Noise suppression over bi-level graphical documents by sparse representation

T-H. Do*[1], S. Tabbone*, O. Ramos-Terrades[2]**

** LORIA - UMR 7503*
Campus scientifique - BP 239 - 54506 Vandoeuvre-les-Nancy, France
{ha-thanh.do, antoine.tabbone}@loria.fr

*** Universitat Autonoma de Barcelona - Computer Vision Centre*
08193 Bellaterra (Cerdanyola) Barcelona, Espanya

oriolrt@cvc.uab.cat

RÉSUMÉ. *Dans cet article, nous explorons l'utilisation de l'algorithme d'apprentissage (K-SVD) pour construire des dictionnaires adaptés aux documents graphiques. En plus, dans notre modèle, nous avons également modélisé l'énergie du bruit à partir de la fonction de la corrélation croisée normalisée entre les documents bruités et non bruités définis dans notre base d'apprentissage. Nous avons évalué cette méthode sur la base de données Grec2005. Les résultats expérimentaux démontrent la robustesse de notre approche en comparant à des méthodes de l'état de l'art.*

ABSTRACT. *In this paper, we explore the use of learning algorithm (K-SVD) for building dictionaries adapted to the image properties. In addition, in our model, we also modeled the energy of the noise basing on the function of the normalized cross-correlation between noised and non noised documents identified in training set. We have evaluated this method on the Grec2005 dataset. The experimental results demonstrate the robustness of our approach by comparing it with state-of-the-art methods.*

MOTS-CLÉS : *Représentation parcimonieuse, dictionnaire d'apprentissage, K-SVD, suppresion le bruit, documents graphiques.*

KEYWORDS: *Sparse representation, learned dictionary, K-SVD, noise removal, graphical documents.*

1. This work is partially supported by the European project Eureka SCANPLAN.
2. Partially supported by Spanish projects TIN2011-24631, TIN2009-14633-C03- 01, and CONSOLIDER-INGENIO 2010 (CSD2007-00018)

1. Introduction

Images often contain noise, which may arise from the printing, photocopying, and scanning processes. Noise not only gives an image a generally undesirable appearance, but also has an influence on the performance of the symbol recognition process as some features of symbols are covered and reduced.

Modelling this kind of noise has been an object of research in the field of document analysis. For instance, in 1993, *Kanungo et al* (Kanungo *et al.,* 1993) introduced a statistical model for document degradation showing quite realistic results of document degradation. Kanungo gave three main reasons justifying research on it. First of all, noise modelling allows to study pattern recognition algorithms in general, as a function of the perturbation of the input data. Secondly, it permits the evaluation of any algorithm depending on the degradation level. Third, a knowledge of the degradation model can enable us to design algorithms for image restoration. More recently, in (Barney, 2008) has proposed an alternative noise modelling (called Noise Spread model) for document degradation.

Solutions to denoising problem in the binary case can be obtained by the use of general algorithms such as median filter (Davies, 1990), morphological filters (Maragos *et al.,* 1987), or curvelet transform (Starck *et al.,* 2002). Median filter replaces each pixel in the noisy image by the median of pixels in a neighborhood of that pixel, while morphological filter carries out dilation and erosion operations as the denoise operators. In addition, morphological filtering can discriminate between positive and negative noise spikes, whereas median filter cannot (Maragos *et al.,* 1987). Both filters, median and morphological, are appealing because they are easy to implement and perform well with the presence of impulse noise. However, neither of them are efficient for other types of noise, such as additive Gaussian noise (white noise), Noise Spread or Kanungo noise.

Recent researches obtain state-of-the-art performance in denoising gray-scale images with white noise using Multi-Resolution Analysis (MRA) methods (Starck *et al.,* 2002). Moreover, sparse transforms like a curvelet transform, has also been successfully applied to edge noise removing in bi-level graphical document images with Noise Spread, showing than sparse representation can effectively be used for denoising purposes (Hoang *et al.,* 2011). Sparse transforms represent images as linear combinations of (atom) functions of a given particular family of functions (dictionary). Nevertheless, sparse-based methods are constrained to the choice of these pre-defined basis functions (dictionary), e.g. curvelet, contourlets, wedgelets, bandelets, and the steerable wavelets and the overall performance of the denoising method depend on the *a priori* knowledge about images as well as the kind of noise affect them.

In this paper, we overcome this difficulty, i.e. the choice of suitable dictionaries, and take advantage of sparse representation for image denoising. More precisely, a learning algorithm is used to find out the proper set of atom functions adapted to the main characteristics of data and also adapted to the kind of noise for bi-level graphical images.

The remaining of this paper is organized as follows : Section 2 summarizes document degradation models and briefly describes the noise model proposed in (Kanungo *et al.,* 1993). Theoretical framework of sparse representation and the learning algorithm used for noise removal is presented in Section 3. Experimental results are discussed in Section 4 and, conclusions and further work are drawn in Section 5.

2. Document degradation model

In 2008, Barney proposed a degradation model inspired on the physics of image acquisition process (Barney, 2008). According to this model the acquired image is obtained as a result of convolving the source image with the sensor function (the Point Spread Function) and adding white noise. Some years before (Kanungo *et al.,* 1993, Kanungo *et al.,* 2000) proposed a statistical model for local distortions. This model uses six parameters as a function of the pixel distance to shape boundaries to degrade a binary image.

One of the advantages of degradation models is that they permit to generate degraded images controlled by parameter models. The qualitative results of these images ranges from quite realistic noisy images to unrealistic, or even highly degraded images in function of the parameters set up. For instance, in Figure [1], symbol images (b) and (d) provide more realistic symbol degradation than (c), (e), (f) and (g). Moreover, the Kanungo model has been applied to generate evaluation datasets used in each edition of symbol recognition and symbol location contests held during the GREC Workshops. Although, the Noise Spread model is simpler than the Kanungo model in terms of number of parameters, the disponibility of evaluation datasets for this noise make this model more suitable than the Noise Spread model.

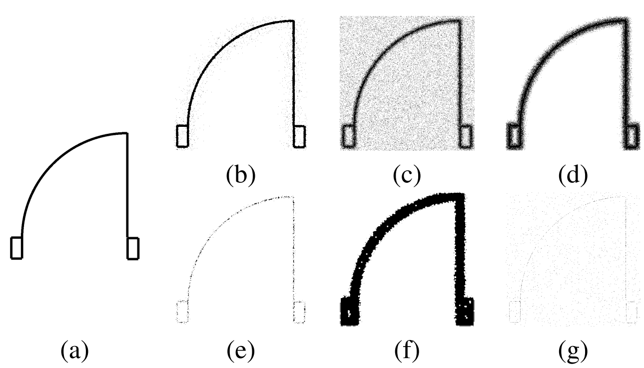


Figure 1. *(a) : original binary symbol ; from (b) to (g) examples of six levels of Kanungo noise of the GREC 2005 dataset. Observe than (b) and (d) symbol images provide more realistic symbol degradation than (c), (e), (f) and (g).*

Bi-level graphic images are represented by white (0) background pixels and black (1) foreground pixels representing the different entities of the document. The Kanungo model needs six parameters : α_0, α, β_0, β, η and k. α and α_0 to control the probability of flipping a foreground pixel to a background pixel or, in other words, these two parameters provide the probability of changing a black pixel to a white one. Similarly, the β and β_0 parameters control the probability of changing a background (white) pixel with a foreground (black) pixel. In addition, these two probabilities exponentially decay on the distance of each pixel to the nearest boundary pixel. In contrast, the η parameter is a constant value added to all pixels regardless their relative position to shape boundaries. Finally, the last parameter k is the size of the disk used in the morphological closing operation. The whole process of image degradation can be summarized in the following three steps :

1) Use standard distance transform algorithms to calculate the distance d of each pixel from the nearest boundary pixel.

2) Each foreground pixel and background pixel is flipped with probability $p(0|1, d, \alpha_0, \alpha) = \alpha_0 e^{-\alpha d^2} + \eta$, and $p(1|0, d, \beta_0, \beta) = \beta_0 e^{-\beta d^2} + \eta$

3) Use a disk structuring element of diameter k to perform a morphological closing operation.

3. Noise suppression using sparsity in learning dictionary

Sparse transforms and Multi-Resolution Analysis methods are applied to a wide range of image processing problems as image compression, image restoration and image denoising (Starck *et al.,* 2002). These methods have proven to perform well in terms of Mean Square Error (MSE) measure as well as peak signal-to-noise ratio (PSNR) measure with gray images where noise essentially is white (additive and following a Gauss distribution). Nevertheless, as we have already explained in the previous section, bilevel images of graphic documents suffer from other types of noise than white noise.

Recently, curvelet transform has also been applied in document denoising on bilevel graphical documents with a relative high degree of success (Hoang *et al.,* 2011). In that approach, they take advantage of directional properties of curvelet transform when it is applied to graphic document with the Noise Spread model. The results obtained in that work show an improvement in removing noise for graphic document images comparing with other state-of-the-art methods. However, neither of these methods take into account the informations of images to be denoised.

The main idea of the proposed method is to find a dictionary adapted to the properties of the data which will allow us to obtain a denoised version of the original degraded images.

In this Section, we will first review the basic of sparse representation methods. Then in the next subsection, we explain how to learn an optimal dictionary adapted to data. Finally, at the end of this section the denoising scheme is explained in detail.

3.1. *Sparse representation*

Sparse representation means to represent a signal like a linear combination of a few atoms of a given dictionary. Mathematically, given a dictionary Φ and a signal y, we consider the underdetermined linear system of equations $y = \Phi x$, with $\Phi = \phi_1, \phi_2, \ldots, \phi_m \in R^{n \times m}$, $y \in R^n$, $x \in R^m$, $m \gg n$. If Φ is a full$-$rank matrix, there will be infinitely many different sets of values for the x_i's that satisfy all equations simultaneously. The sets of such these x can be described using mathematical language. However, from the application point of view, one of the main tasks in dealing the above system of equations is to find the proper x that can 'explain' y well comparing with others. To gain this well-defined solution, a function $f(x)$ is added to assess the desirability of a would-be solution x, with smaller values being preferred :

$$(P_f) : \hat{f} = \arg\min_x f(x) \text{ subject to } \Phi x = y \quad [1]$$

If $f(x)$ is the l_0 pseudo-norm $\|x\|_0$ (number nonzero elements in vector x), then the problem (P_f) becomes finding the sparse representation x of y satisfying :

$$(P_0) : \hat{x} = \arg\min_x \|x\|_0 \text{ subject to } \Phi x = y \quad [2]$$

In general, solving equation [2] is often difficult (NP-hard problem) and one of the choices is to look for an approximate solution using greedy algorithms. Donoho proposed to substitute this problem by a convex relaxation instead (Donoho *et al.,* 2003) :

$$(P_1) : \hat{x} = \arg\min_x \|W^{-1}x\|_1 \text{ subject to } \Phi x = y \quad [3]$$

The matrix W is a diagonal positive-definite matrix, defined by $w(i,i) = 1/\|\phi_i\|_2$ [1]. In case of matrix Φ has the normalized columns, i.e. $W \equiv I$, (P_1) can be written as following

$$(P_1) : \hat{x} = \arg\min_x \|x\|_1 \text{ subject to } \Phi x = y \quad [4]$$

The equation [4] is the classic basis pursuit format and can be solved effectively by some existing numerical algorithms, such as Basis Pursuit by linear programming, IRLS (*Iterative Reweighed Least Squares*), LARS (*Least Angle Regression Stagewise*)

1. $\|x\|_2 \triangleq (\sum_{i=1}^m |x_i|^2)^{1/2}$, with $x \in R^m$

(Elad, 2010). If there exists some appropriate conditions on Φ and x, like $\|x\|_0 = k_0 \leq$ *spark*$(\Phi)/2$ [2], or

$$k_0 \leq \frac{1}{2}(1 + \frac{1}{\max_{i \neq j} \frac{|\phi_i^T \phi_j|}{\|\phi_i\|_2 \|\phi_j\|_2}})$$

then the solution of [4] is unique and also the unique solution of (P_0) .

Sometimes, the exact constraint $y = \Phi x$ is changed by relaxed one $\|\Phi x - y\|_2 \leq \epsilon$, with $\epsilon \geq 0$ is the error tolerance :

$$(P_1^\epsilon) : \hat{x} = \arg\min_x \|x\|_1 \text{ subject to } \|\Phi x - y\|_2 \leq \epsilon \qquad [5]$$

We can see one of advantages of this change through noise removal. Assuming that signal y has noise e with finite energy $\|e\|_2^2 \leq \epsilon^2$, $y = \Phi x + e$. Solving (P_1^ϵ) can help us to find the solution $\hat{x}$ that bases on it we can find the unknown denoised signal $\hat{y}$ by $\hat{y} = \Phi \hat{x}$. In fact, (P_1^ϵ) is known in the literature as *basis pursuit denoising*.

3.2. *Learned methodology for dictionary and K-SVD algorithm*

The effect in solving (P_1^ϵ) directly depends on a dictionary Φ. Naturally, if Φ is contructed by using a learning algorithm on a training database, then it can adapt better to new models of noise. So in this section, we review the score of learning methodology for constructing dictionary Φ as well as describe one of the learning algorithms, the K-SVD algorithm.

In a general learning methodology, a family l signals $\{y_j\}_{j=1}^l$ is considered as the training database. Our goal is to find a dictionary Φ in which each signal $y_j \in R^n$ has an 'optimally' sparse approximation $\bar{y}_j \backsimeq \Phi x_j$ satisfying $\|\bar{y}_j - y_j\|^2 \leq \epsilon$, or finding :

$$\min_{\Phi, x_j} \sum_{j=1}^{l} \|x_j\|_1 \text{ subject to } \|y_j - \Phi x_j\|_2 \leq \epsilon, \text{ for all } j = 1, .., l \qquad [6]$$

This dictionary can be obtained by the learning process. This process iteratively adjusts Φ via two main steps. In each iteration, first of all, all sparse representations $X = \{x_j\}_{j=1}^l \in R^{m \times l}$ of $Y = \{y_j\}_{j=1}^l \in R^{n \times l}$ are found by solving equation [5], on condition that Φ is fixed. Then, an updating rule is applied to optimize the sparse representation of all examples. In general, a way of updating the dictionary in the second step is different from each learning algorithm to others.

In K-SVD algorithm, proposed by (Aharon *et al.,* 2006), the updating rule is to make a modification on dictionary's columns. At this step, we handle to update se-

2. *spark*(Φ) : is the smallest number of columns that are linearly-dependent.

quentially each columns ϕ_{j_0} of Φ such that the residual error [7] is minimized, where X and $\{\phi_1, ...\phi_{j_0-1}, \phi_{j_0+1}, ..., \phi_m\}$ are fixed,

$$\begin{aligned}\|Y - \Phi X\|_F^2 &= \|Y - \sum_{j=1}^{m} \phi_j x_j^T\|_F^2 \\ &= \|(Y - \sum_{j \neq j_0} \phi_j x_j^T) - \phi_{j_0} x_{j_0}^T\|_F^2 \\ &= \|E_{j_0} - \phi_{j_0} x_{j_0}^T\|_F^2\end{aligned} \quad [7]$$

In this description [7], $x_{j_0}^T \in R^l$ is the kth row in X and the notation $\|.\|_F$ stands for the Frobenius norm. Because X and all the columns of Φ are fixed excepted the column ϕ_{j_0}, so $E_{j_0} = Y - \sum_{j \neq j_0} \phi_j x_j^T$ is fixed. It means that the minimum error $\|Y - \Phi X\|_F^2$ depends only on the optimal ϕ_{j_0} and $x_{j_0}^T$. This is the problem of approximating a matrix E_{j_0} with another matrix which has a rank 1 based on minimizing the *Frobenius* norm. The optimal solutions $\tilde{\phi}_{j_0}, \tilde{x}_{j_0}^T$ can be given by the SVD (*Singular Value Decomposition*) of E_{j_0} of rank r_1, namely

$$\tilde{\phi}_{j_0} \tilde{x}_{j_0}^T = Q_1 \tilde{U} Q_2$$

where $Q_1 = \{q_1^1, ..., q_n^1\}, U = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_{r_1}), Q_2 = \{q_1^2, ..., q_l^2\}$ is SVD of $E_{j_0} : E_{j_0} = Q_1 U Q_2$; and $\tilde{U}$ is the same matrix as U except that it contains only one singular values σ_1 (the other singular values are replaced by zero). This means $\tilde{\phi}_{j_0} = q_1^1$ and $\tilde{x}_{j_0}^T = \sigma_1 q_1^2$. However, the new vector $\tilde{x}_{j_0}^T$ is very likely to be filled, implying that we increase the number of non-zeros in the representation of X, or the condition about the sparsity of X can be broken.

This problem can be overcome as follows. Define the group of indices where $x_{j_0}^T$ is nonzero :

$$\omega_{j_0} = \{i | 1 \leq i \leq l, x_{j_0}^T(i) \neq 0\}.$$

and a matrix $\Omega_{j_0} \in R^{l \times |\omega_{j_0}|}$ is defined $\Omega_{j_0}(\omega_{j_0}(i), i) = 1$ and zeros elsewhere. Let

1) $x_{j_0}^R = x_{j_0}^T \Omega_{j_0}, x_{j_0}^R \in R^{|\omega_{j_0}|}$
2) $E_{j_0}^R = E_{j_0} \Omega_{j_0}, E_{j_0}^R \in R^{n \times |\omega_{j_0}|}$

and return to equation [7], this equation is equivalent to the minimization of

$$\|E_{j_0} \Omega_{j_0} - \phi_{j_0} x_{j_0}^T \Omega_{j_0}\|_F^2 = \|E_{j_0}^R - \phi_{j_0} x_{j_0}^R\|_F^2 \quad [8]$$

Note that the solution of [8] $\tilde{x}_{j_0}^R$ has the same support as the original $\tilde{x}_{j_0}^T$, and the optimal values $\tilde{x}_{j_0}^R, \tilde{\phi}_{j_0}$ can be obtained by finding SVD of a subset of the columns of the error matrix $E_{j_0}^R$ of rank $r_2 : E_{j_0}^R = SDV^T$. The solution for $\tilde{\phi}_{j_0}$ is defined as the first column of S, and the coefficient vector $\tilde{x}_{j_0}^R$ as the first column of V multiplied by d_1, with $D = \text{diag}(d_1, d_2, ..., d_{r_2})$. More detail about the K-SVD algorithm can be found in algorithm [1]

Algorithm 1 Learning algorithm K-SVD

INPUT : $\Phi_{(0)} \in R^{n\times m}$; $\{y_i\}_{i=1}^{l}$; $k = 0$;
1. Initialize : Normalization the columns of matrix $\Phi_{(0)}$;
2. Main Iteration
$k = k + 1$;
while ($\|Y - \Phi_{(k)} X_{(k)}\|_F^2$ is not small enough) **do**
  - Solve (P_1^{ϵ}) to find all sparse representation $\{x_i\}_{i=1}^{l}$ of $\{y_i\}_{i=1}^{l}$
  for ($j_0 = 1$ to m) **do**
    - Define : $\omega_{j_0} = \{i|1 \leq i \leq l, x_{j_0}^T(i) \neq 0\}$
    - Calculate E_{j_0} via equation $E_{j_0} = Y - \sum_{j \neq j_0} \phi_j x_j^T$
    - Let $E_{j_0}^R$ is the submatrix of E_{j_0} on the columns corresponding to ω_{j_0}
    - Calculate SVD of $E_{j_0}^R$: $E_{j_0}^R = SDV^T$.
    - Updating : $\phi_{j_0} = s_1$ and $x_{j_0}^R = d_1 v_1$ with $S = \{s_1, ..., s_n\}$, $V = \{v_1, ..., v_{|\omega_{j_0}|}\}$, $D = \text{diag}(d_1, d_2, ..., d_{r_2})$
  end for
end while
OUTPUT : The result $\Phi_{(k)}$

3.3. *Denoising by learned dictionary*

Denoising images using basis pursuit denoising (Eq.[5]), we need to decide which kind of dictionary Φ is used as well as what is the best value ϵ. From Section 3.2 we know how to learn a dictionary Φ adapted to noisy data. In this Section, we explain how to choose the best value ϵ.

Denoising in MRA methods assumes that images have been corrupted by an additive white noise. For such noised images, the energy of noise ϵ is proportional to both the noise variance and size image. For Noise Spread model, authors empirically found the optimal ϵ depend on the noise spread relation (see (Hoang *et al.,* 2011) for futher details). However, neither of these two criterias can be applied to Kanungo model. The reason is that in Kanungo model pixels near the shape boundaries have higher probability to be affected by noise than pixels far from the boundaries, while the white noise model assumes statistical independence between noise and image. In fact, the probability that a pixel is perturbed by noise depends only on the variance of model and does not depend on the position of pixel itself in the image. In addition, the parameters used in Noised Spread model are different the Kanungo's ones, so we can not use the noise spread relation to decide the value for ϵ.

We proposed one way to evaluate the level of noise according to the Kanungo model using the peak values of the normalized cross-correlation between noised and clean images (Lewis, n.d., Haralick *et al.,* 1992). Define r_i as the peaks of normalized cross-correlation between a noisy image I_i^n and a clean version of it I_i^c (see Figure [2]). The mean of all values $r_i, i = 1, \ldots, t$ is defined by :

$$\bar{r} = \frac{1}{t}\sum_{i=1}^{t} r_i \qquad [9]$$

Alternatively, the parameter threshold ϵ is proportional to $\bar{r}$, $\epsilon = cq\bar{r}$, in which q is the size of extracted patch and c is set by experiment in the interval $[0.4, 1]$.

To train a learned dictionary using K-SVD algorithm, we need to firstly collect some data for training. In fact, in our paper, all patches of the corrupted image y are used as training data. We used a sliding window of size $q \times q$ to scan corrupted image y. The scanning step is set to 1 pixels in both directions. So, if image y belongs to $R^{M\times N}$, then $(M - q + 1)(N - q + 1)$ noised patches are used as training data.

The procedure for denoising symbols by using sparsity in learning dictionary proceeds as follows :

1. *Create training database* : Using a sliding window of size $q \times q$ to scan the corrupted image y. All the obtained patches $\{y_j\}_{j=1}^{l}$, $y_j \in R^{q^2}$ is considered a training database.

2. *Create learned dictionary* : Using K-SVD algorithm to create the learned dictionary Φ from $\{y_j\}_{j=1}^{l}$

3. *Denoising image* : Combining learned dictionary Φ to sparse representation model with the purpose of denoising image y, or

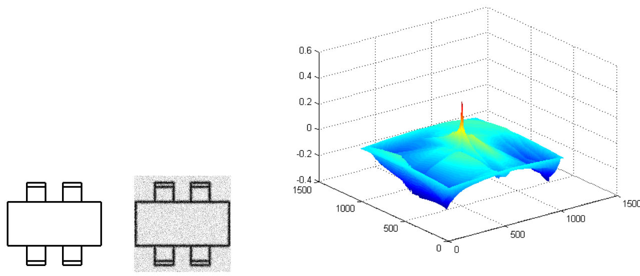
   a. Find the solution of the optimization problem [5] for each patch y_j, we get
   $$\hat{p_j} = \arg\min \|p_j\|_1 \text{ subject to } \|\Phi p_j - y_j\|^2 \leq \epsilon$$
   b. Compute the denoised version of each patch y_j by $\hat{y_j} = \Phi\hat{p_j}$
   c. Merge the denoised patches $\hat{p_j}$ to get the denoised image $\bar{y}$
   d. Binarise $\bar{y}$ to get the final binary denoised image $\hat{y}$

4. Experimental results

We evaluated our algorithm on the GREC 2005 dataset. This dataset has 150 different symbols which have been degraded using the Kanungo's method to simulate the noise introduced by the scanning process. Six sets of parameters are used to obtain six different noise levels as shown in Figure [1]. Each level contains 50 noised symbols.

At each level of noise, the value of $\bar{r}$ is calculated based on Equation [9]. Figure [2] shows the normalized cross-correlation of two images (a) and (b). The maximum value of r_i is 0.4106.

In addition, by using $t = 50$ images for each level of noise, we found the best values of $\bar{r}$ corresponding to each level of degradation as shown in Table [1]. We also



(a) (b) (c)

Figure 2. *Normalized cross-correlation between 2 images*

Level	1	2	3	4	5	6
$\bar{r}$	0.9133	0.4412	0.5629	0.3698	0.4413	0.2006

Table 1. *The value of $\bar{r}$ in accord with six level of noise*

empirically found that the best results in denoising bilevel images are achieved when $\epsilon = cq\bar{r}$, where q is the size of patch, and c belongs to $[0.4, 1]$.

The learning dictionary was produced using K-SVD algorithm with 50 iterations, and the training dataset including all 8×8 patches. Those patches were taken from a corrupted image y. The ratio of the dictionary is 1/4 ($m = 4 \times n$).

To examine the effectiveness of the proposed method, we compare it with three existing methods for denoising binary images : median filtering, opening then closing, and curvelet tranform. The median filtering is performed with a window size of 3×3. The opening then closing uses a 3×3 structuring element. Curvelet transform has been verified upon the best value of $\epsilon = c \times \sqrt{MN} \times \sigma^2$, $\sigma \in [0.02, 0.1]$ for each noise level , where M, N is the size of the image. The criterion to choose these best values σ is the average MSE (*Mean Squared Error*).

$$\text{MSE} = \frac{1}{\sqrt{MN}} \sum_{i=1}^{M} \sum_{j=1}^{N} (y_0(i,j) - \hat{y}(i,j))^2 \qquad [10]$$

Moreover, using the traditional MSE (Equation [10]) to estimate the quality change between the original image $y_0 \in R^{M \times N}$ and the reconstructed image $\hat{y} \in R^{M \times N}$ (defined in equation [10]), all algorithms are evaluated by *Jaccard*'s similarity measure. This measure is computed based on the three values a, b, c as below :

$$S = \frac{a}{a+b+c} \qquad [11]$$

where

$$a = |\{(i,j)|y(i,j) = 0, \hat{y}(i,j) = 0, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

$$b = |\{(i,j)|y(i,j) = 1, \hat{y}(i,j) = 0, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

$$c = |\{(i,j)|y(i,j) = 0, \hat{y}(i,j) = 1, 1 \leq i \leq M, 1 \leq j \leq N\}|$$

a means '*positive matches*', and b, c mean '*mismatches*'.

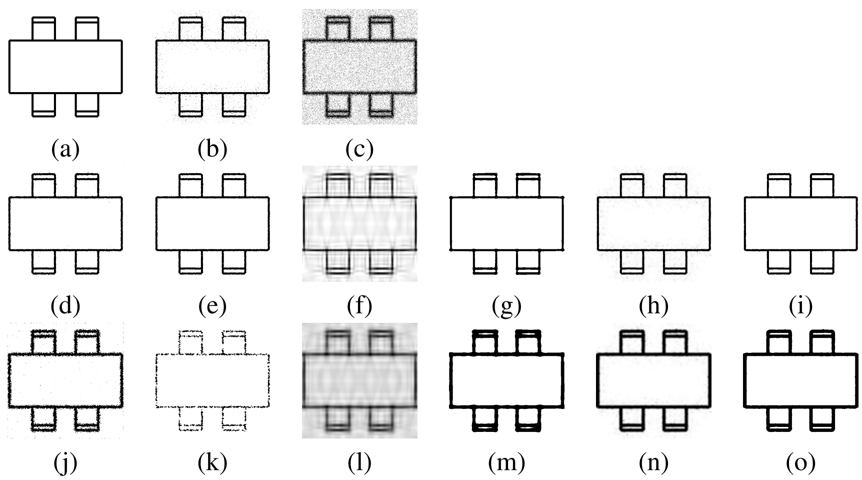


Figure 3. *(a) Original image, (b)-(c) : noised image at level 1, 2 of degradation with Kanungo noise. (d)-(j), (e)-(k), (f)-(l), (h)-(n) are denoised images by Median filter, opening then closing, curvelet transform and learned dictionary in accord with two levels. (g)-(m), (i)-(o) are binarized denoised images of (f)-(l) and (h)-(n)*

Table [2] shows results of denoising on ten classes of symbols with two level of noise (a) and (b) using the different algorithms where (a) and (b) stand for the level 1 and level 2, respectively. The results are evaluated by MSE and Jaccard's measure. The best obtained results for (a) by MSE and Jaccard's measure are in red and green, respectively ; and the best obtained results for (b) are showed in bold font.

Generally, Table [2] shows that our method achieves better results comparing with other methods at both level (a) and (b) of degradation. At level (b), the error values of our method are lower than others' for the MSE criterion. However, for level (a), the results of our method are quite similar with Curvelet method using MSE and Jaccard's measure, 1.482, 0.9609 in comparing with 1.6094 and 0.9589, respectively. So, to check whether the difference between the results obtained by our method and the ones obtained by the other methods is significant, we perform a paired Wilcoxon signed test with a significance level of 5%.

Table [3] and [4] show the average results obtained by the four methods on 6 levels of noise that are respectively evaluated by MSE and *Jaccard*'s measure. In these

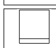
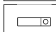
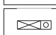
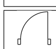
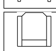
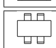
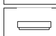
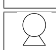
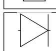
Symbols	MSE								Jaccard's measure							
	Median		OC		Curvelet		Proposed method		Median		OC		Curvelet		Proposed method	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
	2.8887	22.1582	2.1374	33.1523	0.5534	29.1777	1.0202	**20.6354**	0.9299	0.6358	0.9467	0.1726	0.9863	0.5785	0,9747	**0.6592**
	1.9473	15.5039	1.5107	22.2402	1.0967	22.7988	0.8789	**13.6172**	0.9295	0.6257	0.9441	0.1744	0.9607	0.5414	0.9676	**0.6630**
	3.9206	27.3984	3.6126	39.6582	2,9382	38.8574	2.7702	**25.7852**	0.9195	0.6242	0.9248	0.1643	0.9413	0.5496	0.9428	**0.6468**
	2.7702	15.6582	2.7552	24.1665	2.3132	21.9902	1.9297	**14.3384**	0.9200	0.6437	0.9305	0.1801	0.9437	0.5726	0.9517	**0.6716**
	2.0674	15.1836	1,7935	22.2949	1.2637	21.4375	1.2319	**13.3223**	0.9230	0.6216	0.9320	0.1465	0.9534	0.5492	0.9536	**0.6611**
	4.0371	29.5938	3.3633	44.4590	1.2529	38.6289	1.6123	**29.3594**	0.9284	0.6414	0.9390	0.1913	0.9775	0.5871	0.9709	**0.6512**
	3.6221	27,2002	2.8418	42.0830	2.6196	37.1270	1.4966	**26.0410**	0.9324	0,6500	0.9458	0.1949	0.9515	0.5845	0.9716	**0.6667**
	2.0710	16.3333	1.6361	23.8281	0.6393	23.3164	0.7285	**14.6536**	0.9300	0.6304	0.9434	0.1749	0.9782	0.5532	0.9749	**0.6626**
	2.4600	17.7080	2.3516	24.7490	2.2441	25.0537	1.7637	**16.4609**	0.9205	0.6199	0.9232	0.1797	0.9302	0.5462	0.9428	**0.6455**
	2.7253	19.4909	2,5020	28.2539	1.1725	27.1602	1,4297	**17.9434**	0.9217	0.6258	0.9272	0.1673	0.9658	0.5552	0.9584	**0.6528**
Average	2.8972	20.6229	2.45041	30.4885	1.6094	28.5548	**1.4862**	**19.2157**	0.9255	0.5677	0.9357	0.1555	0.9589	0.5030	**0.9609**	**0.5929**

Table 2. *Summary of the denoising results in GREC 2005. (a), (b) are level 1, level 2 of noise, respectively.*

	Median	OC	Curvelet	Proposed method
Level 1	0.9264 (-)	0.9368 (-)	0.9632 (=)	**0.9633**
Level 2	0.6302 (-)	0.1766 (-)	0.5649 (-)	**0.6551**
Level 3	0.4279 (-)	0.7871 (-)	0.3889 (-)	**0.8258**
Level 4	0.0878 (-)	0.0011 (-)	**0.4589** (+)	0.1348
Level 5	0.2610 (-)	0.2680 (-)	0.2629 (-)	**0.2844**
Level 6	0.0014 (-)	0.0000 (-)	0.0594 (=)	**0.0604**

Table 3. *Average value gained by Jaccard's similarity measure.*

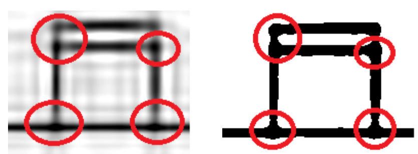
tables, an entry mark by $(-)$ indicates that the corresponding method performs worst than our method. Similarly, an entry marked by $(+)$ indicates that the corresponding method outperforms the proposed method, and an entry marked by $(=)$ indicates that results obtained by the two methods are not significantly different.

Table [3] and [4] also show that at level 5 and level 6, none of the four methods is good enough but other methods are worse than ours. We further examine the set of noised images at level 4 and we found that the set of noised patches p_i of the corrupted image y cannot provide the good training data. Most of patches p_i is trivial (zeros value), making not enough discrimination between $\Phi_{(0)}$ and $\Phi_{(k)}$ (Algorithm 1). This can explain why the curvelet transform method is better than the proposed method at this level of noise. At level 1, although the Wilcoxon signed test indicates that results

	Median	OC	Curvelet	Proposed method
Level 1	2.8956 (-)	2.4543 (-)	**1.4032** (=)	1.4230
Level 2	21.0147(-)	30.8649 (-)	28.4205 (-)	**19.6873**
Level 3	53.6900 (-)	8.8996 (-)	57.7709 (-)	**8.0082**
Level 4	33.5986 (-)	36.8191 (-)	**19.8148** (+)	31.8805
Level 5	111.6315 (-)	107.3583 (-)	110.2861 (-)	**98.2124**
Level 6	37.8629 (-)	37.9154 (-)	**35.6315** (=)	35.7717

Table 4. *Average values gained by MSE*

obtained by the curvelet and our method are not significantly different. However, when we zoom the denoised images we found that the intersection of edges are not well restored with the Curvelet as shown in Figure [4]. Therefore, to confirm the relative advantages of the proposed method, we check with the R-signature (Tabbone *et al.*, 2006) the similarity between the original image and the reconstructed ones (Figure [5]).



(a) (b)

Figure 4. *A zoom of denoised image by curvelet.*

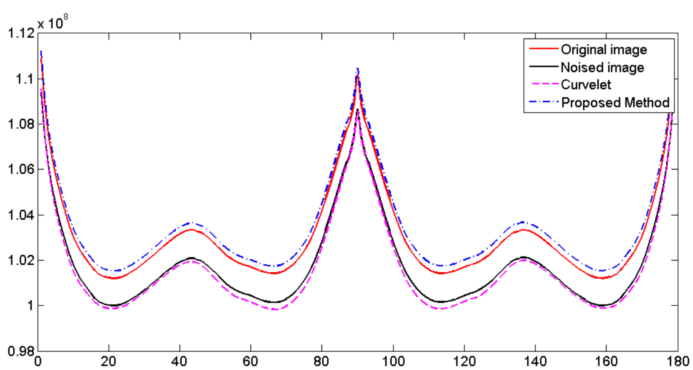


Figure 5. *R-signature of images in figure 3(a), 3(b), 3(g) and 3(i)*

The plot demonstrates that the proposed method give a denoised image that is closest to the original image.

5. Conclusions

A novel algorithm for denoising bilevel graphical document images by using learning dictionary based on sparse representation has been presented in this paper. Learning method starts by building a training dababase from corrupted images, and constructing an empirically learned dictionary by using sparse representation. This dictionary can be used as a fixed dictionary to find the solution of the basis pursuit denoising problem. In addition, we provide a way to define the best value of ϵ based on a measure of fidelity between two images. The efficiency of ϵ has been also aproved experimentally on GREC 2005 dataset. All experimental results, that are evaluated using MSE, and *Jaccard*'s binary similarity measure, as well as R-signature show that our method outperforms existing ones in most of cases.

6. Bibliographie

Aharon M., Elad M., Bruckstein A., « K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation », *IEEE Transactions on signal processing*, vol. 54 (11), p. 4311-4322, 2006.

Barney E., « Modeling Image Degradations for Improving OCR », *European Conference on Signal Processing*, p. 1-5, 2008.

Davies E., *Machine Vision : Theory, Algorithms and Practicalities*, Academic Press, 1990.

Donoho D., Elad M., « Optimally sparse representation in general (nonorthogonal) dictionaries via ell1 minimization », *PNAS*, vol. 100 (5), p. 2197-2202, 2003.

Elad M., *Sparse and redundant representation : From theory to applications in signal and images processing*, Springer, 2010.

Haralick R.-M., Shapiro L.-G., *Computer and Robot Vision (Volume I)*, Addison-Wesley Publishing Company, 1992.

Hoang V.-T., Smith E. B., Tabbone S., « Edge noise removel in bilevel graphical document images using sparse representation », *IEEE international conference on Image Processing - ICIP 2011*, 2011.

Kanungo T., Haralick R., Baird H., Stuezle W., Madigan D., « A statistical, nonparametric methodology for document degradation model validation », *IEEE Transactions on PAMI*, vol. 22(11), p. 1209-1223, 2000.

Kanungo T., Haralick R.-M., Phillips I. T., « Global and local document degradation models », *IEEE*p. 730-734, 1993.

Lewis J. P., « Fast Normalized Cross-Correlation », *Industrial Light & Magic*, n.d.

Maragos P., Schafer R., « Morphological Filters, Part 2 : Their Relations to Median, Order-Statistic, and Stack Filters », *IEEE Transactions on acoustics, speech, and signal processing*, vol. ASSP-35 (8), p. 87-134, 1987.

Starck J.-L., E.J.Candes, Donoho D., « The curvelet transform for Image denoising », *IEEE Transactions on image processing*, vol. 11(6), p. 670-684, 2002.

Tabbone S., Wendling L., Salmon J.-P., « A new shape descriptor defined on the Radon transform », *Computer Vision and Image Understanding*, vol. 102, p. 42-51, 2006.