Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 4 Number 4 (2011) pp. 435-445
© Research India Publications
http://www.ripublication.com/
https://dx.doi.org/10.37622/ACST/4.4.2011.435-445

Optimal Feature Selection using Independent Component Analysis and Significant Feature Processing Applied to Human Brain MR Images

[1]Gayatri Mirajkar, [2]Balaji Vithalrao Barbadekar and [3]Yogesh Sonavane

[1]Shivaji University, Kolhapur, India
[2]Research Guide, Shivaji University, Kolhapur, India
[3]Chalmers University of Technology, Gothenburg, Sweden

Abstract

The optimal feature selection has been done using ICA and significant feature processing applied to human brain MR images. Implementation in MR image segmentation has been studied significantly with kmeans and fuzzy c –means algorithm using MATLAB 2009a. The input to the kmeans algorithm and fuzzy c-means algorithm is the final feature image obtained by processing the outputs of Gabor wavelets and the number of desired classes in which the input image must be segmented. The result obtained from the clustering algorithm is the tumor region delineated from the unaffected brain tissue. Both ICA as well as significant feature processing gives the resultant feature image that contains the tumor and edema highlighted from the surrounding brain tissue. It has been observed that both methods are effective even in the presence of complex structures. ICA provides an effect that is similar to smoothing of the images resulting in absolute segmentation. The statistical method provides sharp segmentation results.

Clustering, magnetic resonance image, independent component analysis, significant feature processing, kmeans, fuzzy c-means.

Introduction

Image segmentation aims to cluster pixels into salient image regions i.e., regions corresponding to individual surfaces, objects or natural parts of objects. It is an especially important operation in biomedical image processing since it is used to isolate physiological and biological structures of interest. A well-known segmentation problem within MRI is the task of labeling voxels according to their tissue type which includes white matter, gray matter, cerebrospinal fluid, and sometimes pathological tissues like tumor etc. Computer-aided detection of abnormal growth of tissues is primarily motivated by the necessity of achieving maximum possible accuracy [1]. From the viewpoint of exploratory data analysis, segmentation can be formulated as a classification problem in which pixels or small image patches are classified based on local image feature information. A crucial problem in image segmentation is to select and extract image features. MR images of tissues contain a lot of microscopic information that may not be assessed visually, and texture analysis provides the means for obtaining this information [2].

Texture is an image feature that corresponds to both brightness value and pixel locations. Texture features are mathematical patterns which are obtained by applying texture analysis methods. These are used to discriminate the different textures present in images. The extracted image features often contain irrelevant or redundant features that can complicate the classification process. Even if the set of feature images contain the features that are sufficient for correct classification, due to the size of the feature vector, or dimension of the feature space, a wrong prediction is made. Therefore, one of the primary concerns of classification is the reduction of the dimensionality of feature space retaining the features that are significant to the classification process and ignoring the remaining features. Such a type of redundant information analysis is rarely reported in MR image segmentation [3].

In this research, Independent component analysis (ICA) has been employed to reduce the dimensionality of the feature space and to extract image features. ICA involves finding a linear representation of nongaussian so that the components are statistically independent or maximally independent as possible. Such a representation seems to capture the essential structure of the data in many applications including feature extraction and signal separation [4]. ICA is a technique that exploits higher-order statistical structure in the data. ICA generalizes PCA, and like PCA, has proven a useful tool for finding structure in data.

This paper proposes an ICA-based feature extraction method which reduces the dimensionality of the feature space and compresses redundant information without degrading classifier performance. At the same time, reduction of dimensionality of feature space has also been also achieved by a newly proposed statistical method that delivers results that are at par with those provided by ICA.

The classification is done using kmeans and fuzzy c-means algorithms. Simulations have been done in MATLAB 2009a. The feature vector is obtained by the application of Gabor wavelets to the wavelet coefficients acquired by applying undecimated wavelets to the MR images [5]. The region with the highest intensity in the tumor has been labeled as the darkest in the kmeans algorithm. For fuzzy c-means, the candidate pixels showing the greatest tendency of belonging to the particular class are highlighted with the darkest colors. The efficacy of both methods is verified for MR images containing tumors. In fuzzy c-means clustering, at every level of decomposition, the tumor has been perfectly segmented out showing its presence as an object at all levels of approximation which is definite advantage over kmeans clustering algorithm.

Both methods of reducing dimensionality of feature vectors work well even in the presence of complex structures. ICA provides an effect that is similar to smoothing of the images which results in absolute segmentation. The statistical method provides sharp segmentation results that are presented in section 6. The Gabor wavelets implemented are presented in section 2.

Gabor Wavelets

The efficient segmentation of tumor mass from the MRI images is achieved by a multi-channel filtering approach of texture analysis. The 2D spatial functions proposed by Daugman are local spatial bandpass filters that achieve the theoretical limit for conjoint resolution of information in the 2D spatial and 2D Fourier domains [6]. The original Gabor elementary functions, in the form proposed by Gabor [7] are generated with a fixed Gaussian, while the frequency of the modulating wave varies.

A signal can be encoded by its projection onto these elementary functions. This decomposition is equivalent to the Gaussian – windowed Fourier transform. However, from recent neurophysiological evidence [8], it is evident that the spatial structure of the receptive fields of simple cells having different sizes is virtually invariant. It has been proposed [9], that an ensemble of simple cells is best modeled as a family of 2D Gabor wavelets sampling the frequency domain in a log-polar manner. A particular Gabor elementary function can be used as a mother wavelet to generate a whole family of Gabor wavelets.

We consider the following 2D family of Gabor wavelets as shown in equation 1.

$$\psi(x,y,\omega_0,\theta) = \frac{\omega_0}{\sqrt{2\pi}\kappa} e^{-\frac{\omega_0^2}{8\kappa^2}\left(4(x\cos\theta+y\sin\theta)^2+(-x\sin\theta+y\cos\theta)^2\right)\cdot\left[e^{j(\omega_0 x\cos\theta+\omega_0 y\sin\theta)}-e^{-\frac{\kappa^2}{2}}\right]} \tag{1}$$

Here, ω_0 is the radial frequency in radians per unit length, and θ is the wavelet orientation in radians. The Gabor wavelet is centered at $(x = 0, y = 0)$ and the normalization factor is such that $\langle\psi,\psi\rangle = 1$, i.e., normalized by L^2 norm, κ is a constant, with $\kappa = \pi$ for a frequency bandwidth of one octave and $\kappa = 2.5$ for a frequency bandwidth of 1.5 octaves. In our algorithm, we have used Gabor wavelets with $\kappa = \pi$. The filter center frequency is determined by the method described in [10 – 11].

The Gabor wavelet filter bank implemented contains forty filters resulting in forty filter outputs. These forty feature images must be processed to obtain a single feature image that can be given as an input to the clustering algorithms to produce the final segmented result. In this paper, two methods are proposed to reduce the dimensionality of feature space and for compression of redundant information without degrading classifier performance. One method uses ICA to obtain a single feature image from the set of feature images and second one is based on image statistics.

Independent Component Analysis

Independent Component Analysis (ICA) attempts to linearly transform the data so as obtain statistically independent components. There has been a considerable amount of research on algorithms for performing ICA. But Hyvarinen's fast and robust fixed-point algorithm is the most popular algorithm because of its better performance. Assume the set of features, $X = [x_1, x_2, \dots, x_n]$, where column vector x_i represents a sample and the total number of samples is *N*. The general model of ICA can be described as follows,

$$X = AS \tag{2}$$

Where, $S = [s_1, s_2, \dots, s_n]$ is the mutually independent source signal. *A* is a square mixing matrix and its column vectors are basis functions. The ICA is to find a separating matrix *W*, so that $Y = WS$, approximates the independent source signal *S*, possibly permuted and rescaled while assuming as little as possible about the natures of *A* and the component original signal *S*. To solve this problem, the key assumption used in ICA is that the original signals S are as statistically independent as possible.

Hyvarinen's Fast and Robust Fixed-point learning algorithms for W can be summarized as follows:

Choose an initial (e.g., random) weight vector *w*

$$w^{+} = \{xg(w^{T}x)\} - E\{g'(w^{T}x)\}w \tag{3}$$

Let $w = \frac{w^{+}}{\|w^{+}\|}$
If not converged, go back to 2)
Where $g(x)$ would be either,

$$g1(x) = \tanh(a_1 x) \tag{4}$$

or

$$g2(x) = x.\exp\left(\frac{-x^2}{2}\right) \tag{5}$$

For most of the ICA algorithms, a preprocessing algorithm known as whitening or sphering is required. The transformed data is zero-mean, decorrelated data.

$$W_p E\{XX^T\} W_p^T = I \tag{6}$$

Where I is the identity matrix. Independent Component Analysis is employed to extract image features, which are mutually independent or as independent as possible described in section 4.

Optimal Feature Selection

The feature images displaying the largest variance are selected out of the forty Gabor filtered outputs by using a simple peak finding algorithm. These selected images are then given as inputs to ICA as a matrix of size $M \times N$ with each row of the matrix consisting of the selected image. The result obtained from ICA is a single image in which the tumor and edema region are highlighted from the surrounding brain tissue.

Thus, instead of processing forty filter outputs only one feature image containing the relevant regions highlighted is obtained. This feature image is given as an input to the clustering algorithms. Thus, optimal feature selection is achieved using ICA for dimensionality reduction of feature vector.

In significant feature processing method, the second to fourth central moments, entropy and coefficients of variation for all the Gabor outputs at all levels of decomposition are computed. The peaks within the array of feature vector are then determined using a simple peak-finding algorithm. The images satisfying this criterion are further compared to obtain the minimum values from each image. Finally, this image consisting of the collected minimum values is given as an input to kmeans and fuzzy c-means clustering algorithm. The results of the clustering algorithms are the final segmented image.

Clustering

Clustering or Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense [12]. Clustering algorithms partition data objects (patterns, entities, instances, observations, units) into a certain number of clusters (groups, subsets, or categories) [13]. Several definitions of clustering have been summarized in [14] and are as follows:

A cluster is "an aggregate of points in the test space such that the distance between any two points in the cluster is less than the distance between any point in the cluster and any point not in it.", "Clusters may be described as continuous regions of this space (d-dimensional feature space) containing a relatively high density of points, separated from other such regions by regions containing a relatively low density of points." Clustering is also referred as unsupervised classification or exploratory data analysis. No labeled data are available [15], [16] in clustering. The goal of clustering is to separate a finite, unlabeled data set into a finite and discrete set of " natural, " hidden data structures, rather than to provide an accurate characterization of unobserved samples generated from the same probability distribution [17], [18]. It is clear from the above discussion that a direct reason for unsupervised clustering comes from the requirement of exploring the unknown natures of the data that are integrated with little or no prior information. Sometimes, the process of labeling data samples may become extremely expensive and time consuming, which also makes clustering a good choice considering the great savings in both cost and time. In addition, cluster analysis provides a compressed representation of the data and is useful in large - scale data analysis.

Kmeans is one of the most popular and well-known clustering algorithms [19], [20], [21] also known as Isodata or c-means algorithm. It can be viewed as a special case of the generalized hard clustering algorithmic scheme when point representatives are used and the squared Euclidean distance is adopted to measure the dissimilarity between vectors x_i and cluster representatives θ_j.

Fuzzy c-means clustering algorithm is a method that is frequently used in pattern recognition. It has the advantage of giving good modeling results in many cases, although it is not capable of specifying the cluster by itself. Fuzzy c-means is a method of clustering which allows one piece of data to belong to two or more clusters. The reason of using fuzzy models is that formulated problem may not be easier to solve computationally. The difficulty associated with probabilistic algorithm is the involvement of the pdf's, for which a suitable model has to be assumed [13].

A fuzzy clustering of the data set X into the m clusters is as follows:

A fuzzy *m*-clustering of data set *X* is defined by a set of functions $u_j: X \to A, j = 1, \dots\dots, m,$ where $A = [0, 1]$. In the case where $A = \{0, 1\}$, a hard *m*-clustering of *X* has been defined. In this case, each vector belongs exclusively to a single cluster. It has been assumed that the number of clusters as well as their shape is known a priori.

θ_j is the parameterized representative of the *j*-th cluster, $\theta = [\theta_1^T, \dots\dots, \theta_m^T]^T$, U is an $N \times m$ matrix whose (i, j) element equals $u_j(x_j)$, $d(x_j, \theta_j)$ is the dissimilarity between x_i and θ_j, and q (> 1) is a parameter called a fuzzifier. Most of the well-known fuzzy clustering algorithms are those derived by minimizing a cost function of the form,

$$J_q(\theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij}^{q}\, d(x_i, \theta_j) \tag{7}$$

With respect to θ and U, subject to the constraints

$$\sum_{j=1}^{m} u_{ij} = 1, \qquad i = 1, \dots, N \tag{8}$$

Where

$$u_{ij} \in [0,1], \qquad i = 1, \dots, N, \qquad j = 1, \dots, m \tag{9}$$

$$0 < \sum_{i=1}^{N} u_{ij} < N, \qquad j = 1, 2, \dots, m \tag{10}$$

The grade of membership of x_i in the *j*-th cluster is related to the grade of membership of x_i to the rest $m - 1$ clusters through eq. Different values of *q* in eq. bias $J_q(\theta, U)$ toward either the fuzzy or the hard clusterings. More specifically, for fixed θ, if $q = 1$, no fuzzy clustering is better than the best hard clustering in terms of $J_q(\theta, U)$. However, if $q > 1$, there are cases in which fuzzy clustering lead to lower value of $J_q(\theta, U)$ than the best hard clustering.

Fuzzy clustering algorithm has been implemented. The generalized fuzzy algorithmic scheme is presented below. The algorithm begins with a initial estimate for θ_j. The parameter updating has been estimated using equation (10). The results are calculated until a termination criterion is satisfied.

Experimental Results

Study of detection of tumor in human brain has been done with the help of T2-weighted MRI image [23] showing metastatic bronchogenic carcinoma. The MR image is showing the axial section of the human brain. The MR image that has been selected for analysis exhibits a diffuse structure in the left occipital lobe. Due to complexity of diffuse structure it is difficult to differentiate the pathological tissue from normal tissue.

The brain image in fig. 1 shows a large mass with surrounding edema, and compression of adjacent mid-brain structures. The MR demonstrates the tumor as an area of high signal intensity on T2-weighted images in a large left temporal region. Also, the T2-weighted image shows the left hemisphere to be greatly swollen. The thorough analysis of given image has been performed with kmeans algorithm with MATLAB 2010b. Various observations have been made. These are discussed as follows.

The kmeans clustered output of fig. 2 is obtained from the feature image that is a result of the selective feature processing algorithm. The tumor region is clearly segmented out from the surrounding edema. Even though, the edema is diffuse from the surrounding brain tissue, it is captured in the final segmentation. Fig. 3 is showing the kmeans clustered result obtained from the feature image obtained from ICA. The result obtained from ICA is smooth as compared to that from significant feature processing.

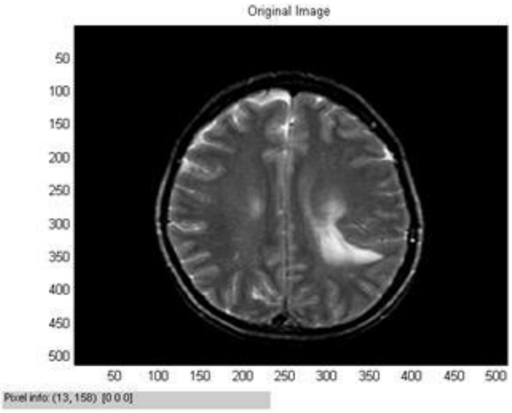

Figure 1: Original Image.

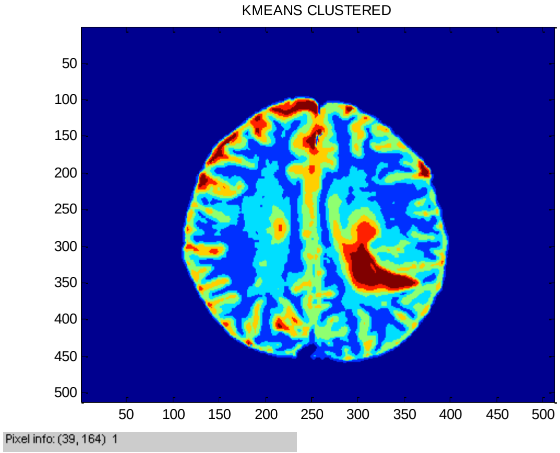

Figure 2: Kmeans clustered image of feature image obtained from significant feature processing

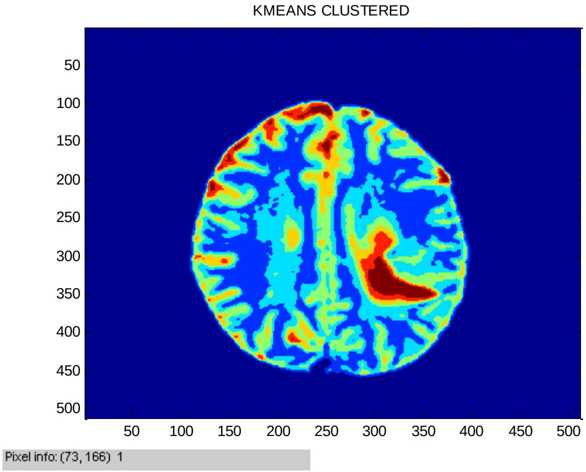

Figure 3: Kmeans clustered image of feature image obtained from ICA

The results of applying fuzzy c-means clustering to the MR image of fig. 1 showing metastatic bronchogenic carcinoma are presented in figures 4 and 5. Fig. 4 is showing the fuzzy c-means clustered output by giving the feature image obtained from significant feature processing and fig. 5 is showing the clustered output obtained from the feature image that is a result of ICA. The clustering result from significant feature processing shows the tumor region with sharp features. Fig. 5 shows tumor region as a smooth region as compared to fig. 4. Both results show the tumor region clearly segmented out from the surrounding brain tissue.

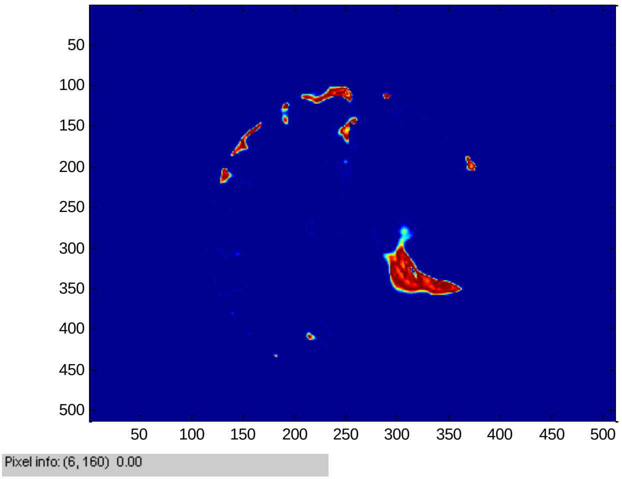

Figure 4: Fuzzy c-means clustered image obtained from feature image from significant feature processing

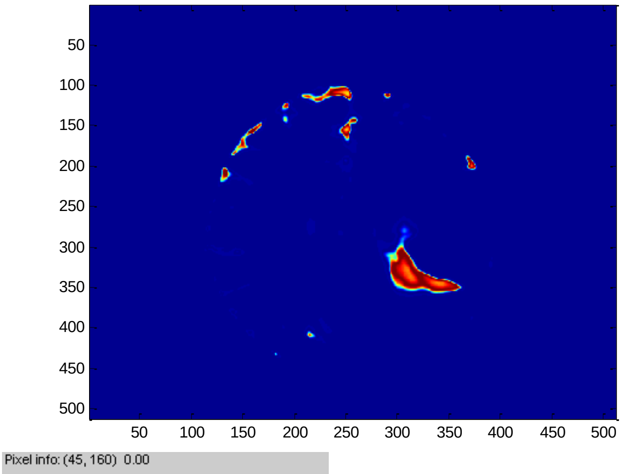

Figure 5: Fuzzy c-means clustered image obtained from feature image from ICA

Rigorous analysis of the optimal feature selection schemes along with clustering has been thoroughly analyzed and well explained in [24]. Instead of using the blob detector popularly used in texture segmentation methods the proposed techniques give accurate segmentation results.

Conclusion

Optimal selection of features done using ICA and significant feature processing gives good segmentation results. The output of the clustering algorithm is a delineation of tumor region from the unaffected brain tissue. Applying the expectation-maximization algorithm directly to the MR image of human brain does not provide satisfactory results. Same is the case with fuzzy c-means clustering. The feature image obtained by dimensionality reduction has the tumor region and its surrounding edema highlighted thus enhancing the performance of the clustering algorithms. As the outputs of the clustering algorithms show, the tumor region is clearly segmented out in the MR image. Comparison of the results from ICA and significant feature processing show that ICA produces smoother results as compared to significant feature processing.

References

[1] M. Masroor Ahmed & Dzulkifli, “Segmentation of brain MR images for tumor detection by combining kmeans clustering and Perona-Malik anisotropic diffusion model,” International Journal of Image Processing, vol. 2, issue 1, pp. 27 – 34, 2008

[2] K. K. Holli, et al., “Texture analysis of MR images of patients with mild traumatic brain injury,” BMC Medical Imaging, vol. 10, no. 8, pp. 1 – 10, 2010

[3] Z. L. S. Nie & Z. Chen, “Application of independent component analysis in MR image segmentation,” 2nd International Conf. on Bioinformatics and Biomedical engg., pp. 2686 – 2689, May 2008

[4] A. Hyvärinen, E. Oja, “Independent component analysis: algorithms and applications,” Neural Networks, vol. 13, pp. 411 – 430, 2000

[5] G. Mirajkar & B. Barbadekar, “Automatic segmentation of brain tumors from MR images using undecimated wavelet transform and gabor wavelets,” 17th IEEE International Conf. on Electronics, Circuits, and Systems, pp. 702 – 705, December 2010

[6] T. S. Lee, “Image representation using 2D Gabor wavelets,” IEEE Trans. On Pattern Anal. Machine Intell., vol. 18, no. 10, pp. 959 – 971, October 1996

[7] D. Gabor, “Theory of communication,” J. IEE, vol. 93, pp. 429 – 459, 1946

[8] R. L. De Valois & K. K. De Valois, “Spatial vision,” Oxford Univ. Press, New York, 1988

[9] J. G. Daugman, “Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression,” IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, no. 7, pp. 1169 – 1179, 1988

[10] J. Krumm & S. A. Shafer, “Local spatial frequency analysis of image texture,” Proceedings on Third International Conference on Computer Vision, pp. 354 – 358, 1990

[11] T. Ezzat, J, Bouvrie, and T. Poggio, “AM-FM demodulation of spectrograms using localized 2D max-Gabor analysis,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, 2007

[12] S. Thilagamani and N. Shanthi, “Literature survey on enhancing cluster quality,” International Journal on Computer Science and Engineering, vol. 02, no. 06, pp. 1999 – 2002, 2010

[13] R. Xu and D. C. Wunsch, II, “Clustering,” IEEE Press, Wiley, NJ, 2009

[14] B. Everitt, “Cluster analysis,” Social Science Research Council, London, 2nd edition, 1980

[15] B. Everitt, S. Landue, M. Leesse, “Cluster analysis,” Arnold, London, 4th edition, 2001

[16] A. Jain and R. Dubes, “Algorithms for clustering data,” Prentice Hall, Englewood Cliffs, NJ, 1988

[17] V. Cherkassy and F. Mulier, “Learning from data: Concepts, theory, and methods,” John Wiley & Sons, New York, NY, 1998

[18] A. Baraldi and E. Alpaydin, “Constructive feedforward ART clustering networks – Part I and II,” IEEE Transactions on Neural Networks, vol. 13, no. 3, pp. 645 – 677, 2002

[19] R. Duda, P. Hart, and D. Stork, “Pattern classification,” John Wiley & Sons, New York, NY, 2nd edition, 2001

[20] G. H. Ball and D. J. Hall, “A clustering technique for summarizing multivariate data,” Behavioral Science, vol. 12, pp. 153 – 155, March 1967

[21] S. P. Lloyd, “Least squares quantization in PCM,” IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129 – 137, March 1982

[22] S. Theodoridis and K. Koutroumbas, “Pattern Recognition,” Academic Press, CA, USA, 2006

[23] K. A. Johnson & J. A. Becker, The whole brain atlas, Harvard Medical School Lecture Notes: Introduction to Neuroimaging

[24] G. S. Mirajkar, “Some investigations into deterministic object detection using wavelets and morphological techniques along with a case study on radar images,” Shivaji University, Kolhapur, India, Ph. D Thesis submitted July 2010