Blind Source Separation Using Hessian Evaluation

Jyothirmayi M mjyothiprasanna@gmail.com
Department of Instrumentation
M.S.Ramaiah Institute of Technology,
Karnataka, Bangalore-560054, India

Elavaar Kuzhali S kuzhalisubbiah @yahoo.com
Department of Instrumentation
M.S.Ramaiah Institute of Technology,
Karnataka, Bangalore-560054, India

Sethu Selvi S selvi@msrit.edu
HOD, Department of Electronics and Communication
M.S.Ramaiah Institute of Technology,
Karnataka, Bangalore-560054, India

Abstract

This paper focuses on the blind image separation using sparse representation for natural images. The statistics of the natural image is based on one particular statistical property called sparseness, which is closely related to the super-gaussian distribution. Since natural images can have both gaussian and non gaussian distribution, the original infomax algorithm cannot be directly used for source separation as it is better suited to estimate the super-gaussian sources. Hence, we explore the property of sparseness for image representation and show that it can be effectively used for blind source separation. The efficiency of the proposed method is compared with other sparse representation methods through Hessian evaluation.

Keywords: Blind Source Separation, Infomax Algorithm, Sparseness, Hessian Evaluation.

1. INTRODUCTION

Blind source separation (BSS) is the process of extracting the underlying sources from the mixed images or observed signals, and since no *a priori* knowledge of the mixed sources is known or very little information is available, it is called *blind*. Independent component analysis (ICA) is most widely used technique to solve the blind source separation [1-7, 20] problem. BSS is based on the assumptions that source signals are independent and non gaussian. The problem of BSS can be stated as follows.

Given M linear mixtures of N original images mixed via a $M \times N$ mixing matrix, these mixtures can be represented as a linear equation of the form

$$X = AS \tag{1}$$

S is the original sources to be extracted, X is the observation random vector, and A is a full rank $M \times N$ mixing matrix. The task is to estimate the mixing matrix A and then recover the source images S. The complexity of source separation is based on the dimensions M and N. If $M=N$, the mixing matrix A is a square matrix and the original sources can be estimated by a linear transformation. If $M>N$, the mixing matrix is a full rank over complete matrix and the original sources can be estimated using linear transformation involving pseudoinverse matrix.

After estimating the matrix A using BSS algorithm, the inverse of A is computed, $W=A^{-1}$ called separator and the independent sources are obtained simply by

$$U = WX \tag{2}$$

There are several methods [5], [6] to separate the independent components from the original data. Bell and Sejnowski [7] developed a neural learning algorithm for separating the statistically independent components of a dataset through unsupervised learning. The algorithm is based on the principle of maximum information transfer between sigmoid neurons. The features obtained were not very interesting from a neural modeling viewpoint, and the mixing matrix reduces the sparsity of the original images which motivates us to find better models like exploiting the property of sparseness, Zibulevsky et al.(2002).The advantage of this property is that, two or more sources being active at the same time is low. Thus, sparse representations provide good separability as most of the energy is confined to a single source, at a given time instant.

Sparse coding is a method for finding suitable representation of data in which the components are rarely active. It has been shown [9, 10-17] that this sparse representation can be used to solve the BSS problem. ICA algorithms i.e., FASTICA uses kurtosis as a sparseness measure and since kurtosis is sensitive to the outliers as it applies more weight on heavy tails rather than on zero, the sparseness measure is mostly unreliable. When the sources are locally very sparse the matrix identification algorithm is much simpler. Sparse representation of image matrix can be performed using clustering algorithms [12, 13, 15], Gradient Ascent Learning [14], Laplacian Prior [16, 7], wavelet [17] , Finite difference method (FDM)[29] etc.

In this paper, the sparsity representation of the image mixture is exploited using Hessian transformation and the sparseness is measured using l_0 norm Donoho [26]. The Hessian transform method provides a powerful approach to solve differential equations, non-linear problems and is widely used in the field of applied sciences. The proposed BSS algorithm is more efficient and leads to improved separation quality which is measured in terms of Signal to Noise ratio (SNR), Mean squared error (MSE), Structural Similarity Index Measure (SSIM) and estimating the Mutual information (MI) of the separated images with original source images.

The rest of the paper is organized as follows: Section 2 deals with the method used for sparse representation of data, sparse measure using l_0 norm and the algorithm used for separation. Section 3 illustrates the results where we compare the separated images with original images and Section 4 gives the conclusion.

2. SPARSE REPRESENTATION

In order to represent the sparsity of the original images that solves the BSS problem, several methods are proposed [5, 6, 12, 13, 14, 15, and 21]. A simpler and efficient method to make the image sparse is by calculating the Hessian matrix of the image. The problem of estimating the Hessian matrix is to derive the second order derivative of a given image. The equation for the second order can be derived using Taylor series expansion [23].

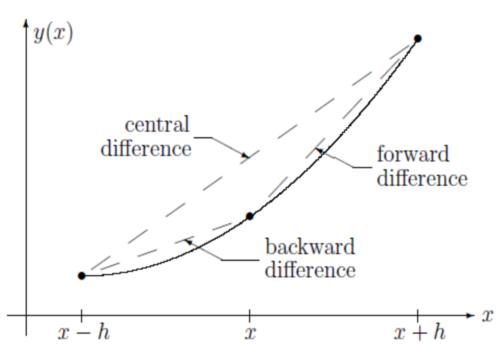


FIGURE 1: Finite difference approximations to derivatives

$$y(x+h) = y(x) + hy'(x) + \frac{h^2}{2}y''(x) + \frac{h^3}{6}y'''(x) + O(h^4) \tag{3}$$

Finding the first derivative and second derivative

$$y'(x) = \frac{y(x+h) - y(x)}{h}$$

$$y''(x) = \frac{y'(x+h) - y'(x)}{h}$$

$$y''(x) = \frac{y(x+2h) - 2y(x+h) - y(x)}{h^2} \tag{4}$$

Similarly, by replacing h by –h in the Taylor series of Equation (3) and deriving the second derivative, the backward difference equation is obtained as ,

$$y''(x) = \frac{y(x) - 2y(x-h) + y(x-2h)}{h^2} \tag{5}$$

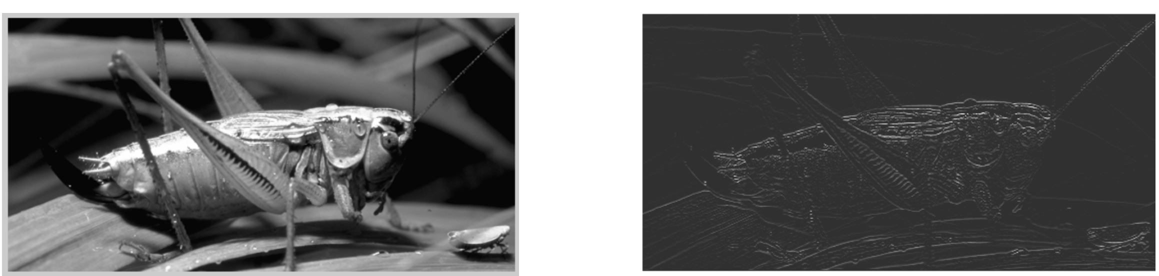
The average of the two Equations (4) and (5) results in a central difference approximation (Hessian), of the form

$$y''(x) = \frac{y(x+h) - 2y(x) + y(x-h)}{h^2} \tag{6}$$

as represented in Figure 1.Hence for the given image matrix X, the Hessian matrix obtained will be of the form

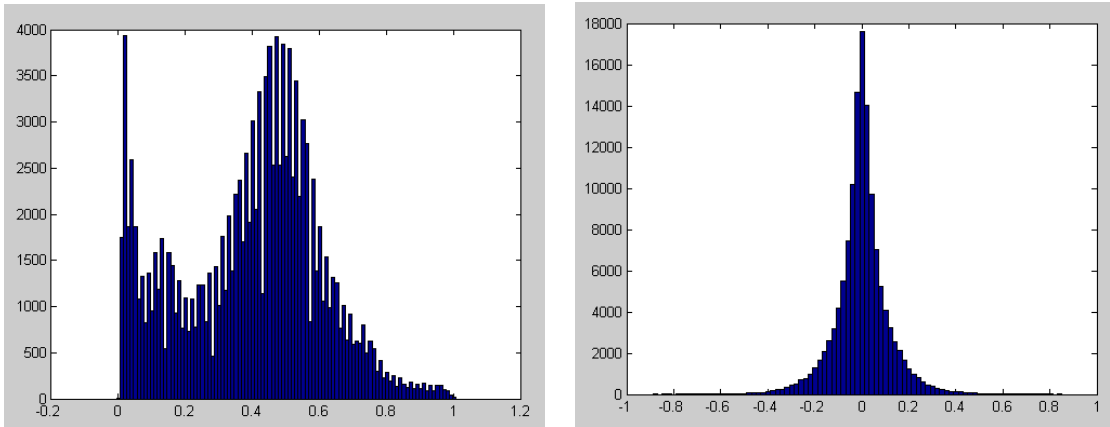
$$H(x) = \begin{bmatrix} \frac{\partial^2 x}{\partial x_1^2} & \frac{\partial^2 x}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 x}{\partial x_1 \partial x_n} \\ \frac{\partial^2 x}{\partial x_2 \partial x_1} & \frac{\partial^2 x}{\partial x_2^2} & \cdots & \frac{\partial^2 x}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 x}{\partial x_m \partial x_1} & \frac{\partial^2 x}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 x}{\partial x_n^2} \end{bmatrix} \tag{7}$$

The Hessian evaluation method can be used for sparse representation of the image since it acts as an edge detector which provides a two-level image, the edges and the homogeneous background. By using this method, the separation matrix estimated to separate the image mixture is similar to that of the method used (FASTICA) for image separation. In Figure 2, shows the natural image as well as the image obtained from the above method. Figure 3, represents the histograms of the original and the Hessian transformed image from which the sparsity of the latter can be inferred.



(a) (b)

FIGURE 2: a) Original Image and **b)** Hessian Transformed Image



(a) (b)

FIGURE 3: (a) Original Image Histogram and **(b)** Hessian Transformed Image Histogram

2.1 Sparsity Measure

Sparsity measures are used to calculate a number that describes the sparsity of an image vector $C = [c_1 c_2 .... c_N]$. The most common sparsity measure is the ℓ_p norm defined by

$$\left\| \vec{C} \right\|_p = \left(\sum_j C_j^p \right)^{1/p} \text{ for } 0 \le p \le 1 \qquad (8)$$

The simplest is the l_0 norm that calculates the number of non-zero coefficients in the image vector

$$\left\| \vec{C} \right\|_0 = \# \{ c_j \neq 0,\ j = 1....N \}$$

Since this traditional method is unsuited for many practical scenarios [15,24,25],a modified approximation method suggested by Donoho [26] is being used where a threshold is used .

$$\left\| \vec{C} \right\|_0 = \# \{ c_j \neq \varepsilon,\ j = 1....N \}, \qquad (9)$$

where ε is some threshold value.

Under this measure, the sparse solution is obtained by finding the number of non-zero elements in a block. Bronstein et al. [5], discusses that the use of block partitioning, which is the natural way of handling mixing coefficients in an image that vary spatially and also gives a better refinement to the sparseness. Hence the block having the maximum sparse is selected to estimate the separation matrix *W* Equation (2). Figure 4 shows sparseness measure values using l_0 norm for different sparse functions.

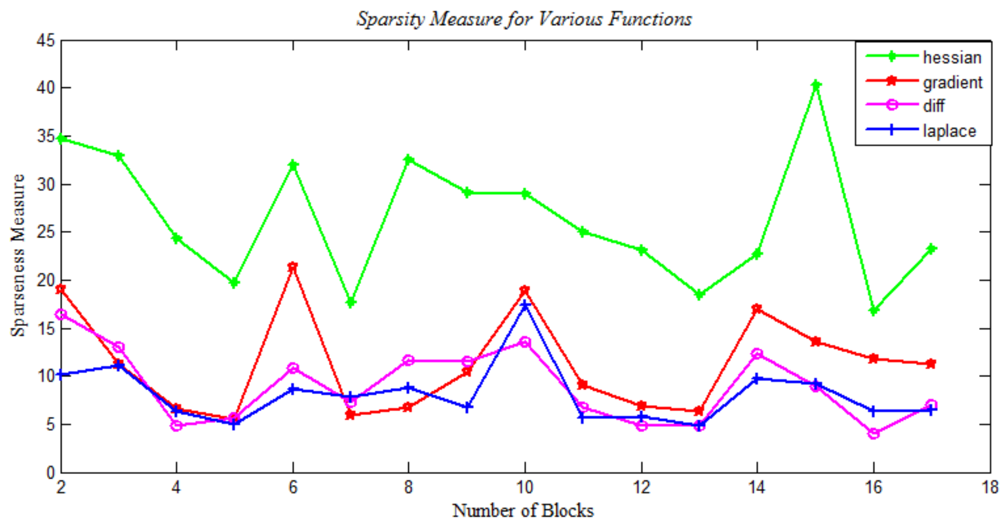


FIGURE 4: Sparsity measure plot obtained for different functions

2.2 Algorithm

Figure 5 shows the flow diagram for the representation of sparse matrix using second order differentiation (Hessian) and the selection of the best block by evaluating the quality factor for the generation of the separation matrix. Infomax ICA algorithm [10] is used for solving BSS problem.

Steps for Hessian Transformation and selection of best sparse block for BSS

1. Normalize and mean removed data is formed from N images.
2. Mixed images are obtained by linearly mixing with a random matrix.
3. The second order derivative (Equation 6) is applied for each of the mixed images to get sparse images. The component in the X direction and Y direction is considered.(Figure 5)
4. The sparse images along X and Y directions are divided into blocks of equal size.
5. The blocks having same spatial location are considered for evaluation of the sparseness (l_0 norm).
6. The blocks having maximum sparseness is considered for estimating the separation matrix using Infomax algorithm.

2.3 ICA-BSS Algorithm

The block having maximum sparseness can be represented as a linear equation of the form

$$X = \sum_{i=1}^{n} a_i s_i = AS \quad (10)$$

where A is the unknown fixed matrix $M \times N$ of full rank, called *mixing matrix* and its columns are known as *ICA basis vectors*. The fundamental problem of ICA is then to recover the underlying causes S from the mixture X. Since only X is observed the problem changes to estimating W.

$$S = (A^T A)^{-1} A^T X = WX \quad (11)$$

In order to determine *W*, we use the Sejnowski infomax algorithm [10] that maximizes the information for ICA. A feedforward neural network Figure 6 is used to train the weights of training images. As shown in the Figure 6, *X* is the input to the neural network, the output U is the summed output of the weights connected to neuron and the input *X*.

$$u_i = \sum_{j=1}^{m} w_{ij} x_j \quad \text{i.e} \quad U = WX \qquad (12)$$

ICA works by adjusting the unmixing coefficients of W in order to maximize the uniformity (entropy) of the distribution $Y = f(u)$ where f is cumulative density function (cdf). Therefore, if W is optimal, then Y will have maximum entropy and are therefore independent which ensures that the extracted sources are also independent. Hence the gradient update rule for the weight matrix, *W* used is as follows:

$$\Delta W \ \alpha \ \nabla_W H(Y) W^T W = (I + Y' U^T) W \qquad (13)$$

where $Y' = (1 - 2Y_1)$

$$Y_i' = f_i''(U_i) / f_i'(U_i),$$

$$f_i(u) = 1 \big/ 1 + e^{-u}$$

ICA finds a matrix *W*, such that the rows of $X = AS$ are statistically independent as possible. Hence we find the sources S=*WX*, where $W = A^{-1}$.

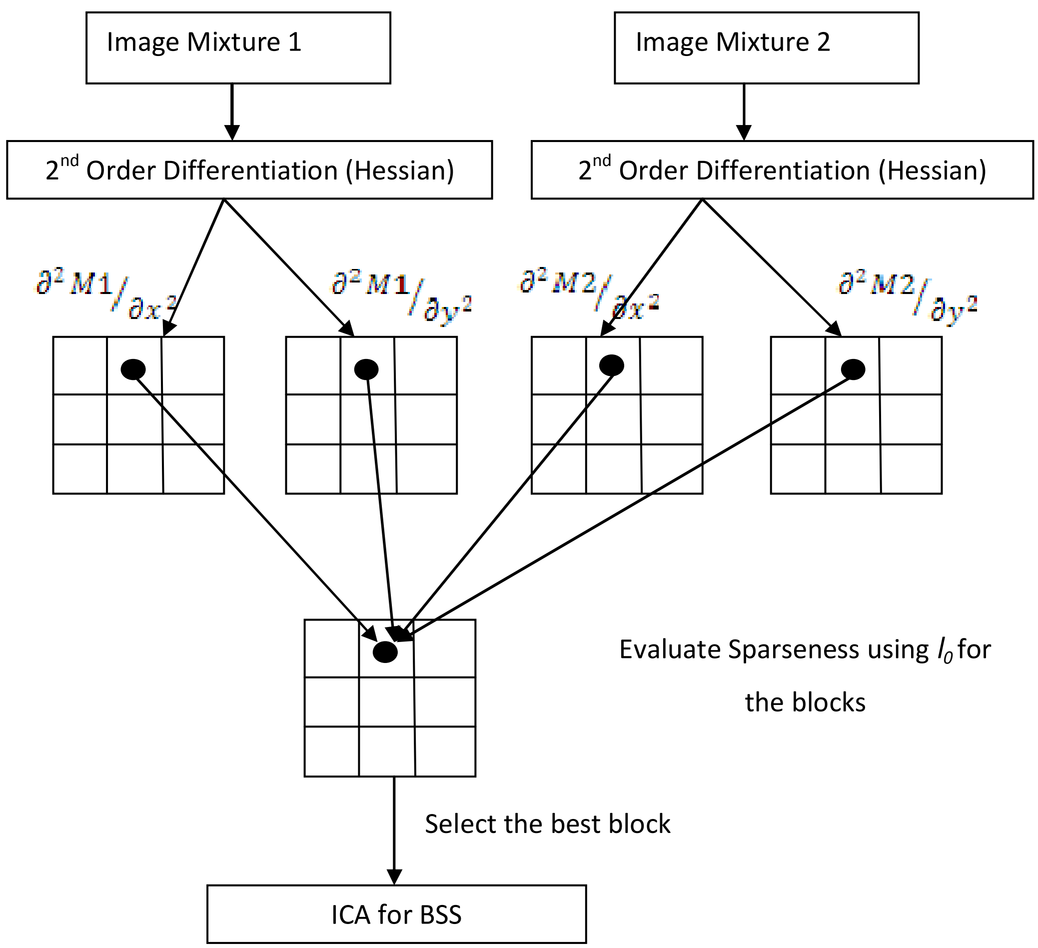


FIGURE 5: Block schematic describing the algorithm for BS

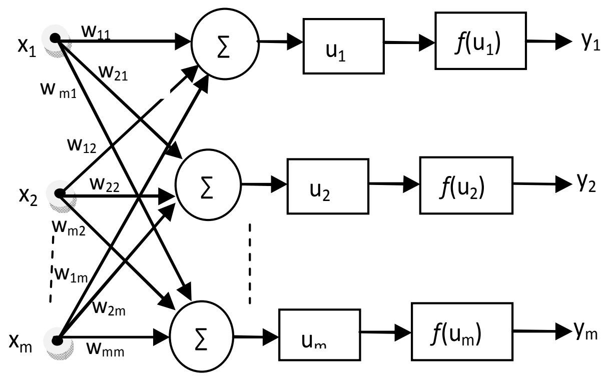


FIGURE 6:. Neural Network to train the weights

3. SIMULATION RESULTS

Simulation experiments are conducted to demonstrate the feasibility of the proposed BSS method. All simulations are carried on natural images[1] of dimension 256 x 512. The algorithms are developed on MATLAB environment. The images are mixed with a random matrix 4 x 4 (example Equation (14)) and the Hessian transform as in Equation (6) is applied on it to get sparse images. The resultant of Hessian derivative along the X direction and Y direction are considered.

$$M = \begin{bmatrix} -0.0716 & 0.3296 & -2.6350 & -0.3276 \\ -2.4146 & 0.5985 & 0.0281 & -1.1582 \\ -0.6943 & 0.1472 & -0.8763 & 0.5801 \\ -1.3914 & -0.1014 & -0.2655 & 0.2398 \end{bmatrix} \quad (14)$$

The sparse images are then divided into blocks of equal size 64 x 128 for which the l_0 norm, Equation (9), is applied to evaluate the sparseness measure. Figure 4 shows the sparseness measure values for different blocks and for different sparse representation methods. The blocks which has maximum value of the sparseness (e.g for hessian, block 15 from figure 4) is considered for evaluating the separation matrix. The separation matrix is obtained by using Infomax algorithm indicated in Equations (10-13). The results shown in Figure 7 and Figure 8 can be used as a subjective quality assessment. The histogram plot shows that the separated images are similar to original. Mean Squared Error (MSE), Signal to Noise Ratio (SNR) are used as objective image quality measures. The Structural Similarity Index Measure (SSIM) a well-known quality metric is used to measure the similarity between two images. It was developed by Wang et al. [16], and is considered to be correlated with the quality perception of the Human visual system (HVS). The Mutual Information between the extracted source and the corresponding original source is also considered as a quality metric. Here, 5000 samples (pixels) are randomly selected from the separated image and the original source to form pixel pairs. The Mutual Information between these 5000 pixel pairs was estimated using $I^{(1)}$ estimator described by Kraskov et.al.(2004) with k=2, (k recommended can be between 2 and 4 which is the nearest neighbor order)[27,28].

A comparison of proposed method with other existing sparse representation methods such as Gradient, Laplace [5] and Finite Difference Method [29] for assessing the amount of separation achieved is tabulated in Tables 1 to 4. The lower value for MSE as in Table 1 indicates, better quality separation. The higher values of SNR, SSIM, MI in Table 2, 3 and 4a implies that the

[1] http://www.cis.hut.fi/projects/ica/data/images/

separated images are closer to original source images, whereas Table 4b, the low values corresponds to the comparison of extracted image with the wrong source image. The values in the last column of the Tables 1-4 show clearly that the hessian representation method has an advantage over other sparse representation methods.

Figure	FDM	Laplace	Gradient	Hessian
S1	1.82	7.96	2.65	1.49
S2	0.83	0.009	2.32	0.065
S3	4.58	65.18	6.76	3.53
S4	0.38	18.87	1.25	1.31

TABLE 1: MSE values for the four separated images

Figure	FDM	Laplace	Gradient	Hessian
S1	38.54	23.10	36.89	39.37
S2	41.23	60.79	36.77	52.32
S3	34.82	23.29	33.12	33.28
S4	44.45	27.54	39.30	39.76

TABLE 2: SNR values (in dB) for the four separated images

Figure	FDM	Laplace	Gradient	Hessian
S1	0.8590	0.8570	0.9998	0.9979
S2	0.9883	0.9884	0.9994	0.9999
S3	0.9986	0.9986	0.9999	0.9981
S4	0.8739	0.8744	0.9931	0.9993

TABLE 3: SSIM values for the four separated images

Figure	FDM	Laplace	Gradient	Hessian
S1	5.33	2.9	5.15	5.4
S2	5.46	4.39	4.60	5.73
S3	4.63	2.85	4.13	4.97
S4	5.4	2.07	5.03	5.59

TABLE 4a: MI values for the four separated images with corresponding source image.

Figure	FDM	Laplace	Gradient	Hessian
S1	0.31	0.18	0.28	0.18
S2	0.13	0.09	0.11	0.10

TABLE 4b: MI values for the four separated images with other source image

4. CONCLUSIONS

For the solution to source separation, natural images are considered, as the natural scenario provides various mixtures of images. The proposed sparse representation method using block partitioning approach is more suitable for both spatially variant and invariant images. Experiments conducted on the natural image data set shows the efficiency of this approach and its advantages over previously-proposed sparse representation methods. We observe that there is improvement in MSE, SNR, MI and SSIM values. These values can be considered as objective evaluation for the image quality. Figure 8 shows the histogram for the original and separated images which clearly show that separated images are equal to the original images.

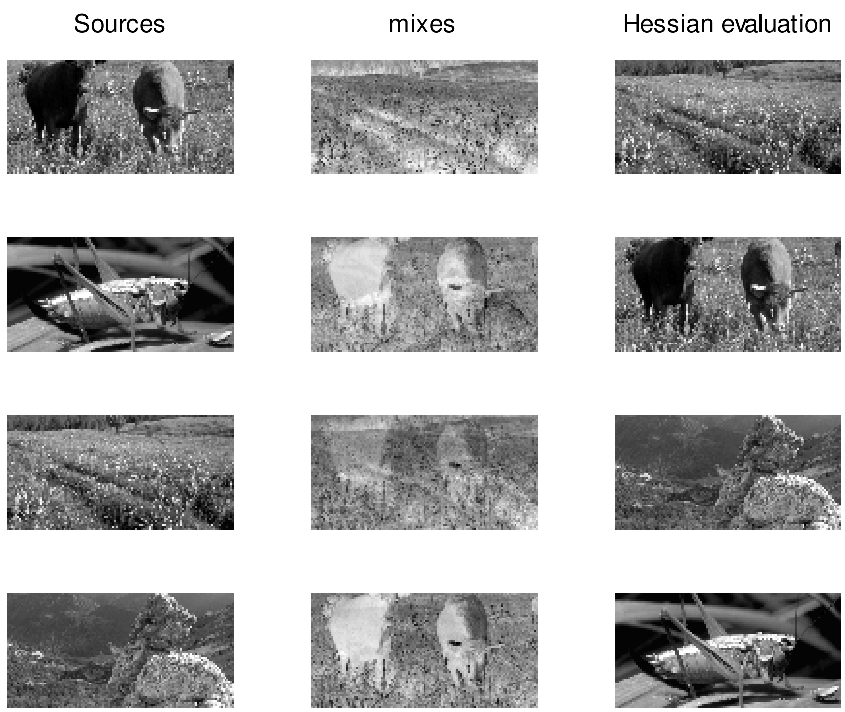


FIGURE.7: Original images, Mixed and Separated images for the proposed Method

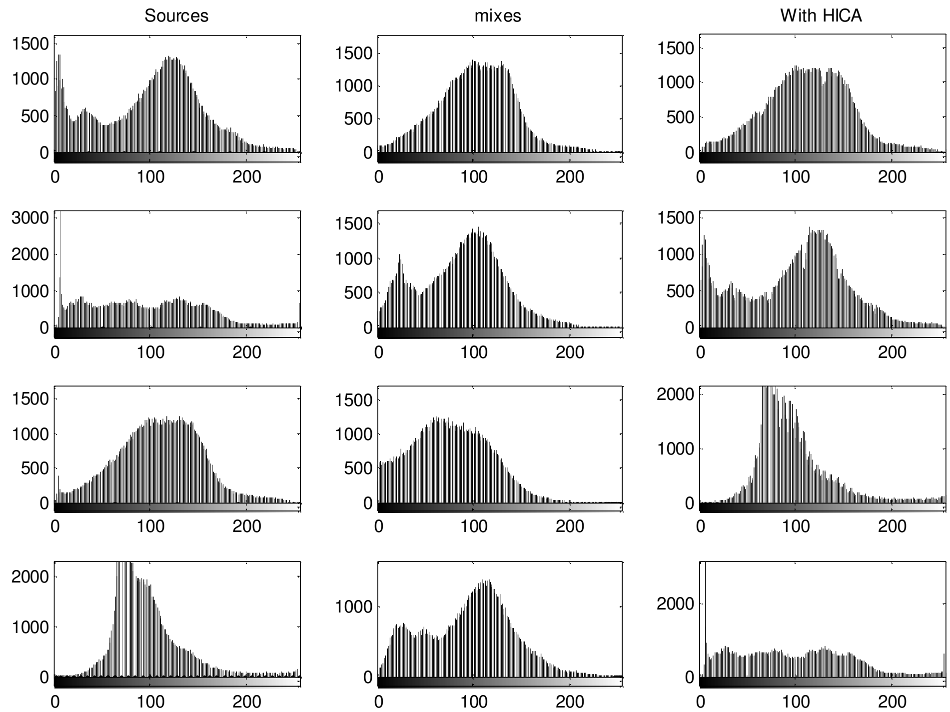


FIGURE 8: Histogram plot for original, mixed and separated images

5. REFERENCES

[1] Jutten.C, Herault. J.Blind separation of sources part I: "'A adaptive algorithm based on neuromimetic architecture", Signal Process, ,24:1-10,1991

[2] Girolami. M. "An alternative perspective on adaptive independent component analysis algorithms". Neural Computation, 10 (8) :2103-2114, 1998

[3] A. Hyvärinen and E. Oja., A Fast Fixed-Point Algorithm for Independent Component Analysis, Neural Computation, 9(7):1483-1492,1997

[4] Aapo Hyvärinen, Jarmo Hurri, Patrik O. Hoyer, Natural Image Statistics-A Probabilistics approach to early computational vision, springer 2009.

[5] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Separation of reflections via sparse ICA," in *Proc. International Conference on Image Processing ICIP*, September 2003, vol. 1, pp. 313–316.

[6] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in signal dictionary," *Neural Computation*,vol. 13, no. 4, pp. 863–882, 2001.

[7] Pando Georgiev, Fabian Theis, and Andrzej Cichocki, "Sparse Component Analysis and Blind Source Separation of Underdetermined Mixtures" IEEE Transactions on Neural Networks, VOL. 16, NO. 4, July 2005

[8] P Comon. "Independent component analysis –a new concept".Signal Processing, 36:287-314, 1994.

[9] Hyvarinen A. "Independent Component Analysis:Algorithms and Applications". Neural Networks, 2000, 13:411-430.

[10] Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129-1159(1995)

[11] Zibulevsky M., Pearlmutter B. A., Boll P., & Kisilev P. (2000) Blind source separation by sparse decomposition in a signal dictionary. In Roberts, S. J. and Everson, R. M. (Eds.), *Independent Components Analysis: Principles and Practice*, Cambridge University Press

[12] Yuanqing Li, Andrzej Cichocki, Shun-ichi Amari, Sergei Shishkin , *Sparse Representation and Its Applications in Blind Source Separation* NIPS-2003

[13] W. Souidene, A. A¨ıssa-El-Bey, K. Abed-Meriam, and A. Beghdadi, "Blind image separation using sparse presentation," in Proc. IEEE International Conference on Image Processing ICIP, September 2007, vol. 3, pp. 125–128.

[14] Hyvarinen A. "Survey on Independent Component Analysis". Neural Computing Surveys, 1999, 2:94-128.

[15] Yuanqing Li, Andrzej Cichocki, Shun-ichi Amari, Sergei Shishkin, Jianting Cao, Fanji Gu, *Sparse Representation and Its Applications in Blind Source Separation* Seventeenth Annual Conference. on. Neural Information Processing Systems 2003.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity" IEEE Transactios on Image Processing, vol. 13, no. 1, Jan. 2004

[17] V.K.Ananthashayana and Jyothirmayi M, " Blind Source Separation Using Modified Gaussian Fastica", International Conference on Signal Processing, Communications and Networking, Sept 2009.

[18] J. K. Lin, D. G. Grier, and J. D. Cowan. "Feature extraction approach to blind source separation".In *IEEEWorkshop on Neural Networks for Signal Processing (NNSP)*, pages 398.405, 1997.

[19] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski. "Blind source separation of more sources than mixtures using overcomplete representations". *IEEE Signal Processing Letters*, 4(5):87.90, 1999.

[20] A. Jourjine, S. Rickard, and O. Yilmaz."Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures". In *IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP2000)*, volume 5, pages 2985.2988, June 2000.

[21] P. Bo_ll and M. Zibulevsky."Blind separation of more sources than mixtures using the sparsity of the short-time fourier transform". In *2nd International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 87.92, Helsinki, Finland, June 19.20, 2000.

[22] M. Zibulevsky and B. A. Pearlmutter."Blind source separation by sparse decomposition in a signal dictionary",*Neural Computation*, 13(4):863.882, Apr. 2001

[23] Gordon D. Smith, 'Numerical solution to partial differential equation: finite difference methods', Oxford University Press, third edition, 2004.

[24] J. Karvanen and A. Cichocki, "Measuring sparseness of noisy signals,"in *Proc. ICA03*, 2003.

[25] Niall Hurley and Scott Rickard, ”Comparing Measures of sparcity”,in IEEE Transactions on Information Theory , VOL. 55, NO. 10, OCTOBER 2009

[26] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Trans.Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[27] A. Kraskov, H. Stögbauer and P. Grassberger, "Estimating Mutual Information", Physical Review E, vol. 69, pp. 066138, 2004.

[28] L. B. Almeida, "Separating a Real-Life Nonlinear Image Mixture", *Journal of Machine Learning Research*, vol. 6, pp. 1199–1232, 2005.

[29] Jyothirmayi M, Elavaar Kuzhali S and Sethu Sevi S, “Blind Image Separation using Forward Difference Method”,Signal and Image Processing :An International Journal, vol2,No 4, pp121-127, Dec2011.