(x, y)$  and the orientation,  $\theta(x, y)$  are pre-calculated by using the differences of pixel:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= \text{atan2}\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \end{aligned}$$

The histogram of orientation is formed starting from the orientations of gradients of the simple points inside the area around the key point (interest).

Finally the description of the point of interest SIFT is made by calculation magnitude of the gradient and the orientation of each simple point of image in an area around the localization of the point of interest. These points are balanced by a Gaussian window and then accumulated in a histogram of orientation of 8 elements (directions) which synthesizes contained 4x4 under areas, with like length of each column, the sum magnitudes of gradient close to this orientation inside the area. The size of the final descriptor is of 4x4x8=128 and the procedure of standardization is applied to the descriptor to allow Invariance with the changes of illumination

### III. Texture description

Each image is treated to extract the descriptors of texture by employing the most popular methods of extraction of texture descriptors, which are 1) spectral filtering of Gabor [ 3 ], 2) laws of energy filtering [ 4.5 ], and the 3) 'Wavelet transform'. These methods were largely used by researchers and are carried out very well for various tasks of segmentation and classification of image. The 'Wavelet transform' used in this study offer space information, of frequency and very rich scale for the descriptors of texture and is a robust and effective method for the image processing.

### IV. Dimension Reduction Techniques.

We present in this section a panorama on the traditional methods for dimension reduction, and synthetically the global methods based on the spectral decomposition. We

conclude by an evaluation of the contributions of these approaches within the meaning of the metric properties. Reducing a dimension means to find a configuration  $\Phi$ , which draws up the maps of data of the starting space  $X$  with their original description such as:

$$X = \{x_1, x_2, \dots, x_m\}, \text{ Where } x_i \in \mathbb{R}^n.$$

To a new space  $Y$  where their description is considered simplest.

$$Y = \{y_1, y_2, \dots, y_m\} \text{ with } y_i \in \mathbb{R}^d; \text{ and } d \ll n.$$

Thus, the configuration  $\Phi$  will be mentioned like a fixing of the space of high dimension  $X$  in another lower space of dimension and more expressive  $Y$ :

$$\Phi: X \rightarrow Y$$

Among these configurations, we distinguish the linear methods and the nonlinear methods.

#### IV.1. The Linear methods.

For the linear methods, we find the algorithm of analysis in principal components (ACP), consisting in projecting the samples on the axes of maximum variance of the data. In other words, the ACP aims at identifying the linear dependences subjacent with a stochastic observation (random) in order to obtain a description of lower dimension. The use of this algorithm is legitimate if the data observed are roughly located on a linear variety.

In the same method, one finds the algorithm of multidimensional positioning (MDS) which carries out a linear reduction of dimension indirectly, starting from square distances. Clearly, multidimensional positioning (MDS) makes it possible to build a configuration of  $m$  not in  $\mathbb{R}^n$  starting from the distances between  $m$  objects with which these points are associated. In this context, one does not observe any more the points directly in space of origin but rather the  $m(m-1)/2$  distances corresponding to these points. The distances provided to MDS should not necessarily be Euclidean and can consequently be to represent by the other metric ones.

Unfortunately, conflicts of these methods resident in the fundamental character of spaces of data observed in nature which very often exhibit highly nonlinear characteristics. If it is supposed that the data have an intrinsic dimension 'd' weaker than the dimension of observations ( $n$ ) and if the data are rather drawn from a nonlinear variety, the ACP will be unable to express the characteristics of this variety and consequently will not be able to create a representation in lower dimension respecting the properties of the variety.

In the same order of idea, if the distances provided to MDS are the Euclidean distances, this algorithm will not

be able to either generate a positioning of the adequate points to the nature of the variety.

#### IV.2. The Nonlinear methods.

Deficiencies of the linear methods involved the development of the configurations being able to approximate the data by nonlinear varieties. The kernels provide nonlinear measurements of similarity easily being able to be incorporated in already existing linear algorithms. Due to this, spectral methods of estimation of nonlinear varieties were introduced.

These algorithms are based, for a set of vectors of entry  $x_i; i = 1, \dots, m; x_i \in \mathbb{R}^n$  on a matrix of similarity  $\mathbb{K}^{m \times m}$ , and drive to calculate its principal's eigen vectors and eigen values. The representation in small dimension of every vector  $x_i$  in entry is gotten while using the  $j - s$  first eigenvectors of  $\mathbb{K}$ . (Isomap, LLE 'Locally Linear Embedding Algorithm').

The ISOMAP method builds a graph from the  $K$  nearest neighbors and then performs Multi-Dimensional Scaling MDS (Michael Lee).

### V. Spectral Clustering.

#### V.1. Notations.

The notation for spectral graph theory foots on graph theory and appeals to intuition.

Except for minor differences in the names of matrices, variables, and counters, this notation seems pretty standard.

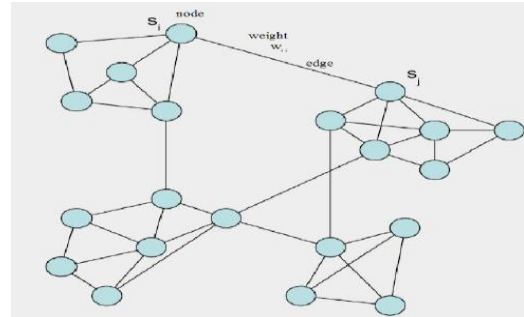


Fig.1: Example of graph

In This section we present briefly the mathematical notations of the algorithms that will be presented.

A graph or undirected graph  $G$  can be written as an ordered pair  $G := (V, E)$ , where  $V$  is a set of *vertices* or *nodes*, and  $E$  is a set of pairs (unordered) of distinct vertices, called *edges* or *lines*. The vertices belonging to an edge are called the *ends*, *endpoints*, or *end vertices* of the edge. [21]

The *adjacency matrix* (also *similarity matrix* or *weight matrix*) of a finite directed or undirected graph  $G$  on  $n$  vertices is the  $n \times n$  matrix where the no diagonal entry  $w_{ij}$  is the number of edges from vertex  $i$  to vertex  $j$ , and the diagonal entry  $w_{ii}$  is either twice the number of loops at vertex  $i$  or just the number of loops (usages differ, depending on the mathematical needs; this report

is not concerned with reflexive connections). The adjacency matrix is symmetric for undirected graphs.[18]

In the following, we assume  $w_{ij} = w_{ji} \geq 0$ .

Given  $A \in V$ , its complement,  $V \setminus A$  will be denoted as  $\bar{A}$ .  $i \in A$  shall be shorthand for indices of  $\{i | v_i \in A\}$ .  $|A|$  will denote the number of vertices in  $A$ .

$vol(A)$  measures the size of  $A$  by the weights of its edges, i.e.  $vol(A) = \sum_{i \in A} \deg(v_i)$

The *degree* (or *valency*) of a vertex is the number of edge endpoints to the vertex. Loops are counted twice. The *degree matrix*  $D$  for  $G$  is a  $n \times n$  matrix defined as [21].

$$d_{i,j} := \begin{cases} \deg(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Intuitively, a subset  $A \in V$  is *connected*, if paths between any two points in  $A$  need only points in  $A$ .  $A$  is called *connected component* with respect to  $\bar{A}$ , if  $A$  is connected, and there are no edges between vertices in  $A$  and  $\bar{A}$ . Subsets  $A_1, A_2, \dots, A_k$  represent a *partitioning* of  $V$ , if, for all  $1 \leq j, x \leq k$ ,  $A_i \cap A_j = \emptyset$ . And  $\bigcup_{i=1}^k A_i = V$  (Usually also defined nonempty).

## V.2. Clustering.

Generally speaking, clustering means partitioning of a graph, so that the edges between different groups have low similarity (low distance) and the edges within groups have high weight (low distance). The first requirement for the partitioning can be stated as the *minicut criterion*, which has to be minimized:

$$\begin{aligned} cut(A_1, \dots, A_k) &= \sum_{i=1}^k cut(A_i, \bar{A}_i), \text{ where } cut(A_i, \bar{A}_i) \\ &= \sum_{i \in A_i, j \in \bar{A}_i} w_{ij} \end{aligned}$$

Or following the *minicut* requirement, often only a little group of points is isolated.

For this reason, tweaks have been introduced:

- $RadioCut(A_1, \dots, A_k) = \sum_{i=1}^k \frac{cut(A_i, \bar{A}_i)}{|A_i|}$
- $NCut(A_1, \dots, A_k) = \sum_{i=1}^k \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}$

As for the second requirement, within-cluster similarity means optimizing

$$\sum_{i=1}^k \sum_{i,j \in A_i} w_{ij}$$

Within-cluster similarity is maximized if  $cut(A, \bar{A})$  is small and  $vol(A)$  is big.

Therefore, *Ncut* implements the second criterion. *Ncut* can be interpreted as cutting through edges rarely transitions by a random walk. *RadioCut*, by maximizing  $|A_i|$ , as within-cluster similarity is not related to the number of vertices in  $A$ , does not implement this requirement.[5]

## V.3. Laplacians and their properties

The spectral algorithms presented here foot on eigenvectors of Laplacians, which are a combination of the weight and the degree matrix. For a more thorough and broader discussion of mathematical properties of Laplacians refer to [10] and [1].

The (unnormalized) graph *Laplacian* is defined as  $L = D - W$ .

For a graph  $G$  and its admittance matrix  $L$  with eigen values  $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ , these properties are important, in the context of spectral clustering. [20].

- $L$  is always positive-semi definite ( $\forall i \lambda_i \geq 0$ ).
- The multiplicity of '0' as an eigen values of  $L$  is the number of connected components of  $G$ .
- $\lambda_1$  is called the algebraic connectivity.
- The smallest non-trivial eigen values of  $L$  is called the spectral gap.

Noteworthy is furthermore, that matrices with identical non-diagonal elements have the same unnormalized Laplacians.

The *normalized graph Laplacian* is defined in two distinct ways:

- $L_{sym} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$  (symmetric matrix).
- $L_{rw} := D^{-1} L = I - D^{-1} W$  (random walk).

Some properties are:

- $L_{rw}$  and  $L_{sym}$  are positive semi-definite and have  $n$  non-negative real-valued eigen values  $0 = \lambda_1 \leq \dots \leq \lambda_n$
- The multiplicity  $k$  of the eigen values 0 of both  $L_{rw}$  and  $L_{sym}$  equals the number of connected components  $A_1, \dots, A_k$  in the graph.

Refer to [1] for an in-depth discussion of normalized Laplacians.

## V.4. Clustering algorithms

Spectral clustering became popular with, among others, [15] and [13]. Its efficiency is mainly based on the fact that it does not make any assumptions on the form of the clusters. This property comes from the mapping of the original space to an Eigen space. Algorithms differ basically in the number of eigenvectors they use for partitioning.

Algorithms can be categorized based on the number of eigenvectors they use: [16]

- One eigenvector – recursively uses a single eigenvector on partitions (**recursive**).
- Many eigenvectors – Directly computes a **multi-way partition** of the data.
- **Non spectral** – Grouping algorithms that can be used in conjunction with multi-way spectral algorithms.

### V.5. Spectral clustering algorithms

For the algorithms we assume data points  $x_1, \dots, x_n$  and their similarities

$s_{ij} = s(x_i, x_j)$ , where  $1 \leq i, j \leq n$  and  $s \in S$ .

$S$  is symmetric and non-negative. They are rather similar, except for the different Laplacians they use. The change of representation from  $x$  to  $u$  with the help of the Laplacians enhances the cluster properties of the data. [5].

Algorithms introduced here are *unnormalized spectral clustering*, ([15], [4], [8]).

Unnormalized spectral clustering proceeds as in the following:

Inputs:  $k$ , the number of desired clusters,  $X$ , and  $S$ .

- Following the steps in section “similarity graph” construct  $W$  (graph matrix),
- then calculate ‘admittance matrix)  $L$ ,
- then the first  $k$  eigenvectors  $v_1, \dots, v_k$  of  $L$ .
- Let  $V \in \mathbb{R}^{n \times k}$  contain  $v_1, \dots, v_k$  as columns and  $y_i \in \mathbb{R}$ , with  $i = 1, \dots, n$  correspond to the  $i$ -th row of  $V$ . Cluster the points  $y_i$  with the  $k$ -means algorithm into clusters  $C_1, \dots, C_k$ .

Output: Clusters  $A_1, \dots, A_k$  with  $A_i = \{j | y_j \in C_i\}$ .

Note that similar to *RadioCut*, unnormalized spectral clustering does not optimize within-cluster similarity. [6]

#### Algorithm of Unnormalized Spectral Clustering:

**Input:** Similarity matrix  $S \in \mathbb{R}^{n \times n}$ , number  $k$  of clusters to construct

- Construct a similarity graph by one the way described in section 5.
- Let  $W$  be its weighted adjacency matrix.
- Compute the unnormalized Laplacian  $L$ .
- Compute the first  $k$  eigenvectors  $v_1, \dots, v_k$  of  $L$ .
- Let  $V \in \mathbb{R}^{n \times k}$  be the matrix containing the vectors  $v_1, \dots, v_k$  as columns.
- For  $i=1, \dots, n$ , let  $y_i \in \mathbb{R}^k$  be the vector corresponding to the  $i$ -th row of  $V$ .
- Cluster the points  $(y_i)_{i=1, \dots, n}$  in  $\mathbb{R}^k$  with the  $k$ -means algorithm into clusters  $C_1, \dots, C_k$ .

**Output:** clusters  $A_1, \dots, A_k$  with  $A_i = \{j | y_j \in C_i\}$ .

#### Algorithm of Normalized Spectral Clustering:

**Input:** Similarity matrix  $S \in \mathbb{R}^{n \times n}$ , number  $k$  of clusters to construct

- Construct a similarity graph by one the way described in section 5.
- Let  $W$  be its weighted adjacency matrix.
- Compute the unnormalized Laplacian  $L$ .
- Compute the first  $k$  eigenvectors  $v_1, \dots, v_k$  of the generalized eigenproblem  $L_v = \lambda D_v$ .
- Let  $V \in \mathbb{R}^{n \times k}$  be the matrix containing the vectors  $v_1, \dots, v_k$  as columns.
- For  $i=1, \dots, n$ , let  $y_i \in \mathbb{R}^k$  be the vector corresponding to the  $i$ -th row of  $V$ .
- Cluster the points  $(y_i)_{i=1, \dots, n}$  in  $\mathbb{R}^k$  with the  $k$ -means algorithm into clusters  $C_1, \dots, C_k$ .

**Output:** clusters  $A_1, \dots, A_k$  with  $A_i = \{j | y_j \in C_i\}$ .

#### Algorithm of Normalized Spectral Clustering:

**Input:** Similarity matrix  $S \in \mathbb{R}^{n \times n}$ , number  $k$  of clusters to construct.

- Construct a similarity graph by one the way described in section 5.
- Let  $W$  be its weighted adjacency matrix.
- Compute the unnormalized Laplacian  $L_{sym}$ .
- Compute the first  $k$  eigenvectors  $v_1, \dots, v_k$  of  $L_{sym}$ .
- Let  $V \in \mathbb{R}^{n \times k}$  be the matrix containing the vectors  $v_1, \dots, v_k$  as columns.
- Form the matrix  $U \in \mathbb{R}^{k \times k}$  from  $V$  normalizing the row sums to have norm 1, that is

$$u_{ij} = v_{ij} / \left( \sum_k v_{ik}^2 \right)^{1/2}$$

- For  $i=1, \dots, n$ , let  $y_i \in \mathbb{R}^k$  be the vector corresponding to the  $i$ -th row of  $V$ .
- Cluster the points  $(y_i)_{i=1, \dots, n}$  in  $\mathbb{R}^k$  with the  $k$ -means algorithm into clusters  $C_1, \dots, C_k$ .

**Output:** clusters  $A_1, \dots, A_k$  with  $A_i = \{j | y_j \in C_i\}$ .

## VI. SIMILARITY GRAPHS.

Given a set of points,  $x_1 \dots x_n$ , and their distances  $d_{jk} \in D$  not related to the degree), there are several constructions to obtain a graph (i.e. construct  $W$ ), which are regularly used in spectral clustering.

The  $\varepsilon$ -neighborhood graph is the simplest possibility. All connections with distances below a threshold are set to '1'.

i.e.

$$\sum_{\langle ij \rangle} w_{ij} = 1, \text{ with } \{ \langle ij \rangle \mid d_{ij} \in D, d_{ij} < \varepsilon \}.$$

$k$ -nearest neighbor graphs lead to a directed graph. The  $k$ -shortest distances from  $i$  are connected. In order to make the graph undirected, directions can be ignored, e.g. constructing a  $k$ -nearest neighbor graph, then doing either:  $A \cup A^T$ .

(Usually this is called  $k$ -nearest neighbor graph) or  $A \cap A^T$  (mutual  $k$ -nearest neighbor graph).

A fully connected graph results from connecting all points with positive similarity with each other.

## VII. CONCLUSION

The spectral algorithms are simple methods simple and efficient for the classification.

The preprocessing is an important stage and can help meaningfully to raise results of classification.

The choice of a likeness graph can be non trivial and can require preprocessing more extended. However, there is a graph of likeness once, the problem is linear and the spectral methods don't endure the local optimum problem intrinsically. [16], [15].

Besides, the objective of this paper is the automatic classification of the input image in a database image, so the determination of the number of  $k$  classes, and selections / weight of the different measurements of the weight matrix that permits to refine to the maximum the classification on the robust and invariant descriptors to changes of scales and noises, of this survey, we disqualify *RadioCuts* because they don't optimize intra - classes likeness, and according to the study of the state of the art as well as advantages and inconveniences of several other algorithms we conclude that algorithms of spectral clustering give some superior results has those of the no - spectral methods, particularly their properties of non restriction to the convex regions of likeness and the hardness to noises make them more attractive.

## REFERENCES

- [1]. Chung, F. R. K., Spectral graph theory. Providence, RI: American Mathematical Society. 01, 1997.
- [2]. Dy, J. G., Brodley, C. E., Feature selection for unsupervised learning. J. Mach. Learn. Res., 5, pp. 845–889, 2004.
- [3]. Hagen, L., Kahng, A., New spectral methods for radio cut partitioning and clustering. IEEE Transactions on Computer-Aided Design, II(9), pp. 1074–1085, 1992.
- [4]. Kannan, R., Vempala, S., Vetta, A., On clusterings: Good, bad and spectral. J. ACM, 51(3), pp. 497–515, 2004.
- [5]. Luxburg, U. von. A., tutorial on spectral clustering. In (PI- Technical Reports No. 149. Tübingen: Max Planck Institute for Biological Cybernetics, august 2006.
- [6]. MacKay, D. J. C., Information theory, inference, and learning algorithms, <http://www.cambridge.org/0521642981>: Cambridge University Press. (Available from <http://www.inference.phy.cam.ac.uk/mackay/itila/>), 2003.
- [7]. Meila, M. Comparing clusterings. In (Technical Report 418). Washington University of Washington, Statistics department, 2002.
- [8]. Meila, M., Shi, J., Learning segmentation by random walks. NIPS, pp. 873–879, 2000.
- [9]. Meila, M., Shi, J. A., random walks view of spectral segmentation. AISTATS 2001.
- [10]. Mohar, B., Some applications of Laplace eigenvalues of graphs, [NATO ASI Series C]. G. Hahn and G. Sabidussi, editors, Graph Symmetry: Algebraic Methods and Applications, 497, pp. 227–275, 1997.
- [11]. Moore, A., The anchors hierarchy: Using the triangle inequality to survive high dimensional data, (Tech. Rep. No. CMU-RI-TR-00-05). Pittsburgh, PA: Robotics Institute, Carnegie Mellon University, February 2000.
- [12]. Newman, M., Watts, D., Strogatz, S., Random graph models of social networks, Proc. Natl. Acad. Sci. USA, 99, pp. 2566–2572, 2002.
- [13]. Ng, A. Y., Jordan, M., Weiss, Y., On spectral clustering: analysis and an algorithm, Advances in Neural Information Processing Systems, 14, 2002.
- [14]. Shamir, R., Sharan, R., Algorithmic approaches to clustering gene expression data. Current Topics in Computational Molecular Biology, T.Jiang, T. Smith, Y. Xu, M. Q. Zhang (editors), pp. 269–300, 2002.
- [15]. Shi, J., Malik, J., Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2000.
- [16]. Verma, D., Meila, M. A., comparison of spectral clustering algorithms, University of Washington, uw-cse-03-05-01, 2003.
- [17]. Ward, J. H., Hierarchical grouping to optimize an objective function, Journal of American Statistical Association, 58(301), pp. 236–244, 1963.
- [18]. Wikipedia. Adjacency matrix — wikipedia, the free encyclopedia, 2006a.
- [19]. Wikipedia. K-means algorithm — wikipedia, the free encyclopedia, 2006b.
- [20]. Wikipedia. (2006c). Laplacian matrix — wikipedia, the free encyclopedia.
- [21]. Wikipedia. Degree matrix — wikipedia, the free encyclopedia, 2007a.
- [22]. Wikipedia. Graph (mathematics) — wikipedia, the free encyclopedia, 2007b.
- [23]. Cottier, J., Extraction et appariements robustes des points d'intérêts de deux images non étalonnés, A representation for visual information. Phd thesis, Carnegie Mellon University, USA, 1994.
- [24]. Crowley, J. and Parker, A. A representation for shape based on peaks and ridges in the difference of low pass



- transform. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(2), pp. 156–170, 1984.
- [25]. Förstner, W. and Gülch, E. A., fast operator for detection and precise location of distinct points, corners and centres of circular features, In Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305, 1987.
  - [26]. Förstner, W. A., framework for low level feature extraction, In Proceedings of the 3<sup>rd</sup> European Conference on Computer Vision, Stockholm, Sweden, pp. 383–394, 1994.
  - [27]. Freeman, W. and Adelson, E., The design and use of steerable filters, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(9), pp. 891–906, 1991.
  - [28]. Goodrum, A., Image Information Retrieval, An Overview of Current Research. In Informing Science, Volume 3, No 2, 2000.
  - [29]. Harris, C. and Stephens, A combined corner and edge detector, In Alvey Vision Conference, pp. 147–151, 1988.
  - [30]. Heitger, F., Rosenthaler, L., Von der Heydt, R., Peterhans, E., and Kuebler, O., Simulation of neural contour mechanism: From simple to end-stopped cells, Vision Research, 32(5), pp. 963–981, 1992.
  - [31]. Lowe, D.G., Distinctive image features from scale invariant keypoints., International Journal of Computer Vision 20, pp. 91-100 (2004).
  - [32]. Mikolajczyk, K., Schmid, C. A., performance evaluation of local descriptor IEEE trans. Pattern Anal. Mach. Intell 27(10) pp. 1615-1630, 2005.
  - [33]. Ke, Y., Sukthankar, R., Pca-Sift: A more distinctive representation for local image descriptors. In : CVPR (2), pp. 506-513, 2004.
  - [34]. S. Lafon. Diffusion maps and geometric harmonics PhD thesis, Yale University, Dept of Mathematics & Applied Mathematics, 2004.



Mr. Houari Kamel is an assistant professor in University of Oum El Bouaghi, Algeria.



Dr Youssef Chahir is Master of conferences (Associate professor) in University of Caen, France. He is member of Research Group in Data processing, Picture, Instrumentation of Caen (GREYC - URA CNRS 6072).



Dr. Mohamed-Khireddine KHOLLADI is a Doctor of state in Data processing (graduate Doctor Of the INSA of Lyon). He is currently a head of Computer science department of the Faculty of Science of the engineer of the Mentouri university of Constantine and president of association sciences and development of the university. He is a member of the vision team & computer graphics of the research laboratory LIRE of the computer sciences department. He is member of the network Cassini in France and member of network ACIT International.