Towards dictionary learning from images with non Gaussian noise

Pierre Chainais

▶ **To cite this version:**

Pierre Chainais. Towards dictionary learning from images with non Gaussian noise. IEEE Int. Workshop on Machine Learning for Signal Processing, Sep 2012, Santander, Spain. hal-00749035

HAL Id: hal-00749035
https://hal.science/hal-00749035

Submitted on 6 Nov 2012

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.

2012 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 23–26, 2012, SATANDER, SPAIN

TOWARDS DICTIONARY LEARNING FROM IMAGES WITH NON GAUSSIAN NOISE

Pierre Chainais

Ecole Centrale Lille - LAGIS CNRS UMR 8219 - INRIA Lille-Nord Europe, SequeL
Avenue Paul Langevin - BP48 59651 Villeneuve d'Ascq Cedex France

ABSTRACT

We address the problem of image dictionary learning from noisy images with non Gaussian noise. This problem is difficult. As a first step, we consider the extreme sparse code given by vector quantization, i.e. each pixel is finally associated to 1 single atom. For Gaussian noise, the natural solution is K-means clustering using the sum of the squares of differences between gray levels as the dissimilarity measure between patches. For non Gaussian noises (Poisson, Gamma,...), a new measure of dissimilarity between noisy patches is necessary. We study the use of the generalized likelihood ratios (GLR) recently introduced by Deledalle et al. in [1] to compare non Gaussian noisy patches. We propose a K-medoids algorithm generalizing the usual Linde-Buzo-Gray K-means using the GLR based dissimilarity measure. We obtain a vector quantization which provides a dictionary that can be very large and redundant. We illustrate our approach by dictionaries learnt from images featuring non Gaussian noise, and present preliminary denoising results.

Index Terms— clustering, dictionary learning, denoising, patch similarity.

1. INTRODUCTION

In presence of additive Gaussian noise, the most basic approach to dictionary learning is K-means clustering of patches (e.g. of size $w = 7 \times 7$) using the sum of the squares of differences between gray levels as the dissimilarity measure. For non Gaussian noises (Poisson, Gamma,...), dictionary learning remains a difficult problem. Even the generalization of K-means clustering is not that simple. In particular, a new measure of dissimilarity between noisy patches is necessary. In this work, we propose to use the generalized likelihood ratios recently introduced in [1]. Based on this adapted dissimilarity measure, we propose a recursive version of the Linde-Buzo-Gray (LBG) K-medoids algorithm [2] to cluster patches and obtain a vector quantization (VQ) which provides the most fundamental dictionary. An optimized and accelerated version of NL-means was proposed in [3] using similar ideas. Even though simplistic, this dictionary may be useful if sufficiently redundant as argued in [4] where very good results are obtained from VQ dictionaries. The true question is not to find the optimal dictionary or the optimal encoding on a dictionary but to find the best compromise between sparsity, adaptivity and redundancy [5, 6]. Thus, we think that even though VQ may seem not sophisticated enough, it diserves some study, at least as the natural first step toward dictionary learning with non Gaussian noise.

We present and study an adapted clustering method toward dictionary learning from noisy images. Then we illustrate our study by numerical experiments on Poisson and Gamma noisy images in particular. We present an application to denoising without aiming at state of the art results. Our purpose is to show that interesting results may be obtained even from the most basic approach. Noticeably, we will observe promising results as far as textured images are considered. We use the powerful GLR based NL means (non Gaussian) denoising technique as a reference [1] for comparisons.

Section 2 presents the notion of generalized likelihood ratios. Section 3 describes the proposed algorithm. Section 4 briefly recalls the principle of NL-means for denoising and presents the VQ dictionary based denoising technique used here. Section 5presents experimental results obtained from natural and textured images. Section 6 gathers concluding comments and perspectives.

2. GENERALIZED LIKELIHOOD RATIOS

2.1. Definition and main properties

Generalized Likelihood Ratios (GLR) have recently been introduced by Deledalle et al. in [1] to efficiently compare noisy patches. The main purpose of this profound work was to propose an NL-means algorithm in presence of non Gaussian noise, Poisson and Gamma noises in particular. As explained in the introduction, we precisely need a relevant measure of similarity (or equivalently dissimilarity) between noisy patches. GLR share many good properties compared to other possible choices, see [1] for a detailed presentation. We briefly recall their definition and main properties.

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ denote two noisy patches seen as the realizations of independent random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ following the same parametric distribution of parameter $\boldsymbol{\theta}_1$ and

978-1-4673-1026-0/12/$31.00 ©2012 IEEE

$\boldsymbol{\theta}_2$, the $\boldsymbol{\theta}$s denoting noise-free patches. GLR are built on an extension of a hypothesis test:

$$\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{12} \qquad \text{(null hypothesis)} \tag{1}$$
$$\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \qquad \text{(alternative hypothesis)} \tag{2}$$

Neyman-Pearson theorem leads to the optimal criterion which is given by the likelihood ratio test:

$$\mathcal{L}(\mathbf{x}_1, \mathbf{x}_2) = \frac{p(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}_{12}, \mathcal{H}_0)}{p(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathcal{H}_1)} \tag{3}$$

Such a criterion requires the knowledge of the parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_{12}$ which correspond to the unavailable noise-free patches. The generalized likelihood ratio (GLR) approach consist in replacing the noise-free patches by their maximum likelihood estimates $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_{12}$ under each hypothesis:

$$GLR(\mathbf{x}_1, \mathbf{x}_2) = \frac{p(\mathbf{x}_1, \boldsymbol{\theta}_1 = \hat{\mathbf{t}}_{12}) p(\mathbf{x}_2, \boldsymbol{\theta}_2 = \hat{\mathbf{t}}_{12})}{p(\mathbf{x}_1, \boldsymbol{\theta}_1 = \hat{\mathbf{t}}_1) p(\mathbf{x}_2, \boldsymbol{\theta}_2 = \hat{\mathbf{t}}_2)} \tag{4}$$

GLR have all the most desirable properties of a similarity measure (symmetry, maximal self-similarity, equal self-similarities, identity of indiscernibles, invariance...). An important one is that there exists explicit expressions for usual distributions like Poisson and Gamma. The dissimilarity measure is given by $D = -\log GLR$. In the Gaussian case one recovers the usual Gaussian kernel; dissimilarity is then simply the usual L^2 distance.

$$GLR_{Gauss}(\mathbf{x}_1, \mathbf{x}_2) = \exp[-(\mathbf{x}_2 - \mathbf{x}_1)^2] \tag{5}$$
$$D_{Gauss}(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_2 - \mathbf{x}_1)^2 \tag{6}$$

2.2. GLR for Poisson and Gamma noises

Speckle noise is encountered in synthetic aperture radar (SAR) imagery and is generally modelled by Gamma noise. Poisson noise appears in photon-limited imagery (astronomy, microscopy...). For two Poisson distributed random variables, the GLR based similarity is defined by:

$$GLR_{\mathcal{P}}(x_1, x_2) = \frac{(x_1 + x_2)^{x_1 + x_2}}{2^{x_1 + x_2} x_1^{x_1} x_2^{x_2}} \tag{7}$$

Assuming independent Poisson noise, this yields the dissimilarity measure between two patches of size w denoted by $\mathbf{x} = (x(i))_{1 \leq i \leq w}$ and $\mathbf{y} = (x(i))_{1 \leq i \leq w}$:

$$D_{\mathcal{P}}(\mathbf{x}, \mathbf{y}) = \textstyle\sum_i x(i) \log x(i) + y(i) \log y(i) \\ +(x(i) + y(i)) \log 2 - (x(i) + y(i)) \log(x(i) + y(i)) \tag{8}$$

For two Gamma distributed random variables with parameter L, the GLR based similarity is defined by:

$$GLR_{\mathcal{G}}(x_1, x_2) = 2^{2L} \left(\frac{x_1 x_2}{(x_1 + x_2)^2} \right)^L \tag{9}$$

Assuming independent Poisson noise, this yields the dissimilarity measure between two patches $\mathbf{x}$ and $\mathbf{y}$:

$$D_{\mathcal{G}}(\mathbf{x}, \mathbf{y}) = -\textstyle\sum_i \log x_(i) + \log y(i) \\ +2 \log(x(i) + y(i)) \tag{10}$$

3. LBG K-MEDOIDS

Since the problem of dictionary learning from noisy images in the presence of non Gaussian noise is a difficult one, our purpose is to first consider the most simple way of learning a dictionary, namely vector quantization. In the Gaussian case, K-means clustering is the natural answer since the denoising problem is linked to the minimization of the usual L^2 norm (6). The corresponding similarity measure is described by the Gaussian kernel (5). For non Gaussian noises, another similarity measure is needed and GLR has all the desired properties. The generalization of K-means to general dissimilarity measures like (8) or (10) in place of the usual L^2 norm is called K-medoids [2]. We recall below the basic structure of this algorithm as well as its recursive Linde-Buzo-Gray (LBG) version. We also give its detailed implementation when using GLR.

3.1. K-medoids

We recall the principle of the K-medoids algorithm:

Algorithm : K-medoids clustering

---

1. For a given cluster assignment C, find the centers m_k minimizing total distance to points in each cluster:

$$m_k = \underset{m}{\operatorname{argmin}} \sum_{i:C(i)=k} D(m, x_i) \tag{11}$$

Then m_k, $k = 1, 2, ..., K$ are the current estimates of the cluster centers.

2. Given a current set of cluster centers $\{m_1, ..., m_K\}$, minimize the total error by assigning each observation to the closest (current) cluster center:

$$C(i) = \underset{1 \leq k \leq K}{\operatorname{argmin}} D(x_i, m_k), \qquad 1 \leq k \leq K \tag{12}$$

3. Iterate steps 1 and 2 until the assignments do not change.

K-medoids is equivalent to K-means when $D(\mathbf{x}_1, \mathbf{x}_2) = ||\mathbf{x}_2 - \mathbf{x}_1||^2$. In K-means, computing cluster centers requires $O(N_k)$ operations, for N_k data assigned to class k. For other more complex dissimilarities, the computation to solve the optimization problem (11) increases as $O(N_k^2)$ so that K-medoids is far more computationally intensive than K-means. As a consequence, K-medoids is not a computational

panacea. It should be used only if really needed to get better results than by using K-means, that is likely in presence of a relatively high level of noise.

3.2. Recursive LBG K-medoids

Since we have no prior information on the distribution of patches and taking into account that we are working in a space of large dimension (typically 7×7) space, we will use a Linde-Buzo-Gray (LBG) recursive version of K-medoids. By using , we get a multiresolution Voronoi tessellation of the distribution of patches.

At each recursion, each cluster is splitted in 2 new clusters until the desired number K of clusters is reached. In practice, we do not choose precisely the final number of clusters, but only the number of levels of the recursion p so that we get at most $K = 2^p$ clusters. To ensure that no patch remain alone, we stop splitting when a cluster reaches some minimal size N_{min} (typically 8 or 16).

Algorithm LBG K-medoids (recursive)

1. Compute the centroid μ_0 of the training set, as well as the standard deviation σ_0.

2. Perturb μ_0 to get μ_1 and μ_2: we use $\mu_i = \mu_0 + (-1)^i \varepsilon_i \sigma_0$ where ε_i is a uniform random variable in $[0, 1]$.

3. Apply the 2-medoids algorithm starting from μ_1 and μ_2 to get an optimum binary codebook and two clusters $\mathcal{C}_1$ and $\mathcal{C}_2$.
Remark: if the size of one of the new clusters is below N_{min}, then stop splitting.

4. Apply steps 2 and 3 to subsets $\mathcal{C}_1$ and $\mathcal{C}_2$ until the depth of the binary splitting reaches p.

In the end, the cluster barycenters (not the centroids) are used to build the desired dictionary. It contains at most 2^p atoms, each of them being a denoised patch obtained by averaging at least N_{min} patches to ensure both its reperesentativity and sufficient denoising. Let us emphasize that in the present case, the recursive LBG version is expected to be faster and more stable than a direct K-medoids clustering. Indeed, at each level of the recursion, new clusters are obtained as subclusters of the previous ones. As a consequence, the resulting Voronoi tessellation is in some sense a multiresolution clustering. At each level, the most different patches are separated in 2 classes. It is expected that such an implementation be therefore more robust to outliers.

We usually work with patches of typical size 7×7 so that we work in dimension $w = 49$. One may first use a principal component analysis (PCA) as dimension reduction pre-processing of patches. PCA is also known to have good denoising properties, which may help to that purpose. However, we must keep in mind that PCA has close link with the asumption of additive Gaussian noise while we precisely want to consider non Gaussian noises. We do not use PCA in the present study.

3.3. Newton iterates for medoids

We now explain how to solve the optimization problem (11) in K-medoids thanks to Newton iterates when some explicit dissimilarity measure other than the quadratic distance is chosen. We get an explicit Newton formulation with respect to $\{m_1, ..., m_K\}$. The quantity to be minimized is:

$$E\left((m_k)_{1\leq k\leq K}\right) = \sum_{k=1}^{K} \sum_{i:C(i)=k} \sum_{j=1}^{w} D(x_i(j), m_k(j)) \quad (13)$$

For Poisson noise, the GLR based dissimilarity D is given by (8). As a consequence, to determine the zeros of $\frac{\partial E}{\partial m_k(j)}$ for each component $m_k(j)$ is equivalent to find the zeros of

$$F(m_k(j)) = \log m_k(j) - \frac{1}{N_k} \sum_{i:C(i)=k} \log\left(\frac{m_k(j) + x_i(j)}{2}\right) \quad (14)$$

which can be quickly estimated by using Newton-Raphson algorithm. For Gamma noise, the GLR based dissimilarity D is given by (10). To determine the zeros of $\frac{\partial E}{\partial m_k(j)}$ for each component $m_k(j)$ is then equivalent to find the zeros of

$$F(m_k(j)) = \frac{2}{N_k} \sum_{i:C(i)=k} \frac{m_k(j)}{m_k(j) + x_i(j)} - 1. \quad (15)$$

Again Newton algorithm is used. In both cases, Poisson and Gamma, we use the usual quadratic distance barycenter to initialize Newton iterates. In practice, only 3 Newton iterations are then sufficient to accurately estimate the optimal $(m_k)_{1\leq k\leq K}$. Note that this can be done in parallel for all components $(m_k(j))$, $1 \leq k \leq K$ and $1 \leq j \leq w$.

4. NL-MEANS AND VQ DENOISING

NL-means method [7] and its generalizations [1, 8] yield state of the art denoising results. They are exploiting the similarities present in images to estimate denoised information. While theoretically non-local, NL-means use spatialy localized neighborhoods in practice. Dictionary learning is an alternative to learn typical features that may be present anywhere in the image. Once the dictionary is estimated, the encoding projection of each patch on this dictionary must be estimated (in place of the similarity coefficients used by NL-means). Many approaches propose an alternate optimization of the dictionary and the encoding coefficients [6]. However, it has recently been argued in [4] that the dictionary learning and the encoding steps might be decoupled and yield interesting results for a specific task. Here we explore the most

simple dictionary learning approach, namely vectorial quantization by clustering. To test the relevance of the dictionary only (and not of the encoding strategy), for sake of severity, we will compare a very simple encoding strategy for denoising to optimized NL-means.

4.1. NL-means denoising

Let x be the central noisy pixel of some noisy patch $\mathbf{x}$. Let $\theta(x)$ and $\boldsymbol{\theta}(\mathbf{x})$ denote their noise-free values. Then, for some (well chosen) scale parameter h, NL-means defines the estimate $\hat{\theta}(x)$ as an average over other patches weighted by similarities:

$$\hat{\theta}(x) = \frac{\sum_{\mathbf{y}} GLR(\mathbf{x},\mathbf{y})^{1/h}\, y}{\sum_{\mathbf{y}} GLR(\mathbf{x},\mathbf{y})^{1/h}} \tag{16}$$

In practice, for computational reasons, the patches $\mathbf{y}$ similar to some patch $\mathbf{x}$ are looked for within some spatial neighborhood of the pixel under consideration. As a consequence, practical NL-means are not truly non-local over the whole image, but only over a 'large' neighborhood around each pixel. Each denoised pixel is estimated from a weighted average of several (sometimes many) similar pixels. In some sense, NL-means consider similar neighboring patches as the dictionary, and simillarities as the coefficients. NL-means yields close to state of the art denoising results for additive Gaussian noise when its parameters are optimized. Moreover, in the case of non Gaussian noises, the GLR based NL-means introduced in [1] yield state of the art results.

4.2. Vectorial quantization based denoising

Dictionary learning, even with a very simple approach as vector quantization, allows to work in a really non local manner by learning representative atoms from a large set of patches from the whole image. During the encoding step, each pixel can thus be denoised by using information from everywhere in the image (or family of images). This also proves useful when dealing with a set of images since the same dictionary can be used for all of them after only one learning stage. Part of the success of a dictionary based denoising process is then due to the encoding procedure, e.g. sparse coding.

As an example of an application, we will consider non Gaussian denoising from our VQ dictionary learnt thanks to the GLR based K-medoids or K-means clustering described above. To study the efficiency of the learning procedure on its own, we will use one of the most simple encoding procedure: each pixel is simply replaced by the average of the values of pixels from denoised overlapping patches. To make the comparisons severe, we will compare our denoising results using simple VQ based denoising to the GLR based NL-means [1] on some images.

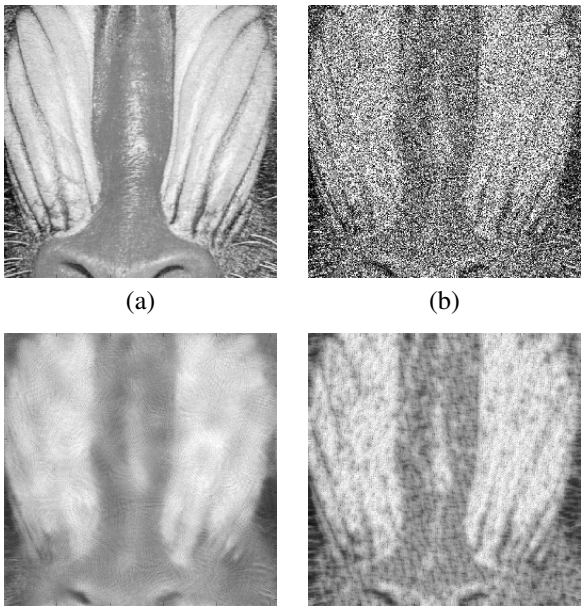


Fig. 1. (a) Example of a 256×256 natural image ; (b) affected by strong Poisson noise with PSNR = 9.49 dB ; (c) NL-means denoised with optimal parameters as in [1], PSNR = 22.9 dB; (d) VQ denoised using K-medoids dictionary learning, PSNR = 22.4 dB.

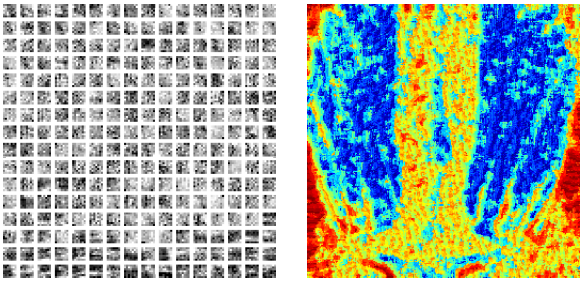
5. EXPERIMENTAL RESULTS

5.1. Setting

As an illustration, we will show two types of results. First, we show the appearance of learnt dictionaries, which is always indicative of its ability to capture representative content. Second, we show some results of denoising by the very simple procedure described in previous section. Note that these denoising results are all but optimal due to the chosen encoding procedure: each pixel is simply replaced by the average of the values of pixels from denoised overlapping patches. As discussed before, aiming at better denoising, some sparse encoding method or thresholding strategy should be used. This is the subject of ongoing work. Our main purpose is to study whether even a simple vector quantization, with a high level of redundancy, can yield interesting dictionaries, and permit to take into account the presence of non Gaussian noise.

We have processed a wide set of images but display the representative results obtained from 2 images in particular. The first image is the 256×256 central part of Mandrill (the famous one) to illustrate what happens on some natural image, while the second one is a scale invariant texture aimed at modeling an image of the Sun in the extreme UV [9].

To make the comparison of denoising results with NL-means as severe as possible, the results of NL-means are ob-



(a) (b)

Fig. 2. A 256×256 piece of Mandrill image: (a) 256 most frequently used atoms in the dictionary; (b) repartition of pixels cluster indices is clearly non local over the whole image .

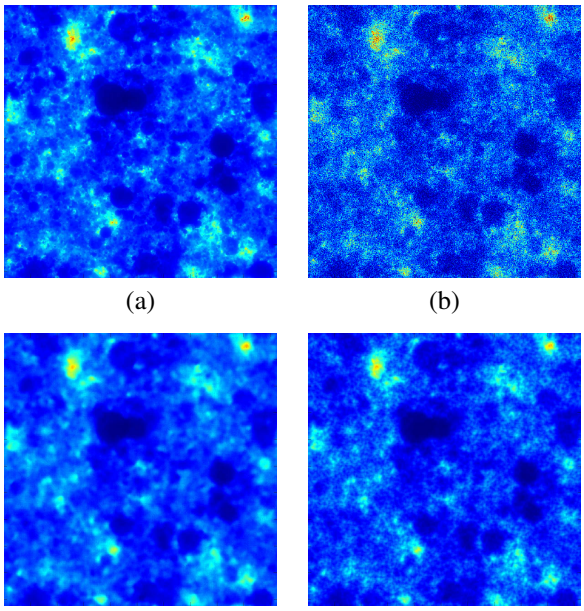
tained using GLR as described in [1] and using the optimal scale parameter h in (16) (this is possible thanks to a Newton descent knowing the original image). Note that in true life, the original image is not known and such an optimization is not possible: the choice of the scale parameter h is critical. On the contrary, apart from the depth of the hierarchical clustering procedure, there is no such parameter in the clustering dictionary learning approach.

5.2. Example 1: Mandrill

The image considered in fig.1(a) is a piece of Mandrill. In this experiment, a strong Poisson noise is applied with a PSNR of 9.49 dB, fig.1(b). The learnt dictionary is presented in fig.1(a) and looks satisfactory. Fig.1(b) represents the spatial origin of pixels which contributed to each cluster. Note how they are spatially distributed over the whole image, obeying to its original symmetry. Turning to the denoising experiment, the optimal result from the GLR based NL-means [1] yields fig.3(c) with a PSNR of 22.9 dB. The K-medoids VQ based denoising method yields fig.1(d) with a PSNR of 22.4 dB, which is 0.5 dB below NL-means and seems worse. However, we remark that NL-means has led to a very good PSNR by smoothing some small scale details (while we also optimized the sizes of patches and neighbohoods to limit this smoothing) while VQ based has been able to preserve some details

noise (PSNR dB)	K-means	K-medoids	NL-means
texture 10.0/15.2	27.5/28.3	25.2/29.7	**28.0/30.4**
20.3/20.3	**33.1**/32.8	33.0/**32.9**	**33.1**/32.0
30.0/30.0	**37.8/36.8**	**37.7**/36.2	36.8/36.5
mandrill 15.2/15.0	**25.5**/25.3	**25.5/25.5**	25.2/25.4
256×256 20.1/20.2	27.0 /**27.1**	27.0/ **27.1**	**27.1/27.1**

Table 1. Denoising results in terms of PSNR of scale invariant texture for various Poisson/Gamma noise levels.



(a) (b)

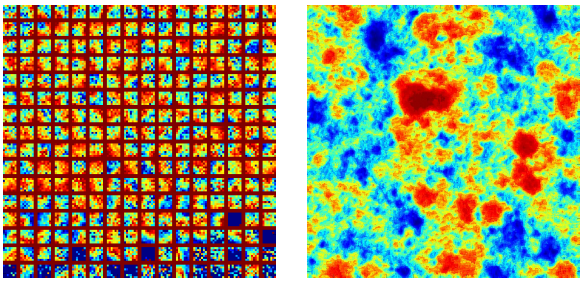
(c) (d)

Fig. 3. (a) A disordered and textured image ; (b) with Poisson noise, PSNR = 20.34 dB ; (c) NL-means denoised (optimized parameters as before), PSNR = 33.1 dB; (d) VQ denoised using K-medoids dictionary learning, PSNR = 33.0 dB.

and looks sharper. We also tried K-means VQ based denoising and got a PSNR of 22.0 dB (not displayed), still preserving small scales. Averaging K-medoids VQ and NL-means yields a PSNR of 23.3 dB which confirms their complemenarity. To conclude, the true non locality of a redundant dictionary learning in itself brings a lot of advantages, even when using a crude encoding and the most basic dictionary learning approach, namely vector quantization.

5.3. Example 2: scale invariant texture

The image considered in fig.3 is a random scale invariant (multifractal) texture, fig.3(a). Such images are common in astronomy for instance [9]. In this experiment, a Poisson noise is applied with a PSNR of 20.34 dB, fig.3(b). The learnt dictionary is presented in fig.4(a) and satisfactorily shows both smooth and very irregular atoms (the image is monochrome; the colormap is artificial for better contrast). Fig.4(b) represents the spatial origin of pixels which contributed to each cluster. Note how they are spatially distributed over the whole image. This is certainly key to the good denoising results. Indeed, turning to the denoising experiment, the optimal result from the GLR based NL-means [1] yields fig.3(c) with a PSNR of 33.1 dB. The K-medoids VQ based denoising method yields fig.3(d) with a PSNR of 33.0 dB, nearly the same as NL-means. Above all, we emphasize that NL-means has led to a very good PSNR by



(a) (b)

Fig. 4. A very disordered and textured image: (a) 256 most frequent atoms in the dictionary ; (b) Repartition of pixels cluster indices is clearly non local over the whole image.

smoothing some small scale details (we optimized parameters as before) while VQ has remarkably well preserved these details and looks sharper. We also tried K-means VQ based denoising and got a PSNR of 33.1 dB (not displayed), that is the same PSNR as NL-means and moreover preserving small scales. In both cases, the textured aspect of the VQ denoised image looks visually more satisfying.

An interesting observation is that denoising results competing with optimized NL-means are observed on non Gaussian noisy randomly textured images. This has to be confirmed by more complete and detailed experiments, but gives an encouraging indication. Moreover, while similar PSNR may be achieved by both methods, which may appear somewhat surprising with such a basic vector quantization, the visual aspect compares in favor of VQ based denoising since small scale details are not smoothed out. Note that, when looking at scale invariant textures (with a decreasing power-law Fourier spectrum), most of the energy is concentrated at larges scales while the eye remains very sensitive to small scales which content the visual essence of the texture. Therefore we note that this 'small scale preserving' property may be useful, even though the PSNR may not be optimal. This can even be seen on Mandrill for which the PSNR of NL-means is clearly bettter, while the contours and details are better seen in the VQ denoised version.

6. CONCLUSION AND PERSPECTIVES

In the present contribution, we have presented a first study of a simple approach towards dictionary learning from noisy images with non Gaussian noise. To this aim a recursive binary LBG K-medoids clustering yields a vector quantization of patches (typically 7×7). The generalized likelihood ratios introduced in [1] were used as dissimilarity measures between noisy patches, adapted to the type of noise. We have shown that relevant dictionaries can be learnt even with a rather high level of Poisson or Gamma noise. We have illustrated the interest of such dictionaries for denoising on some examples, usual images and more textured ones, see tab. 1. Despite the use of the most simple encoding strategy, vector quantization, the results can reasonably be compared to GLR based NL means results which are close to state of the art. Ongoing work deals with the use of other encoding, e.g. sparse coding. A point which calls for further study concerns the interest of K-medoids compared to K-means since an important limitation of the LBG K-medoids approach is it numerical cost mainly due to the optimization step to determine cluster centroids. This may be a difficulty common to any attempt of dictionary learning with non Gaussian noise. Interestingly, we often observed denoising performances similar to optimized NL-means, see tab. 1, with often an even better rendering at small scales. This may be interesting when denoising textured images. The average of both results appears to be even better. Note that there is nearly no adjustable parameter (except the depth of the recursive clustering and the minimum size of clusters) while NL-means (like other methods) needs to choose the critical scale parameter h in (16).

7. REFERENCES

[1] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? Patch similarity beyond gaussian noise," *Int. J. of Comp. Vis.*, vol. 99, pp. 86–102, 2012.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning (2nd ed.)*, Springer-Verlag, 2009.

[3] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1083 – 1092, 2008.

[4] A. Coates and A.Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proc. of the 28th ICML*, 2011.

[5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311 –4322, 2006.

[6] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Sig. Proc. Mag.*, vol. 28, no. 2, pp. 27 –38, 2011.

[7] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[8] P. Chatterjee and P. Milanfar, "Patch-based near-optimal denoising," *IEEE Trans. on Image Processing*, vol. 21, no. 4, pp. 1635–1649, 2012.

[9] P. Chainais, E. Kœnig, V. Delouille, and J.-F. Hochedez, "Virtual super resolution of scale invariant textured images using multifractal stochastic processes," *J. of Math. Im. & Vis.*, vol. 39, no. 1, pp. 28–44, 2011.