Image Denoising Based On Sparse Representation In A Probabilistic Framework

Mohamed EL-Sayed Waheed *mewahed@yahoo.com*
Faculty of Computers & Informatics / Department of Computer Science
Suez Canal University
Ismailia, Egypt

Hassan Ahmad Khalil *h.a.khalil@gmail.com*
Faculty of Science / Department of Mathematics
Zagazig University
Zagazig, Egypt

Osama Farouk Hassan *Osamafarouk4@gmail.com*
Faculty of Science / Department of Mathematics
Suez Canal University
Ismailia, Egypt

Abstract

Image denoising is an interesting inverse problem. By denoising we mean finding a clean image, given a noisy one. In this paper, we propose a novel image denoising technique based on the generalized k density model as an extension to the probabilistic framework for solving image denoising problem. The approach is based on using overcomplete basis dictionary for sparsely representing the image under interest. To learn the overcomplete basis, we used the generalized k density model based ICA. The learned dictionary used after that for denoising speech signals and other images. Experimental results confirm the effectiveness of the proposed method for image denoising. The comparison with other denoising methods is also made and it is shown that the proposed method produces the best denoising effect.

Keywords: Sparse Representation, Image Denosing, Independent Component Analysis, Dictionary Learning.

1. INTRODUCTION

Being a simple inverse problem, the denoising is a challenging task and basically addresses the problem of estimating a signal from the noisy measured version available from that. A very common assumption is that the present noise is additive zero-mean white Gaussian with standard deviation σ . In this paper, we only consider the contaminated source, noise, of natural images. In other words, the purpose of image denoising is to restore the original image with noise-free. This problem appears to be very simple however that is not so when considered under practical situations, where the type of noise, amount of noise and the type of images all are variable parameters, and the single algorithm or approach can never be sufficient to achieve satisfactory results.

Many solutions have been proposed for this problem based on different ideas, such as statistical modeling [1], spatial adaptive filters, diffusion enhancement [2], transfer domain methods [3,4], order statistics [5], independent component analysis (ICA) and standard sparse coding (SC) shrinkage proposed by Alpo Hyvärinen in 1997 [6,7], and yet many more. Among these methods, methods based on sparse and redundant representations has recently attracted lots of attentions [8]. Many researchers have reported that such representations are highly effective and promising toward this stated problem [8]. Sparse representations firstly examined with unitary wavelet dictionaries leading to the well-known shrinkage algorithm [5]. A major motivation of using overcomplete representations is mainly to obtain translation invariant property [9]. In this respect, several multiresolutional and directional redundant transforms are introduced and applied to denoising such as curvelets, contourlets, wedgelets, bandlets and the steerable wavelet [5,8].

Moreover, the Ref. [10] gave an important conclusion: when ICA is applied to natural image data, ICA is equivalent to SC. However, ICA emphasizes independence over sparsity in the output coefficients, while SC requires that the output coefficients must be sparse and as independent as possible. Because of the sparse structures of natural images, SC is more suitable to process natural images than ICA. Hence, SC method has been widely used in natural image processing [10,11].

The now popular sparse signal models, on the other hand, assume that the signals can be accurately represented with a few coefficients selecting atoms in some dictionary[12]. Recently, very impressive image restoration results have been obtained with local patch-based sparse representations calculated with dictionaries learned from natural images [13,14]. Relative to pre-fixed dictionaries such as wavelets [1], curve lets [15], and band lets [16], learned dictionaries enjoy the advantage of being better adapted to the images, thereby enhancing the sparsity.

However, dictionary learning is a large-scale and highly non-convex problem. It requires high computational complexity, and its mathematical behavior is not yet well understood. In the dictionaries aforementioned, the actual sparse image representation is calculated with relatively expensive non-linear estimations. Such as ℓ_1 or matching pursuits [17,18]. More importantly, as will be reviewed, with a full degree of freedom in selecting the approximation space (atoms of the dictionary), non-linear sparse inverse problem estimation may be unstable and imprecise due to the coherence of the dictionary [19].

Structured sparse image representation models further regularize the sparse estimation by assuming de- pendency on the selection of the active atoms. One simultaneously selects blocks of approximation atoms, thereby reducing the number of possible approximation spaces [20,21]. These structured approximations have been shown to improve the signal estimation in a compressive sensing context for a random operator. However, for more unstable inverse problems such as zooming or deblurring, this regularization by itself is not sufficient to reach state-of-the-art results. Recently some good image zooming results have been obtained with structured sparsity based on directional block structures in wavelet representations [19]. However, this directional regularization is not general enough to be extended to solve other inverse problems.

In this paper we show that the over complete basis dictionary which learning by using the ICA probabilistic technique can capture the main structure of the data used in learning the dictionary, which used to represent the main component of the image. The results show that our technique is as the state-of-the-art in a number of imaging inverse problems, at a lower computational cost. The paper is organized as follows. In sections 2 and 3, we briefly introduce ICA and sparse representation. In section 4, we briefly present modeling of the scenario in decomposing a signal on an overcomplete dictionary in the presence of noise and discuss our algorithm in the real image denoising task. In section 5, we discuss the results of using our algorithm in image denoising. At the end we conclude and give a general overview to future's work.

2. INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) is a higher order statistical tool for the analysis of multidimensional data with inherent data addictiveness property. The noise is considered as Gaussian random variable and the image data is considered as non-Gaussian random variable. Specifically the Natural images are considered for research as they provide the basic knowledge for understanding and modeling of human vision system and development of computer vision systems.

In Gaussian noise, each pixel in the image will be changed from its original value by a (usually) small amount. A histogram, a plot of the amount of distortion of a pixel value against the frequency with which it occurs, shows an estimation of the distribution of noise. While other distributions are possible, the Gaussian (normal) distribution is usually a good model, due to the central limit theorem that says that the sum of independent noises tends to approach a Gaussian distribution. The case of Additive White Gaussian Noise (AWGN) will be considered. The acquired image is expressed in this case in the following form:

$$x = s + n \tag{1}$$

where x is the observed/acquired image, s is the noiseless input image and n is the AWGN component.

Estimating x requires some prior information on the image, or equivalently image models. Finding good image models is therefore at the heart of image estimation.

Some ICA algorithm such as FastICA [6] can be extended to overcomplete problems [22].

In information-theoretic ICA methods [23,24] statistical properties (distributions) of the sources are not precisely known. The learning equation $W \cong A^{-1}(y = Wx)$ has the form:

$$W(\mathrm{k}+1) = W(\mathrm{k}) + \eta[I - E\{\varphi(x)x^T\}]\, W(\mathrm{k}) \tag{2}$$

where is the score function by obtain from:

$$\varphi_i = (\frac{-1}{p_i})(\frac{dp_i}{dx_i}) \tag{3}$$

The unknown density functions p_i can be parameterized, as Generalized K Density (GKD), which is characterized by the following probability density function [25]

$$p(x \mid \alpha, \beta, k) = \frac{\alpha\beta x^{\alpha-1} \exp_k(-\beta x^{\alpha})}{\sqrt{1 + k^2 \beta^2 x^2 \alpha}} \tag{4}$$

where the generalized exponential function $\exp_k(x)$ given by

$$\exp_k(x) = (\sqrt{1 + k^2 x^2} + kx)^{\frac{1}{k}} \tag{5}$$

where $\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale and $k \in [0,1)$ measures the heaviness of the right tail.

The ICA algorithm in the framework of fast converge Newton type algorithm, is derived using the parameterized generalized k distribution density model. The nonlinear activation function in ICA algorithm is self-adaptive and is controlled by the shape parameter of generalized k distribution density model. To estimate the parameters of such activation function we use an efficient method based on maximum likelihood (ML). If generalized k probability density function is inserted in the optimal form for score function the expression for flexible nonlinearity is obtained:

$$\phi_i = -\frac{(\alpha-1)}{x} + \frac{1}{k}\left(\frac{\frac{\alpha k^2\beta^2 x^{2\alpha-1}}{\sqrt{1+k^2\beta^2x^{2\alpha}}} - \alpha\beta k x^{\alpha-1}}{\sqrt{1+k^2\beta^2x^{2\alpha}} - \beta k x^{\alpha}}\right) + \frac{\alpha k^2\beta^2 x^{2\alpha-1}}{1+k^2\beta^2x^{2\alpha}} \quad (6)$$

The maximum likelihood estimators (MLEs) [26,27] is

$$L(x/\alpha,\beta,k) = \log\prod_{i=1}^{N} p_{x_i}(x_i/\alpha,\beta,k) = (\alpha\beta)^N \prod_{i=1}^{N}\frac{x_i^{\alpha-1}\exp(-\beta x_i^{\alpha})}{\sqrt{1+k^2\beta^2 x_i^2\alpha}} \quad (7)$$

Normally, ML parameter estimates are obtained by first differentiating the log-likelihood function in equation(7) with respect to the generalized k-distribution parameters and then by equating those derivatives to zero (e.g. see [28]). Instead, here we choose to maximize the ML equation in equation (7) by resorting to the Nelder-Mead (NM) direct search method [27] . The appeal of the NM optimization technique lies in the fact that it can minimize the negative of the log-likelihood objective function given in equation (7) essentially without relying on any derivative information. Despite the danger of unreliable performance (especially in high dimensions), numerical experiments have shown that the NM method can converge to an acceptably accurate solution with substantially fewer function evaluations than multi-directional search or steeps descent methods [27]. Good numerical performance and a significant improvement in computational complexity for our estimation method are also insured by obtaining initial estimates from the method of moments. Therefore, optimization with the NM technique to produce the refined ML shape estimates $\hat{\alpha}$ and $\hat{k}$ can be deemed as computationally efficient. Also, an estimate for parameter $\hat{\beta}$ can be calculated for known $\hat{\alpha}$ and $\hat{k}$

$$\hat{\beta} = \frac{1}{2k}\left[\frac{\Gamma(\frac{1}{\alpha})\Gamma(\frac{1}{2k}-\frac{1}{2a})}{k+\alpha\Gamma(\frac{1}{2k}+\frac{1}{2a})}\right]^{\alpha} \quad (8)$$

3. SPARSE REPRESENTATION AND DICTIONARY LEARNING

Sparse representations for signals become one of the hot topics in signal and image processing in recent years. It can represent a given signal $x \in R^n$ as a linear combination of few atoms in an overcomplete dictionary matrix $A \in \square^{n\times k}$ that contains k atoms $\{a_i\}_{i=1}^{k}$ (k>n). The representation of x may be exact $x = As$ or approximate, $x \approx As$,satisfying $\| x - As \|_p \le \varepsilon$, where the vector s is the sparse representation for the vector x .To find s we need to solve either

$$(P_0)\ \min_s \|s\|_0 \text{ subject to } x = As \quad (9)$$

Or

$$(P_{0,\varepsilon})\ \min_s \|s\|_0 \text{ subject to } \|x - As\|_2 \le \varepsilon \quad (10)$$

where $\| \|_0$ is the ℓ_0 norm, the number on non-zero elements.

In this paper we use an ICA based algorithm to learn the basis of an overcomplete dictionary. Like the known K-SVD algorithm but instead of using the SVD decomposition for dictionary atoms update we used the FastICA algorithm with nonlinearity from the Generalized K Distribution for sparse representation for the data matrix. Also we choose the Gabor dictionary as an initial dictionary.

4. ICA FOR OVERCOMPLETE DICTIONARY LEARNING

ICA can be efficient in dictionary learning. Because ICA is most often applied for solving instantaneous Blind Source Separation (BSS) problem:

$$x = As \ , \ A \in R^{N \times M} , s \in R^{N \times T} \tag{11}$$

Classical ICA methods solve complete (determined and over-determined) BSS problems: $M \leq N$. That was one of the main arguments against using ICA for dictionary learning. Overcomplete dictionary is of practical interest because results in denoising can be better when dictionary is overcomplete (a frame).

In comparison with the probabilistic framework to basis learning in [29], that in partis also based on the use of ICA, the use of ICA proposed here is motivated by two reasons:

1. It extends the probabilistic framework to learn the overcomplete basis, this is achieved through the use of the FastICA algorithm, [12], that works in sequential mode.

2. In regard to the probabilistic framework to basis learning presented in [29], the adopted ICA approach is more flexible, this is due to the fact that proper selection of the nonlinear functions (that are related to parameterized form of the probability density functions of the representation) enables basis learning that is tied with a representation with the pre-specified level of sparseness without affecting the structure of the basis learning equation (by ICA the basis inverse is actually learned).

As opposed to that, in the Bayesian paradigm to the basis learning presented in [29], the structure of the basis learning equation depends on the choice of what was previously imposed on the probability density function of the sparse representation coefficients. We suppose that the linear model $y = D\ x$ is valid; where y and x are random vectors (we interpret columns of the data matrix Y , denoted as y_i , as realizations of y), and D is the basis matrix we want to estimate. For now we consider only the complete case (D is a $n \times n$ square matrix, and y and x are n dimensional).Hence, the basis D is what in blind source separation is referred to as a mixing matrix. Extraction of the code matrix X (also referred to as a source matrix in blind source separation) can be performed by means of the ICA algorithms.

Herein, we are interested in the ICA algorithm that:

1. Can be casted into the probabilistic framework tied with the linear generative model as in[29].

2. Can be extended for learning the overcomplete basis.

When blind source separation problem, $y = D\ x$, the minimization of the mutual information $I(x)$ is used:

$$I(x) = \sum_{i=1}^{n} H(x_i) - H(y) - \log\left|\det D^{-1}\right| \tag{12}$$

where $H(x_i)$ stands for the differential entropy of the representation and $H(\mathrm{y})$ stands for the joint entropy of the data.

The ICA algorithms that maximize information flow through nonlinear network (Infomax algorithm), maximize likelihood (ML) of the ICA model $y = D\ x$, or minimize mutual information between components of $x = D^{-1}y$, are equivalent in a sense that all minimize $I(x)$ and yield the same learning equation for D^{-1}.

$$D^{-1}(i+1) \leftarrow D^{-1}(i) + \eta[I - \phi(x(k)x(i)^T]D^{-1}(i) \tag{13}$$

If the generalized k probability density function is inserted in the optimal form for score function the expression for flexible nonlinearity is obtained by Equation (6).This enables learning the basis matrix D that gives sparse representation for y_i.For learning an overcomplete dictionary basis we used the FastICA algorithm with the nonlinearity obtained from the GKD. Thus, nonlinear function in the FastICA algorithm can be also chosen to generate sparse distribution of the representation x_i. In the experiments we have used the nonlinearity comes from the GKD, which models sparse or super-Gaussian distributions.

In the sequential mode of the FastICA, basis vectors are estimated one at a time. After every iteration, the basis vector is orthogonalized with respect to previously estimated basis vectors using the Gram-Schmidt orthogonalization. This idea can be extended to over complete case as follows:

$$d_i \leftarrow d_i - \alpha \sum_{j=1}^{i-1} (d_i^T d_j) d_j \tag{14}$$

and the dictionary updated using equation (13), where ϕ_i represents the score function defined as equation(6).

Reconstruction: reconstruct the denoised image $\hat{x} = D^{-1}y$.

5. EXPERIMENTS AND RESULTS

In this work, the underlying dictionary was trained with the new ICA technique, we used an overcomplete Gabor dictionary as an initial dictionary of size 64×256 generated by using Gabor filter basis of size 8×8, each basis was arranged as an atom in the dictionary. The dictionary then learned and updated by using the proposed algorithm in section 4. We applied the algorithm to images, mainly of size 256×256 and 512×512 with different noise levels, "Lena" and “Barbra” images. The results showed that using the overcomplete dictionary learned by using the FastICA gave a good results. To evaluate our method we calculate the PSNR for denoised BARBRA and LENA images using our method, K-SVD method, and Clustered-based Sparse Representation (CSR) [30]. The comparison results between the three methods are shown in figure 1 and figure 2. The results of the overall algorithm for the images ”Barbara” and “Lena” for $\|n\|_2$ = 20 is shown in Table 1, as it is seen, when the level of noise grows, our approach outperforms K-SVD with OMP and CSR methods. We can conclude that the mentioned algorithms are suitably designed for noisy cased with known low energy.

FIGURE 1: From left to right: original image, noisy image with zero-mean white Gaussian noise of $\|n\|_2$ = 20, the cleaned image via ICA based sparse representation described.

FIGURE 2: From left to right: original image, noisy image with zero-mean white Gaussian noise of $\|n\|_2 = 20$, the cleaned image via ICA based sparse representation described.

Sigma	BARBARA			LENA		
	K-SVD	ICA based	CSR	K-SVD	ICA based	CSR
5	38.08	37.41	37.52	38.60	38.18	38.56
10	34.42	34.51	34.35	35.52	35.42	35.38
15	32.36	32.79	32.45	33.69	33.88	33.62
20	30.83	32.02	30.94	32.38	33.46	32.56
25	29.62	31.05	30.02	31.32	32.72	31.47

TABLE 1: The PSNR computed for Barbra image and Lena image with different noise variance level (sigma).

6. DISCUSSION AND CONCLUSION

ICA-learned dictionary yields good or favorable results when compared against other methods. Yet, the ICA-based dictionary learning is faster than those by competing methods. It appears that ICA-learned dictionary is less coherent than the dictionary learned by K-SVD and the sparsity based structural clustering (CSR) on the same training set.

In this paper a simple algorithm for denoising application of an image was presented leading to state-of-the-art performance, equivalent to and sometimes outperform recently published leading alternative. We addressed the image denoising problem based on sparse coding over an overcomplete dictionary. Based on the fact that the ICA can capture the most important component of real data, which implies on real images. We presented our algorithm, which used the technique of learning the dictionary to be suitable for representing the important component in the image by using the FastICA technique that uses the nonlinearity induced from the Generalized K Distribution (GKD) for updating the dictionary in the learning process. Experimental results show satisfactory recovering of the source image. Moreover, for our technique, the larger the noise level is, the better the effect on the denoising results is.

7. REFERENCES

[1] Mihcak, M.K., Kozintsev, I., Ramchandran, K., Moulin, P., "Low complexity image denoising based on statistical modeling of wavelet coefficients", *IEEE Signal Processing Letters* ,6(12), pp. 300–303 ,1999.

[2] Gilboa, G., Sochen, N., Zeevi, Y., "Forward-and-Backward Diffusion Processes for Adaptive Image Enhancement and Denoising", *IEEE Trans. on Image Processing* ,11(7), pp. 689–703 , 2002.

[3] Grace Chang S., Yu B., Vetterli M., "Adaptive wavelet thresholding for image denoising and compression", *IEEE Trans. on Image Processing*, 9(9), pp. 1532–1546 , 2000.

[4] Starck J., Candes E., Donoho D.L., "The curvelet transform for image denoising", *IEEE Trans on Image Processing*, 11(6), pp. 670–684 , 2002.

[5] Donoho D.L. " De-noising by soft thresholding", *IEEE Trans. Inf. Theory* , 41(3), pp. 613–627, 1995.

[6] A. Hyvärinen, E. Oja, "A fast fixed-point algorithm for independent component analysis", *Neural Comput.* , 9, pp. 1483-1492, 1997.

[7] A. Hyvärinen, "Sparse coding shrinkage: denoising of nongaussian data by maximum likelihood estimation", *Neural Computation*, 11 , pp. 1739–1768, 1997.

[8] Aharon M., Elad M., Bruckstein A. "K-SVD: An Algorithm for Designing OvercompleteDictionaries for Sparse Representation", *IEEE Trans. on Signal Processing*, 54(11),November 2006.

[9] Coifman, R., Donoho, D.L., "Translation invariant de-noising", *In: Wavelets andStatistics. Lecture Notes in Statistics*, pp. 125–150. Springer, New York ,1995.

[10] P. Hoyer, A. Hyvärinen, E. Oja, "Sparse coding shrinkage for image denoising", *in neural networks proceedings, IEEE World Congress on Computation an intelligence* ,2, pp.859–864, 1998.

[11] D.M. Chandler, D.J. Field., "Estimates of the Information content and dimensionality of natural scenes from proximity distributions", *Journal of the Optical Society of America A* ,24(4) , pp. 922–941, 2007.

[12] MallatS., "A Wavelet Tour of Signal Proc.: The Sparse Way", 3rd edition, Academic Press, 2008.

[13] EladM. and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Trans. On Image Proc.,* 15(12): pp.3736–3745, 2006.

[14] MairalJ., M. Elad, and G. Sapiro, "Sparse representation for color image restoration", *IEEE Trans. on Image Proc.*, pp. 17, 2008.

[15] E.J. Candes and D.L. Donoho,"New tight frames of curvelets and optimal representations of objects with C2 singularities", *Comm. Pure Appl. Math.,* 56: pp.219–266, 2004.

[16] MallatS. and G. Peyr´e, "Orthogonal band let bases for geometric images approximation", *Comm. Pure Appl. Math.,* 61(9): pp. 1173– 1212, 2008.

[17] DaubechiesI., M. Defrise, and C. De Mol., "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", *Comm. Pure Appl. Math.*, 57: pp. 1413–1457, 2004.

[18] EfronB., T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression". *Annals of Statistics*, 32(2): pp. 407–451, 2004.

[19] MallatS. and G. Yu., "Super-resolution with sparse mixing estimators", *Accepted to IEEE Trans. on Image Proc.*, 2010.

[20] BaraniukR.G., V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing", *Preprint*, 2008.

[21] Huang J., T. Zhang, and D. Metaxas, "Learning with structured sparsity", *In Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 417–424. ACM, 2009.

[22] A. Hyvärinen, R. Cristescu, E. Oja, "A fast algorithm for estimating overcomplete ICA bases for image windows*", in:Proc. Int. Joint Conf. on Neural Networks, Washington, D.C.*, pp. 894-899, 1999.

[23] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution”, *Neural Comp.*, 7, pp. 1129-1159, 1995.

[24] D. Erdogmus, K. E. Hild II, Y. N. Rao and J.C. Principe, “Minimax Mutual Information Approach for Independent Component Analysis”, *Neural Computation*, vol. 16, No. 6, pp. 1235-1252

[25] Clementi F, Gallegati M, Kaniadakis G, "A new model of income distribution: The k-generalized distribution", *Empir Econ* , 39: pp. 559–591, 2011.

[26] Diversi R.,Guidorzi R. ,Soverini U. , "Blind identification and equalization of two-channel FIR systems in unbalanced noise environments", *Signal Processing* , vol. 85, pp. 215–225, 2005.

[27] Lagarias J. C. ,Reeds J.A.,Wright M.H., Wright P.E., "Convergence properties of the Nelder-Mead simplex method in low dimensions", *SIAM J. Optim.* ,9(1) ,pp. 112-147, December 1998.

[28] Shin J. W. Chang J. H., Kim N. S., "Statistical modeling of speech signals based on generalized gamma distribution", *IEEE Signal Process. Lett.*, 12 (3) , pp. 258-261, March 2005.

[29] Neelamani R., H. Choi, and R. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems", *IEEE Trans. on Signal Proc.*, 52(2): pp. 418–433, 2004.

[30] Weisheng D., Xi L., Lei Z., Guangming S., "Sparsity-based Image Denoising via Dictionary Learning and Structural Clustering" , *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*, pp. 457 – 464, 2011.