Visual Analysis and Representations of Type-2 Fuzzy Membership Functions

Desh Raj*‡, Kenil Tanna*‡, Bhuvnesh Garg†‡, and Frank Chung-Hoon Rhee‡, *Member, IEEE*
*Department of Computer Science and Engineering
Indian Institute of Technology Guwahati, Guwahati, India 781039
Email: {r.desh,kenil}@iitg.ac.in
†Department of Mathematics
Indian Institute of Technology Guwahati, Guwahati, India 781039
Email: g.umang@iitg.ac.in
‡Department of Electronics and Communication Engineering
Hanyang University, Ansan-Si, Korea 15588
Email: frhee@fuzzy.hanyang.ac.kr, Contact: +82.31.400.5296, Fax: +82.31.436.8152

***Abstract*—This paper presents several pictorial and graphical techniques that may be used for effectively visualizing type-2 fuzzy membership functions (T2 FMFs). In our first proposed technique, two-dimensional data sets have been modeled using grayscale entropies to make the uncertainty interpretation easier. Next, the concept of a vertical drill and a primary membership drill has been introduced to obtain information from a T2 FMF representing a multi-dimensional data set. Further, the generation of general T2 FMFs with secondary membership functions in the form of asymmetric Gaussian distributions, referred to as "snaky" surfaces, has been discussed as an extension of symmetric Gaussian T2 FMF. These graphical techniques may be applied for making inferences or predictions about the uncertainty level of a T2 FMF in applications such as data clustering, computing with words (CWW), and logic control for robots, to name a few.**

I. Introduction

Data sets in one and two dimensions have conventionally been represented in the form of scatter plots, when no uncertainty is associated with them. However, if the data is considered to have some degree of uncertainty [1], [2], such as those commonly found in pattern recognition [3], accurately describing the data along with its grade of uncertainty may only be possible by tabulating the complete set. While this approach may provide a precise representation of the data, it may be impractical to apply it to large data sets, in which gathering, storing and visualizing the data can become complex, and in some cases, impossible.

To alleviate this problem, normalized two and three-dimensional histograms have been constructed by partitioning the data sets into grids of appropriate sizes [4]. To obtain higher accuracy with the normalization, grid sizes may be reduced, which results in an increase in the number of generated histograms. This further gives rise to the issue of storing the heights of the histograms, which in itself may enumerate to high orders of magnitude. If the grid sizes are increased to reduce the number of histograms, it makes storing easier at the expense of decreasing accuracy of representation, since more data points are grouped under one average grade of uncertainty. It is, therefore, a tradeoff between accuracy and efficiency, similar to most approximation algorithms [5].

The issue of storing data may be addressed efficiently by approximating the discrete data with continuous paramteric functions, such as a Gaussian fitted over the histograms. In addition to describing the data using minimum number of parameters, these functions are visually descriptive and may be interpreted better than the raw data. Grayscale entropies achieve these metrics and are also easier to represent than the three-dimensional surfaces.

Type-2 fuzzy membership functions (T2 FMFs) have also been represented using the vertical slice, horizontal slice and wavy slice [6]. While the visual concepts of horizontality, verticality and waviness are appropriate for data sets in two dimensions, the same may not be valid when the data is spread across multiple dimensions such that the T2 FMF itself is a hypergeometric surface. Such a scenario necessitates the introduction of the concept of a primary membership (PM) "drill," as a representation method for T2 FMFs.

In the last two decades, much progress has been made with respect to interval type-2 fuzzy sets (IT2 FSs) [7], [8]. They have been applied to fuzzy C-means (FCM) clustering [9], encoding words [10], and logic controllers for autonomous robots [11] with satisfactory results. We now introduce another kind of T2 FMF, which adds asymmetric Gaussian type uncertainty to the primary membership. The resulting secondary MFs obtained has been visualized as a "snaky" surface, which is not uniformly distributed across the footprint of uncertainty (FOU), such as IT2 FSs.

The remainder of this paper is organized as follows. In Section 2, we describe grayscale entropies that may be used for visualizing the distribution of uncertainty in data sets. Section 3 introduces the concept of a PM drill and its application in mathematically describing multi-dimensional uncertain data. In Section 4, "snaky" T2 FMFs are discussed. We conclude by discussing the use of these visual representation techniques in algorithms that incorporate T2 fuzzy sets.

II. Grayscale Entropy Representation of Data Distribution

In this section, we address some of the common types of data sets in two dimensions, where the data points involved have a primary membership for uncertainty but may or may not have higher degrees of uncertainty. In the case of existence of uncertainties greater than T1, we will be restricting our study to T2 FMFs [12], [13], [14], [2].

As an illustration, consider the two-dimensional data sets *A* and *B* that are represented by scatter plots shown in Fig. 1(a) and 1(b). From the scatter representation of data, it may be considered possible to visualize its distribution, if the data is certain. However, since the data in this case has one level of uncertainty in the form of a T1 fuzziness, this information cannot be interpreted solely from the scatter plots. Fig. 1(c) and 1(d) denote the corresponding grayscale entropies, which include the degree of uncertainty for the data points shown in Fig. 1(a) and 1(b), respectively.

From the grayscale entropies, it may be convenient to observe that the data sets *A* and *B* are distributed in the form of overlapping square and circular regions, respectively. Such a situation may occur in problems like clustering of data points consisting of two features into a finite number of clusters, such that the membership of each data element in each of the clusters is uncertain [15], [16], [17]. Fig. 1(c) and 1(d) may further be used to visualize the primary memberships of the data points, since a brighter grayscale entropy corresponds to a higher degree of membership of a data point in a particular cluster, and vice versa. This interpretation may be extended to T2 FMFs, as shown in Fig. 1(e) and 1(f).

A T2 FMF assigns uncertainty to primary memberships. If the primary membership representing a data set is in the form of a smooth pyramidal or Gaussian function, the corresponding secondary memberships may ideally be a skewed surface having some positive or negative deviation from the T1 FMF. As such, Fig. 1(e) and 1(f) may easily be interpreted as T2 FMF for the data sets shown in Figs. 1(a) and 1(b), respectively. If the brightness of the entropies is viewed to vary uniformly at random from the entropies corresponding to the T1 FMF, we can further observe that they may be best represented as an IT2 FMF [1], [2], [7], [8], [18]. If the universe of discourse is denoted by X, a T2 FS, denoted by $\widetilde{A}$, is characterized by MF $\mu_{\widetilde{A}}\left(\left(x_1, x_2, ..., x_n\right), u\right)$, where $x_i \in X$ and $u \in J_{(x_1,x_2,...,x_n)} \subseteq [0, 1]$, i.e.,

$$\begin{aligned}\widetilde{A} = \{ & \left(\left(x_1, x_2, ..., x_n\right), u\right), \mu_{\widetilde{A}}\left(\left(x_1, x_2, ..., x_n\right), u\right) \mid \\ & \forall x_i \in X \; and \; u \in J_{(x_1,x_2,...,x_n)} \subseteq [0, 1] \}\end{aligned} \tag{1}$$

where $0 \leq \mu_{\widetilde{A}}\left(\left(x_1, x_2, ..., x_n\right), u\right) \leq 1$

For an IT2 FMF, the secondary membership $\mu_{\widetilde{A}}(x, u)$ is mathematically expressed as follows.

$$\mu_{\widetilde{A}}(x, u) = \begin{cases} 1 & \forall x \in X, \forall u \in J_x \subseteq [0, 1] \\ 0 & otherwise \end{cases} \tag{2}$$

While the grayscale entropies for a two-dimensional data set may be visually interpreted, those for data sets in higher dimensions may only be intelligible to a computer. However, since actual computation of multidimensional data such as those used in bioinformatics, satellite imagery, and finance, is complex and time-consuming, this method of pattern matching through the use of entropy maps may be useful for estimating the similarity of data before performing the actual computation. This may even reduce the requirement of performing complex computations if data sets are found to be very similar and results have already been obtained for one such data set. With increased use of distributed computing, such pattern recognition may be performed accurately and efficiently.

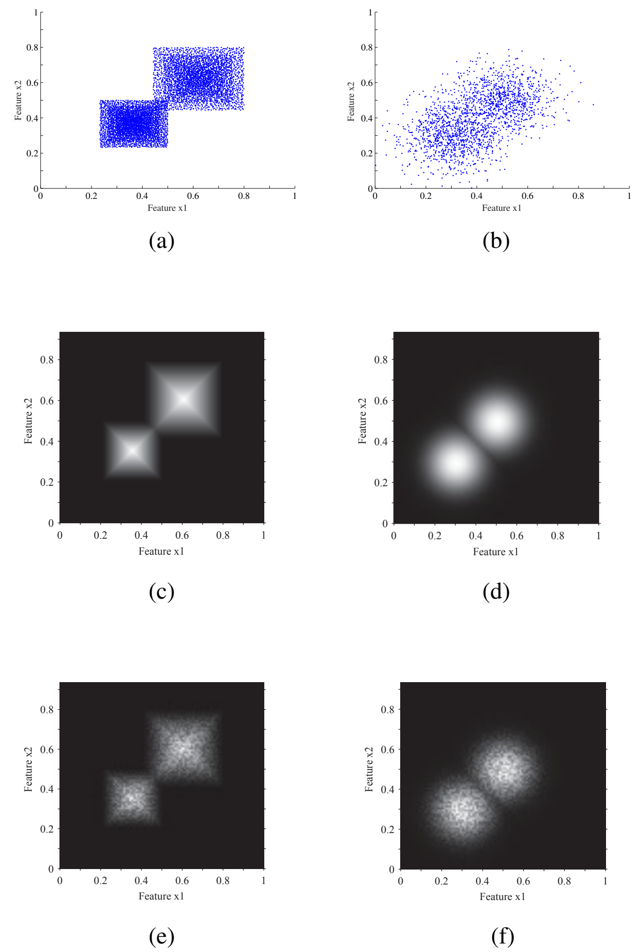

Fig. 1. Grayscale entropy representations of T1 and T2 FMFs for different types of data sets: (a) Scatter plot for data set *A*, (b) Scatter plot for data set *B*, (c) Grayscale entropy for T1 FMF of data set *A*, (d) Grayscale entropy for T1 FMF of data set *B*, (e) Grayscale entropy for T2 FMF of data set *A*, and (f) Grayscale entropy for T2 FMF of data set *B*.

III. Primary Membership Drill

T2 FMFs have been studied extensively and usually represented through one of three methods vertical slice, horizontal slice, and wavy slice [6]. In this section, we extend the definition of the *vertical slice* to incorporate multidimensional data sets, by introducing the concept of a primary membership

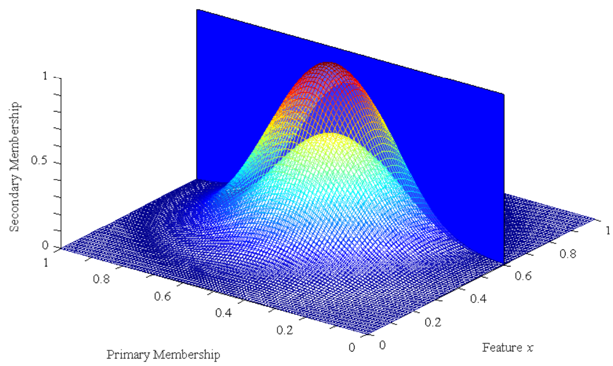

Fig. 2. Vertical slice representation of secondary membership.

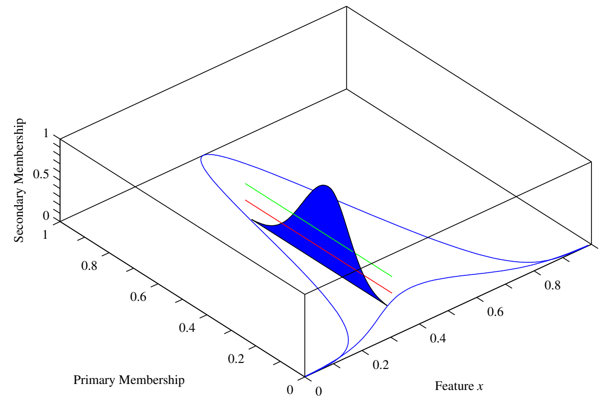

Fig. 3. Vertical drill for one-dimensional data.

(PM) drill. First we restate the vertical slice definition of a T2 FMF [1].

The *vertical slice* representation of T2 FS $\widetilde{A}$ focuses on each value of the primary variable x, and expresses (1) as the union of all its secondary memberships as

$$\widetilde{A} = \int_{x \in X} \widetilde{A}(x) \ / \ x, \tag{3}$$

where

$$\widetilde{A}(x) = \int_{\forall u \in [0,1]} \mu_{\widetilde{A}(x)} \ / \ u. \tag{4}$$

An example of a vertical slice representation in the case of one-dimensional data is shown in Fig. 2.

While the *vertical slice* representation of T2 FMFs is accurate for data represented in one dimension, the notions of verticality and slicing may be unfounded for hypergeometric spaces. To remedy this issue, it may be beneficial to define a new representation in terms of the primary membership of the fuzzy set of multiple dimensions.

Let us consider a universe of discourse X, defined by $X = \left\{x : x \in \mathbb{R}^{n+2}\right\}$, i.e. every data point in this universe consists of n independent features. The primary membership of data points is defined by the MF $u_A\left(x_1, x_2, ..., x_n, 0, 0\right)$, i.e.,

$$A = \left\{\left(\left(x_1, x_2, ..., x_n, 0, 0\right), u_A\left(x_1, x_2, ..., x_n, 0, 0\right)\right)\right\}, \tag{5}$$

and the T2 FMF is given by (1). Note that, since the data points have n features, the last two values in the n-tuple denoting the data point must be 0. For instance, if the data set has only one feature, the points representing the data will be along the x axis, and as such, the values for y and z will be 0. Also, note that the T1 membership is in the $(n+1)^{th}$ dimension and the T2 membership is in the $(n+2)^{th}$ dimension. Suppose the lower membership function (LMF) and upper membership function (UMF) for this fuzzy set is defined by the functions given in (6) and (7), respectively, where

$$\begin{aligned} LMF\left(\widetilde{A}\right) &= \underline{\mu}_{\widetilde{A}}(x) \\ &= inf\left\{u \mid u \in [0,1] \ \ and \ \mu_{\widetilde{A}}(x,u) > 0\right\} \end{aligned} \tag{6}$$

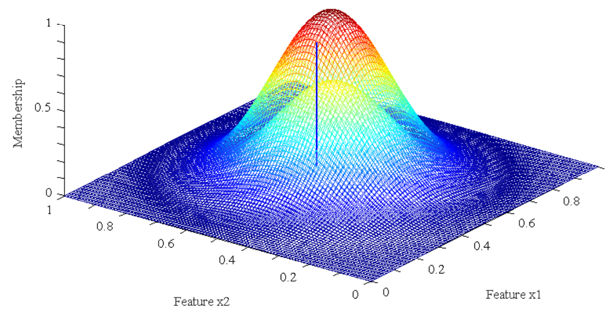

Fig. 4. Vertical drill for two-dimensional data.

and

$$\begin{aligned} UMF\left(\widetilde{A}\right) &= \overline{\mu}_{\widetilde{A}}(x) \\ &= sup\left\{u \mid u \in [0,1] \ \ and \ \mu_{\widetilde{A}}(x,u) > 0\right\}. \end{aligned} \tag{7}$$

The region enclosed by the LMF and UMF is called the footprint of uncertainty (FOU) [19]. Suppose we drill through the FOU of the fuzzy set at any arbitrary point such that the drill is parallel to the unit vector along the direction of the primary membership, and orthogonal to unit vectors in all other directions. The parametric equation of this drill, called the primary membership (PM) drill, is given by

$$x_1 = t_1, x_2 = t_2, ..., x_n = t_n, x_{n+1} \in \mathbb{R}, x_{n+2} = \mu \tag{8}$$

Now, if we fix the first n parameters $x_1, x_2, \ldots, x_n$ of the PM drill and start varying the last parameter μ between 0 and 1, we can obtain all the information about the T1 and T2 fuzzy memberships of the data sample corresponding to the tuple $(t_1, t_2, \ldots, t_n, 0, 0)$. Suppose μ = 0, then the intersection of the PM drill with the functions representing the LMF and the UMF gives the range of uncertainty (or fuzziness) of the sample in the set. As we increase the value of μ, the range of values u through which the PM drill passes represents the primary membership values that are at least as certain (i.e., memberships are greater than or equal) as μ. This is shown for one and two-dimensional data sets in Fig. 3 and 4, respectively. Note that for a two-dimensional data set, the vertical drill shown in Fig. 4 is graphically representable only for $\mu = 0$. For higher values of μ, the PM drill extends to four dimensions, and cannot be represented visually.

Extracting information regarding the range of uncertainty corresponding to various values of secondary membership may be convenient in many application that use the FOU of T2 FMFs, such as optimization methods for T2 fuzzy inference systems [20], and T2 fuzzy logic controllers [21], to name a few.

IV. Snaky Type-2 Fuzzy Membership Functions

In this section, we introduce a general T2 FMF that incorporates asymmetric Gaussian distributions for the primary as well as secondary memberships. For a one-dimensional data set, such a fuzzy set may be graphically shown to be a smooth wavy surface, hence the name "snaky" is given.

To generate such a "snaky" general T2 FMF from a one-dimensional data set, a histogram-based approach is first followed for obtaining the LMF and UMF. Choi and Rhee [4] outline this method as follows:

(i) A histogram is constructed and smoothened for each labeled class of the sampled data.
(ii) Polynomial function (PF) fitting is performed to obtain the approximate parameter values.
(iii) Gaussian function (GF) fitting is performed using the values obtained in step (ii).
(iv) GF fitting is performed for the upper and lower histogram values with respect to the GF in step (iii).
(v) UMF is determined by normalizing the height of the upper GF and LMF by proportionally scaling the lower GF obtained in step (iv).

The T1 FMF is defined by the GF obtained in step (iii). The functions obtained after the above procedure are illustrated in Fig. 5. We require the secondary memberships of the fuzzy set to be in the form of asymmetric Gaussians (or univariate skewed normal distribution) [22], [23]. Mathematically, such a distribution is defined as explained below.

Consider first a continuous random variable X having probability density function of the following form

$$f(x) = 2\phi(x)\,\Phi(\alpha x) \tag{9}$$

where α is a fixed arbitrary number (explained later), and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \Phi(\alpha x) = \int_{-\infty}^{\alpha x} \phi(t)\,dt \tag{10}$$

denote the standard normal (Gaussian) density function and its distribution function (the latter evaluated at point αx), respectively. The component α is called the *shape parameter* since it regulates the shape of the density function. Now consider the linear transform

$$Y = \xi + \omega X \tag{11}$$

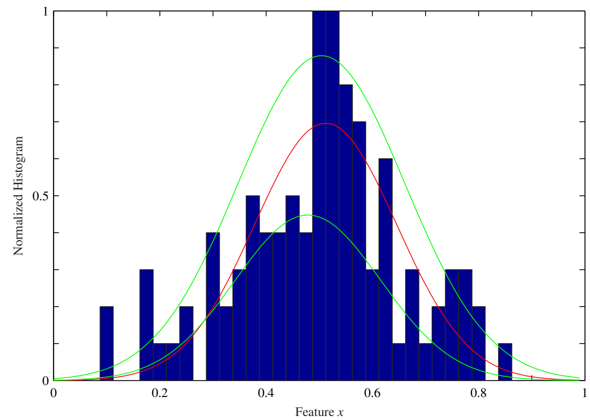

Fig. 5. Obtaining the LMF and UMF for one-dimensional data by Gaussian fitting on the histogram.

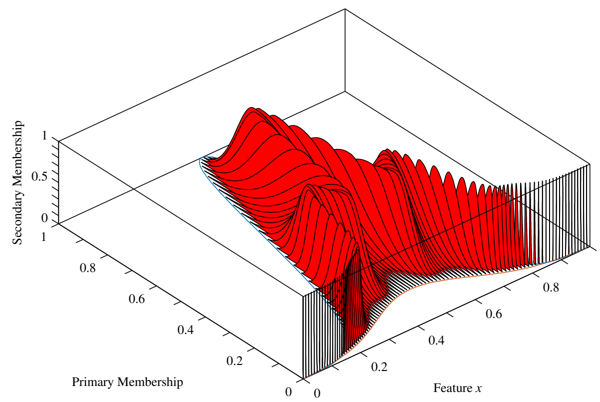

Fig. 6. "Snaky" type-2 fuzzy membership function for one-dimensional data.

which is then said to have a skew-normal distribution with parameters (ξ, ω, α), and write

$$Y \sim SN\left(\xi, \omega^2, \alpha\right) \tag{12}$$

We refer to ξ, ω, and α as the location, scale, and shape parameters, respectively. With this definition of Y, the cumulative distribution function (CDF) of Y, denoted by Ψ can be written as

$$\Psi(x) = \Phi\left(\frac{x-\xi}{\omega}\right) - 2T\left(\frac{x-\xi}{\omega}, \alpha\right), \tag{13}$$

where $T(h, a)$ is called Owens T-function [24] and is defined as

$$T(h, a) = \int_0^a \frac{e^{-\frac{1}{2}h^2\left(1+x^2\right)}}{1+x^2}\,dx. \tag{14}$$

To precisely describe a skew-normal, we need to specify the value of these parameters. With this definition of Y, the mean value $\mathbb{E}(Y)$ is defined by

$$\mathbb{E}\{Y\} = \xi + \omega\sqrt{\frac{2}{\pi}}\delta, \tag{15}$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$.

Since the mean of the secondary membership is actually obtained by superposing a sinusoidal function on the T1 FMF, $\mathbb{E}(Y)$ can be rewritten in terms of fuzzy parameters as

$$\mathbb{E}\left\{Y\right\} = u\left(x\right) + \left(\overline{\mu}_{\widetilde{A}}\left(x\right) - \underline{\mu}_{\widetilde{A}}\left(x\right)\right)\sin x. \tag{16}$$

Since (15) and (16) both define the same quantity $\mathbb{E}(Y)$, we can equate the RHS of both equations to obtain

$$\xi + \omega\sqrt{\frac{2}{\pi}}\delta = u\left(x\right) + \left(\overline{\mu}_{\widetilde{A}}\left(x\right) - \underline{\mu}_{\widetilde{A}}\left(x\right)\right)\sin x. \tag{17}$$

Without loss of generality, we may assume that the FOU of the fuzzy set is analogous to the confidence interval of the normal distribution. Since confidence intervals are usually stated at the 95% confidence level, we end up with the following

$$\Psi\left(\underline{\mu}_{\widetilde{A}}\left(x\right)\right) = 0.025 \; and \tag{18}$$

$$\Psi\left(\overline{\mu}_{\widetilde{A}}\left(x\right)\right) = 0.975. \tag{19}$$

Solving (17), (18), and (19) simultaneously, we can obtain the required value of parameters ξ, ω, and α and hence precisely define the asymmetric Gaussian which characterizes the secondary membership of the T2 FMF. Integrating over all values of x in the sample space, we finally obtain the "snaky" surface describing the T2 fuzzy set, as illustrated in Fig. 6.

V. Conclusion

In this paper, we introduced three techniques for the visualization and representation of T2 fuzzy sets. The grayscale entropy method made use of brightness levels to quantify the degree of uncertainty of a fuzzy data set. Although this method may not provide accurate values of uncertainty, it may be beneficial to initially obtain an understanding of the distribution pattern and fuzziness of data. The second technique proposed was the PM drill, which is effectively an extension of the *vertical slice* representation of T2 FMFs. As discussed, the PM drill may be useful in extracting information about the domain of uncertainty (DOU) of a fuzzy set [6]. The third technique introduced was the snaky T2 FMF, which may be understood as a more general form of the symmetric Gaussian T2 FMF, wherein a periodically varying function is superimposed on the primary memberships to obtain deviation.

Increasing use of T2 fuzzy sets in diverse applications like classification/clustering, autonomous navigation and computing with words may find the grayscale entropy and PM drill visualizations useful, since the computations involved in such applications are generally complex in nature. Our proposed "snaky" T2 FMFs may be applied to existing algorithms that have already been modelled with IT2 FMFs. Depending on the data, an improvement over results may be achieved using "snaky" T2 FMFs as opposed to IT2 FMFs. In particular, finite data may be modelled using fuzzy sets and data may then be generated back from these FMFs. We may further compare the similarity among these data sets to determine which MF best represents a particular data distribution. This is currently under investigation.

Acknowledgment

This work was supported by the Technology Innovation Program of the Korea Institute for Advancement of Technology(KTA) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 2015-122).

References

[1] J. Mendel, *Uncertain rule-based fuzzy logic system: Introduction and new directions*. Prentice Hall PTR, 2001.

[2] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*. Prentice Hall New Jersey, 1995, vol. 4.

[3] J. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[4] B. Choi and F. C. H. Rhee, "Interval type-2 fuzzy membership function generation methods for pattern recognition," *Information Sciences*, vol. 179, no. 13, pp. 2102–2122, 2009.

[5] H. Liu, T. Hong, M. Herman, T. Camus, and R. Chellappa, "Accuracy vs efficiency trade-offs in optical flow algorithms," *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 271–286, 1998.

[6] J. Mendel, "General type-2 fuzzy logic systems made simple: a tutorial," *Fuzzy Systems, IEEE Transactions on*, vol. 22, no. 5, pp. 1162–1182, 2014.

[7] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *Fuzzy Systems, IEEE Transactions on*, vol. 8, no. 5, pp. 535–550, 2000.

[8] J. Mendel, R. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *Fuzzy Systems, IEEE Transactions on*, vol. 14, no. 6, pp. 808–821, 2006.

[9] C. Hwang and F. C. H. Rhee, "Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means," *Fuzzy Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 107–120, 2007.

[10] F. Liu and J. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 1503–1521, 2008.

[11] H. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *Fuzzy Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 524–539, 2004.

[12] N. Karnik, J. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *Fuzzy Systems, IEEE Transactions on*, vol. 7, no. 6, pp. 643–658, 1999.

[13] J. Mendel and R. John, "Type-2 fuzzy sets made simple," *Fuzzy Systems, IEEE Transactions on*, vol. 10, no. 2, pp. 117–127, 2002.

[14] N. Karnik and J. Mendel, "Operations on type-2 fuzzy sets," *Fuzzy Sets and Systems*, vol. 122, no. 2, pp. 327–348, 2001.

[15] J. Bezdek, R. Ehrlich, and W. Full, "Fcm: the fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.

[16] F. C. H. Rhee and C. Hwang, "A type-2 fuzzy c-means clustering algorithm," in *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, vol. 4, pp. 1926–1929.

[17] N. Pal, K. Pal, J. Keller, and J. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *Fuzzy Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 517–530, 2005.

[18] F. C. H. Rhee, "Uncertain fuzzy clustering: insights and recommendations," *Computational Intelligence Magazine, IEEE*, vol. 2, no. 1, pp. 44–56, 2007.

[19] H. Mo, F. Wang, M. Zhou, R. Li, and Z. Xiao, "Footprint of uncertainty for type-2 fuzzy sets," *Information Sciences*, vol. 272, pp. 96–110, 2014.

[20] D. Hidalgo, P. Melin, and O. Castillo, "An optimization method for designing type-2 fuzzy inference systems based on the footprint of uncertainty using genetic algorithms," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4590–4598, 2012.

[21] R. Sepulveda, P. Melin, A. Díaz, A. Mancilla, and O. Montiel, "Analyzing the effects of the footprint of uncertainty in type-2 fuzzy logic controllers," *Engineering Letters*, vol. 13, no. 2, pp. 138–147, 2006.

[22] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, pp. 171–178, 1985.

[23] ——, "Further results on a class of distributions which includes the normal ones," *Statistica*, vol. 46, no. 2, pp. 199–208, 1986.

[24] R. Sowden and J. Ashford, "Computation of the bi-variate normal integral," *Applied Statistics*, pp. 169–180, 1969.