ICA Based Dictionary Learning for Image Denoising

M. EL-SayedWaheed[1], H. Ahmad Khalil[2] and O. Farouk Hassan[3]

[1.] Department of Computer Science, Faculty of Computers & Informatics, Suez Canal University, Ismailia, Egypt
[2.] Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, Egypt, P.O. Box 44519
[3.] Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia, Egypt
osamafarouk4@gmail.com

Abstract: Image denoising problem can be addressed as an inverse problem. An extension to the probabilistic framework for solving Image denoising problem is introduced in this paper. The approach is based on using over complete basis dictionary for sparsely representing the signal under interest. To learn the over complete basis, we used Generalized Gamma Distribution based ICA. We used the FastICA algorithm that works in sequential mode. The learned dictionary used after that for denoising speech signals. The results shows that our algorithm produced either equal or often significantly better than some of the best-known algorithms in signal denoising.
[M. EL-Sayed Waheed, H. Ahmad Khalil and O. Farouk Hassan. **ICA Based Dictionary Learning for Image Denoising.** *Life Sci J* 2014;11(3):112-115]. (ISSN:1097-8135). http://www.lifesciencesite.com. 15

Keywords: Sparse Representation, Image Denosing, Non-Negative Matrix Factorization, Dictionary Learning**.**

1. Introduction

Image denoising has been a well-studied problem in the image processing community and continues to attract researchers with an aim to perform better restoration in the presence of noise. Denoising often requires solving an inverse problem. An ideal image x is measured in the presence of an additive zero-mean white and Gaussian noise, e, with standard deviation s . The measured image y is, thus

y = **x**+**e**.

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.

Indeed, numerous contributions in the past decays addressed this problem from many and diverse points of view. Statistical estimators of all sorts, spatial adaptive filters, stochastic analysis, partial differential equations, transform-domain methods, splines and other approximation theory methods, morphological analysis, order statistics, and more, are some of the many directions explored in studying this problem.

Mixture models are often used as image priors since they enjoy the flexibility of signal description by assuming that the signals are generated by a mixture of probability distributions [1]. Gaussian mixture models (GMM) have been shown to provide powerful tools for data classification and segmentation applications, however, they have not yet been shown to generate state-of-the-art in a general class of inverse problems. Ghahramani and Jordan have applied GMM for learning from incomplete data, i.e., images degraded by a masking operator, and have shown good classification results, however it does not lead to state-of-the-art in painting [2]. Portilla *et al.* have shown image denoising impressive results by assuming Gaussian scale mixture models (deviating from GMM by assuming different scale factors in the mixture of Gaussians) on wavelet representations [3], and have recently extended its applications on image deblurring [4]. Recently, Zhou *et al.* have developed an nonparametric Bayesian approach using more elaborated models, such as beta and Dirich let processes, which leads to excellent results in denoising and in painting [5].

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 [6]. Recently, very impressive image restoration results have been obtained with local patch-based sparse representations calculated with dictionaries learned from natural images [7,8]. Relative to pre-fixed dictionaries such as wavelets [6], curve lets [9], and band lets [10], 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 [11,12]. More importantly, as will be reviewed in Section III-A, 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 [13].

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 [14-19]. These structured approximations have been shown to improve the signal estimation in a compressive sensing context for a random operator **U**. 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 [20]. However, this directional regularization is not general enough to be extended to solve other inverse problems [21-23].

This work shows that the over complete basis dictionary which learning by using the ICA technique can capture the main structure of the data used in learning the, lead to results in the same ballpark as the state-of-the- art in a number of imaging inverse problems, at a lower computational cost [25-28].

2. 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 over complete dictionary matrix $A \in \mathbb{R}^{n\times k}$ 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

$$(P_0) \quad \min_s \|s\|_0 \text{ subject to } x = As \qquad (1)$$

Or

$$(P_{0,\varepsilon}) \quad \min_s \|s\|_0 \text{ subject to } \|x - As\|_2 \le \varepsilon \qquad (2)$$

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

In this paper we use an algorithm to learn the basis of an over complete 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 Gamma Distribution for sparse representation for the data matrix,. Also we choose the Gabor dictionary as an initial dictionary instead of the DCT dictionary used on the K-SVD.

3. ICA for over complete dictionary learning

In comparison with the probabilistic framework to basis learning in [18], that in part is 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 over complete basis, this is achieved through the use of the FastICA algorithm, [6], that works in sequential mode
2. In regard to the probabilistic framework to basis learning presented in [18], 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[18], 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 [18].
2. Can be extended for learning the over complete basis.

When blind source separation problem,

$y = Dx$, the minimization of the mutual information I(x) is used:

$$I(x) = \sum_{i=1}^{n} H(x_i) - H(y) - \log\left|\det D^{-1}\right| \qquad (3)$$

where $H(x_i)$ stands for the differential entropy of the representation and H(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}.

$$D^{-1}(i+1) \leftarrow D^{-1}(i) + \eta[I - \phi(x(k)x(i)^T]D^{-1}(i) \qquad (4)$$

where

$$f_i = -\frac{1}{p_i}\frac{dp_i}{dx_i}$$

is the score function.

The unknown density functions p_i can be parameterized, as Generalized Gamma Density (GGD), which is characterized by the following probability density function

$$p(x \mid c, \alpha, \beta, \gamma) = \frac{\gamma\beta^{\alpha}|x-c|^{\alpha\beta-1}}{2\Gamma(\alpha)}\exp(-\beta|x-c|^{\gamma}) \qquad (5)$$

where c is the location parameter, $\beta \ge 0$ is the scale parameter, $\alpha \ge 0$ is the shape/power parameter and $\gamma \ge 0$ is the shape parameter.

If generalized Gamma probability density function is inserted in the optimal form for score function the expression for flexible nonlinearity is obtained:

$$f_i = \left(\frac{-1}{p_{y_i}(y_i)}\right)\frac{\gamma\beta^{\alpha}\,sign(y_i - c)|y_i - c|^{(\alpha\gamma-1)}\exp(-\beta|y_i - c|^{\gamma})}{2\Gamma(\alpha)}[(\alpha\gamma - 1) - \beta\gamma|y_i - c|^{\gamma}] \qquad (6)$$

This enables learning the basis matrix D that gives sparse representation for y_i .For learning an over complete dictionary basis we used the FastICA algorithm with the nonlinearity obtained from the GGD. 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 GGD, 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

$$d_i \leftarrow d_i - \alpha \sum_{j=1}^{i-1} (d_i^T d_j) d_j \qquad (7)$$

and the dictionary updated using equation (4)

where f_i represents the score function defined as

$$f_i = - \frac{1}{p_i} \frac{dp_i}{dx_i}$$

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

4. Experiments and Results

In this work, we used an over complete Gabor dictionary as an initial dictionary of size 64x256 generated by using Gabor filter basis of size 8x8, each basis was arranged as an atom in the dictionary. The dictionary then learned and updated by using the FastICA algorithm. We applied the algorithm to Lena image, and a speech signal.

The results showed that using the over complete dictionary learned by using the FastICA gave a good results. We used that method for image denoising and evaluate our method by calculating the PSNR and compare our results with the K-SVD methods, which showed that our method gave a better results over the K-SVD specially with low level noise energy.

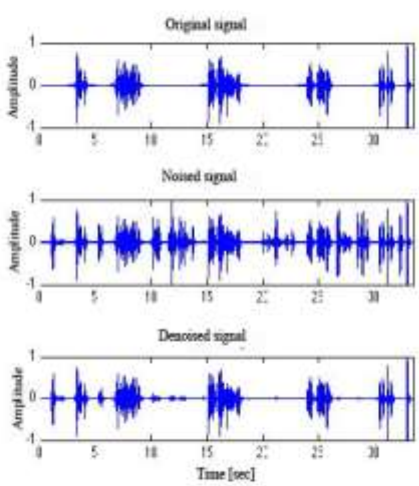

Fig. 1. (a) The original image. (b) The noised image by adding Gaussian noise with sigma=30. (c) The denoised image by using A-CMF algorithm and (d) the denoised image by using K-SVD.
Fig. 1 From upper to down: the original image, the noised signal, and the cleaned signal by using the over complete dictionary learned using ICA

Fig. 2 The original image in a, the image with additive noise in b, the denoised image by using ICA_DL in c, and the denoised image by using k_SVD in d.

Table 1. The PSNR computed for speech signal and Lena image with different noise variance level (sigma).

Sigma	Speech		Lena	
	KSVD	ICA_D	KSVD	ICA_D
10	32.1236	31.8481	33.3948	37.0749
15	30.1523	30.0193	31.1033	32.6425
25	28.0565	28.2762	28.4547	28.8607
30	25.4521	25.5401	27.2819	27.5758

5. Discussion and Conclusion

In this paper, we address the image denoising problem based on sparse coding over an over complete dictionary. Based on the fact that the ICA can capture the most important component of real data, which implies on real images. We presented an algorithm ICA_LD, 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 GGD for updating the dictionary in the learning process. Experimental results show satisfactory recovering of the source image. Future theoretical work on the general behavior of this algorithm is on our further research agenda.

Corresponding Author:
Osama Farouk
Department of Mathematics, Faculty of Science, Suez Canal University, Egypt.
Email Address: osamafarouk4@gmail.com

References

1. McLachlan G.J. and K.E. Basford. Mixture Models: Inference and Applications to Clustering. New York: Dekker, 1988.
2. Saba,T., Rehman, A., Al-Dhelaan, A., Al-Rodhaan, M. (2014) Evaluation of current documents image denoising techniques: a comparative study , Applied Artificial Intelligence, vol.28 (9), pp. 879-887, doi. 10.1080/08839514.2014.954344

3. Portilla J., V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wave let domain. IEEE Trans. on Image Proc., 12(11):1338–1351, 2003.
4. Harouni, M., Rahim,MSM., Al-Rodhaan,M., Saba, T., Rehman, A., Al-Dhelaan, A. (2014) Online Persian/Arabic script classification without contextual information, The Imaging Science Journal, vol. 62(8), pp. 437-448, doi. 10.1179/1743131X14Y.0000000083
5. MSM Rahim, SAM Isa, A Rehman, T Saba (2013) Evaluation of Adaptive Subdivision Method on Mobile Device 3D Research 4 (2), pp. 1-10
6. Mallat S.. A Wavelet Tour of Signal Proc.: The Sparse Way, 3rd edition. Academic Press, 2008.
7. Elad M. and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. On Image Proc., 15(12):3736–3745, 2006.
8. Mairal J., M. Elad, and G. Sapiro. Sparse representation for color image restoration. IEEE Trans. on Image Proc., 17, 2008.
9. E.J. Candes and D.L. Donoho. New tight frames of curvelets and optimal representations of objects with C2 singularities. Comm. Pure Appl. Math, 56:219–266, 2004.
10. JWJ Lung, MSH Salam, A Rehman, MSM Rahim, T Saba (2014) Fuzzy phoneme classification using multi-speaker vocal tract length normalization, IETE Technical Review, vol. 31 (2), pp. 128-136, doi. 10.1080/02564602.2014.892669
11. Daubechies I., M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm. Pure Appl. Math, 57:1413–1457, 2004.
12. Nodehi, A. Sulong, G. Al-Rodhaan, M. Al-Dhelaan, A., Rehman, A. Saba, T. (2014) Intelligent fuzzy approach for fast fractal image compression, EURASIP Journal on Advances in Signal Processing, doi. 10.1186/1687-6180-2014-112.
13. Ahmad, AM., Sulong, G., Rehman, A., Alkawaz,MH., Saba, T. (2014) Data Hiding Based on Improved Exploiting Modification Direction Method and Huffman Coding, Journal of Intelligent Systems, vol. 23 (4), pp. 451-459, doi. 10.1515/jisys-2014-0007
14. Baraniuk R.G., V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. Preprint, 2008.
15. Rehman, A. Mohammad, D. Sulong, G. Saba, T. (2009). Simple and effective techniques for core-region detection and slant correction in offline script recognition Proceedings of IEEE International Conference on Signal and Image Processing Applications (ICSIPA'09), pp. 15-20.
16. Muhsin; Z.F. Rehman, A.; Altameem, A.; Saba, A.; Uddin, M. (2014). Improved quadtree image segmentation approach to region information. the imaging science journal, vol. 62(1), pp. 56-62, doi. http://dx.doi.org/10.1179/1743131X13Y.0000000063.
17. Neamah, K. Mohamad, D. Saba, T. Rehman, A. (2014). Discriminative features mining for offline handwritten signature verification, 3D Research vol. 5(3), doi. 10.1007/s13319-013-0002-3
18. Saba T, Al-Zahrani S, Rehman A. (2012) Expert system for offline clinical guidelines and treatment Life Sci Journal, vol.9(4), pp. 2639-2658.
19. Mundher, M. Muhamad, D. Rehman, A. Saba, T. Kausar, F. (2014) Digital watermarking for images security using discrete slantlet transform, Applied Mathematics and Information Sciences Vol 8(6), PP: 2823-2830, doi.10.12785/amis/080618.
20. Norouzi, A. Rahim, MSM, Altameem, A. Saba, T. Rada, A.E. Rehman, A. and Uddin, M. (2014) Medical image segmentation methods, algorithms, and applications IETE Technical Review, Vol.31(3), doi. 10.1080/02564602.2014.906861.
21. Selamat, A. Phetchanchai, C. Saba, T. Rehman, A. (2010). Index financial time series based on zigzag-perceptually important points, Journal of Computer Science, vol. 6(12), 1389-1395, doi. 10.3844/jcssp.2010.1389.1395
22. Rehman, A. Kurniawan, F. Saba, T. (2011) An automatic approach for line detection and removal without smash-up characters, The Imaging Science Journal, vol. 59(3), pp. 177-182, doi. 10.1179/136821910X12863758415649
23. Saba, T. Rehman, A. Sulong, G. (2011) Cursive script segmentation with neural confidence, International Journal of Innovative Computing and Information Control (IJICIC), vol. 7(7), pp. 1-10.
24. Rehman, A. Alqahtani, S. Altameem, A. Saba, T. (2014) Virtual machine security challenges: case studies, International Journal of Machine Learning and Cybernetics vol. 5(5), pp. 729-742, doi. 10.1007/s13042-013-0166-4.
25. Meethongjan, K. Dzulkifli, M. Rehman, A., Altameem, A., Saba, T. (2013) An intelligent fused approach for face recognition, Journal of Intelligent Systems, vol. 22 (2), pp. 197-212, doi. 10.1515/jisys-2013-0010,
26. Joudaki, S. Mohamad, D. Saba, T. Rehman, A. Al-Rodhaan, M. Al-Dhelaan, A. (2014) Vision-Based Sign Language Classification: A Directional Review, IETE Technical Review, Vol.31 (5), 383-391, doi. 10.1080/02564602.2014.961576.
27. Rehman, A. and Saba, T. (2014). Evaluation of artificial intelligent techniques to secure information in enterprises, Artificial Intelligence Review, vol. 42(4), pp. 1029-1044, doi. 10.1007/s10462-012-9372-9.

2/3/2014