

# Multiresolution Laplacian Sparse Coding Technique for Image Classification

Intidhar Jemel, Ridha Ejbali and Mourad Zaied

REsearch Team in Intelligent Machines,

University of Gabes, ENIG, Street Omar Ibn EL Khattab, Zrig Gabes, 6029, Gabes, Tunisia

## Article history

Received: 04-04-2018

Revised: 08-05-2018

Accepted: 22-05-2018

Corresponding Author:

Intidar Jemel

REsearch Team in Intelligent  
Machines, University of Gabes,  
ENIG, Street Omar Ibn EL  
Khattab, Zrig Gabes, 6029,  
Gabes, Tunisia

Email: intidhar66jemel@gmail.com

**Abstract:** Sparse coding is a set of techniques used for learning a collection of over-complete bases to represent data efficiently. This technique has been used in different domain such as feature quantization and image classification. Despite its capacity of modeling, it could not represent similarity of the image coding which cause a poor performance in locality. The cause of this limitation is the features separation of the representation. To surmount the limitations of these techniques, we propose a new approach that is able to calculate similarity by taking into account the image's spatial neighborhood of pixels. This approach is based on the integration of Kullback-Leibler distance and wavelet decomposition in the domain of image. The association of the Kullback-leibler distance and wavelet decomposition is robust to small deformations (translation, dilation and rotation). It improves the representation of locality by considering each element of an image and its neighbors in similarity calculation. Results show clear improvements in performance compared to the above techniques.

**Keywords:** Sparse Coding, Laplacian Sparse Coding, Representation of Similarity of the Image Coding, Image's Spatial Neighborhood of Pixels, Kullback-Leibler Distance, Wavelet Decomposition

## Introduction

Sparse coding techniques marked a great revolution in the field of computer vision and its applications. Those techniques suffer from the inability to model the locality and the similarity among the instances to be encoded owing to the over complete codebook and the independent coding process. To overcome these limitations, Gao proposed an approach called Laplacian Sparse Coding (Gao *et al.*, 2010a). This approach exploits the dependence among local features. He recommended the use of histogram intersection based on K Nearest Neighbor (KNN) method to build a Laplacian matrix. This technique characterizes the similarity of local features. To maintain the consistence in sparse representation of those features, a matrix was incorporated into the function of sparse coding. In a second work, Gao enhanced the Kernel Sparse Representation technique (Gao *et al.*, 2010b).

The third approach proposed by Gao is Hypergraph Laplacian Sparse Coding techniques (Gao *et al.*, 2013). Within the same hyperedge, it simultaneously extracts the similarity between the instances and also composes their sparse codes.

In this paper, we suggest an enhancing of the Laplacian sparse coding technique by altering the way of similarity

calculation. In our case, the computation of similarity in the image domain is based on the divergence of Kullback-Leibler and wavelet decomposition. This idea comes from its ability to take into account neighbors' similarity.

This paper is organized as follows: In section1, we introduce the Laplacian sparse coding technique. Section 2 describes the kernel sparse representation. We explain our approach in section 3. Our approach is evaluated in the last section.

## Laplacian Sparse Coding

In order to alleviate the problem of hard quantization, researchers have proposed a new technique called sparse coding. This settles the problem by offering a sparse linear combination of basis vectors for each image feature. Sparse coding seeks a linear reform of a signal  $x, (x \in \mathbb{R}^d)$  using the bases in the codebook  $U = (u_1, u_2, \dots, u_k), U \in \mathbb{R}^{d \times k}$ . Sparse codes are defined by  $V = (v_1, v_2, \dots, v_n), v_i \in \mathbb{R}^{k \times 1}$  and  $v_{ik}$  are the weight of the vector  $x_i$  in the basis vector  $u_k$ . The optimization difficulty of sparse coding can be summarized as follows:

$$\min_{U, V} \|x - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 \quad \text{subject to} \quad \|u_j\| \leq 1; \forall j = 1, \dots, K \quad \lambda$$

is the tradeoff parameter used to stabilize the sparsity and the rebuilding error because of the independent encoding feature resulting from an over complete or sufficient codebook. Supposing that  $x = (x_1, x_2, \dots, x_n)$  is the vector of features and  $W$  is the matrix of similarity having  $W_{ij}$  the measuring of similarity of the pair  $(x_i, x_j)$ .  $D$  the matrix of degree defined by  $D_{ii} = \sum_{j=1}^n W_{ij}$  is a diagonal matrix.

The Laplacian Sparse Coding, as detailed in (Gao *et al.*, 2010a; 2010b) considers the similarity between features. The term of this approach is described by:

$$\min_{v_1, \dots, v_n} \sum_i \|x_i - Uv_i\|_F^2 + \lambda \sum_i \|v_i\|_1 + \frac{\beta}{2} \sum_{ij} \|v_i - v_j\|_{ij}^2 W_{ij} \quad (1)$$

It can be written as:

$$\min_V \sum_i \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \beta \text{tr}(VLV^T) \quad (2)$$

The definition of the Laplacian is  $L = D - W$  (Luxburg, 2007).

Since the codebook  $U$  is not optimal, the expression can be rewritten as follows:

$$\min_{U, V} \sum_i \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 + \beta \text{tr}(VLV^T) \quad (3)$$

### Kernel Sparse Representation

Gao proposed another approach called Kernel Sparse Representation to ameliorate the technique of features representation using sparse coding. He used the kernel trick because he noticed that this technique can identify the nonlinear similarity of features. This approach is fundamentally the sparse coding technique in an elevated dimensional feature space traced by tacit mapping function (Gao *et al.*, 2010a; 2010b).

$\Phi: IR^d \rightarrow IR^k, (d < k)$  is a feature mapping function. In the same condition of sparse coding, this function is defined by: the same condition of sparse coding, this function is defined by:  $\Phi: IR^d \rightarrow IR^k, (d < k)$  with  $x \rightarrow \Phi(x), U = (u_1, u_2, \dots, u_k) \rightarrow U_\Phi, U_\Phi = (\Phi(u_1), \Phi(u_2), \dots, \Phi(u_k))$ .

Using this expression, the formulation of Kernel Sparse Coding is written as follow:

$$\min_{U, v} \|\Phi(x) - U_\Phi v\| + U_\Phi \|v\|_1 \quad (4)$$

Gao adopted the Gaussian kernel because of its satisfactory performance in many works (Chen *et al.*, 2010; Donoho, 2006).

## Proposed Approach

### A. General Context of Multiresolution Wavelet Decomposition

As we all know, multiresolution formalism enables the decomposition of a signal over several scales. The signal is constructed using the most suitable approximation at each level. Also, the wavelet decomposition gives details and approximations threshold coefficients (Hassairi *et al.*, 2015; 2016; 2018). Multiresolution wavelet decomposition studies each signal in frequency and time domains. For lower frequency, it gives better frequency resolution and poorer time resolution. While for higher frequency, it offers better time resolution and poorer frequency resolution (Ejbali *et al.*, 2017). Fortunately, this condition is suitable for real applications; as signals have low frequency for longer time and high frequency for very short intervals.

A set of sub spaces  $L^2(IR)$  noted  $(V_j)_{n \in \mathbb{Z}}$  have the following properties:

$$V_j = \left\{ \sum_{n \in \mathbb{Z}} a_n^j \Phi_{j,n} : a_n^j \in IR \right\}, V_{j+1} \subset V_j, \quad (5)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \bigcup_{j \in \mathbb{Z}} \bar{V}_j = L^2(IR)$$

These sub spaces are the formulation of a multiresolution analysis.

Expression (5) confirms that  $(V_j)_{n \in \mathbb{Z}}$  is a space engendered by the family  $(\Phi_{j,n})_{n \in \mathbb{Z}}$ . Its description depends on the selected topology for the efficient space. We can describe it more firmly as the union of the restricted space of linear groupings of functions  $\Phi_{j,n}$ . Thus the estimation of a signal  $f$  on the space  $V_j$  is:

$$A_j = \sum_n a_n^j \Phi_{j,n} \quad (6)$$

Coefficients  $a_n^j$  are computed by performing a scalar product signal with the family features  $\Phi_{j,n}$ :

$$a_n^j = \langle f, \Phi_{j,n} \rangle \quad (7)$$

Instinctively, we can see that all the functions of  $V_j$  are richer or denser than those of  $V_{j+1}$ , which does not mean the inclusion relationship. The same expression (5) requires that the wavelets appear as a natural way to write the difference between two consecutive spaces  $V_j$  and  $V_{j+1}$ . We construct  $W_{j+1}$  to complete this space:

$$V_{j+1} \subset V_j \Rightarrow V_j = V_{j+1} \oplus W_{j+1} \quad (8)$$

The space  $W_{j+1}$  is yielded by a function  $\psi_{j,n}$ :

$$W_j = \left\{ \sum_{n \in \mathbb{Z}} d_n^j \psi_{j,n} : d_n^j \in IR \right\} \quad (9)$$

$\psi_{j,n}$  have values on the space  $W_j$  that is opposite to  $V_j$  in  $V_{j+1}$ . We have the same translational properties, expansion on  $\psi_{j,n}$  as on  $\Phi_{j,n}$ . The set of functions  $\psi_{j,n}$  is named space details. Thus, the detail of the signal  $f$  in the space  $W_j$  is computed as follows:

$$D_j = \sum_n d_n^j \psi_{j,n} \quad (10)$$

and the coefficients of details  $d_n^j$  are calculated by the expression (11):

$$d_n^j = \langle f, \psi_{j,n} \rangle \quad (11)$$

The signal of which is supposed to be exposed on a basis of  $V_j$ . Applying the wavelet transforms to the  $k \in IN$  scale expresses the signal from a personalized base to the direct sum:

$$V_k \oplus W_k \oplus W_{k-1} \oplus \dots \oplus W_{j+1} \quad (12)$$

This algorithm results in replacing the representation in a component  $A$  by a representation on  $V_{j+1} \oplus W_{j+1}$ . We sequentially pass from one space to another by different breakdowns on these direct sums:

$$\begin{aligned} &V_{j+1} \oplus W_{j+1} \\ &V_{j+2} \oplus W_{j+2} \oplus W_{j+1} \\ &\vdots \\ &V_k \oplus W_k \oplus W_{k-1} \oplus \dots \oplus W_{j+1} \end{aligned} \quad (13)$$

The sets of spaces  $V_j$  being fitted and following any function  $f \in L^2(IR)$  of size  $n$ , can be decomposed into the basis of wavelets and scaling functions:

$$f = \sum_{k=1}^{n/2^j} a_k^j \Phi_{j,k} + \sum_{i=1}^j \sum_{k=1}^{n/2^i} d_k^i \psi_{i,k} \quad \text{with } j \leq m$$

If we complete the analysis to the last level,  $f$  will be written as follow:

$$f = A_n + D_n + \dots + D_2 + D_1 \quad (15)$$

### B. Multiresolution Laplacian Sparse

In his three works (Gao *et al.*, 2010a; 2010a; 2013), Gao preserved the similarity by adding the Laplacian capacity to the sparse coding technique. Moreover, he added the hypergraph technique to the Laplacian sparse coding to ameliorate it. In this case, the similarity among the instances is defined by a hypergraph. As shown in Equation 4, the similarity is captured by this technique among the instances within the same hyperedge simultaneously and also makes their sparse codes similar to each other (Ben Said *et al.*, 2017).

In spite of these contributions, the proposed technique is still incapable to cover all similarities between features. It analyses images spatially and does not focus on the details of each object. We can therefore say that the analysis is carried out in a superficial way. Which is why, we propose the multiresolution Laplacian sparse coding to deepen these analyses.

In the case of Multiresolution Laplacian sparse coding, the variations of the neighbors of each object of an image are taken into account when modeling an image (Jemel *et al.*, 2016). This capacity of modeling is based on the strength of the divergence of Kullback-Lebleir and wavelet decomposition.

### C. Wavelet and Kullback-Leibler Divergence

The analysis of an image  $I$  by a family of functions  $\{\psi_{j,k}\}_{j,k}$  is called wavelet transform. This transformation is based on dilation and translation of a mother wavelet  $\psi$ . The wavelet coefficient  $w(I)_{j,k} = \langle \psi_{j,k}, I \rangle$  of the localization properties in space and frequency gives information about the content of the image  $I$  around a point  $k$  and in a frequency band near the scale  $j$ . The wavelet transform localizes the majority of the spatio-frequency information of the image into a few large amplitude coefficients when the image is reasonably smooth (Piro *et al.*, 2008).

As an initial estimation, these coefficients are uncorrelated which requires a treatment by thresholding and denosing the wavelet coefficients which is very effective in image compression. In fact, the wavelet coefficients scales are correlated at each scale. A discontinuity along a curve is converted into a point on this curve  $k_0$  by large coefficients at all scales  $\forall j, w(I)_{j,k_0} \lim_{x \rightarrow \infty}$ .

Dependency models between all coefficients have been suggested to improve the spatial structures (Goria *et al.*, 2005; Huber, 1981). Notably, there is a dependency between a wavelet coefficient  $w(I)_{j,k}$  and its closest neighbor's ladder  $(w(I)_{j-1,k})$ .

Banerjee *et al.* (2005) demonstrated that wavelet coefficient vectors statistics lead to Equation 6 which is used to distinguish the spatial structures of a very different kind:

$$w(I)_{j,k} = \left( w(I)_{j,k}, w(I)_{j-1,k}, w(I)_{j,k \pm 1_x}, w(I)_{j,k \pm 1_y} \right) \quad (16)$$

To do this, it simply regulates a Gaussian mixture model for each occurrence to express the joint probability of these vectors. In this case, it is uncertain what types of structures are present in the submissions; it cannot therefore set a model. However, it is desirable that the distribution of these vectors will be illustrative of spatial structures present in the image. Accordingly, it is essential to describe a dimension taking into account the joint probability of neighborhood vectors wavelet  $w(I)_{j,k}$ .

Given the variability of spatial structures that can be encountered in the residue, the choice of a parameterization

would be difficult to justify. We propose to introduce similarity metrics without valid parameterization of the distribution of neighborhoods: The metrics derived from the information theory such as residual entropy neighborhoods, mutual information or the Kullback-Leibler divergence between distribution neighborhoods of wavelet coefficients of the two images.

Suppose a neighborhood  $w(I)_{j,k}$  containing  $d$  coefficients. Distribution of all neighborhoods of the image  $I$  is denoted by  $p_{w(I)}$  and checks  $p_{w(I)}:IR \rightarrow IR$  and  $\int_{IR^d} p_{w(I)}(x)dx = 1$ . The differential entropy of Shannon  $p_{w(I)}$  is defined by:

$$H(p_{w(I)}) = - \int_{IR^d} p_{w(I)}(x) \log p_{w(I)}(x) dx \quad (17)$$

It measures the amount of information contained in this distribution. The Kullback-Leibler is a measure of similarity between the distributions  $p_w(I_1)$  and  $p_w(I_2)$ :

$$D_{KL}(p_w(I_1) \| p_w(I_2)) = \int_{IR^d} p_w(I_1)(x) \log \frac{p_w(I_1)(x)}{p_w(I_2)(x)} dx \quad (18)$$

Based on Equations 17 and 18, the Kullback-Leibler distance is expressed as a difference of entropies:

$$D_{KL}(p_w(I_1) \| p_w(I_2)) = H_x(p_w(I_1), p_w(I_2)) - H(p_w(I_1))$$

Knowing that the cross-entropy is defined as follows (Piro *et al.*, 2008):

$$H_x(p_w(I_1), p_w(I_2)) = \int_{IR^d} p_w(I_1)(x) \log p_w(I_2)(x) dx \quad (20)$$

The use of these dimensions on the distributions of the intensity of pixels of an image gives excellent results in the field of segmentation and image realignment (Banerjee *et al.*, 2005; Fukunaga, 1990; Kozachenko and Leonenko, 1987). A Kullback distance in wavelet space was also advanced for the indexing problem in (Collins *et al.*, 2005; Leonenko *et al.*, 2008). Particularly, in these two articles, the authors parameterize the allocation of the wavelet coefficients for each scale  $j$  by a generalized Gaussian and sum the Kullback distances obtained at each scale for the similarity between the two images.

We suggest studying similar measures to establish the similarity between two images, but with two major differences. First, the wavelet coefficients at different scales are not independent. Now summing the Kullback distances

at each scale matches the supposed independence. We therefore consider the associated entropy coefficients, in particular that of the previously illustrated neighborhoods. On the other hand, we do not parameterize distributions diversion. We suggest determining the similarity between images  $I_1$  and  $I_2$  as follows (Piro *et al.*, 2008):

$$S(I_1, I_2) = \sum_j \alpha_j D_{KL}(p_{w_j}(I_1) \| p_{w_j}(I_2)) \quad (21)$$

$p_{w_j}(I_1)$  is the non-parametric allocation of the coefficients of neighborhoods wavelet image  $I_1$  to scale  $j$  (Piro *et al.*, 2008).

$\alpha_j > 0$  is normalization weight according to attach redundancy wavelet system used in (Piro *et al.*, 2008).

Based on the definition of the sparse coding and expression (2), we establish the rule of multiresolution sparse coding:

$$\min_{v_1, \dots, v_n} \sum_i \|x_i - Uv_i\|_F^2 + \lambda \sum_i \|v_i\|_1 + \frac{\beta}{2} \sum_{ij} \|v_i - v_j\|^2 < S_{ij} > \quad (22)$$

Based on the same expression (2), matrix  $W$  implemented by Gao *et al.* (2013) is fulfilled by the coefficients of similarity of Kullback-Leibler  $S$ . Using expression (21) in completion, Boltz *et al.* (2006) suggested an estimator of the Kullback-Leibler as follows:

$$D_{KL}(\tilde{T}, \tilde{R}) \stackrel{KNN}{=} \log \frac{N_{\tilde{R}}}{N_{\tilde{T}} - 1} + d\mu_{\tilde{T}}(\log \rho_k(\tilde{R})) - d\mu_{\tilde{T}}(\log \rho_k(\tilde{T})) \quad (23)$$

$\tilde{T}$  and  $\tilde{R}$  are a set of data.  $N_{\tilde{T}}$  and  $N_{\tilde{R}}$  are the number of samples.  $\rho_k(s)$  is a radius equal to the distance to the  $k^{th}$  nearest neighbor of  $s$  excluding  $s$  itself. This evaluation is based on KNN.

This estimator of the Kullback-Leibler distance can be computed relatively quickly whatever the size of samples. It is more robust to the choice of the number of  $k$  nearest neighbors.

Using the definition of Laplacian as in (Gao *et al.*, 2013) we obtain the same Equation 3.

## Experiments Results

We evaluated our approach on different well known datasets. The results are evaluated based on global classification rates. The first database is the UIUC sport dataset (Li and Fei-Fei, 2007) the second is Corel 10 dataset (Lu and Horace, 2009) and the third is Scene 15.

**Table 1:** Description of UIUC sport dataset

Type	Rowing	Badminton	Polo	Bocce	Snowboarding	Croquet	Sailing	Climbing
Number	250	200	182	137	190	236	190	194

### A. UIUC Sport Dataset

This dataset is composed of 8 classes (Li and Fei-Fei, 2007) (Table 1):

### B. Corel 10

Corel 10 dataset contains 1000 images. Those images are divided into 10 classes of 100 images (Lu and Horace, 2009). The ten classes are beach, skiing, tigers, buildings, owls, flowers, elephants, horses, food and mountains.

### C. Scene 15

This dataset is composed of 4485 images categorized into 15 classes. Each class consists of 200 to 400 Images. It contains indoor scenes such as bedroom, kitchen, as well as outdoor scenes, such as Buildings and landscapes.

## Results

To compare our approach to those contrasted to Gao's *et al.* (2013), we selected the same basis and the same number of chosen images. The results are summarized in Table 2 and 3.

The evaluation of classification rates obtained with different methods is depicted in Table 2. Our technique gives better results than the LScSPM (Gao *et al.*, 2013) approach in the context of classification applied to UIUC sport dataset.

**Table 2:** Classification rate based on UIUC sport dataset

Method	Classification rate
HIK+one Class SVM (Wu and Rehg, 2009)	83.54±1.13
ScSPM (1024) (Yang <i>et al.</i> , 2009)	82.74±1.46
ScSPM (2048) (Yang <i>et al.</i> , 2009)	82.94±1.60
LLC (1024) (Wang <i>et al.</i> , 2010)	83.09±1.30
LLC (2048) (Wang <i>et al.</i> , 2010)	82.50±1.27
LScSPM (1024) (Gao <i>et al.</i> , 2013)	85.18±0.46
LScSPM (2044) (Gao <i>et al.</i> , 2013)	85.27±1.14
Our approach	86.55

**Table 3:** Classification rate based on corel 10

Method	Classification rate
Spatial Mismatch Kernel (Lu and Horace, 2009)	90.0
Spatial Markov Model (Lu and Ip, 2009)	77.9
ScSPM (Yang <i>et al.</i> , 2009)	86.6±1.01
LLC (Wang <i>et al.</i> , 2010)	87.93±1.04
LScSPM (Gao <i>et al.</i> , 2013)	88.76±0.76
LScSPM+CM (Gao <i>et al.</i> , 2013)	91.86±0.89
Our approach	92.91

**Table 4:** Classification rate based on scene 15 dataset

Method	Classification rate
SPM (Jacob <i>et al.</i> , 2009)	81.40±0.50
ScSPM (Yang <i>et al.</i> , 2009)	84.30±0.50
HIK+one class SVM (Wu and Rehg, 2009)	84.00±0.46
LLC (Wang <i>et al.</i> , 2010)	81.50±0.87
LScSPM (Gao <i>et al.</i> , 2013)	89.78±0.40
Our approach	91.00

Table 3 illustrates a comparison of our technique to six other techniques in the case of classification application. As can be seen from this table, the multiresolution Laplacian sparse coding outperforms all the other techniques in the case of image classification application.

Table 4 illustrates a comparison of our technique to four other techniques in case of classification application. These results show that the multiresolution Laplacian sparse coding is the best technique in the case of image classification application.

## Conclusion

In this study, we proposed a perfect approach of image classification based on Laplacian sparse coding and the divergence of Kullback Leibler and wavelet decomposition. The moderation of similarity is computed between images which combines the concepts of information theory and wavelet transform. The principle of this method is to sum the Kullback distances of each scale distribution called neighborhood vectors of wavelet coefficients. The neighborhood coefficients, containing not only spatial locations but also relative scales, get the spatial dependencies and inter-scale coefficients which can identify finer spatial structures. The Kullback distance on these vectors is predicted in a non-parametric method despite their higher dimension, thanks to entropy estimators of the nearest neighbors.

## Acknowledgment

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

## Authors Contributions

**Intidhar Jemel:** Participate in all experiments, coordinate the data-analysis and contributed to the writing of the manuscript.

**Ridha Ejbali:** Contribute to the writing of the manuscript.

**Mourad Zaied:** Coordinate and supervise the words.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and there are no ethical issues involved.

## References

- Banerjee, A., S. Merugu, I. Dhillon and J. Ghosh, 2005. Clustering with bregman divergences. *J. Mach. Learn. Res.*, 6: 1705-1749.

- Ben Said, A., I. Jemel, R. Ejbali and M. Zaied, 2017. A hybrid approach for image classification based on sparse coding and wavelet decomposition. Proceedings of the International Conference on Computer Systems and Applications, Oct. 30-Nov. 3, IEEE Xplore Press, Hammamet, Tunisia, pp: 63-68. DOI: 10.1109/AICCSA.2017.117
- Boltz, S., E. Wolsztynski, E. Debreuve, E. Thierry and M. Barlaud *et al.*, 2006. A minimum-entropy procedure for robust motion estimation. Proceedings of the International Conference Image Processing, Oct. 8-11, IEEE Xplore Press, Atlanta, GA, USA, pp: 1249-1252. DOI: 10.1109/ICIP.2006.312552
- Chen, X., Q. Lin, S. Kim, J. Pena and J.G. Carbonell *et al.*, 2010. An efficient proximal-gradient method for single and multi-task regression with structured sparsity.
- Collins, R., X. Zhou and S.K. The, 2005. An open source tracking testbed and evaluation web site. Proceedings of the International Workshop Performance Evaluation Tracking Surveillance, (ETS' 05).
- Donoho, D., 2006. For most large underdetermined systems of equations, the minimal  $\ell_1$ -norm near-solution approximates the sparsest near-solution. Commun. Pure Applied Math., 59: 907-934. DOI: 10.1002/cpa.20131
- Ejbali, R. and M. Zaied, 2017. A dyadic multi-resolution deep convolutional neural wavelet network for image classification. Multimedia Tools Applic., 77: 6149-6163. DOI: 10.1007/s11042-017-4523-2
- Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. 2nd Edn., Academic Press Professional, Inc., ISBN-10: 0122698517, pp: 592.
- Gao, S., I.W. Tsang and L.T. Chia, 2010b. Kernel sparse representation for image classification and face recognition. Proceedings of the 11th European Conference Computer vision, Sep. 05-11, Springer-Verlag Berlin, Heraklion, Crete, pp: 1-14.
- Gao, S., I.W. Tsang and L.T. Chia, 2013. Laplacian sparse coding, hypergraph laplacian sparse coding and applications. I. Trans. Patt. Anal. Mach. Intell., 35: 92-104. DOI: 10.1109/TPAMI.2012.63
- Gao, S., Z. Wang, L.T. Chia and I.W.H. Tsang, 2010a. Automatic image tagging via category label and web data. Proceedings of the 18th ACM International Conference Multimedia, Oct. 25-29, ACM, Firenze, Italy, pp: 1115-1118. DOI: 10.1145/1873951.1874164
- Goria, M.N., N.N. Leonenko, V.V. Mergel and P.L.N. Inverardi, 2005. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. J. Nonparametr. Stat., 17: 277-297. DOI: 10.1080/104852504200026815
- Hassairi, S., R. Ejbali and M. Zaied, 2015. A deep convolutional neural wavelet network to supervised Arabic letter image classification. Proceedings of the 15th International Conference Intelligent Systems Design Applications, Dec. 14-16, IEEE Xplore Press, Marrakech, Morocco, pp: 207-212. DOI: 10.1109/ISDA.2015.7489226
- Hassairi, S., R. Ejbali and M. Zaied, 2016. Sparse wavelet auto-encoders for image classification. Proceedings of the International Conference Digital Image Computing: Techniques Applications, Nov. 30-Dec. 2, IEEE Xplore Press, Gold Coast, QLD, Australia, pp: 1-6. DOI: 10.1109/DICTA.2016.7797085
- Hassairi, S., R. Ejbali and M. Zaied, 2018. A deep stacked wavelet auto-encoders to supervised feature extraction to pattern classification. Multimedia Tools Appli., 77: 5443-5459. DOI: 10.1007/s11042-017-4461-z
- Huber, P., 1981. Robust Statistics. 1st Edn., John Wiley Sons.
- Jacob, L., G. Obozinski and J.P. Vert, 2009. Group lasso with overlap and graph lasso. Proceedings of the 26th Annual International Conference Machine Learning, Jun. 14-18, ACM, Montreal, Quebec, pp: 433-440. DOI: 10.1145/1553374.1553431
- Jemel, I., R. Ejbali and M. Zaied, 2016. Multiresolution laplacian sparse coding technique for image representation. Proceedings of the International Conferences Computer Graphics, Visualization Computer Vision, (VCV' 16).
- Kozachenko, L. and N. Leonenko, 1987. On statistical estimation of entropy of random vector. Problems Inform. Transmiss, 23: 95-101.
- Leonenko, N., L. Pronzato and V. Savani, 2008. A class of rényi information estimators for multidimensional densities. Ann. Stat., 36: 2153-2182.
- Li, L.J. and L. Fei-Fei, 2007. What, where and who? classifying events by scene and object recognition. Proceedings of the 11th International Conference Computer Vision, Oct. 14-21, IEEE Xplore Press, Rio de Janeiro, Brazil, pp: 1-8. DOI: 10.1109/ICCV.2007.4408872
- Lu, Z. and H.H. Ip, 2009. Image categorization by learning with context and consistency. Proceedings of the Conference Computer Vision Pattern Recognition, Jun. 20-25, IEEE Xplore Press, Miami, FL, USA, pp: 2719-2726. DOI: 10.1109/CVPR.2009.5206851
- Lu, Z. and H.S.I. Horace, 2009. Image categorization with spatial mismatch kernels. Proceedings of the Conference Computer Vision Pattern Recognition, Jun. 20-25, IEEE Xplore Press, Miami, FL, USA, pp: 397-404. DOI: 10.1109/CVPR.2009.5206861
- Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput., 17: 395-416. DOI: 10.1007/s11222-007-9033-z

- Piro, P., S. Anthoine, E. Debreuve and M. Barlaud, 2008. Image retrieval via Kullback-Leibler divergence of patches of multiscale coefficients in the KNN framework. Proceedings of the International Workshop Content-Based Multimedia Indexing, Jun. 18-20, IEEE Xplore Press, London, UK, pp: 230-235. DOI: 10.1109/CBMI.2008.4564951
- Wang, J., J. Yang, K. Yu, F. Lv and Y. Gong, 2010. Locality-constrained linear coding for image classification. Proceedings of the Conference Computer Vision Pattern Recognition, Jun. 13-18, IEEE Xplore Press, San Francisco, CA, USA, pp: 3360-3367. DOI: 10.1109/CVPR.2010.5540018
- Wu, J. and J.M. Rehg, 2009. Beyond the euclidean distance: creating effective visual codebooks using the histogram intersection kernel. Proceedings of the 12th International Conference Computer Vision, Sept. 29-Oct. 2, IEEE Xplore Press, Kyoto, Japan, pp: 630-637. DOI: 10.1109/ICCV.2009.5459178
- Yang, J., K. Yu, Y. Gong and T. Huang, 2009. Linear spatial pyramid matching using sparse coding for image classification. Proceedings of the Conference Computer Vision Pattern Recognition, Jun. 20-25, IEEE Xplore Press, Miami, FL, USA, pp: 1794-1801. DOI: 10.1109/CVPR.2009.5206757