Multiresolution Laplacian sparse coding technique for image representation

Intidar Jemel
REsearch Laboratory in Intelligent Machines
University of Sfax, ENIS
BP 1173, Sfax, 3038, Sfax, Tunisia
intidhar_jemel@yahoo.fr

Ridha Ejbali
REsearch Laboratory in Intelligent Machines
University of Sfax, ENIS
BP 1173, Sfax, 3038, Sfax, Tunisia
ridha_ejbali@ieee.org

Mourad Zaied
REsearch Laboratory in Intelligent Machines
University of Sfax, ENIS
BP 1173, Sfax, 3038, Sfax, Tunisia
mourad.zaied@ieee.org

ABSTRACT

Sparse coding techniques have given good results in different domains especially in feature quantization and image representation. However, the major weakness of those techniques is their inability to represent the similarity between features. This limitation is due to the separate representation of features. Although the Laplacian sparse coding doesn't focus on the spatial similarity in the image space, it preserves the locality of the features only in the data space. Due to this, the similarity between two local features belong to the similarity of their spatial neighborhood in the image. To overcome this flaw, we propose the integration of similarity based on Kullback-Leibler and wavelet decomposition in the domain of an image. This technique may surmount those limitations by taking into account each element of an image and its neighbors in similarity calculation. Classifications rates given by our approach show a clear improvement compared to those cited in this article.

Keywords

Sparse coding, features quantization, image representation, Laplace sparse coding, Kullback-Leibler , wavelet decomposition.

1. INTRODUCTION

Computer vision applications have experienced a great revolution with the integration of sparse coding techniques. Unfortunately, those techniques have not be able to model the locality and the similarity among the instances to be encoded owing to the overcomplete codebook and the independent coding process. Several approaches have been proposed to overcome this limitation. In [Gao10a], Gao proposed a method called Laplacian Sparse Coding which exploits the dependence among local features. Specifically, he suggested using histogram intersection based K-NN method to build a Laplacian matrix, which will characterize the similarity of local features. Furthermore, Laplacian matrix will be incorporated into the function of sparse coding to maintain the consistence in sparse representation of those features. In [Gao10b], Gao improved the technique of Kernel Sparse Representation. It is essentially the sparse coding technique in a high dimensional feature space mapped by implicit mapping function. In 2013, he proposed the Hypergraph Laplacian Sparse Coding techniques [Gao13]. In this case, he extracts the similarity between the instances within the same hyperedge simultaneously and also composes their sparse codes similar to each other.

In this paper, we propose an amelioration of the Laplacian sparse coding technique by changing the manner of similarity computation. In our case, the calculation of similarity in the image domain is based on the divergence of Kullback-Leibler and wavelet decomposition. This idea comes from its capacity to take into account neighbors' similarity.

This paper is composed of four sections. In section1, we introduce the Laplacian sparse coding technique. Section 2 describes the kernel sparse representation. We explain our approach in section3 and in the last one, we evaluate our approach.

2. Laplacian sparse coding

Sparse coding technique was proposed in order to reduce the problem of hard quantization. It solves the problem by proposing a sparse linear combination of basis vectors for each image feature. Sparse coding looks for a linear reconstruction of a signal $x,(x \in IR^d)$ using the bases in the codebook $U=(u_1,u_2,...,u_k),(U \in IR^{d \times k})$.

The matrix of the sparse codes is $V=(v_1,v_2,...,v_n)$ where $v_i \in IR^{K \times 1}$ and v_{ik} is the weight of the vector x_i in the basis vector u_k , the optimization problem of sparse coding can be reformulated as follows:

$$\min \|v\|_0 \; subject\,to\, x = UV \; or \min_{U,V} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1$$

$$\min_{U,V} \|X - UV\|_F^2 + \lambda \sum_i \|v_i\|_1 \; subject\,to \|u_j\| \le 1; \forall j = 1,...,K$$

λ is the tradeoff parameter used to balance the sparsity and the reconstruction error.

Because of the independent encoding feature resulting from an overcomplete or sufficient codebook. Asuming that $X = (x_1, x_2, ..., x_n)$ the vector of features, 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 $Dii = \sum_{j=1}^{n} W_{ij}$ is a diagonal matrix.

The Laplacian Sparse Coding, proposed in [Gao10a], takes into account the similarity between images both in features and image domains. The expression of Laplacian Sparse Coding is as follows:

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

This expression is defined by

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

Taking into account the Laplacian definition $L = D - W$ [Lux07].

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

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

3. KERNEL SPARSE CODING

To ameliorate the technique of features representation using sparse coding, Gao proposed another approach called Kernel Sparse Representation. He noticed that kernel trick can pick up the nonlinear similarity of features. Kernel Sparse Representation is basically the sparse coding approach in a high dimensional feature space mapped by implicit mapping function [Gao10a] [Gao10b].

With the same consideration of sparse coding, we assume that there exists a feature mapping function $\Phi : IR^d \to IR^k, (d < k)$ with

$x \to \Phi(x)$,

$U = (u_1, u_2, ..., u_k) \to U_\Phi = (\Phi(u_1), \Phi(u_2), ..., \Phi(u_k))$.

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

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

Gao used Gaussian kernel due to its excellent performance in many works [Chen10] [Don04].

4. PROPOSED APPROACH

4.1. General context of multiresolution wavelet decomposition

Multiresolution wavelet decomposition analyses an image in time and frequency domains together. For lower frequency, it offers poor time resolution and better frequency resolution. Whereas, for higher frequency, it offers poor frequency resolution and better time resolution.

A multiresolution analysis is a family of nested sub spaces $L^2(IR)$ noted $(V_j)_{j \in Z}$ which have the following properties:

$$Vj = \left\{ \sum_{n \in Z} a_n^j \Phi_{j,n} : a_n^j \in IR \right\},$$

$$V_{j+1} \subset V_j, \bigcap_{j \in Z} V_j = \{0\}, \quad (5)$$

$$\bigcup_{j \in Z} \overline{V}_j = L^2(IR)$$

Hypothesis (5) means that $(V_j)_{j \in Z}$ is a space generated by the family $(\Phi_{j,n})_{n \in Z}$. Its definition depends on the chosen topology for the functional space. We can define it more strictly as the adhesion of the finite space of linear combinations of functions $\Phi_{j,n}$. Thus the approximation of a signal f on the space V_j is:

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

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

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

To write the difference between two consecutive spaces V_j and V_{j+1}, the space W_{j+1} is generated by a function $\psi_{j,n}$:

$$W_j = \left\{ \sum_{n \in Z} d_n^j \psi_{j,n} : d_n^j \in IR \right\} \quad (8)$$

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

$$D_j = \sum_n d_n^j \psi_{j,n} \quad (9)$$

And the coefficients of details d_n^j are calculated by the following formula:

$d_n^j = < f, \psi_{j,n} >$ (10)

The signal of which is assumed to be represented on a basis of V_j. Apply wavelet transform to the $k \in IN$ scale returns to representing the signal to a base adapted to the direct sum:

$V_k \oplus W_k \oplus W_{k-1} \oplus ... \oplus W_{j+1}$ (11)

The series 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:

$$f = \sum_{k=1}^{n/2^j} a_k^j \Phi_{j,k} + \sum_{i=1}^{j} \sum_{k=1}^{n/2^j} d_k^i \psi_{i,k} \quad with \quad j \le m \quad (12)$$

If we perform an analysis to the last level, f becomes:

$f = A_n + D_n + ... + D_2 + D_1$ (13)

4.2. Multiresolution Laplacian sparse coding

Gao in [Gao10a] [Gao10b] [Gao13] tried to preserve similarity by adding the Laplacian capacity to sparse coding technique. Furthermore, he ameliorated his technique by adding the hypergraph technique to the Laplacian sparse coding where the similarity among the instances is defined by a hypergraph. In this case, this technique captures the similarity among the instances within the same hyperedge simultaneously, and also makes their sparse codes similar to each other as shown in (4).

Despite these contributions, the modeling technique based on sparse coding remained unable to cover all similarities between features. This Laplacian sparse coding approach analyses images spatially and does not focus on the details of each object. We can therefore say that the analysis is done in a superficial way. For this, we propose the multiresolution Laplacian sparse coding to deepen these analyses.

Based on multiresolution Laplacian sparse coding, the modeling of images takes into account the modification of the neighbors of each object of an image. This idea came from the modeling capacity based on the divergence of Kullback-Lebleir and wavelet decomposition.

4.3. Wavelet and Kullback-Leibler divergence

Wavelet transform of an image I is the analysis of the image by a family of functions $\{\psi_{j,k}\}_{j,k}$. It consists of a dilated and translated ψ function called mother wavelet. Because of the localization properties in space and frequency of the mother wavelet, the wavelet coefficient $w(I)_{j,k} = < \psi_{j,k}, I >$ provides information about the content of the image I around point k and in a frequency band near the scale j. If the image is relatively smooth, then the wavelet transform concentrates most of the spatio-frequency information of the image into a few large amplitude coefficients [Pir08].

As a first approximation, these coefficients are uncorrelated which leads to processing by thresholding and denosing the wavelet coefficients which is very effective in image compression. But in reality, the wavelet coefficients scales are correlated at different scale. For example the presence of a discontinuity along a curve is translated into a point of this curve k_0 by large coefficients at all scales $\forall j, w(I)_{j,k_0}$.

Dependency models between different coefficients have been proposed to improve the description of spatial structures [Gor05] [Hub81]. In particular 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. showed that coefficient vectors statistics wavelet shape (equation 14) is used to characterize the spatial structures of a very different kind [Ban05].

$w(I)_{j,k} = (w(I)_{j,k}, w(I)_{j-1,k}, w(I)_{j,k\pm1_x}, w(I)_{j,k\pm1_y})$ (14)

To do this, it simply adjusts a Gaussian mixture model for each phenomenon to describe the joint probability of these vectors. In this case, it is unclear what types of structures will be present in the submissions, it cannot therefore set a model. However, it is hoped that the distribution of these vectors will be representative of spatial structures present in the image. Consequently, it is important to define a measurement taking into account the joint probability neighborhoods vectors wavelet $w(I)_{j,k}$.

Given the variability of spatial structures that can be encountered in the residual, 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 is denoted by $p_{w(I)}$ and checks $p_{w(I)} : IR \to IR$ et $\int_{IR^d} p_{w(I)}(x)\,dx = 1$.

The differential entropy of Shannon $p_{w(I)}$ is defined by:

$$H(p_{w(I)}) = -\int_{IR^d} p_{w(I)}(x)\log p_{w(I)}(x)dx \quad (15)$$

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 $p_w(I_2)$.

$$D_{KL}(p_w(I_1)\,\|\,p_w(I_2)) = \int_{IR^d} p_w(I_1)(x)\log\frac{p_w(I_1)(x)}{p_w(I_2)(x)}dx \quad (16)$$

Based on equations (15) and (16), the Kullback-Leibler distance is expressed as a difference of entropies:

$$D_{KL}(p_w(I_1)\,\|\,p_w(I_2)) = H_x(p_w(I_1), p_w(I_2)) - H(p_w(I_1)) \quad (17)$$

Knowing that the cross-entropy is defined as follows [Pir08]:

$$H_x(p_w(I_1), p_w(I_2)) = \int_{IR_d} p_w(I_1)(x)\log p_w(I_2)(x)dx \quad (18)$$

The use of these measurements on the distributions of the intensity of pixels of an image gives good results in the field of segmentation and image realignment [Ban05] [Fuk90] [Koz87]. A Kullback distance in wavelet space was also proposed for the indexing problem in [Col05] [Leo05]. Specifically, in these two articles, the authors parameterize the distribution 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 propose to study similar measures to determine 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 corresponds to the supposed independence. We therefore consider the accompanying entropy coefficients, in particular those of the previously described neighborhoods. On the other hand, we do not parameterize distributions game. We suggest measuring the similarity between images $I1$ and I2 as follows [Pir08]:

$$S(I1, I2) = \sum_j \alpha_j D_{KL}(p_{w_j}(I_1)\,\|\,p_{w_j}(I_2)) \quad (19)$$

- ✓ $p_{w_j}(I_1)$ is the non-parametric distribution of the coefficients of neighborhoods wavelet image I_1 to scale j [Pir08].
- ✓ $\alpha_j > 0$ is normalization weight according to attach redundancy wavelet system used [Pir08].

Based on the expression of the sparse coding and equation (2), we introduce the formula of multiresolution sparse coding:

$$\min_{v_1,\ldots,v_n} \sum_i \|x_i - Uv_i\|_F^2 + \lambda\sum_i \|v_i\|_1 + \frac{\beta}{2}\sum_{ij}\|v_i - v_j\|^2 < S_{ij} > \quad (20)$$

Based on the same equation (2), matrix W adopted by Gao in [Gao13] is filled by the coefficients of similarity of Kullback-Leibler S.

Using equation (19) in implementation, Sylvain Boltz in [Bol06] proposed an estimator of the Kullback-Leibler as follows:

$$D_{KL}(\tilde{T},\tilde{R}) \overset{KNN}{=} \log\frac{N_{\tilde{R}}}{N_{\tilde{T}}-1} + d\mu_{\tilde{T}}(\log\rho_k(\tilde{R})) - d\mu_{\tilde{T}}(\log\rho_k(\tilde{T})) \quad (21)$$

$\tilde{T}$ 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 estimation is based on k nearest neighbors K-NN.

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 in [Gao13]. We get the same equation (3).

5. EXPERIMANTS RESULTS

5.1. UIUC-Sport Dataset

This dataset consist of 8 classes [Li07]. Each class contains a set of images described in the following table:

Type	Rowing	Badminton	Polo	Bocce
Number	250	200	182	137

Snowboarding	Croquet	Sailing	Climbing
190	236	190	194

Table 1. Description of UIUC-Sport Dataset

5.2. Corel 10 Dataset

Corel 10 dataset is composed of 10 classes [Lu09]. Each class contains 100 images. The ten classes are skiing, beach, buildings, tigers, owls, elephants, flowers, horses, mountains, and food.

5.3. Results

To compare our results with those of Gao in [Gao13], we have chosen the same basis and the same number of selected images. The following tables resume all results.

Method	Classification rate
HIK+one Class SVM [Wu09]	83.54±1.13
ScSPM (1024) [Yang09]	82.74±1.46
ScSPM (2048) [Yang09]	82.94±1.60
LLC (1024) [Wang10]	83.09±1.30
LLC (2048) [Wang10]	82.50±1.27
LScSPM (1024) [Gao13]	85.18±0.46
LScSPM (2044) [Gao13]	85.27±1.14
Our approach	86.55

Table 2. Classification rate based on of UIUC-Sport Dataset

Table 2 evaluate classification rates obtained with different methods. Results of comparison have shown that our approach performs better results than the LScSPM [Gao13] approach in the context of classification applied to UIUC sport dataset.

Method	Classification rate
Spatial Mismatch Kernel [Lu09]	90.0
Spatial Markov Model [Lu09]	77.9
ScSPM [Yang09]	86.6±1.01
LLC [Wang10]	87.93±1.04
LScSPM [Gao13]	88.76±0.76
LScSPM+CM [Gao13]	91.86±0.89
Our approach	92.91

Table 3. Classification rate based on of Corel 10 Dataset

In table 3, we compared our technique to six other techniques in case of classification application. Results illustrated in the same table showed that multiresolution Laplacian sparse coding is the best technique of image representation for classification application.

6. CONCLUSION

In this paper, we suggested an improved method of image representation based on Laplacian sparse coding and the divergence of Kullback Leibler and wavelet decomposition. This measure of similarity is calculated between images which combine the concepts of information theory and wavelet transform. The principle of this approach is to sum the Kullback distances of each scale distributrion called neighborhood vectors of wavelet coefficients. The neighborhood coefficients, containing not only spatial locations but also relative scales, capture the spatial dependencies and inter-scale coefficients which can detect finer spatial structures. The Kullback distance on these vectors is estimated in a non-parametric manner despite their higher dimension, thanks to entropy estimators of nearest neighbors.

7. ACKNOWLEDGMENTS

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.

8. REFERENCES

[Gao10a] Gao ,S. Wang, Z. Chia, L.T. and Tsang, I.W.H. Automatic Image Tagging via Category Label and Web Data, Proc. ACM Int'l Conf. Multimedia, 2010.

[Gao10b] Gao, S. Tsang, I.W. and Chia, L.T. Kernel Sparse Representation for Image Classification and Face Recognition, Proc. European Conf. Computer Vision, 2010.

[Gao13] Gao, S. Tsang, I.W. and Chia, L.T. Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 35, NO. 1, 2013

[Lux07] Luxburg, U.V. A Tutorial on Spectral Clustering, Statistics and Computing, vol. 14, no. 7, pp. 395-416, 2007.

[Chen10] Chen, X. Lin, Q. Kim, S. Pen, J. Carbonell, J.G. and Xing, E.P. An Efficient Proximal-Gradient Method for Single and Multi-Task Regression with Structured Sparsity, technical report, arXiv:1005.4717, 2010.

[Don04] Donoho, D. For Most Large Underdetermined Systems of Linear Equations, the Minimal l1-Norm Near-Solution Approximates the Sparsest Near-Solution, technical report, Stanford Univ., 2004.

[Pir08] Piro, P., Anthoine, S., Debreuve, E., Barlaud, M., Image retrieval via Kullback-Leibler divergence of patches of multiscale coefficients in the KNN framework, International Workshop on Content-Based Multimedia Indexing, PP:230-235, 2008.

[Gor05] Goria, M.N. Leonenko, N. N. Mergel, V. V. and Inverardi, P. L. N. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. J. Nonparametr. Stat., 17(3):277–297, 2005.

[Hub81] Huber, P. Robust Statistics. John Wiley and Sons, 1981.

[Ahm76] Ahmad, I. and Lin, P. E. A nonparametric estimation of the entropy for absolutely continuous distributions. IEEE Trans. Inform. Theory, 22(3):372– 375, 1976.

[Ban05] Banerjee, A. Merugu, S. Dhillon, I. and Ghosh, J. Clustering with bregman divergences. J. Mach. Learn. Res., 6:1705– 1749, 2005.

[Fuk90] Fukunaga, K. Introduction to statistical pattern recognition (2nd ed.). Academic Press Professional, Inc., 1990.

[Koz87] Kozachenko, L. and Leonenko, N. On statistical estimation of entropy of random vector. Problems Infor. Transmiss., 23(2):95–101, 1987.

[Col05] Collins, R. Zhou, X.and The, S. K. An open source tracking testbed and evaluation web site. In IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), Breckenridge, CO, http://www.vividevaluation.ri.cmu.edu/, 2005

[Leo05] Leonenko, N. Pronzato, L. and Savani, V. A class of Renyi information estimators for multidimensional densities. PASCAL archive: http://eprints.pascalnetwork.org/archive/0000103 1/, 2005.

[Bol06] Boltz, S. Wolsztynski, E. Debreuve, E. Thierry, E. Barlaud, M. and Pronzato, L. A minimum-entropy procedure for robust motion estimation. In ICIP, Atlanta, GA, October 2006.

[Li07] Li, L.J. and Fei-Fei, L. What, Where and Who? Classifying Events by Scene and Object Recognition, Proc. IEEE Int'l Conf. Computer Vision, 2007.

[Lu09] Lu, Z and Ip, H.H. Image Categorization with Spatial Mismatch Kernels, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[Wu09] Wu, J. and Rehg, J.M. Beyond the Euclidean Distance: Creating Effective Visual Codebooks Using the Histogram Intersection Kernel, Proc. IEEE Int'l Conf. Computer Vision, 2009.

[Yang09] Yang, J. Yu, K. Gong, Y. and Huang, T. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.

[Wang10] Wang, J. Yang, J. Yu, K. Lv, F. and Gong, Y. Locality-Constrained Linear Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.

[Lu09] Lu, Z. and Ip, H.H. Image Categorization by Learning with Context and Consistency," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.