Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 636858, 14 pages
doi:10.1155/2010/636858

Research Article

Validity-Guided Fuzzy Clustering Evaluation for Neural Network-Based Time-Frequency Reassignment

Imran Shafi,[1] Jamil Ahmad,[1] Syed Ismail Shah,[1] Ataul Aziz Ikram,[1] Adnan Ahmad Khan,[2] and Sajid Bashir[3]

[1] *Information and Computing Department, Iqra University, Islamabad Campus, Sector H-9, Islamabad 44000, Pakistan*
[2] *Electrical Engineering Department, College of Telecommunication Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan*
[3] *Computer Engineering Department, Centre for Advanced Studies in Engineering, Islamabad 44000, Pakistan*

Correspondence should be addressed to Imran Shafi, imran.shafi@gmail.com

Received 1 March 2010; Revised 21 May 2010; Accepted 15 July 2010

Academic Editor: Srdjan Stankovic

This paper describes the validity-guided fuzzy clustering evaluation for optimal training of localized neural networks (LNNs) used for reassigning time-frequency representations (TFRs). Our experiments show that the validity-guided fuzzy approach ameliorates the difficulty of choosing correct number of clusters and in conjunction with neural network-based processing technique utilizing a hybrid approach can effectively reduce the blur in the spectrograms. In the course of every partitioning problem the number of subsets must be given before the calculation, but it is rarely known apriori, in this case it must be searched also with using validity measures. Experimental results demonstrate the effectiveness of the approach.

1. Introduction

Clustering is important for pattern recognition, classification, model reduction, and optimization. Cluster analysis plays a pivotal role in solving practical issues related to image and signal processing, bioengineering, medical science, and psychology [1]. The problem of clustering is to partition the data in a given finite data set into a number of appropriate relevant groups. The data can be quantitative, qualitative, or a mixture of both. In classical cluster analysis, these groups are required to form a partition such that the degree of the association is strong for the objects falling in a particular group than to members of other groups. The term "association" or "similarity" is mathematical similarity, measured in some well-defined sense [2]. Moreover, finding out the appropriate number of groups for a particular data set is also a quantitative task. Different classifications based on the algorithmic approach of the clustering techniques, include the partitioning, hierarchical, graph-theoretic, and objective function-based methods [3].

Localized neural processing is considered important due to numerous reasons. Firstly it is a well-known fact that different parts of the human brain are designated to perform different tasks [4]. The nature of the task imposes certain structure for the region resulting in a structure-function correspondence. Also, different regions in the brain compete to perform a task and the task is assigned to the winning region. Mimicking the behavior of brain, artificial neural networks (ANNs) may also be employed based on these arguments. An image contains structural information with low and high-frequency contents with a blurred version losing most of its high-frequency information. The objective of any deblurring system is to restore this information by gaining sufficient knowledge about the blur function. However, information is generally lost at various scales in different regions, which must be taken into account [5]. For example, the edges and the flat regions are blurred simultaneously but at the different rate. This favours the idea of subdividing the data into appropriate groups. A second reason is the problem of overtraining for the ANN which causes loss of the generalization ability. If only a single ANN is used, it may end up memorising the training data and may adjust its weights to any noise. Yet another reason is specific to the case of image processing, that is, if an ANN is

trained by an entire image containing different distribution characteristics for data corresponding to different structures in the image. It may attempt to represent different structures by finding a common ground between the different data distributions and thus limits the recognition ability of the network. This forces one network to learn distant input patterns, causing training to slow down in attempting to represent input data that are significantly different [6].

During the last decade there has been spectacular growth in the volume of research on studying and processing the signals with time-dependant spectral content. For such signals we need techniques that can show the variation in the frequency of the signal over time. Although some of the methods may not result in a proper distribution, these techniques are generally known as time-frequency distributions (TFDs). The TFDs are aimed to obtain the temporal and spectral information of the nonstationary signals with high-resolution without any potential interference [7]. These characteristics are necessary for an easy visual interpretation and a good discrimination between known patterns for nonstationary signal classification tasks [8]. They were partly addressed by the development of the Choi-Williams distribution (CWD) [9], followed by many other advanced techniques. Concept of scale is also used by some authors as another time-varying signal analysis tool rather than frequency, such as the scalogram [10], the affine smoothed pseudo-Wigner-Ville distribution (WVD) [11], or the Bertrand distribution [12]. Some TFDs are proposed to adapt to the signal time-frequency (t-f) changes. The example of such adaptive TFDs includes the classical work by Flandrin et al. in the form of the reassigned TFDs [13], and by Jones et al. in the form of the high-resolution TFD [14], the signal-adaptive optimal-kernel TFD [15], and the optimal radially Gaussian kernel TFD [16]. For the analysis of signals with varying IF, higher-order distributions are used [17, 18]. There are some newer techniques based on nonparametric snakes for the reassignment of TFDs [7], neural networks [19], sparsity constraint of energy distribution [20], and t-f autoregressive moving-average spectral estimation [21] to improve the resolution in the t-f domain. A comparison of high-resolution TFDs for test signals can be found in [22]. In order to provide an accurate IF estimation even when the signal phase varies significantly within a few signal samples, the distributions with complex lag argument have been introduced [23–25] and improved [26, 27].

The neural network-based method fundamentally involves training and selection of a set of suitably chosen ANNs that provide the improved TFDs (NTFDs) in the testing phase [28]. The vectors from the training t-f images are required to be clustered. The determination of the optimum cluster number is important due to localized neural processing for the reasons mentioned earlier. The goal of this paper is to evaluate fuzzy clustering to achieve this task automatically based on cluster validity measures and more efficiently by checking quality of clustering results. Fuzzy clustering methods allow objects to belong to several clusters simultaneously, with different degrees of membership. It is believed that, in many factual situations, fuzzy clustering is more intuitive choice than hard clustering. It is so because data vectors on the boundaries between two clusters are assigned membership degrees between 0 and 1 indicating their partial memberships. On the contrary, the analytic functions defined for hard clustering methods are not differentiable due to their discrete nature, causing analytical and algorithmic intractability of algorithms. A detailed treatment of the subject can be found in the classical attempt by Bezdek [29], Hopner [2], and Babuska [30].

The objective of this work is to explore the effectiveness of the fuzzy clustering for Bayesian regularized neural network model to obtain high-resolution reassigned TFDs. No assumption is made about any prior knowledge about the components present in the signal. The goal of the proposed neurofuzzy reassignment method is to get a high-resolution TFD which can provide an easy visual interpretation and a good discrimination between known patterns for nonstationary signal classification tasks. The rest of the paper is structured as follows. Section 2 gives a brief review of some popular related fuzzy clustering algorithms, various scalar validity measures, and some information theoretic criteria. We also suggest a modification in an existing instantaneous concentration measure that can provide TFDs' performance in a more efficient manner. Section 3 introduces the method proposed in this paper, combining fuzzy clustering with neural networks to achieve high concentration and good resolution on the t-f plane. This hybrid method enables us to determine the optimal number of clusters for localized neural network processing searched using various cluster validity measures and checking the quality of clustering results. Section 4 presents the results of applying the proposed method to both synthetic and real-life signals. The discussion on the determination of the optimal number of the cluster using the validity measures is also given in this section. Finally, Section 5 concludes the paper and discusses the major contribution.

2. Background

The main potential of clustering is to detect the underlying structure in data, not only for classification and pattern recognition, but for model reduction and optimization. For this reason data vectors are divided into clusters such that similar vectors belong to the same cluster. The resulting data partitioning is expected to improve data understanding by the ANN by avoiding learning distant input patterns. Fuzzy clustering approaches assign different degrees of membership to data vectors associating them to several clusters simultaneously. In real applications there is hardly a sharp boundary between clusters, and fuzzy clustering is often better suited for the data. In this way, data on the boundaries between several clusters are not forced to belong to one of the clusters.

2.1. Fuzzy Clustering Algorithms. The objective of clustering is to partition the finite data set $\mathbb{Q} = [q_1, q_2, \ldots, q_N]$ into c clusters where $2 \leq c < N$. The value of c is assumed to be known a priori, or it is a trial value to be validated [29]. The

structure of the partition matrix $\Lambda = [\lambda_{ik}]$:

$$\Lambda = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,c} \\ \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N,1} & \lambda_{N,2} & \cdots & \lambda_{N,c} \end{pmatrix}. \tag{1}$$

Fuzzy partition allows λ_{ik} to attain real values in $[0, 1]$. A $N \times c$ matrix represents the fuzzy partitions, with the following conditions:

$$\begin{aligned} &\lambda_{ik} \in [0,1], \quad 1 \le i < N,\ 1 \le k < c, \\ &\sum_{k=1}^{c} \lambda_{ik} = 1, \quad 1 \le i < N, \\ &0 < \sum_{i=1}^{N} \lambda_{ik} < N, \quad 1 \le k < c. \end{aligned} \tag{2}$$

The fuzzy partitioning space for $\mathbb{Q}$ is defined to be the set

$$F_{fc} = \left\{ \Lambda \in \mathbb{R}^{N \times c} \mid \lambda_{ik} \in [0,1], \forall i, k; \ \sum_{k=1}^{c} \lambda_{ik} = 1, \forall i;\ 0 < \sum_{i=1}^{N} \lambda_{ik} < N, \forall k \right\}. \tag{3}$$

The ith column of Λ contains values of the membership function of the ith fuzzy subset of $\mathbb{Q}$.

2.1.1. Fuzzy c-Means Algorithm. The most prominent fuzzy clustering algorithm is the fuzzy c-means, a fuzzification of K-means hard partitioning method. It is based on the minimization of an objective function called c-means functional, defined by [31]

$$\begin{aligned} &\Gamma(\mathbb{Q}; \Lambda, V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\lambda_{ik})^m \|q_k - v_i\|_A^2, \\ &\text{with } V = [v_1, v_2, \ldots, v_c], \quad v_i \in \mathbb{R}^n, \end{aligned} \tag{4}$$

where A_i is a set of data vectors in the ith cluster and V is a vector of cluster prototypes or cluster centers such that $v_i = (\sum_{k=1}^{N_i} q_k)/N_i$, $q_k \in A_i$, is the mean for data vectors over cluster i with N_i being the number of data vectors in A_i. Here the vector of cluster prototypes have to be computed, and $D_{ikA}^2 = \|q_k - v_i\|_A^2 = (q_k - v_i)^T A (q_k - v_i)$ is a squared inner-product distance norm. The c-means functional given by (4) is a measure of the total variance of q_k from v_i. The minimization of (4) is a nonlinear optimization case that can be solved by various methods like group coordinate minimization, over-simulated annealing and genetic algorithms. The fuzzy c-means algorithm solves it by a simple Picard iteration method through the first-order conditions for stationary points of (4).The fuzzy c-means algorithm computes with the standard Euclidean distance norm, which induces hyperspherical clusters. Hence it can only detect clusters with the same shape and orientation.

2.1.2. The Gustafson-Kessel Algorithm. Gustafson and Kessel extended the standard fuzzy c-means algorithm by employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set [32, 33]. Each cluster has its own norm-inducing matrix A_i, which yields slightly, different inner-product norm:

$$D_{ikA_i}^2 = (q_k - v_i)^T A_i (q_k - v_i), \quad 1 \le i < c,\ 1 \le k < N. \tag{5}$$

Here A_i are used as optimization variables in the c-means functional, thus allowing each cluster to adapt the distance norm to the local, topological structure of the data. Let $A = [A_1, A_2, \ldots, A_c]$ denote a c-tuple of the norm-inducing matrices. The objective functional of the Gustafson-Kessel algorithm is defined by

$$\Gamma(\mathbb{Q}; \Lambda,\ V, A) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\lambda_{ik})^m D_{ikA_i}^2. \tag{6}$$

It is important to highlight that Γ can be minimized by simply making A_i less positive definite. This is accomplished by allowing the matrix A_i to vary with its determinant fixed, that is, $\|A_i\| = \rho_i$ with ρ_i being fixed for each cluster. The expression for A_i can be expanded by use of Lagrange multiplier method as

$$A_i = [\rho_i \det(F_i)]^{1/n} F_i^{-1}, \tag{7}$$

where F_i is the fuzzy covariance matrix of the ith cluster defined by

$$F_i = \frac{\sum_{k=1}^{N} (\lambda_{ik})^m (q_k - v_i)(q_k - v_i)^T}{\sum_{k=1}^{N} (\lambda_{ik})^m}. \tag{8}$$

2.2. Validation Measures. Cluster validity measures are used to confirm whether a given fuzzy partition fits to the data all. There are various scalar validity measures proposed in the literature; however, none of them is perfect by oneself. Therefore, several measures have been used, which are described below.

2.2.1. Partition Coefficient (PC). It measures the amount of "overlapping" between cluster, defined as follows [29]:

$$\mathrm{PC}(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \left(\lambda_{ij}\right)^2, \tag{9}$$

where λ_{ij} is the membership of data point j in cluster i. The disadvantage of PC is the lack of direct connection to some property of the data themselves. The optimal number of the cluster is at the maximum value.

2.2.2. Classification Entropy (CE). It is similar to the PC that measures the fuzziness of the cluster partition, defined by

$$\mathrm{CE}(c) = -\frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \lambda_{ij} \log\left(\lambda_{ij}\right). \tag{10}$$

2.2.3. Partition Index (SC). It is a sum of individual cluster validity measures normalized through the division by the fuzzy cardinality of each cluster [3]. A lower value of SC indicates a better partition, mathematically defined as

$$\mathrm{SC}(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{N} \left(\lambda_{ij}\right)^{m} \left\| q_j - v_i \right\|^2}{N_i \sum_{k=1}^{c} \left\| v_k - v_i \right\|^2}. \tag{11}$$

2.2.4. Seperation Index (S). On the contrary of above measure, this index uses a minimum-distance separation for partition validity, defined as [3]

$$\mathrm{S}(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} \left(\lambda_{ij}\right)^{2} \left\| q_j - v_i \right\|^2}{N_{\min_{i,k}} \left\| v_k - v_i \right\|^2}. \tag{12}$$

2.2.5. Xie and Beni's Index (XB). It aims to quantify the ratio of the total variation within clusters and the separation of clusters defined by [34]

$$\mathrm{XB}(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} \left(\lambda_{ij}\right)^{2} \left\| q_j - v_i \right\|^2}{N_{\min_{i,j}} \left\| q_j - v_i \right\|^2}. \tag{13}$$

A lower value of XB indicates a better partition and the optimal number of clusters.

2.2.6. Dunn's Index (DI). This is proposed to identify compact and well-separated clusters and the result of clustering has to be calculated again and again. Due to this, Dunn's index is not very popular because as c and N increase calculation becomes computationally very expensive. It is defined as [31]

$$\mathrm{DI}(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{k \in c} \left\{ \max_{x, y \in c} d(x, y) \right\}} \right\} \right\}, \tag{14}$$

where $d(x, y)$ is the dissimilarity function between two cluster.

2.2.7. Alternative Dunn Index (ADI). Here the dissimilarity function $d(x, y)$ between two clusters is rated in value from beneath by the triangular nonequality $d(x, y) \geq |d(y, v_j) - d(x, v_j)|$ with an aim to simplify the calculation of original DI. It is defined as

$$\mathrm{ADI}(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x_i \in c_i, x_j \in c_j} \left| d\left(y, v_j\right) - d\left(x, v_j\right) \right|}{\max_{k \in c} \left\{ \max_{x, y \in c} d(x, y) \right\}} \right\} \right\}, \tag{15}$$

where v_j is the cluster center of the jth cluster.

2.3. TFDs' Information Theoretic Criteria. The estimation of signal information and complexity in the t-f plane is quite challenging. A criterion for comparison of time-frequency distributions may be defined in various ways [8]. An orderly way is to assume that the "ideal" TFD is the one producing the Dirac pulse at the IF of an arbitrary frequency modulated signal; elsewhere the value of the distribution should be zero [35]. However, this requires well-defined mathematical representations of various TFDs. Alternatively for a monocomponent signal, performance of its TFD is conventionally defined in terms of its energy concentration about the signal IF. To measure distribution concentration for monocomponent signals, some quantities in the statistics were the inspiration for defining measures in the form of the distribution energy [16], the ratio of distribution norms [36], and the famous Rényi entropy [37]. Some other measures have been based on the definition of duration of time-limited signals [38] and the combined characteristics of TFDs [39]. Whereas for multicomponent signals, resolution is equally important. The good t-f resolution of the signal components requires a good energy concentration for each of the components and a good suppression of any undesirable artifacts. The resolution may be measured by the minimum frequency separation between the component's main lobes for which their magnitudes and bandwidths are still preserved [39]. Although different concentration and resolution criteria can be found in the literature, but most of them are related to each other. Therefore, we have compiled a compact list of measures that are briefly reviewed as follows.

2.3.1. Normalized Rényi Entropy Measures. The terms entropy, uncertainty, and information are used more or less interchangeably and is the measure of information for a given probability density function. Minimizing the entropy in a TFD is equivalent to maximizing its concentration and resolution [36].

Rényi entropy is a more appropriate way of measuring the t-f uncertainty sidestepping the negativity issue in Shannon entropy. It is derived from the same set of axioms as the Shannon entropy [37], given as

$$E_{\mathrm{RE}_\alpha} = \frac{1}{1 - \alpha} \log_2 \left(\sum_n \sum_\omega Q^\alpha(n, \omega) \right); \tag{16}$$

where α is the order of Rényi entropy, which is taken as 3 being the smallest integer value to yield a well-defined, useful information measure for a large class of signals. However, the Rényi entropy measure with $\alpha = 3$ does not detect zero mean CTs, so normalization either with signal energy or distribution volume is necessary [37].

By definition Rényi entropy normalized by the signal energy is given by

$$E_{N\mathrm{RE}_\alpha} = \frac{1}{1 - \alpha} \log_2 \left(\frac{\sum_n \sum_\omega Q^\alpha(n, \omega)}{\sum_n \sum_\omega Q(n, \omega)} \right), \quad \text{with } a \geq 2. \tag{17}$$

The Rényi entropy normalized by the distribution volume is given by

$$E_{N\mathrm{RE}_\alpha} = \frac{1}{1 - \alpha} \log_2 \left(\frac{\sum_n \sum_\omega Q^\alpha(n, \omega)}{\sum_n \sum_\omega |Q(n, \omega)|} \right), \quad \text{with } a \geq 2. \tag{18}$$

If the distribution contains oscillatory values, then summing them in absolute value means that large CTs will decrease this measure, indicating smaller concentration due to CTs appearance.

2.3.2. Ratio of Norms-Based Measure. Another measure of concentration is defined by dividing the fourth power norm of TFD $Q(n,\omega)$ by its second power norm, given as [37]

$$E_{JP} = \frac{\sum_n \sum_\omega |Q(n,\omega)|^4}{\left(\sum_n \sum_\omega |Q(n,\omega)|^2\right)^2}. \tag{19}$$

The fourth power in the numerator favors a peaky distribution. To obtain the optimal distribution for a given signal, the value of this measure should be the maximum.

2.3.3. Stankovic Measure. This is a simple criterion for objective measurement of TFD concentration that makes use of the duration of time-limited signals [38]. Its discrete form is expressed as

$$J[Q(n,\omega)] \equiv J_\beta^\beta = \left(\sum_n \sum_\omega |Q(n,\omega)|^{1/\beta}\right)^\beta \tag{20}$$

with $\sum_n \sum_\omega Q(n,\omega) = 1$ being the normalized unbiased energy constraint, and $\beta > 1$. The best choice according to this criterion (optimal distribution with respect to this measure) is the distribution that produces the minimal value of $J[Q(n,\omega)]$.

2.3.4. Boashash Performance Measures. The characteristics of TFDs that influence their resolution, such as components concentration and separation and interference terms minimization, are combined to define separate quantitative criterion for concentration and resolution [39].

Instantaneous Concentration Measure. For a given time slice $t = t_0$ of TFD of an n-component signal $z(t) = \sum z_n(t)$, the concentration performance can be quantified by [39]

$$\mathbb{C}_n(t) = \frac{A_{s_n}(t)}{A_{m_n}(t)} \frac{V_{i_n}(t)}{f_{i_n}(t)}, \tag{21}$$

where $\mathbb{C}_n(t)$, $V_{i_n}(t_0)$, $f_{i_n}(t_0)$, $A_{s_n}(t_0)$, and $A_{m_n}(t_0)$ denote, respectively, the concentration measure, instantaneous bandwidth, the IF, the side lobe magnitude, and the main lobe magnitude of the nth component at time $t = t_0$. The instantaneous concentration performance of a TFD will improve if it minimizes side lobe magnitude relative to the main lobe magnitude and main lobe bandwidth about the signal IFs for each signal component.

A Suggested Modification. To account for the effects of TFD parameters like instantaneous bandwidth, IF, side lobe magnitude, and the main lobe magnitude more independently, we suggest a modification in the above mentioned Boashash concentration measure given by (21). For this two terms, $A_{s_n}(t)/A_{m_n}(t)$ and $V_{i_n}(t)/f_{i_n}(t)$, are combined into a sum, rather than a product. This new measure can give a better picture of TFDs' instantaneous concentration performance even for those having no side lobes. The modified instantaneous concentration measure for each signal component of an n-component signal $z(t) = \sum z_n(t)$ for a given time slice $t = t_0$ can be defined as

$$\mathbb{C}_n(t) = \frac{A_{s_n}(t)}{A_{m_n}(t)} + \frac{V_{i_n}(t)}{f_{i_n}(t)}. \tag{22}$$

The good performance of a TFD is characterized by a close to zero value of this measure.

Normalized Instantaneous Resolution Measure. The normalized instantaneous resolution performance measure $\mathbb{R}_i$ is expressed as [39]

$$\mathbb{R}_i(t) = 1 - \frac{1}{3}\left[\frac{A_s(t)}{A_m(t)} + \frac{1}{2}\frac{A_x(t)}{A_m(t)} + (1 - D(t))\right], \quad 0 < \mathbb{R}_i(t) < 1, \tag{23}$$

where $A_m(t) = \sum A_{m_n}(t)/2$, $A_s(t) = \sum A_{s_n}(t)/2$, and $A_x(t)$ denote the average magnitude of the components' main lobes, the average magnitude of the components' side lobes, and the CT magnitude of any two adjacent signal components. $D(t) = 1 - V_i(t)/\Delta f_i(t)$ is a measure of the components' main lobes separation in frequency with $V_i(t) = \sum V_{i_n}/2$ as the components' main lobes average instantaneous bandwidth, and $\Delta f_i(t) = f_{i_{n+1}}(t) - f_{i_n}(t)$ as the difference between the components' IFs. The measure $D(t)$ requires computations for each adjacent pair of components present in the signal indicated by subscript n. The value of the measure $\mathbb{R}_i$ will be close to one for good performing TFDs and zero for poor performing ones (TFDs with large interference terms and components poorly resolved).

3. The Hybrid Neurofuzzy Method

In this paper, we address the concentration and resolution problem in the t-f plane by combining fuzzy clustering and localized neural network processing in a nonstationary setting. The proposed method is composed of two stages for achieving high concentration and good resolution of the image in the t-f plane. The first stage is the optimal fuzzy clustering of vectored image data in the t-f plane. The second stage deals with the localized neural network processing. A self-explanatory block diagram is depicted in Figure 1.

3.1. Time-Frequency Image Vectoring and Fuzzy Clustering. The spectrogram and preprocessed WVD of various known signals constitute the input and target TFDs for the ANN. The ANN may be used to extract mathematical patterns and detect trends in the spectrogram and WVD that are too complex to be noticed by any other technique. The ANN has an ability to learn based on the data given for training and performs well on complicated test cases of a similar nature [4]. We consider a signal containing parallel chirps and another signal containing a sinusoidal modulated FM

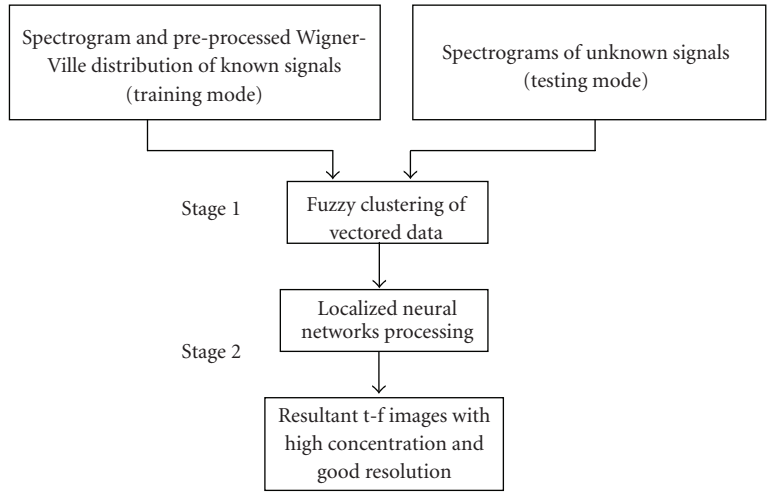

Figure 1: Block diagram of the proposed hybrid neurofuzzy method.

component. The discrete mathematical forms of the training signals are as follows.

$$x_1(n+1) = \exp\big(j\omega_1(n+1)n\big) + \exp\big(j\omega_2(n+1)n\big),$$
$$x_2(n+1) = \exp\left(\left(\left(\frac{j\pi}{2}\right) - j\pi\omega(n+1)\right)n\right), \tag{24}$$

where $\omega_1(n+1) = (\pi n)/4N$, $\omega_2(n+1) = (\pi/3 + (\pi n)/4N)$, and $\omega(n+1) = 0.1\sin(2\pi(n/N))$. Here N refers to the total number of sampling points in these signals ($N = 3000$ for the training signals).

The WVD of these signals suffers from CTs which inhibit its use as target [4]. The CTs are eliminated by multiplying the WVD with the spectrogram of signals. Next, both the spectrogram and preprocessed WVD is converted to 1×3 pixel vectors. This vector size is determined after experimenting with various combinations and ascertaining the effect on the visual quality of the outcome from the trained ANN model. Subsequently the arithmetic means of the vectors from the WVDs are obtained. This is with a view that the IF can be computed by averaging frequencies at each time instant; a definition suggested by many researchers [40, 41]. Vectors from the training spectrograms are grouped in an optimal fashion by the Gustafson-Kessel fuzzy partitioning validated by various objective measures. These vectors are paired with the corresponding average values from the target TFDs for training and subsequent selection of localized neural networks.

3.2. Localized Neural Network Processing. The selected ANN's topology includes 40 hidden units in a single hidden layer with feed-forward back-propagation neural network architecture. The hidden layer consists of sigmoid neurons followed by an output layer of positive linear neurons, respectively. The selected ANN architecture is trained by the Bayesian regularized Lavenberg-Marquardt back-propagation (LMB) algorithm. This choice of the training algorithm and number of hidden neurons and layers are based on some empirical studies [42]. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. The linear output layer lets the network produce values outside the range -1 to $+1$. The LMB training algorithm is the variation of Newton's method that is designed for minimizing sums of squares of nonlinear functions [4]. The Bayesian framework of David Mackay smoothes the network response and avoids overtraining. Also, it helps in determining the optimal regularization parameters in an automated fashion [28].

3.2.1. Multiple Neural Networks Training and Selecting Localized Neural Networks. The spectrogram and preprocessed WVD of the two signals are used to train the multiple neural networks. Fuzzy clustering of the data results in its optimal partitions for which analysis is performed and discussed in the next section. The training vectors from the spectrogram are distributed in different groups by Gustafson-Kessel fuzzy clustering algorithm. They are paired with target values from the preprocessed WVD. It is desired that the ANN does well on data it has not seen before and is not overtrained. For this, data pairs are grouped into separate training and validation sets. The error is monitored on the validation set that does not take part in the training. The training is stopped whenever the ANN tries to learn the noise in the training set.

Under the Bayesian framework, multiple ANNs are trained for each cluster using x_i as the training vector and y_i as its target value. This is advantageous for two main reasons. Firstly, the weights are initialized to random values and may not converge in an optimal fashion. Secondly, an early stopping to avoid overfitting the data may result in poorly trained network [43]. The performance parameters include the mean-square error reached in the last epoch, maximum number of epochs, performance goal, maximum validation failures, and the performance gradient. These can be accessed to find out the most optimally trained ANN out of multiple ANNs for each cluster. These selected ANNs for all clusters are termed as the localized neural networks (LNNs).

3.2.2. Localized Neural Networks' Testing and Data Postprocessing. In the testing phase, the spectrograms of unknown signals are first converted to vectors of specified length. These vectors are fuzzy clustered using Gustafson-Kessel fuzzy clustering algorithm. The test vectors are given as input to the localized neural networks, and the results are obtained. The resultant data is postprocessed to constitute the TFD image. This is achieved by zero padding the resultant scalar values to form the vectors. Next, these vectors are declustered and placed at the appropriate positions to form the two-dimensional image matrix by retrieving their known index values.

4. Results and Discussion

4.1. Cluster Analysis. Using the validity measures described in Section 2.2, both the hard and fuzzy clustering techniques can be compared. For this, a synthetic data set is used to demarcate the index values. However, these experiments and evaluations are not the proposition of this work and will be discussed elsewhere. On the score, of the values of these validity measures for fuzzy clustering the Gustafson-Kessel clustering has the very best results. The Gustafson-Kessel fuzzy clustering algorithm forces each cluster to adapt the distance norm to the local, topological structure of the data points. It uses the Mahalanobis distance norm. There are two numerical problems with this algorithm. When an eigenvalue is zero or when the ratio between the maximal and the minimal eigenvalue is very large, the matrix is nearly singular. As a result, the normalization to a definite volume fails, as the determinant becomes zero. The problem is solved if the ratio between the maximal and minimal eigenvalue is kept smaller than some predetermined threshold. Another problem appears if the clusters are vastly extended in the direction of the largest eigenvalues. In this case, the computed covariance matrix cannot estimate the underlying data distribution, so a scaled identity matrix can be added to the covariance matrix to resolve the issue.

In the course of partitioning the data vectors, fuzzy Gustafson-Kessel algorithm is applied and the optimal number of subsets is searched with using validity measures before the localized neural network processing stage. During this optimization process, all parameters are fixed to the default values and number of clusters are varied such that $c \in [2\ \ 14]$. The values of the validity measures depending from the number of the cluster are plotted and embraced in Table 1. It is important to mention that no single validation index is perfect and reliable only by itself. The optimal value can be only detected with the comparison of all the results. We choose a number of clusters so that adding another cluster does not add sufficient information. This means that either marginal gain drops or differences become insignificant between the values of a validation index. The PC and CE suffer from drawbacks of their monotonic decrease with the number of clusters and the lack of direct connection to the data. On the score of Figures 2(a) and 2(b), the number of clusters can be only rated to 3. In Figures 2(c), 2(d) and 2(e), SC and S hardly decreases at the $c = 3$ point. The XB index reaches this local minimum at $c = 10$. However, the optimal number of clusters are chosen to 3 based on the fact that SC and S are more useful, which is confirmed by the Dunn's index too in Figure 2(f). The results of ADI are not validated enough to confirm its reliability.

4.2. Test Cases. There are many advanced techniques proposed in past 15 years attempting to improve the energy concentration in the t-f domain. The results of neural network-based approach have been compared to the results obtained by some traditional as well as recently introduced high-resolution t-f techniques. The list includes the WVD, the CWD, the traditional reassignment method [13], the optimal radially Gaussian kernel method [16], and the t-f autoregressive moving-average spectral estimation method [21]. An empirical judgment on TFDs' performance is possible by objective assessment made by some objective criteria discussed in Section 2.3. We have compiled a compact and meaningful list of objective measures that include the ratio of norms based measure [36], normalized Rényi entropy measure [37], Stankovic measure [38], and Boashash performance measures [39]. The first two multicomponent test cases include two synthetic signals. By using synthetic signals it is verified that the proposed approach produces more accurate representations. Once it is numerically confirmed that the proposed method works more accurately, then it is applied to a real-life example.

4.2.1. Synthetic Test Cases. The first synthetic signal contains two sinusoidal FM components and two chirps intersecting each other. The second test case contains two significantly close parallel chirps to evaluate the TFDs' instantaneous performance by the measures suggested in [39]. The spectrograms of these signals are shown in Figures 3(a) and 4(a), respectively, referred to as test image 1 (TI 1) and test image 2 (TI 2). We consider the first synthetic signal under noisy environment.

The two synthetic signals are used to confirm the proposed scheme's performance at the intersection of the IFs and closely spaced components. This is with a view that estimation of the IF is rather difficult in these situations. The first signal is a four-component signal containing two sinusoidal FM component and two chirps intersecting each other. Its discrete mathematical form is given as

$$x_1(n+1) = \sin\left[\frac{3\pi}{2} + 0.1\pi \sin\left(\frac{2\pi n}{N}\right)n\right] + \exp\left(\frac{j\pi n}{4N}n\right) + \exp\left\{j\left(4\pi - \frac{\pi n}{4N}\right)n\right\}. \tag{25}$$

The additive Gaussian noise of variance 0.01 is added to signal to consider the performance of the algorithm under noise. The noisy spectrogram of the signal is shown in Figure 3(a). The frequency separation is low enough and

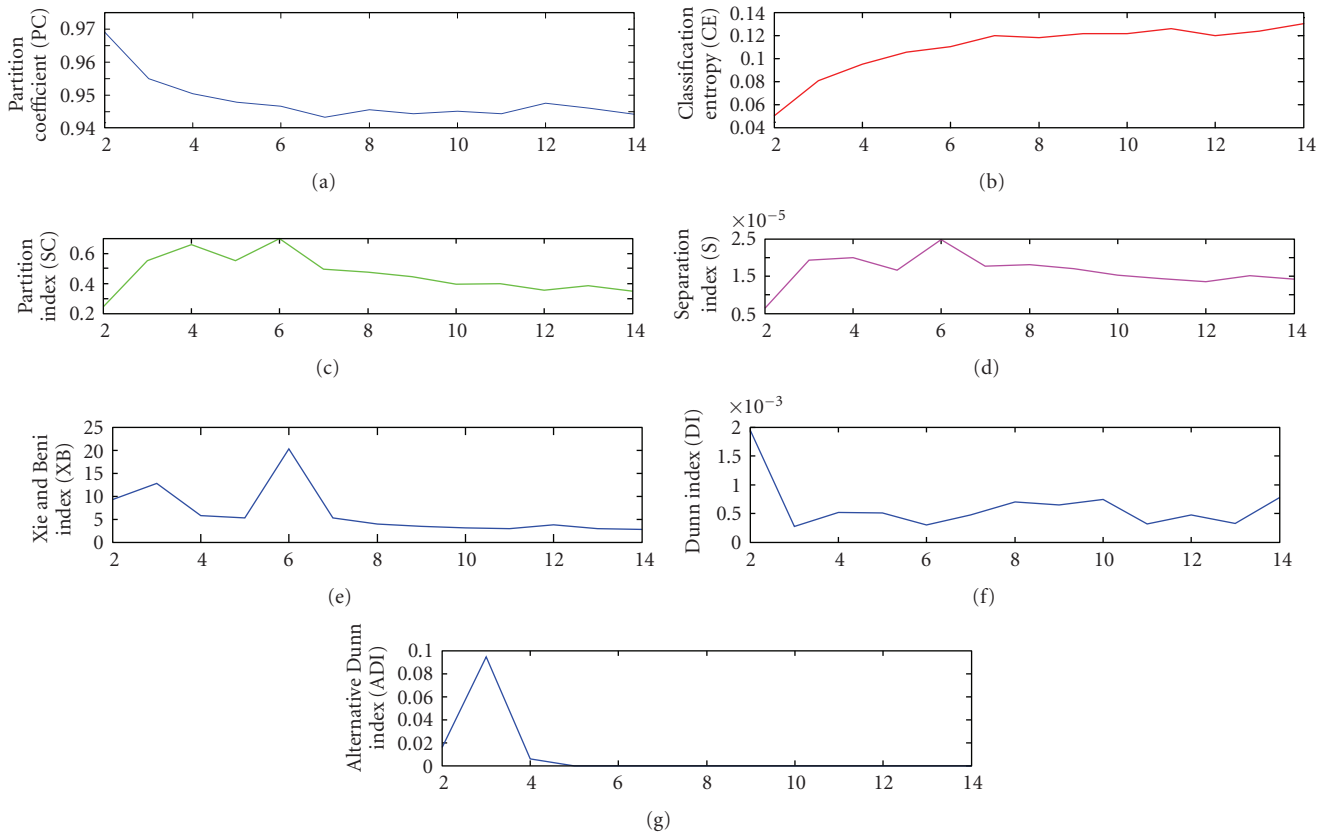

Figure 2: Values of (a) partition coefficient (PC), (b) classification entropy (CE), (c) partition index (SC), (d) separation index (S), (e) Xie and Beni's index (XB), (f) Dunn's index (D), and (g) alternative Dunn index (ADI) for various clusters.

Table 1: Validity measures' values for different clusters.

		Cluster validity measures						
		PC	CE	SC	$S_{(1.0e-004*)}$	XB	DI	ADI
Number of clusters	2	0.9692	0.0502	0.2436	0.0644	9.2561	0.0019	0.0159
	3	0.9549	0.0809	0.5518	0.1926	12.737	0.0003	0.0947
	4	0.9505	0.0951	0.6579	0.1998	5.7947	0.0005	0.0061
	5	0.9479	0.1056	0.5530	0.1666	5.2100	0.0005	0
	6	0.9466	0.1102	0.6968	0.2486	20.314	0.0003	0
	7	0.9433	0.1197	0.4960	0.1769	5.1746	0.0005	0
	8	0.9455	0.1179	0.4736	0.1808	3.9703	0.0007	0
	9	0.9444	0.1215	0.4446	0.1705	3.4385	0.0007	0
	10	0.9451	0.1214	0.3962	0.1527	3.1424	0.0007	0
	11	0.9444	0.1258	0.3976	0.1434	2.9176	0.0003	0
	12	0.9476	0.1197	0.3560	0.1358	3.8053	0.0005	0
	13	0.9461	0.1238	0.3847	0.1508	2.8575	0.0003	0
	14	0.9442	0.1303	0.3493	0.1415	2.7786	0.0008	0

avoids intersection between the two components (sinusoidal FM and chirp components) in between 100–180 Hz and 825–900 Hz near 0.7 second.

The TFDs' instantaneous concentration and resolution performance are evaluated by Boashash instantaneous performance measures using another test case from [39]. The authors in [39] have specifically found the modified B distribution ($\beta = 0.01$) as the best performing TFD for this signal at the middle. The signal is defined as:

$$x_2(n) = \cos\left(2\pi\left(0.15t + 0.0004t^2\right)\right) + \cos\left(2\pi\left(0.2t + 0.0004t^2\right)\right) \quad (26)$$

The spectrogram of the signal is shown in Figure 4(a).

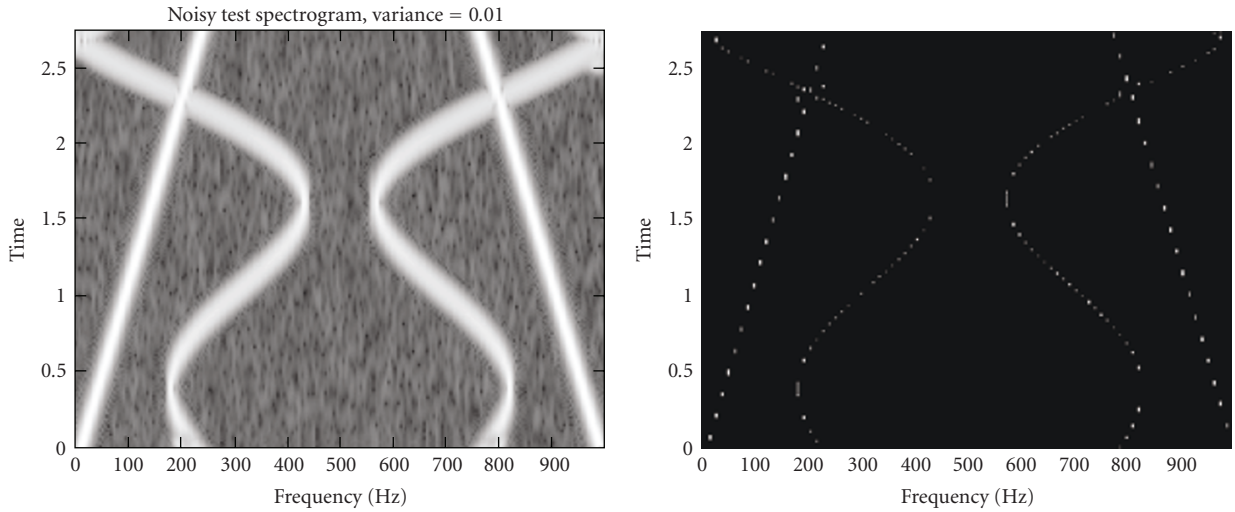

Figure 3: TFDs of a synthetic signal consisting of two sinusoidal FM component and two chirp components. (a) Spectrogram (TI 1) (Hamm, $L = 90$) with additive Gaussian noise and (b) NTFD.

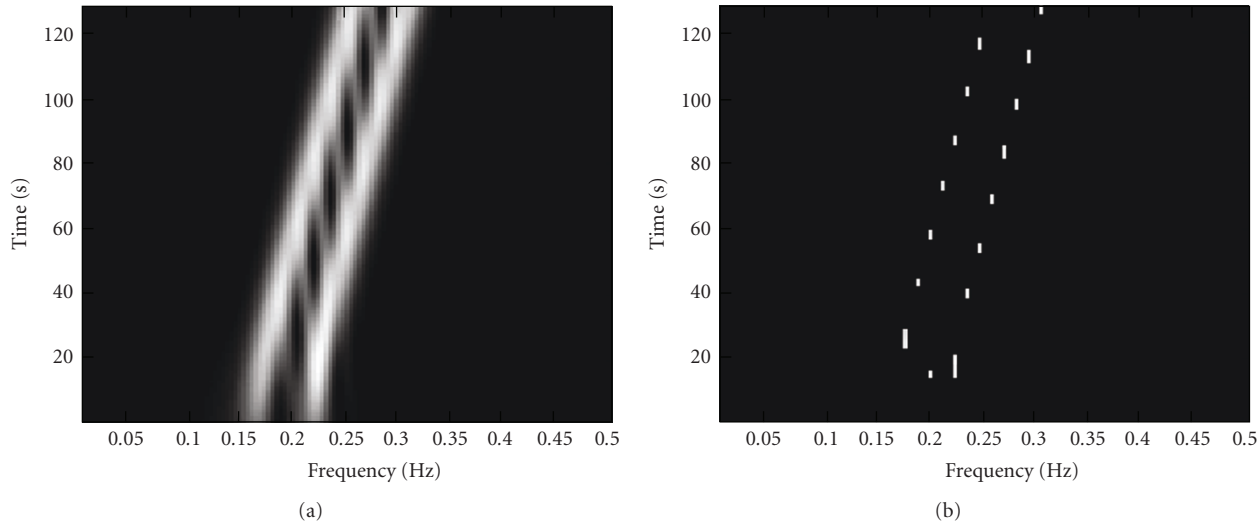

Figure 4: TFDs of a signal consisting of two linear FM components with frequencies increasing from 0.15 to 0.25 Hz and 0.2 to 0.3 Hz, respectively. (a) Spectrogram (TI 2) and (b) NTFD.

The synthetic test TFDs are processed by the proposed hybrid neurofuzzy method and the results are shown in Figures 3(b) and 4(b). Significant improvement in concentration and resolution of these signals in t-f domain can be noticed in these Figures. In order to compare the performance of TFDs by various methods, we quantify the quality of TFDs by objective assessment methods. Such quantitative analysis is presented in Table 2. The results clearly indicate that the proposed hybrid neurofuzzy method achieves the highest resolution and concentration amongst considered methods. The performance deteriorates in the noisy environment for all the considered high-resolution methods. However, the proposed neurofuzzy scheme maintains the best performance. The results are expected to improve further for low SNR values of the signal if the ANN model is trained with the noisy data of similar type.

Boashash instantaneous concentration and resolution measures are computationally expensive because they require calculations at various time instants. To limit the scope, these measures are computed at the middle of the synthetic signal and the results are compared to those reported by the authors in [39]. We take a slice at $t = 64$ and measure the signal components' parameters $A_{m_1}(64)$, $A_{m_2}(64)$, $A_m(64)$, $A_{s_1}(64)$, $A_{s_2}(64)$, $A_s(64)$, $V_{i_1}(64)$, $V_{i_2}(64)$, $V_i(64)$, $f_{i_1}(64)$, $f_{i_2}(64)$, and $\Delta f_i(64)$, as well as the CTs' magnitude $A_x(64)$. The values of the normalized instantaneous resolution measure $\mathbb{R}_i(64)$ and modified concentration performance measure $\mathbb{C}_n(64)$ are recorded in Tables 3 and 4, respectively. A TFD having

Table 2: Objective assessment.

Description	Test TFD	Spec	WVD	CWD	TSE	NTFD	RAM	OKM
Ratio of norm based measure	TI 1	1.98	3.72	2.13	3.14	21	15	4.33
($\times 10^{-4}$)	TI 3	3.81	3.84	2.89	8.13	76	68	8.32
Volume Normalized	TI 1	15.71	10.43	12.59	13.37	7.20	9.99	11.34
Rényi entropy measure	TI 3	12.45	12.02	12.93	13.85	6.21	7.30	11.77
Stankovic measure	TI 1	12.155	10.367	8.052	17.839	0.143	2.396	9.515
($\times 10^{5}$)	TI 3	0.22	3.30	1.06	2.01	0.00019	0.00129	0.63

In this table, the abbreviations for different methods include the spectrogram (spec), Wigner-Ville distribution (WVD), Choi-Williams distribution (CWD), t-f autoregressive moving-average spectral estimation method (TSE), neural network-based TFD (NTFD), reassignment method (RAM), and the optimal radially Gaussian kernel TFD method (OKM).

Table 3: Parameters and the normalized instantaneous resolution performance measure of TFDs for the time instant $t = 64$.

TFD (optimal parameters)	$A_m(64)$	$A_s(64)$	$A_x(64)$	$V_i(64)$	$\Delta f_i(64)$	$D(64)$	$\mathbb{R}(64)$
Spectrogram (Hann, $L = 35$)	0.9119	0.0087	0.5527	0.0266	0.0501	0.4691	0.7188
WVD	0.9153	0.3365	1	0.0130	0.0574	0.7735	0.6199
ZAMD ($a = 2$)	0.9146	0.4847	0.4796	0.0214	0.0420	0.4905	0.5661
CWD ($\sigma = 2$)	0.9355	0.0178	0.4415	0.0238	0.0493	0.5172	0.7541
BJD	0.9320	0.1222	0.3798	0.0219	0.0488	0.5512	0.7388
Modified B ($\beta = 0.01$)	0.9676	0.0099	0.0983	0.0185	0.0526	0.5957	0.8449
NTFD	0.9013	0	0	0.0110	0.0550	0.800	0.9333

the largest positive value (close to 1) of the measure $\mathbb{R}_i$ is the one with the best instantaneous resolution performance. The NTFD gives the largest value of $\mathbb{R}_i$ at time $t = 64$ in Table 3 and hence is selected as the best performing TFD of this signal at $t = 64$.

On similar lines, we have compared the TFDs' concentration performance at the middle of signal duration interval. A TFD is considered to have the best energy concentration for a given multicomponent signal if, for each signal component, it yields the smallest instantaneous bandwidth relative to component IF ($V_i(t)/f_i(t)$) and the smallest side lobe magnitude relative to the main lobe magnitude ($A_s(t)/A_m(t)$). The results in Table 4 indicate that the NTFD gives the smallest values of $\mathbb{C}_{1,2}(t)$ at $t = 64$ and hence is selected as the best concentrated TFD at time $t = 64$.

4.2.2. Real-Life Test Case. The bat echolocation chirp sound provides a perfect real-life multicomponent test case (test image 3 (TI 3)). Its true invariable nature is only obvious from the spectrogram shown in Figure 5(a), but, that is, blurred and difficult to interpret. The results are obtained using other high-resolution t-f methods that include the WVD, the traditional reassignment method, the optimal radially Gaussian kernel method, and the t-f autoregressive moving-average spectral estimation method. These t-f plots are shown in Figures 5(b), 5(d), 5(e), and 5(f), respectively, along with the neural network based reassigned TFD shown in Figure 5(c).

The t-f autoregressive moving-average estimation models are shown to be a t-f symmetric reformulation of time-varying autoregressive moving-average models [21]. The results are achieved for nonstationary random processes using a Fourier basis. This reformulation is physically intuitive because it uses time delays and frequency shifts to model the nonstationary dynamics of a process. The TSE models are parsimonious for the practically relevant class of processes with a limited t-f correlation structure. The simulation result depicted in Figure 5(f) demonstrate the method's ability to improve on the WVD (Figure 5(b)) in terms of resolution and absence of CTs; on the other hand, the t-f localization of the components deviates slightly from that in the WVD.

The traditional reassignment method enhances the resolution in time and frequency of the spectrogram. This is achieved by assigning to each data point a new t-f coordinate that better reflects the distribution of energy in

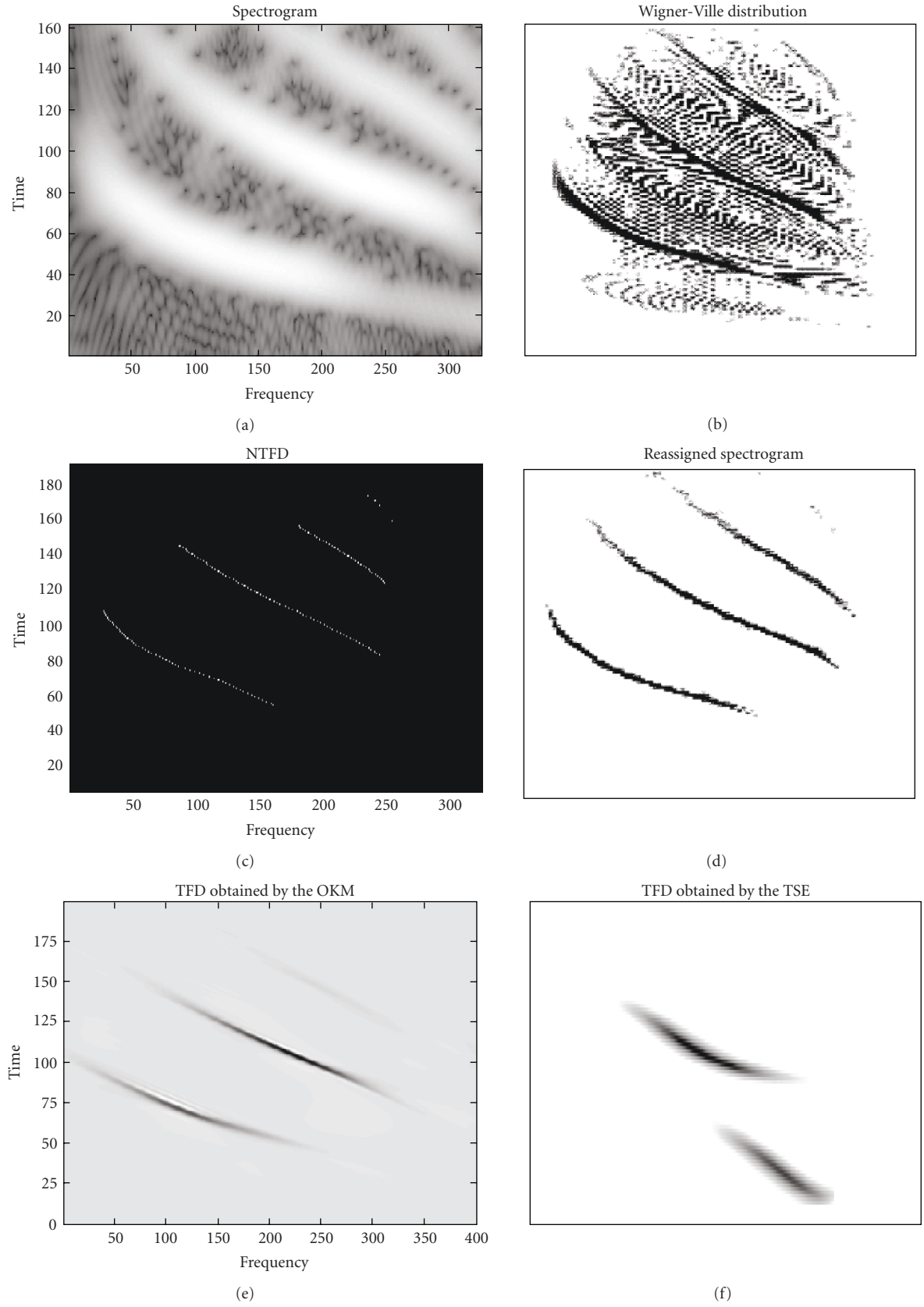

Figure 5: TFDs of the multi-component bat echolocation chirp signal by various high-resolution t-f methods.

Table 4: Parameters and the modified instantaneous concentration performance measure of TFDs for the time instant $t = 64$.

TFD(optimal parameters)	$A_{s_1}(64)$	$A_{s_2}(64)$	$A_{m_1}(64)$	$A_{m_2}(64)$	$V_{i_1}(64)$	$V_{i_2}(64)$	$f_{i_1}(64)$	$f_{i_2}(64)$	$\mathbb{C}_1(64)$	$\mathbb{C}_2(64)$
Spectrogram (Hann, $L = 35$)	0.0087	0.0087	1	0.8238	0.03200	0.0200	0.1990	0.2500	0.1695	0.0905
WVD	0.3365	0.3365	0.9153	0.9153	0.0130	0.013	0.1980	0.2554	0.4333	0.4185
ZAMD($a = 2$)	0.4848	0.4900	1	0.8292	0.0224	0.0204	0.2075	0.2495	0.5927	0.6727
CWD ($\sigma = 2$)	0.0176	0.0179	1	0.8710	0.0300	0.0176	0.205	0.2543	0.1639	0.0898
BJD	0.1240	0.1204	1	0.8640	0.0270	0.0168	0.2042	0.2530	0.2562	0.2058
Modified B ($\beta = 0.01$)	0.0100	0.0098	1	0.9352	0.0190	0.0180	0.200	0.2526	0.1050	0.0817
NTFD	0	0	0.8846	0.9180	0.0110	0.0110	0.2035	0.2585	0.0541	0.0425

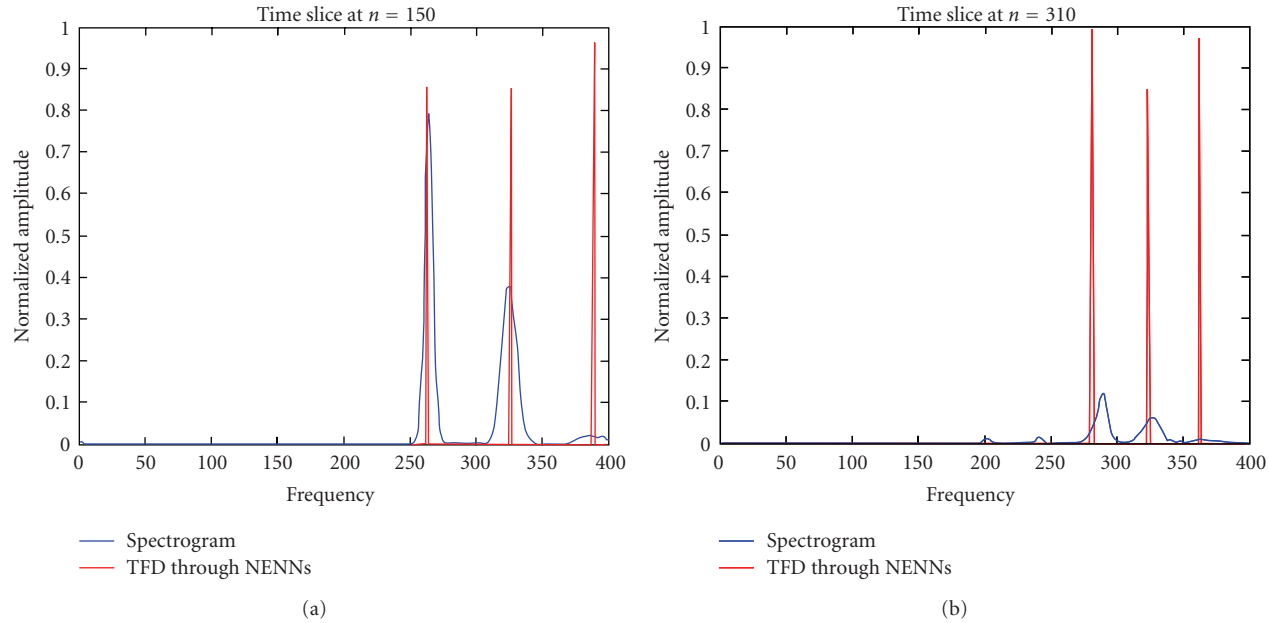

Figure 6: The time slices for the spectrogram (blue) and the NTFD (red) for the bat echolocation chirps signal, at $n = 150$ (a) and $n = 310$ (b).

the analyzed signal [13]. It is shown that this method can be applied advantageously to all the bilinear t-f and time-scale representations, and can be easily computed for the most common ones. The reassigned spectrogram for the bat echolocation chirps signal is shown in Figure 5(d). It shows energy concentration but often can diminish accuracy due to its way of approaching the problem. Also, its performance deteriorates for low signal-to-noise ratio (SNR) values and it contains discontinuities. The evaluation by various objective criteria is presented in Table 2. The analysis indicates that the results obtained by the hybrid neurofuzzy method are significantly better than this approach for all the measures.

On the other hand, the optimal radially Gaussian kernel TFD method proposes a signal-dependent kernel that changes shape for each signal to offer improved t-f representation for a large class of signals based on quantitative optimization criteria [16]. The result by this method is depicted in Figure 5(e). On careful monitoring, it is revealed that it does not recover all the components, thus losing some useful information about the signal. Also, the objective assessment does not point to much significance in achieving energy concentration along the individual components.

The NTFD for the test case is shown in Figure 5(c), which presents satisfactory resolution and is highly concentrated along the individual components. Also, it is more informative as the four components can be clearly identified. For further analysis, slices of the spectrogram and the NTFD are taken at the time instants $n = 150$ and $n = 310$ (recall that

$n = 1, 2, \ldots, 400$). The normalized amplitudes of these slices are plotted in Figure 6. These instants are chosen because visually three chirps can be marked at these instants (see Figure 5(c)). Also, Figure 6 confirms the peaky appearance of three different frequencies at these instants. There are no spurious CTs, and the result indicates much better frequency resolution (i.e., narrower main lobe and no side lobes) in comparison to all other methods and quadratic distributions.

5. Conclusions

The fuzzy framework for neural network based technique is found effective for the TFDs' reassignment using both synthetic and real-life examples. Experimental results demonstrate the effectiveness of the hybrid neurofuzzy approach against some high-resolution t-f methods. This includes distributions known for their high CTs suppression and energy concentration in the t-f domain. The resultant TFDs exhibit high resolution, good concentration, and no interference terms between the signal components. Also, they are found to be better at detecting the correct number of components in a given signal. The performance of the proposed scheme is satisfactory for signals corrupted with additive Gaussian noise with small variance whereas the performance of all other methods deteriorates. These qualities allow an easy visual interpretation and the reassigned TFDs can be used for subsequent classification problems. The trade-off is that these reassigned TFDs do not satisfy some desirable properties such as energy preservation and marginals. Hence the results may not be feasible for certain applications, which may have different preferences and the requirement to the TFDs. However, the results are better or close to the actual TFD images than the spectrogram. Furthermore, several TFDs, especially the adaptive ones like the traditional reassigned TFDs, have discontinuities [40]. The future work will be to adjust the discontinuity phenomenon along the individual components in the reassigned TFDs obtained by the proposed approach. Another direction may be to train the proposed scheme with noisy data and check its performance in very low SNR environment.

References

[1] J. Hardin and D. M. Rocke, "Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator," *Computational Statistics and Data Analysis*, vol. 44, no. 4, pp. 625–638, 2004.

[2] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, Wiley, Chichester, UK, 1999.

[3] A. M. Bensaid, L. O. Hall, J. C. Bezdek et al., "Validity-guided (re)clustering with applications to image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 112–123, 1996.

[4] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*, Thomson Learning, Boston, Mass, USA, 1996.

[5] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1987.

[6] A. E. Ruano, Ed., *Intelligent Control Systems Using Computational Intelligence Techniques*, The IEE Control Series 70, Institution of Engineering and Technology, London, UK, 2005.

[7] E. Sejdić, U. Ozertem, I. Djurović, and D. Erdogmus, "A new approach for the reassignment of time-frequency representations," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '09)*, pp. 2997–3000, Taipei, Taiwan, April 2009.

[8] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Techniques to obtain good resolution and concentrated time-frequency distributions: a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 673539, 43 pages, 2009.

[9] H. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[10] I. Daubechies, "Wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[11] O. Rioul and P. Flandrin, "Time-scale energy distributions: a general class extending wavelet transforms," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1746–1757, 1992.

[12] J. Bertrand and P. Bertrand, "Time-frequency representations of broad-band signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '88)*, pp. 2196–2199, New York, NY, USA, April 1988.

[13] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency reassignment: from principles to algorithms," in *Applications in Time-Frequency Signal Processing*, A. Papandreou-Suppappola, Ed., chapter 5, pp. 179–203, CRC Press, Boca Raton, Fla, USA, 2003.

[14] D. L. Jones and T. W. Parks, "A high resolution data-adaptive time-frequency representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2127–2135, 1990.

[15] D. L. Jones and R. G. Baraniuk, "Adaptive optimal-kernel time-frequency representation," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2361–2371, 1995.

[16] R. G. Baraniuk and D. L. Jones, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Processing*, vol. 32, no. 3, pp. 263–284, 1993.

[17] B. Barkat and B. Boashash, "Design of higher order polynomial Wigner-Ville distributions," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2608–2611, 1999.

[18] G. Viswanath and T. V. Sreenivas, "IF estimation using higher order TFRs," *Signal Processing*, vol. 82, no. 2, pp. 127–132, 2002.

[19] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Computing deblurred time-frequency distributions using artificial neural networks," *Circuits, Systems, and Signal Processing*, vol. 27, no. 3, pp. 277–294, 2008.

[20] P. Borgnat and P. Flandrin, "Time-frequency localization from sparsity constraints," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '08)*, pp. 3785–3788, Las Vegas, Nev, USA, March 2008.

[21] M. Jachan, G. Matz, and F. Hlawatsch, "Time-frequency ARMA models and parameter estimators for underspread nonstationary random processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4366–4381, 2007.

[22] I. Shafi, J. Ahmad, S. I. Shah, A. A. Ikram, A. A. Khan, and S. Bashir, "High resolution time-frequency methods' performance analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 806043, 7 pages, 2010.

[23] S. Stanković and L. Stanković, "Introducing time-frequency distribution with a "complex-time" argument," *Electronics Letters*, vol. 32, no. 14, pp. 1265–1267, 1996.

[24] L. Stanković, "Time-frequency distributions with complex argument," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 475–486, 2002.

[25] C. Cornu, S. Stanković, C. Ioana, A. Quinquis, and L. Stanković, "Generalized representation of phase derivatives for regular signals," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4831–4838, 2007.

[26] S. Stanković, N. Žarić, I. Orović, and C. Ioana, "General form of time-frequency distribution with complex-lag argument," *Electronics Letters*, vol. 44, no. 11, pp. 699–701, 2008.

[27] I. Orović and S. Stanković, "A class of highly concentrated time-frequency distributions based on the ambiguity domain representation and complex-lag moment," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 935314, 9 pages, 2009.

[28] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Evolutionary time-frequency distributions using Bayesian regularised neural network model," *IET Signal Processing*, vol. 1, no. 2, pp. 97–106, 2007.

[29] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, Mass, USA, 1981.

[30] R. Babuska, *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Norwell, Mass, USA, 1998.

[31] J. C. Bezdek and J. C. Dunn, "Optimal fuzzy partitions: a heuristic for estimating the parameters in a mixture of normal distributions," *IEEE Transactions on Computers*, vol. 24, no. 8, pp. 835–840, 1975.

[32] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proceedings of the 17th IEEE Conference on Decision and Control*, pp. 761–766, San Diego, Calif, USA, January 1979.

[33] R. Babuška, P. J. van der Veen, and U. Kaymak, "Improved covariance estimation for Gustafson-Kessel clustering," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 1081–1085, Honolulu, Hawaii, USA, May 2002.

[34] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841–847, 1991.

[35] L. Stankovic and S. Stankovic, "Analysis of instantaneous frequency representation using time-frequency distributions-generalized Wigner distribution," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 549–552, 1995.

[36] D. L. Jones and T. W. Parks, "A resolution comparison of several time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 40, no. 2, pp. 413–420, 1992.

[37] T. Sang and W. J. Williams, "Renyi information and signal-dependent optimal kernel design," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, pp. 997–1000, Detroit, Mich, USA, May 1995.

[38] L. Stankovic, "Measure of some time-frequency distributions concentration," *Signal Processing*, vol. 81, no. 3, pp. 621–631, 2001.

[39] B. Boashash and V. Sucic, "Resolution measure criteria for the objective assessment of the performance of quadratic time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1253–1263, 2003.

[40] L. Cohen, *Time-Frequency Analysis: Theory and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 1995.

[41] B. Boashash, Ed., *Time-Frequency Signal Analysis and Processing*, Elsevier Science, London, UK, 2003.

[42] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Impact of varying neurons and hidden layers in neural network architecture for a time frequency application," in *Proceedings of the 10th IEEE International Multitopic Conference*, pp. 188–193, Islamabad, Pakistan, 2006.

[43] S. I. Shah, I. Shafi, J. Ahmad, and F. M. Kashif, "Multiple neural networks over clustered data (MNCD) to obtain instantaneous frequencies (IFs)," in *Proceedings of the International Conference on Information and Emerging Technologies (ICIET '07)*, pp. 2–7, Karachi, Pakistan, July 2007.