N93-29529
161809
p. 12

A Fuzzy Measure Approach to Motion Frame Analysis for Scene Detection

Albert B. Leigh, McDonnell Douglas Space Systems
Sankar K. Pal, Indian Statistical Institute, Calcutta
NASA Johnson Space Center
Software Technology Branch/PT4
Houston, TX 77058

ABSTRACT

This paper addresses a solution to the problem of scene estimation of motion video data in the fuzzy set theoretic framework. Using fuzzy image feature extractors, a new algorithm is developed to compute the change of information in each of two successive frames to classify scenes. This classification process of raw input visual data can be used to establish structure for correlation. The algorithm attempts to fulfill the need for non-linear, frame-accurate access to video data for applications such as video editing and visual document archival/retrieval systems in multimedia environments.

1. INTRODUCTION

With rapid advancements in multimedia technology, it is increasingly common to have time-varied data like video as computer data types. Existing database systems do not have the capability to search within such information. It is a difficult problem to determine one scene from another because there are no precise markers that identify where they begin and end. And, divisions of scenes can be subjective especially if transitions are subtle. One way to estimate scene transitions is to mathematically approximate the change of information between each of two successive frames by computing the distance between their discriminatory properties. A fuzzy theoretic approach in image processing and pattern recognition provides convenient methods for such ambiguous or uncertainty measure.

1.1 Fuzzy Image Concepts

In classical image processing, given a digital image, which has a M by N dimension with L gray levels, each picture element or pixel is represented as a spatial brightness function or gray information. Using fuzzy notion, an image can be considered as an array of fuzzy singletons, each having a value of membership denoting its degree of brightness relative to some brightness level, l, where $l = 0, 1, 2, \ldots, L-1$. The fuzzy notation can be written as follows:

$$X = \{ \mu_{mn}(x_{mn}) = \mu_{mn} / x_{mn};\ m = 1, 2, \ldots, M;\ n = 1, 2, \ldots, N\}$$

$$\text{or } X = \bigcup_m \bigcup_n \mu_{mn} / x_{mn},\ m = 1, 2, \ldots, M;\ n = 1, 2, \ldots, N$$

where $\mu_X(x_{mn})$ or μ_{mn} / x_{mn}, $(0 \leq \mu_{mn} \leq 1)$ denotes the grade of possessing some property μ_{mn} (e.g., brightness, edginess, smoothness) by the (m,n)th pixel intensity x_{mn}. In other words, a fuzzy subset of an image X is mapping μ from X into [0,1] (Figure 1.1). For any point $p \in X$, $\mu(p)$ is called the degree of membership of p in μ [11].

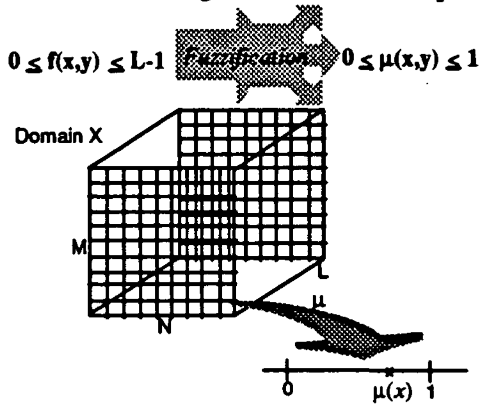

Figure 1.1: Fuzzy representation of an image X

2. IMAGE PROPERTIES

There are many spatial and geometric properties or features that can be measured or extracted from an image. They are used for pattern classifications and scene analysis. There is no trivial solution to selecting optimal features that would provide useful input values to the classifier. The effectiveness of these feature extractors also depends upon scenes. For this paper, six operators for ambiguity and fuzzy geometric measures are selected.

2.1 Ambiguity Measures

Two measures of ambiguities used are second-order local entropy and edginess. They produce a measure of structural information that exists in a given image. The entropy of an image can be defined as a measure of the information (ambiguities) gain in a given image. The edginess measures the coarseness of texture based on the average amount of ambiguity present in a given image.

2.1.1 Second-order Local Entropy

The calculation of the second-order local entropy contains a window that operates on two adjacent pixels. This window is then used to compute the co-occurrence matrix for incorporating the dependency of the spatial distribution of gray levels. In this case, the horizontal co-occurrence matrix is used. Then, the probability of the co-occurrence matrix is calculated with

$$p_{ij} = \frac{c_{ij}}{\sum_{ij} c_{ij}} \quad \text{, where } 0 \leq p_{ij} \leq 1 \;\; [12].$$

$$\mu(X) = -\frac{1}{2} \sum_{ij} p_{ij} \log(p_{ij})$$

The information gain is computed with a logarithmic function. As described in [8], this could be an exponential function. The co-occurrence matrix computation could also be modified with a combination of horizontal and vertical directions for a more accurate measure of the spatial distributions.

2.1.2 Edginess

This image property is a measure of edge information to detect edge intensities in an input image. Note that this is different from the gradient descent edge detectors. It calculates the edge ambiguity using a localized window to find the boundary between the current pixel and neighboring pixels [12].

In the equation

$$\delta(X) = [1 - I(X)]^{\beta},$$

I(X) stands for the ambiguity measure, or the index of fuzziness, and β is a positive constant. The spatial dependent membership function, μ_x, must be computed first.

$$\mu_x(x_{mn}) = \frac{0.5}{1 + \frac{1}{N_1}\sum_{ij} | x_{mn} - x_{ij}|}$$

where N_1 represents the dimensions of the window of i by j, i.e. $N_1 = i*j$. These are neighboring pixels of the point (m, n). As shown in Figure 2.1, the linear index of fuzziness, I(X), can be defined as follows:

$$I(X) = \frac{2}{n} \sum_{i} \min(\mu_X(x_i), 1 - \mu_X(x_i)).$$

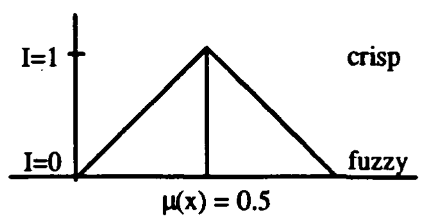

Figure 2.1: The linear index of fuzziness

Other measures of fuzziness, such as the quadratic index of fuzziness [6], fuzzy entropy [2], and index of non-fuzziness (crispness) [12], could also be used for the edginess measure.

2.2 Fuzzy Geometric Measures

Geometric measures define surfaces, shapes, solids, and boundaries of objects. Rosenfeld [13] and Rosenfeld and Haber [15] incorporated the fuzzy theoretic approach to the classical geometric measures and generalized some of the standard geometric properties of the relationships among regions to fuzzy sets [10]. Of these many measures, the primitive measures, such as area and perimeter, orientation measures, and shape measures are applied here.

The remaining methods that were applied, namely fuzzy geometrical properties, were extensions of the traditional geometrical measure concepts to operate in the fuzzy set framework. These measures examine various geometrical properties and relations such as area, perimeter, length, height, breadth, width, compactness, and elongatedness. There are many other topological concepts such as connectedness, major and minor axis, and adjacency, which could have been utilized in this study. These fuzzy measures are the basis for measuring spatial, gray, and region ambiguities.

2.2.1 Area

The area is an integral taken over the fuzzy image subset, i.e. $\int \mu(x)$. For a digital image, it is computed by summing the spatial brightness values of all image pixels. This spatial brightness value function is treated as the fuzzy membership function [11, 14].

$$\text{area}(\mu(x)) = \Sigma\, \mu(x)$$

2.2.2 Perimeter

The perimeter of an image is defined as the circumferential distance around the boundary. Using a faster method of computation, it can be computed as the sum of the product of the co-occurrence matrix and the difference of two adjacent pixels [11].

$$\text{perimeter}(\mu(X)) = \sum_i \sum_j c[i,j]\, |\mu(i)-\mu(j)|$$

where i=1, 2, ..., L and j=1,2, ..., L.

2.2.3 Length

The length of an image is calculated by finding the longest extent in the column direction [11, 14].

$$\text{length}(\mu) = \max_m \left(\sum_n \mu_{mn} \right)$$

2.2.4 Height

The height of an image is another way of measuring its extent by summing the maximum membership values of each row [11, 14].

$$\text{height}(\mu) = \sum_n \max_m \mu_{mn}$$

2.2.5 Breadth

The breadth of an image measures the longest extent in the row direction [11, 14].

$$\text{breadth}(\mu) = \max_n \left(\sum_m \mu_{mn} \right)$$

2.2.6 Width

The width is calculated as the sum of maximum membership values of each column [11, 14].

$$\text{width}(\mu) = \sum_m \max_n \mu_{mn}.$$

2.3 Orientations

The horizontal and vertical orientation of an image can be measured as follows [11]:

If $\dfrac{\text{length}(\mu)}{\text{height}(\mu)} \leq 1$, then vertically oriented.

If $\dfrac{\text{breadth}(\mu)}{\text{width}(\mu)} \leq 1$, then horizontally oriented.

2.4 Shape Measures

Shape measures can be computed using geometrical properties of a given image. These measures can also be defined independently of size measurements [16]. It basically represents the profile and physical structure of an image or image subsets. Two fuzzy measures are used: compactness and index of area coverage.

2.4.1 Compactness

The compactness measures the property of circularity [11].

$$\mathrm{Comp}(\mu) = \frac{\mathrm{area}(\mu)}{(\mathrm{perimeter}(\mu))^2}$$

2.4.2 Index of Area Coverage

The index of area coverage (IOAC) is the fraction of the maximum area (that can be covered by the length and breadth of the image) actually covered by the image [11].

$$\mathrm{IOAC}(\mu) = \frac{\mathrm{area}(\mu)}{\mathrm{length}(\mu) * \mathrm{breadth}(\mu)}$$

3. SCENE ESTIMATION

As discussed in [12], the criterion of a good feature is that it should be invariant within class variation while emphasizing differences that are important in discriminating between patterns of different types. It is difficult to determine an optimal feature space comprising a set of image properties which would produce significant factors influential to classification decision. The approach taken for determining important features is to select image properties, namely ambiguity, size, orientation, and shape measures. Then, it translates all images to this pre-determined feature space.

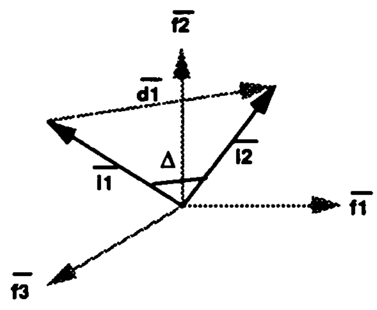

Figure 3.1: Fuzzy Image Feature Vector

Figure 3.1 depicts the sampled feature space having three features

$$\bar{i}_1 = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \qquad \bar{i}_2 = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

C-2

and how the distance, $|\bar{d}_1|$, between two successive frames can be calculated with vector operation, $|\bar{i}_1 - \bar{i}_2|$. Because the goal is to analyze motion, this calculation of change of image constituents from frame to frame in a given time series gives the sampled mean and the sampled variance of all image features. By giving smaller weights to features having larger variance, the important features with small variance have more influence in the decision making process. It is discussed as a useful clustering technique to maximize the inter-set distance or minimize intra-set distance using a diagonal transformation such that features having larger variance are less reliable [12].

3.1 Distance Computation

Before the applied mathematical terms are discussed, the following nomenclatures need to be described.

M	Total number of frames or images
m	Last frame number where m = M-1
N	Total number of features or properties
i	Index to represent current image at *t* where i = 0, 1,....m
k	Index to represent the next image at *t+1* where k = 1,2,...m-1

The sampled mean for the j^{th} feature element is given by,

$$\bar{f}_j = \frac{1}{M} \sum_{i=0}^{m} f_{ij} \quad \text{where } j = 1,2,....,n.$$

Mnemonically, the index of feature element j, where j = 1, 2,....,n, can be represented in the following enumerated terms: edginess, entropy, compactness, ioac, l/h, and b/w, respectively (e.g. $\bar{f}_{entropy}$). To standardize all sampled mean values to be 0.5, the following conversion is performed. This gives equal salience to all features for distance computation [3].

$$f_{ij}^{norm} = 0.5 \; \frac{f_{ij}}{\bar{f}_j}.$$

Consequently, this standardization makes all $\bar{f}_j$ to be set to 0.5. And, the sampled variance for the j^{th} feature element is computed as

$$\sigma_j^2 = \frac{1}{(m-1)} \sum_{i=0}^{m} (f_{ij} - \bar{f}_j)^2 \text{ where } j = 1,2,....,n.$$

The magnitude of the normalized distance between two successive frames i and k is [18],

$$D_{ik}^{norm} = \sqrt{\sum_{j=1}^{n} \frac{(f_{ij} - f_{kj})^2}{\sigma_j^2}} = \sum_{j=1}^{n} \left(\frac{|f_{ij} - f_{kj}|}{\sqrt{\sigma_j^2}} \right).$$

4. EXPERIMENTS

Based on the above formulas, a schematic diagram (Figure 4.1) can be drawn to describe the process of feature selection and frame selection.

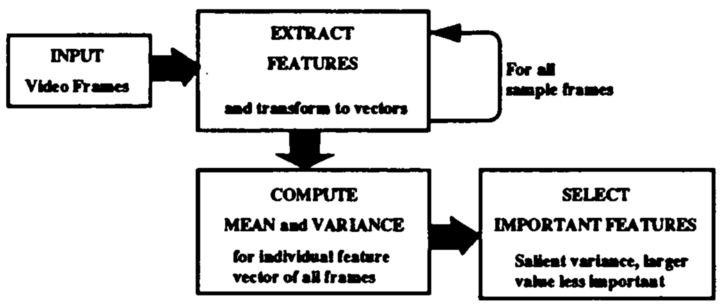

Figure 4.1: Schematic diagram of feature selection process

The distance between two frames in the aforementioned three feature space is computed to check the similarities. If this distance is larger than a predetermined threshold value, then the current video frame is considered to be significantly different from the previous frame, and therefore needs to be registered or stored as one of the abstract keys (Figure 4.2).

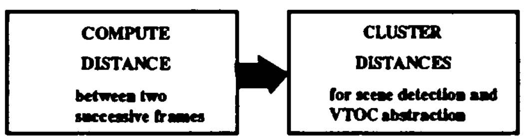

Figure 4.2: Schematic diagram of the frame selection process

4.2 Input Data

Movie film projectors display 24 frames per second whereas NTSC standard television and video devices display 30 frames per second to achieve continuous and fluid full-motion images. The change of inter-frame information is gradual at such high frame rates. For storage conservation and computational efficiency, the simplest way to reduce or abstract video data is to sample it at lower frame rate.

In this paper, a time-suppressed frame rate of one per 5 seconds was assumed. A set of digitized video of previous space shuttle missions obtained from NASA/JSC was used (Figure 4.3). After a pre-processing step, each frame is stored in the CompuServ's Graphic Interchange File (GIF) format for portability.

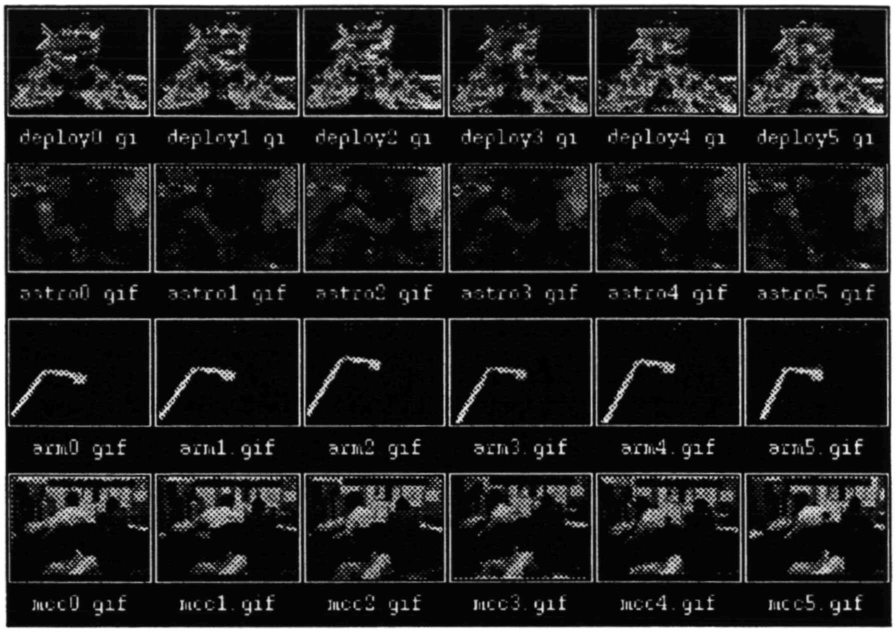

Figure 4.3: Experimental input data

With the fuzzy measures, the resulting distances between each two successive frames are shown in Figures 4.4 through 4.6. The abscissa represents the total number of frame distances in the sampled time series while the ordinate is the computed distance value between two successive images, i.e. $|\bar{i}_j - \bar{i}_{j+1}|$. For example, the abscissa index 0 represents $|\bar{i}_0 - \bar{i}_1|$, 1 represents $|\bar{i}_1 - \bar{i}_2|$, and so on. Each scene consists of six frames, therefore, there is a change of scene at every sixth index on the abscissa. The scene separation is denoted with vertical grid lines. Three sets of detection were experimented as follows:

(1) Entropy, Compactness, L/H (Figure 4.4)
(2) Edginess, IOAC, B/W (Figure 4.5), and
(3) All of the above (Figure 4.6).

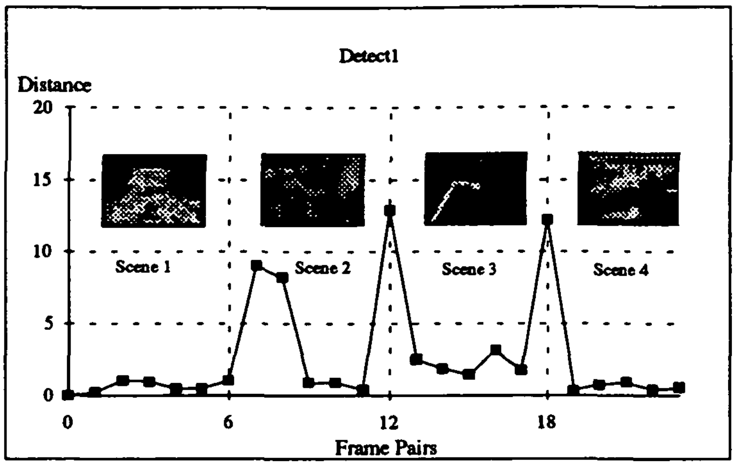

Figure 4.4: Detect 1 - Entropy, Compactness, L/H

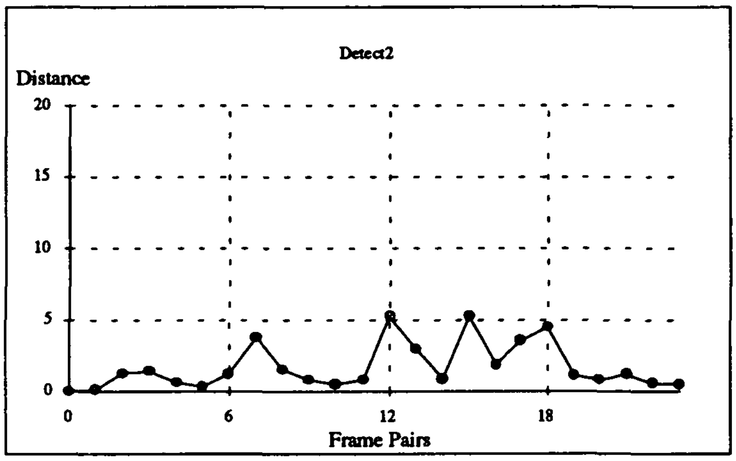

Figure 4.5: Detect 2 - Edginess, IOAC, B/W

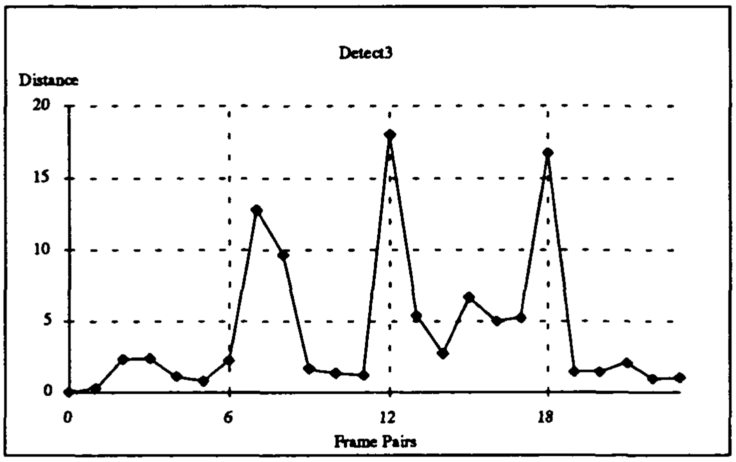

Figure 4.6: Detect 3 - All six features

It is to note that combining all features does not necessarily produce better results just because there are more features. It is not the quantity that is critical, but the discriminatory quality of features.

5. SUMMARY

The technique discussed here needs further improvements. It must have a classifier to correctly cluster the frames to the appropriate scenes. Both statistical and fuzzy approach pattern classifiers are being explored. Video frames that are to be classified are of temporal and dynamic data types, so non-linear classification methods need to be implemented. Scene classification is quite subjective in nature; therefore, the interactive tool developed here can be further extended to provide human interaction in setting problem-dependent criteria for this machine recognition task. Furthermore, the scenes that are detected may not necessarily be different from one another, but rather compose a video segment or document. A hierarchical abstraction scheme that allows for a higher level of abstraction will better suit the visual data management environment.

Finally, in the merging worlds of computers and media, new technologies mix traditional media such as video and publications with computer media as interactive, informational and entertainment software. This trend is rapidly growing at an unprecedented rate. Once digital video becomes a repository of common data on computers, the data needs to be accessed and manipulated just as documents are retrieved and managed by a DBMS. It might be useful to investigate new video inter-referencing strategies in correlating various context from the same event to derive knowledge points. Thus, this automatic abstraction of video index keys for non-linear, frame-accurate access would make information acrchival and retrieval applications more robust and efficient.

6. ACKNOWLEDGMENTS

The authors thank Dr. Robert N. Lea of NASA/JSC for his support in this work. This research was done as a master's thesis by Albert Leigh at the University of Houston-Clear Lake. It was partially supported by a NASA/JSC fellowship, "Graphics Technology in Space Application Program".

7. REFERENCES

1. Bezdeck, J. C. and Pal, S. K., Fuzzy Models for Pattern Recognition, IEEE Press, 1992
2. De Luca, A., and Termini, S., *Information Control*, vol. 20, pp. 301, 1972
3. Duda, R. and Hart, P., Pattern Classification and Scene Analysis, John Wiley & Sons, 1973
4. Gonzalez, R. C. and Wintz, P., Digital Image Processing, Addison-Wesley, 1988
5. Jain, A. K. Fundamentals of Digital Image Processing, Prentice-Hall, 1989
6. Kaufman, A., Introduction to the Theory of Fuzzy Subsets, Academic Press, 1975

7. Klir, G. J. and Folger, T. A., Fuzzy Sets, Uncertainty, and Information, Prentice-Hall, 1988
8. Pal, N. R. and Pal, S. K., "Object-Background Segmentation Using New Definitions of Entropy", *IEE Proceedings*, vol. 136, Pt. E, No. 4, July 1989
9. Pal, N. R. and Pal, S. K., "Entropic Thresholding", *Signal Processing*, vol 16, pp. 97-108, 1989
10. Pal, S. K., "Fuzzy Tools for the Management of Uncertainty in Pattern Recognition, Image Analysis, Vision and Expert systems", *International Journal of Systems Science*, vol. 22, No. 3, pp. 511-549, 1991
11. Pal, S. K., "Fuzziness, Image Information and Scene Analysis", Introduction to Fuzzy Logic Applications and Intelligent Systems, Edited by Yager R., and Zadeh, L., pp. 147-183, 1992
12. Pal, S. K., and Majumder, D. K., Fuzzy Mathematical Approach to Pattern Recognition, Halsted Press, 1986
13. Rosenfeld, A., "The Fuzzy Geometry of Image Subsets", *Pattern Recognition Letters*, vol. 2, pp. 311-317, September 1984
14. Rosenfeld, A., "Fuzzy Digital Topology", *Information Control*, vol. 40, no. 1, pp. 76-87, January 1979
15. Rosenfeld, A., and Haber, S., *Pattern Recognition*, vol. 18, pp 125, 1985
16. Rosenfeld, A. and Kak, A. C., Digital Picture Processing, vols. 1 and 2, Academic Press, Inc., 1982
17. Shannon, C. E., "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27, pp. 379-423, July 1948
18. Tou, J. and Gonzalez, R. C., Pattern Recognition Principles, Addison-Wesley, 1974
19. Zadeh, L. A., "Fuzzy Sets", *Information and Control*, vol. 8, pp. 338-352, 1965