

## An Overview on Dictionary and Sparse Representation in Image Denoising

Mrs.S.Subha, Mr. I.Jesudass, Mrs.S.Jeyasree

Assistant Professor(Sr. Gr.) KLNCE, Pottapalayam, Madurai.

PG KLNCE, Pottapalayam, Madurai.

PG KLNCE, Pottapalayam, Madurai.

---

**Abstract:** The goal of natural image denoising is to estimate a clean version of a given noisy image, utilizing prior knowledge on the statistics of natural images. Noise removal from natural images is a challenging task. Image denoising is an applicable issue for image processing and computer vision problems. There are several existing methods are available for image denoising. A most applicable and expected property of an image denoising is that it should totally remove the noise as well as its preserve edges. This paper represents the review of parameter and algorithms available for image denoising.

**Index Terms:** Image noise, sparse, over-complete dictionary, Redundancy parameters.

---

### I. Introduction

A digital image is an array of real or complex numbers represented by a finite number of bits. Processing of an image involves improvement in its appearance and efficient representation. It involves extraction, analysis and recognition of images and also coding, filtering, enhancement and restoration of the image.

Generally Image processing is divided into three major stages. There are

- (i) Discrimination and Representation,
- (ii) Processing,
- (iii) Analysis.

**(i) Discrimination and Representation:**

It involves conservation of an image into discrete form and approximating them then save it in stored area.

**(ii) Processing:**

It is for improving the image Quality by filtering method and compressing the data to reduce the storage area.

**(iii) Analysis:**

Analysis involves extracting the image and recognition.

### II. Noise Model

**(a) ADDITIVE WHITE GAUSSIAN NOISE (AWGN):**

Noise is present in image either in Additive or multiplicative form. Here we address the image denoising problem, where zero-mean white and homogeneous Gaussian additive noise. Gaussian noise is created in a image by the factors such as electronic circuit noise and sensor noise, which are caused by poor illumination and high temperature.

$$y = x + n \quad (1)$$

Where y is the observed noise image, x is the original image and n is the AWGN noise.[2]

**(b) IMPLUSE NOISE:**

Impulsive noise is common in images which arise at the time of image acquisition and or transmission of images. Impulsive noise can be classified into two categories, namely Salt & Pepper Noise (SPN) and Random Valued Impulsive Noise (RVIN).[1]

**(1) Salt and pepper noise:**

Each pixel in an image has the probability of s/2 (0<s<1) being contaminated by either a white dot (salt) or a black dot (pepper).[2]

$$y = d_{min} \text{ and } y = d_{max} \quad (2)$$

**(2) Random value impulse noise:**

Each noise pixel in a image take any value within the range minimal to maximum.[2]

$$y = d_{ij} \quad (3)$$

Where y is uniformly distributed  $[d_{min}, d_{max}]$

**(c)PROPERTY:**

It is easily traceable and controllable in both Spatial and Frequency domain.[14]

**(1) Spatial domain:**

In spatial domain method, the pixel of images is manipulated directly. Spatial domain process may be expressed as,

$$g(x,y) = T[f(x,y)] \quad (4)$$

where,

$f(x,y)$  - input image

$g(x,y)$  – processed image

T is an operator on ‘f’ defined over some neighborhood of (x,y).

**(2) Frequency domain:**

Frequency domain based on the modifying the Fourier transforms of an image. (i.e.) It is based on convolution theorem. Let  $g(x,y)$  be an image formed by the convolution of an image  $f(x,y)$  and a linear position invariant operator  $h(x,y)$ .

$$g(x,y) = h(x,y) * f(x,y) \quad (5)$$

From convolution theorem,

$$G(u,v) = H(u,v) \cdot F(u,v) \quad (6)$$

Output of an image

$$g(x,y) = F^{-1}[H(u,v) \cdot F(u,v)] \quad (7)$$

**III. Dictionary And Sparse Representation**

Dictionary Learning is a topic in the Signal Processing area, the dictionary is usually used for Sparse Representation or Approximation of signals. A dictionary is a collection of atoms; here the atoms are real column vectors of length  $N$ . A finite dictionary of  $K$  atoms can be represented as a matrix  $D$  of size  $N \times K$ . In a *Sparse Representation* a vector  $X$  is represented or approximated as a linear combination of some few of the dictionary atoms. The approximation  $X\alpha$  can be written as

$$X\alpha = Dw \quad (8)$$

Where  $w$  is a vector containing the coefficients and most of the entries in  $w$  are zero. Dictionary Learning is the problem of finding a dictionary such that the approximations of many vectors, the training set, are as good as possible given a sparseness criterion on the coefficients, i.e. allowing only a small number of non-zero coefficients for each approximation.

**(i)Sparse approximation**

**(a)Noiseless observations:**

Consider a linear system of equations  $x = D\alpha$ , where  $D$  is an underdetermined  $m \times p$  matrix ( $m \ll p$ ) and  $x \in \mathbb{R}^m, \alpha \in \mathbb{R}^p$ .  $D$ , called as the dictionary or the design matrix, is given. The problem is to estimate the signal  $\alpha$ , subject to the constraint that it is sparse. The underlying motivation for sparse decomposition problems is that even though the observed values are in high-dimensional ( $m$ ) space, the actual signal is organized in some lower-dimensional subspace ( $k \ll m$ ). Sparsity implies that only a few components of  $\alpha$  are non-zero and the rest are zero. This implies that  $x$  can be decomposed as a linear combination of only a few  $m \times 1$  vectors in  $D$ , called atoms.  $D$  itself is over-complete ( $m \ll p$ ). Such vectors are called as the basis of  $X$ . However, unlike other dimensionality reducing decomposition techniques such as Principal Component Analysis, the basis vectors are not required to be orthogonal.

The sparse decomposition problem is represented as,

$$\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \text{ such that } x = D\alpha,$$

Where

$\|\alpha\|_0 = \#\{i : \alpha_i \neq 0, i = 1, \dots, p\}$  is a pseudo-norm,  $l_0$ , which counts the number of non-zero components of  $\alpha = [\alpha_1, \dots, \alpha_p]^T$ . This problem is NP-Hard with a reduction to NP-complete subset selection problems in combinatorial optimization. A convex relaxation of the problem can instead be obtained by taking the  $l_1$  norm instead of the  $l_0$  norm, where

$$\|\alpha\|_1 = \sum_{i=1}^p |\alpha_i| \quad (9)$$

The  $l_1$  norm induces sparsity under certain conditions.[6]

**(b) Noisy observations:**

Often the observations  $x$  are noisy. By imposing a  $l_2$  norm on the data-fitting term and relaxing the equality constraint, the sparse decomposition problem is given by,

$$\min_{\alpha \in \mathbb{R}^p} \frac{1}{2} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (10)$$

Where  $\lambda$  is a slack variable and  $\|\alpha\|_1$  is the sparsity-inducing term. The slack variable balances the trade-off between fitting the data perfectly, and employing a sparse solution.[6]

**IV. Algorithms Used For Sparse Representations**

There are several algorithms that have been developed for solving sparse approximation problem.

**(a) Matching pursuit**

Matching pursuit is a greedy iterative algorithm for approximately solving the original  $l_0$  pseudo-norm problem. Matching pursuit works by finding a basis vector in  $D$  that maximizes the correlation with the residual (initialized to  $x$ ), and then recompiling the residual and coefficients by projecting the residual on all atoms in the dictionary using existing coefficients. Matching pursuit suffers from the drawback that an atom can be picked multiple times which is addressed in orthogonal matching pursuit.[7]

**(b) Orthogonal matching pursuit**

Orthogonal Matching Pursuit is similar to Matching Pursuit, except that an atom once picked, cannot be picked again. The algorithm maintains an active set of atoms already picked, and adds a new atom at each iteration. The residual is projected on to a linear combination of all atoms in the active set, so that an orthogonal updated residual is obtained. Both Matching Pursuit and Orthogonal Matching Pursuit use the norm. [6]

**(c) LASSO**

LASSO method solves the norm version of the problem. In LASSO, instead of projecting the residual on some atom as in Matching Pursuit, the residual is moved by a small step in the direction of the atom iteratively.

**(d) Projected Gradient Descent**

Projected Gradient Descent methods operate in a similar fashion with the Gradient Descent: the current gradient provides the information to point to new search directions. Since we are looking for a sparse solution, the putative solutions are projected onto the sparse scaffold of  $k$  vectors.

**V. Total variation denoising**

In signal processing, total variation denoising, also known as ‘total variation regularization’ is a process, most often used in digital image processing, that has applications in noise removal. It is based on the principle that signals with excessive and possibly spurious detail have high total variation, that is, the integral of the absolute gradient of the signal is high. According to this principle, reducing the total variation of the signal subject to it being a close match to the original signal, removes unwanted detail whilst preserving important details such as edges.[18]

This noise removal technique has advantages over simple techniques such as linear smoothing or median filtering, which reduce noise but at the same time smooth away edges to a greater or lesser degree. By contrast, total variation denoising is remarkably effective at simultaneously preserving edges whilst smoothing away noise in flat regions, even at low signal-to-noise ratios.[19]

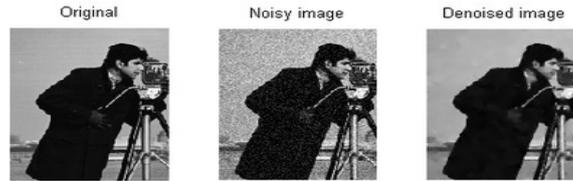


Fig.1 Total variation denoising technique to an image corrupted by Gaussian noise.

**(a)Application**

Total variation can be seen as a non-negative real-valued functional defined on the space of real-valued functions (for the case of functions of one variable) or on the space of integrable functions (for the case of functions of several variables). As a functional, total variation finds applications in several branches of mathematics and engineering, like optimal control, numerical analysis and calculus of variations, where the solution to a certain problem has to minimize its value.

**VI. Dictionary Training:**

Dictionary training is a much more recent approach to dictionary design, and as such, has been strongly influenced by the latest advances in sparse representation theory and algorithms. The most recent training methods focus on  $\ell_0$  and  $\ell_1$  sparsity measures, which lead to simple formulations and enable the use of recently developed efficient sparse-coding techniques. The main advantage of trained dictionaries is that they lead to state-of-the-art results in many practical signal processing applications.

A common setup for the dictionary learning problem starts with access to a training set, a collection of training vectors, each of length  $N$ . This training set may be finite, and then the training vectors are usually collected as columns in a matrix  $X$  of size  $N \times n$ , or it may be infinite. For the finite case the aim of dictionary learning is to find both a dictionary,  $D$  of size  $N \times k$ , and a corresponding coefficient matrix  $W$  of size  $k \times n$  such that the representation error,  $R=X-DW$ , is minimized and  $W$  fulfill some sparseness criterion.

**(a) Method of Optimal Directions**

The Method of Optimal Directions (MOD) was introduced by Egan et al. in 1999 [10], [11], and was one of the first methods to implement what is known today as a sparsification process. Given a set of examples  $X = [x_1, x_2, \dots, x_n]$ , the goal of the MOD is to find a dictionary  $D$  and a sparse matrix  $\Gamma$  which minimize the representation error,

$$\underset{D, \Gamma}{\text{Argmin}} \|X - D\Gamma\|_F^2 \quad \text{Subject To} \quad \|\gamma_i\|_0 \leq T \quad \forall i$$

The MOD typically requires only a few iterations to converge, and is overall a very effective method. The method suffers, though, from the relatively high complexity of the matrix inversion. Several subsequent works have thus focused on reducing this complexity, leading to more efficient methods.

**(b)Union of Orthobasis:**

Training a union-of-orthobasis dictionary was proposed in 2005 by Lesage et al. [12] as a means of designing a dictionary with reduced complexity and which could be more efficiently trained. The process also represents one of the first attempt training a structured over-complete dictionary — a tight frame in this case. The model suggests training a dictionary which is the concatenation of  $k$  orthogonal basis, so  $D = [D_1 D_2 \dots D_k]$  with the  $\{D_i\}$  unitary matrices. Sparse-coding over this dictionary can be performed efficiently through a Block Coordinate Relaxation (BCR) technique [13]. A drawback of this approach is that the proposed model itself is relatively restrictive, and in practice it does not perform as well as more flexible structures. Interestingly, there is a close connection between this structure and the more powerful. Generalized PCA model, described next. As the GPCA model deviates from the classical sparse representation paradigm, identifying such relations could prove valuable in allowing the merge of the two forces.

**(c) Generalized PCA**

Generalized PCA, introduced in 2005 by Vidal, Ma and Sastry [13], offers a different and very interesting approach to over complete dictionary design. The GPCA view is basically an extension of the original PCA formulation, which approximates a set of examples by a low-dimensional subspace. In the GPCA setting, the set of examples is modeled as the union of several low-dimensional subspaces perhaps of unknown number and variable dimensionality and the algebraic geometric GPCA algorithm determines these subspaces and fits orthogonal bases for them. The GPCA viewpoint differs from the sparsity model described in (2), as

each example in the GPCA setting is represented using the atoms corresponding to only one of the subspaces; thus, atoms from different subspaces cannot jointly represent a signal. This property has the advantage of limiting over-expressiveness of the dictionary, which characterizes other over complete dictionaries; on the other hand, the dictionary structure may be too restrictive for more complex natural signals. A unique property of the GPCA is that as opposed to other training methods, it can detect the number of atoms in the dictionary in certain settings. Unfortunately, the algorithm may become very costly this way, especially when the amount and dimension of the subspaces increases. Indeed, intriguing models arise by merging the GPCA viewpoint with the classical sparse representation viewpoint: for instance, one could easily envision a model generalizing (6) where several distinct dictionaries are allowed to co-exists, and every signal is assumed to be sparse over exactly one of these dictionaries.

**(d) The K-SVD Algorithm**

The desire to efficiently train a generic dictionary for sparse signal representation led Aharon, Elad and Bruckstein to develop the K-SVD algorithm in 2005 [15]. The algorithm aims at the same sparsification problem as the MOD, and employs a similar block-relaxation approach. The main contribution of the K-SVD is that the dictionary update, rather than using a matrix inversion, is performed atom-by-atom in a Simple and efficient process. Further acceleration is provided by updating both the current atom and its associated sparse coefficients simultaneously. The result is a fast and efficient algorithm which is notably less demanding than the MOD. The K-SVD algorithm takes its name from the Singular Value-Decomposition (SVD) process that forms the core of the atom update step and which is repeated K times, as the number of atoms. For a given atom k, the quadratic term in (6) is rewritten as

$$\|X - \sum_{j \neq k} d_j \gamma_j^T - d_k \gamma_k^T\|_F^2 = \|E_k - d_k \gamma_k^T\|_F^2 \quad (11)$$

Where  $\gamma_j^T$  are the rows of  $\Gamma$ , and  $E_k$  is the residual matrix. The atom update is obtained by minimizing for  $d_k$  and  $\gamma_j^T$  via a simple rank-1 approximation of  $E_k$ . To avoid introduction of new non-zeros in  $\Gamma$ , the update process is performed using only the examples whose current representations use the atom  $d_k$ .

The K-SVD, as well as the MOD, suffers from a few common weaknesses. The high non-convexity of the problem means that the two methods will get caught in local minima or even saddle points. Also, the result of the training is a nonstructural dictionary which is relatively costly to apply, and therefore these methods are suitable for signals of relatively small size, such as image patches. In turn, in recent years several parametric dictionary training methods have begun to appear, and aim to address these issues by importing the strengths of analytic dictionaries to the world of example based methods.

**VII. Conclusion**

Dictionary design has significantly evolved over the past decades; beginning with simple orthogonal transforms and leading to the complex over-complete analytic and trained dictionaries. Substantial conceptual advancement has been made in understanding the elements of an efficient dictionary design. However, with a wealth of tools already developed, much work remains to be done; indeed, the various components have yet to be neatly merged into a single efficient construct. Many future research directions have been mentioned in the text, and demonstrate the viability and vividness of the field as well as the large number of challenges that still await. Of specific interest, we highlight the strong need for a multi-scale structured dictionary learning paradigm, as well as methods to use such dictionaries in applications, which will clearly be the focus of much research in the near future.

**References**

[1] Jieliang Jiang, Lei Zhang, and Jian Yang, "Mixed Noise Removal by Weighted Encoding With Sparse Nonlocal Regularization", IEEE transactions on image processing, vol. 23, no. 6, June 2014.

[2] Wangmeng Zuo, Lei Zhang, Member in IEEE, Chunwei Song, David Zhang, Fellow, IEEE, and Huijun Gao "Gradient Histogram Estimation and Preservation for Texture Enhanced Image Denoising" IEEE trans. Image process, vol.23, no 6, June 2014.

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

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," Technical Report – Statistics, Stanford, 1995.

[5] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," SIAM Review, vol. 51, no. 1, pp. 34–81, 2009.

[6] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," 1993 Conference Record of The 27th Asilomar Conference on Signals, Systems and Computers, pp. 40–44, 1993.

- [7] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, 1997.
- [8] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," 2007. Submitted.
- [9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [10] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, pp. 2443–2446, 1999.
- [11] K. Engan, B. D. Rao, and K. Kreutz-Delgado, "Frame design using FOCUSS with method of optimal directions (MOD)," *Proc. Norwegian Signal Processing Symposium*, pp. 65–69, 1999.
- [12] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, pp. 293–296, 2005.
- [13] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.
- [14] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, 2005.
- [15] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [16] Donoho, D.L. (2006). "For most large underdetermined systems of linear equations the minimal  $l_1$ -norm solution is also the sparsest solution". *Communications on pure and applied mathematics (Wiley Online Library)* 56 (6): 797–829.
- [17] J. A. Tropp and S. J. Wright, Computational methods for sparse solution of linear inverse problems, *Proceedings of the IEEE*, 98(6):948-958, 2010.
- [18] Rudin, L. I.; Osher, S.; Fatemi, E. (1992). "Nonlinear total variation based noise removal algorithms". *Physica D* 60: 259–268.
- [19] Strong, D.; Chan, T. (2003). "Edge-preserving and scale-dependent properties of total variation regularization". *Inverse Problems* 19: S165–S187.
- [20] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and superresolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [21] A. Buades, B. Coll, and J. Morel, "A review of image denoising methods, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [23] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, Jan. 2010.
- [24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2272–2279.
- [25] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1259–1266.
- [26] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
- [27] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, "Multiresolution histograms and their use for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 831–847, Jul. 2004.