Computational Intelligence in Multimedia Processing: Foundation and Trends

Aboul-Ella Hassanien[1,2], Ajith Abraham[3], Janusz Kacprzyk[4], and James F. Peters[5]

[1] Information Technology Department, FCI
Cairo University
5 Ahamed Zewal Street, Orman, Giza, Egypt
`a.hassanien@fci-cu.edu.eg`
[2] Information System Department, CBA
Kuwait University, Kuwait
`abo@cba.edu.kw`
[3] Center for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology
O.S. Bragstads plass 2E, N-7491 Trondheim, Norway
`ajith.abraham@ieee.org`, `abraham.ajith@acm.org`
[4] Systems Research Institute Polish Academy of Sciences
ul. Newelska 6 01-447 Warsaw, Poland
`kacprzyk@ibspan.waw.pl`
[5] Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba R3T 5V6, Canada
`jfpeters@ee.umanitoba.ca`

Summary. This chapter presents a broad overview of Computational Intelligence (CI) techniques including Neural Network (NN), Particle Swarm Optimization (PSO), Evolutionary Algorithm (GA), Fuzzy Set (FS), and Rough Sets (RS). In addition, a very brief introduction to near sets and near images which offer a generalization of traditional rough set theory and a new approach to classifying perceptual objects by means of features in solving multimedia problems is presented. A review of the current literature on CI based approaches to various problems in multimedia computing such as speech, audio and image processing, video watermarking, content-based multimedia indexing and retrieval are presented. We discuss some representative methods to provide inspiring examples to illustrate how CI could be applied to resolve multimedia computing problems and how multimedia could be analyzed, processed, and characterized by computational intelligence. Challenges to be addressed and future directions of research are also presented.

1 Introduction

Last few decades have seen a new era of artificial intelligence focusing on the principles, theoretical aspects, and design methodology of algorithms gleaned from nature. Examples are artificial neural networks inspired by mammalian neural systems, evolutionary computation inspired by natural selection in biology, simulated annealing inspired by thermodynamics principles and swarm intelligence inspired by collective behavior of insects or micro-organisms, etc., interacting locally with their environment causing coherent functional global patterns to emerge. Computational intelligence is a well-established paradigm, where new theories with a sound biological understanding have been evolving. The current experimental systems have many of the characteristics of biological computers (brains in other words) and are beginning to be built to perform a variety of tasks that are difficult or impossible to do with conventional computers. Defining computational intelligence is not an easy task [95]. In a nutshell, which becomes quite apparent in light of the current research pursuits, the area is heterogeneous as being dwelled on such technologies as neural networks, fuzzy systems, rough sets, evolutionary computation, swarm intelligence, probabilistic reasoning [13] and multi-agent systems. The recent trend is to integrate different components to take advantage of complementary features and to develop a synergistic system. Hybrid architectures like neuro-fuzzy systems, evolutionary-fuzzy systems, evolutionary-neural networks, evolutionary neuro-fuzzy systems, rough-neural, rough-fuzzy, etc., are widely applied for real world problem solving.

Multimedia is any multiple forms of media integrated together at a time. In modern times, the advent of musical accompaniment to silent films was an early form of multimedia. Even the simplest ancient dance forms use multiple media types in the form of sound and vision to convey additional meaning. The currently accepted understanding of multimedia generally involves a variety of media, such as still images, video, sound, music and text, presented using a computer as the storage device, delivery controller and delivery medium. The various media types are usually stored as digital assets and their delivery to the viewer is facilitated by some sort of authoring language. The multimedia technology is one kind development rapid the natural subinformation technique, it changes computer and brings a profound revolution. Multimedia technique will accelerate the development of our live. Even nowadays, most media types are only designed to be perceived by two senses, vision and hearing. Still, incredibly powerful messages can be communicated using just these two senses. A subset of multimedia is interactive multimedia. In this definition the delivery of the assets is dependent on decisions made by the viewer at the time of viewing. Some subject areas lend themselves to interactivity, such as self-paced learning and game play. Other areas are mostly not enhanced by interactivity: here we find the traditional film and storytelling genres, where we are expected to travel in a prescribed direction to perceive the message in a sequential fashion.

Current research of multimedia processing is shifting from coding (MPEG-1,2,4) to automatic recognition (MPEG-7). Its research domain will cover techniques for object-based representation and coding; segmentation and tracking; pattern detection and recognition; multimodal signals fusion, conversion and synchronization, as well as content-based indexing and subject-based retrieval and browsing.

Multimedia processing is a very important scientific research domain with a broad range of applications. The development of new insights and applications results from both fundamental scientific research and the development of new technologies. One of these emerging technologies is computational intelligence, which is a generic term for a specific collection of tools to model uncertainty, imprecision, evolutionary behavior and complex models. This chapter will be a comprehensive view of modern computational intelligence theory in the field of multimedia processing.

The objective of this book chapter is to present to the computational intelligence techniques and multimedia processing research communities the state of the art in the computational intelligence applications to multimedia processing and motivate research in new trend-setting directions. Hence, we review and discuss in the following Sections some representative methods to provide inspiring examples to illustrate how CI techniques could be applied to resolve multimedia problems and how multimedia could be analyzed, processed, and characterized by computational intelligence. These representative examples include (1) Computational Intelligence for speech, audio, image and video processing, (2) CI in audio–visual recognition systems, (3) Computational Intelligence in multimedia watermarking, and (4) CI in multimedia content-based indexing and retrieval.

To provide useful insights for CI applications in multimedia processing, we structure the rest of this chapter into five Sections. Section 2 introduces the fundamental aspects of the key components of modern computational intelligence including neural networks, rough sets, fuzzy sets, particle swarm optimization algorithm, evolutionary algorithm and near sets. Section 3 reviews some past literature in using the computational intelligence in speech, audio, and image processing, as well as in speech emotion recognition and audio–visual recognition systems. A review of the current literature on computational intelligence based approaches in video processing problems such as video segmentation as well as adaptation of c-means clustering algorithm to rough set theory in solving multimedia segmentation and clustering problems is presented in Sect. 4. Section 5 reviews and discuss some successful work to illustrate how CI could be applied to multimedia watermarking problems. Computational intelligence in content-based multimedia indexing and retrieval is reviewed in Sect. 6. Challenges and future trends are addressed and presented in Sect. 7.

2 Computational Intelligence: Foundations

In the following subsection, we present an overview of the modern computational intelligence techniques with their advantages including neural networks, fuzzy sets, particle swarm optimization, genetic algorithm, rough sets and near sets.

2.1 Artificial Neural Networks

Artificial neural networks have been developed as generalizations of mathematical models of biological nervous systems. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the non-linear characteristic exhibited by neurons is represented by a transfer function. There are a range of transfer functions developed to process the weighted and biased inputs, among which four basic transfer functions widely adopted for multimedia processing are illustrated in Fig. 1.

The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. Most applications of neural networks fall into the following categories:

- *Prediction*: Use input values to predict some output
- *Classification*: Use input values to determine the classification
- *Data association*: Like classification but it also recognizes data that contains errors
- *Data conceptualization*: Analyze the inputs so that grouping relationships can be inferred

Mathematical Modeling and Learning in Neural Networks

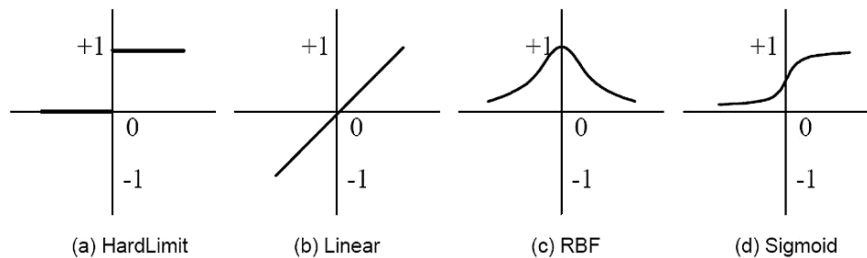A typical multilayered neural network and an artificial neuron are illustrated in Fig. 2.

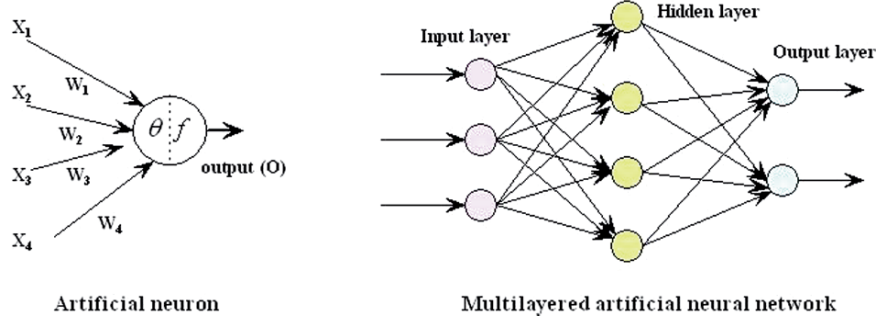
Fig. 1. Basic transfer functions

Fig. 2. Typical multilayered neural network

Each neuron is characterized by an activity level (representing the state of polarization of a neuron), an output value (representing the firing rate of the neuron), a set of input connections, (representing synapses on the cell and its dendrite), a bias value (representing an internal resting level of the neuron), and a set of output connections (representing a neuron's axonal projections). Each of these aspects of the unit is represented mathematically by real numbers. Thus each connection has an associated weight (synaptic strength), which determines the effect of the incoming input on the activation level of the unit. The weights may be positive or negative. Referring to Fig. 2, the signal flow from inputs $\{x1, \ldots, x_n\}$ is considered to be unidirectional indicated by arrows, as is a neuron's output signal flow (O). The neuron output signal O is given by the following relationship:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j \right), \tag{1}$$

where w_j is the weight vector and the function $f(net)$ is referred to as an activation (transfer) function. The variable net is defined as a scalar product of the weight and input vectors

$$net = w^T x = w_1 x_1 + \cdots + w_n x_n , \tag{2}$$

where T is the transpose of a matrix. A typical Gaussian and logistic activation function is plotted in Fig. 3.

Neural Network Architecture

The behavior of the neural network depends largely on the interaction between the different neurons. The basic architecture consists of three types of neuron layers: input, hidden and output layers. In feed-forward networks, the signal flow is from input to output units strictly in a feed-forward direction. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of
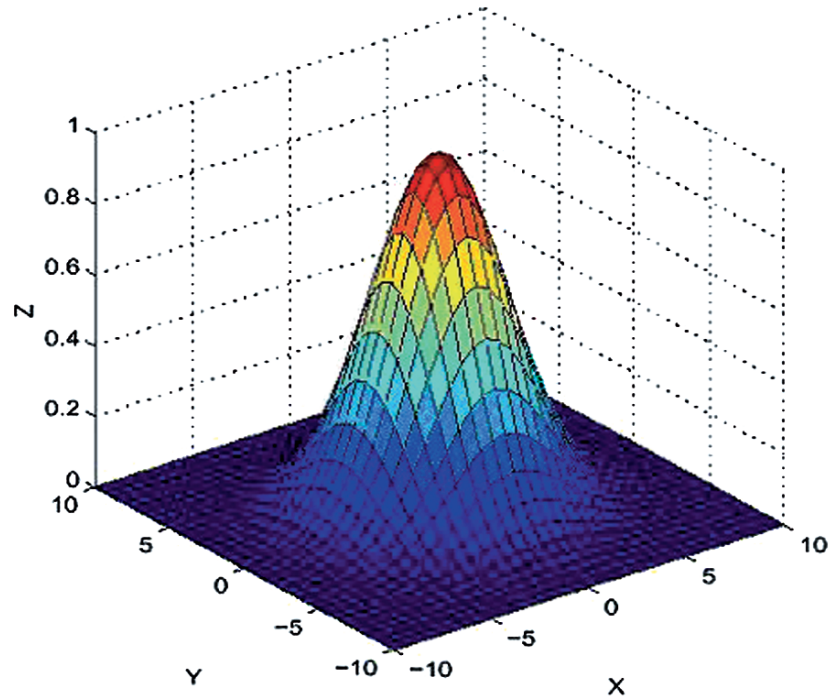
Fig. 3. Typical Gaussian and logistic activation function

units to inputs of units in the same layer or previous layers. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behavior constitutes the output of the network. There are several other neural network architectures (Elman network, adaptive resonance theory maps, competitive networks, etc.) depending on the properties and requirement of the application. Reader may refer to [2] for an extensive overview of the different neural network architectures and learning algorithms. A neural network has to be configured such that the application of a set of inputs produces the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to train the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. The learning situations in neural networks may be classified into three distinct sorts. These are supervised learning, unsupervised learning and reinforcement learning. In supervised learning, an input vector is presented at the inputs together with a set of desired responses, one

for each node, at the output layer. A forward pass is done and the errors or discrepancies, between the desired and actual response for each node in the output layer, are found. These are then used to determine weight changes in the net according to the prevailing learning rule. The term 'supervised' originates from the fact that the desired signals on individual output nodes are provided by an external teacher. The best-known examples of this technique occur in the backpropagation algorithm, the delta rule and perceptron rule. In unsupervised learning (or self-organization) a (output) unit is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli. Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and-error search and delayed reward are the two most important distinguishing features of reinforcement learning.

Major Neural Network Architecture and Learning Models

Via selection of transfer function and connection of neurons, various neural networks can be constructed to be trained for producing the specified outputs. Major neural networks that are commonly used for multimedia applications are classified as feed-forward neural network, feedback network or recurrent, self-organizing map and Adaptive Resonance Theory (ART) networks. The learning paradigms for the neural networks in multimedia processing generally include supervised networks and unsupervised networks. In supervised training, the training data set consists of many pairs in the source and target patterns. The network processes the source inputs and compares the resulting outputs against the target outputs, and adjusts its weights to improve the correct rate of the resulting outputs. In unsupervised networks, the training data set does not include any target information.

Feed-Forward Neural Network

A general Feed-forward network often consists of multiple layers, typically including one input layer, a number of hidden layers, and an output layer. In the feed-forward neural networks, the neuron in each layer are only fully interconnected with the neurons in the next layer, which means signals or information being processed travel along a single direction. Back-propagation (BP) network is a supervised feed-forward neural network and it is a simple

stochastic gradient descent method to minimize the total squared error of the output computed by the neural network. Its errors propagate backwards from the output neurons to the inner neurons. The processes of adjusting the set of weights between the layers and recalculating the output continue until a stopping criterion is satisfied. The Radial basis function (RBF) network is a three-layer supervised feed-forward network that uses a nonlinear transfer function (normally the Gaussian) for the hidden neurons and a linear transfer function for the output neurons. The Gaussian function is usually applied to the net input to produce a radial function of the distance between each pattern vector and each hidden unit weight vector.

Recurrent Networks

Recurrent networks are the state-of-the-art in nonlinear time series prediction, system identification, and temporal pattern classification. As the output of the network at time t is used along with a new input to compute the output of the network at time $t + 1$, the response of the network is dynamic. Time-Lag Recurrent Networks (TLRN) are multi-layered perceptrons extended with short-term memory structures that have local recurrent connections. The TLRN is a very appropriate model for processing temporal (time-varying) information. Examples of temporal problems include time series prediction, system identification and temporal pattern recognition. The training algorithm used with TLRNs (backpropagation through time) is more advanced than standard backpropagation algorithm. The main advantage of TLRNs is the smaller network size required to learn temporal problems when compared to MLP that use extra inputs to represent the past samples (equivalent to time delay neural networks). An added advantage of TLRNs is their low sensitivity to noise.

Self Organizing Feature Maps

Self Organizing Feature Maps (SOFM) are a data visualization technique proposed by Kohonen [3], which reduce the dimensions of data through the use of self-organizing neural networks. A SOFM learns the categorization, topology and distribution of input vectors. SOFM allocate more neurons to recognize parts of the input space where many input vectors occur and allocate fewer neurons to parts of the input space where few input vectors occur. Neurons next to each other in the network learn to respond to similar vectors. SOFM can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. An important feature of SOFM learning algorithm is that it allow neurons that are neighbors to the winning neuron to output values. Thus the transition of output vectors is much smoother than that obtained with competitive layers, where only one neuron has an output at a time. The problem that data visualization attempts to solve is that humans simply cannot visualize high dimensional data. The way SOFM go about reducing dimensions is by producing a map of usually

1 or 2 dimensions, which plot the similarities of the data by grouping similar data items together (data clustering). In this process, SOFM accomplish two things, they reduce dimensions and display similarities. It is important to note that while a self-organizing map does not take long to organize itself so that neighboring neurons recognize similar inputs, it can take a long time for the map to finally arrange itself according to the distribution of input vectors.

Adaptive Resonance Theory

Adaptive Resonance Theory (ART) was initially introduced by Grossberg [5] as a theory of human information processing. ART neural networks are extensively used for supervised and unsupervised classification tasks and function approximation. There are many different variations of ART networks available today [4]. For example, ART1 performs unsupervised learning for binary input patterns, ART2 is modified to handle both analog and binary input patterns, and ART3 performs parallel searches of distributed recognition codes in a multilevel network hierarchy. ARTMAP combines two ART modules to perform supervised learning while fuzzy ARTMAP represents a synthesis of elements from neural networks, expert systems, and fuzzy logic.

2.2 Rough Sets

Rough set theory [75–77,87] is a fairly new intelligent technique for managing uncertainty that has been applied to the medical domain and is used for the discovery of data dependencies, evaluates the importance of attributes, discovers the patterns of data, reduces all redundant objects and attributes, seeks the minimum subset of attributes, recognize and classify objects in medical imaging. Moreover, it is being used for the extraction of rules from databases. Rough sets have proven useful for representation of vague regions in spatial data. One advantage of the rough set is the creation of readable if–then rules. Such rules have a potential to reveal new patterns in the data material; furthermore, it also collectively functions as a classifier for unseen data sets. Unlike other computational intelligence techniques, rough set analysis requires no external parameters and uses only the information presented in the given data. One of the nice features of rough sets theory is that its can tell whether the data is complete or not based on the data itself. If the data is incomplete, it suggests more information about the objects needed to be collected in order to build a good classification model. On the other hand, if the data is complete, rough sets can determine whether there are more than enough or redundant information in the data and find the minimum data needed for classification model. This property of rough sets is very important for applications where domain knowledge is very limited or data collection is very expensive/laborious because it makes sure the data collected is just good enough to build a good classification model without sacrificing the accuracy of the classification model or wasting time and effort to gather extra information about the objects [75–77,87].

In rough sets theory, the data is collected in a table, called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, we assume that the a set of examples with a class label to indicate the class to which each example belongs are given. We call the class label the decision attributes, the rest of the attributes the condition attributes. Rough sets theory defines three regions based on the equivalent classes induced by the attribute values: *lower approximation*, *upper approximation* and *boundary*. Lower approximation contains all the objects, which are classified surely based on the data collected, and upper approximation contains all the objects which can be classified probably, while the boundary is the difference between the upper approximation and the lower approximation. So, we can define a rough set as any set defined through its lower and upper approximations. On the other hand, indiscernibility notion is fundamental to rough set theory. Informally, two objects in a decision table are indiscernible if one cannot distinguish between them on the basis of a given set of attributes. Hence, indiscernibility is a function of the set of attributes under consideration. For each set of attributes we can thus define a binary indiscernibility relation, which is a collection of pairs of objects that are indiscernible to each other. An indiscernibility relation partitions the set of cases or objects into a number of equivalence classes. An equivalence class of a particular object is simply the collection of objects that are indiscernible to the object in question. Here we provide an explanation of the basic framework of rough set theory, along with some of the key definitions. A review of this basic material can be found in sources such as [74–77, 87] and many others.

2.3 Near Sets: Generalization of the Rough Set in Multimedia Processing

Near sets [67, 78–81, 83] offer a generalization of traditional rough set theory [84–88] and a new approach to classifying perceptual objects by means of features [89–94]. The near set approach can be used to classify images that are qualitatively but not necessary quantitatively close to each other. This is essentially the idea expressed in classifying images in [67, 81]. If one adopts the near set approach in image processing, a byproduct of the approach is the separation of images into non-overlapping sets of images that are similar (*descriptively* near to) each other. This has recently led to an application of the near set approach in 2D and 3D interactive gaming with a vision system that learns and serves as the backbone for an adaptive telerehabilitation system for patients with finger, hand, arm and balance disabilities (see, e.g., [100, 101]). Each remote node in the telerehabilitation system includes a vision system that learns to track the behavior of a patient. Images deemed to be 'interesting' (e.g., images representing erratic behavior) are stored as well as forwarded to a rehabilitation center for followup. In such a system, there is a need to identify images that are in some sense near images representing some standard or norm. This research has led to a study of methods of automating image

segmentation as a first step in near set-based image processing. This section is limited to a very brief introduction to near sets and near images useful in image pattern recognition.

Object Description

Perceptual objects that have the same appearance are considered *qualitatively near* each other, i.e., objects with matching descriptions. A *description* is a tuple of values of functions representing features of an object [79]. For simplicity, assume the description of an object consists of one function value. For example, let $w \in I$, $w' \in I'$ be $n \times m$ pixel windows contained in two images I, I' and $\phi(w) =$ information content of pixel window w, where information content is a feature of a pixel window and ϕ is a sample function representing information content defined in the usual way [99]. Then pixel window w is near pixel window w' if $\phi(w) = \phi(w')$.

Near Objects

Objects are known by their descriptions. An *object description* is defined by means of a tuple of function values $\phi(x)$ associated with an object $x \in X$. Assume that $B \subseteq \mathcal{F}$ is a given set of functions representing features of sample objects $X \subseteq \mathcal{O}$. Let $\phi_i \in B$, where $\phi_i : \mathcal{O} \longrightarrow \Re$. In combination, the functions representing object features provide a basis for an *object description* $\phi : \mathcal{O} \longrightarrow \Re^L$, a vector containing measurements (returned values) associated with each functional value $\phi_i(x)$ in (3), where the description length $|\phi| = L$.

Object Description: $\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_i(x), \ldots, \phi_L(x)).$ (3)

The intuition underlying a description $\phi(x)$ is a recording of measurements from sensors, where each sensor is modeled by a function ϕ_i. Then let $\Delta\phi_i$ denote

$$\Delta\phi_i = \phi_i(x') - \phi_i(x),$$

where $x, x' \in \mathcal{O}$. The difference $\Delta\phi$ leads to a definition of the indiscernibility relation $\sim_B$ introduced by Pawlak [86] (see Definition 1).

Definition 1. *Indiscernibility Relation Let* $x, x' \in \mathcal{O}, B \in \mathcal{F}$.

$$\sim_B = \left\{ (x, x') \in \mathcal{O} \times \mathcal{O} \mid \forall \phi_i \in B \centerdot \Delta\phi_i = 0 \right\},$$

where $i \leq |\phi|$ (description length).

Near Sets

The basic idea in the near set approach to object recognition is to compare object descriptions. Sets of objects X, X' are considered near each other if the sets contain objects with at least partial matching descriptions.

Definition 2. *Near Sets Let X, $X' \subseteq \mathcal{O}$, $B \subseteq \mathcal{F}$. Set X is near X' if, and only if there exists $x \in X, x' \in X', \phi_i \in B$ such that $x \sim_{\{\phi_i\}} x'$.*

For example, assume that a pair of images I, I', where a pixel window in image I has a description that matches the description of a pixel window in image I'. The objects in this case are pixel windows. By definition, I, I' are near sets and, from an image classification perspective, I, I' are near images. Object recognition problems, especially in images [67], and the problem of the nearness of objects have motivated the introduction of near sets (see, e.g., [81, 83]).

Near Images

In the context of image processing, the relation $\sim_B$ in Definition 1 is important because it suggests a way to classify images by a number of straightforward steps: (1) identify an image object, e.g., pixel window, (2) select a set B containing functions representing features of an image object such as a pixel window, (3) partition each image using $\sim_B$ and then compare a representative object from a class in each partition. In the case where one discovers that the objects in the selected classes have matching descriptions, then this means the images are near each other at the class level. In effect, if near images are discovered, this means a pair of sample images have been effectively classified. This is important because it leads to effective image segmentation method.

2.4 Fuzzy Sets

Zadeh [115] introduced the concept of fuzzy logic to present vagueness in linguistics, and further implement and express human knowledge and inference capability in a natural way. Fuzzy logic starts with the concept of a fuzzy set. A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership. A Membership Function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the universe of discourse. Let X be the universe of discourse and x be a generic element of X. A classical set A is defined as a collection of elements or objects $x \in X$, such that each x can either belong to or not belong to the set A, $A \sqsubseteq X$. By defining a characteristic function (or membership function) on each element x in X, a classical set A can be represented by a set of ordered pairs $(x, 0)$ or $(x, 1)$, where 1 indicates membership and 0 non-membership. Unlike conventional set mentioned above fuzzy set expresses the degree to which an element belongs to a set. Hence the characteristic function of a fuzzy set is allowed to have value between 0 and 1, denoting the degree of membership of an element in a given set. If X is a collection of objects denoted generically by x, then a fuzzy set A in X is defined as a set of ordered pairs:
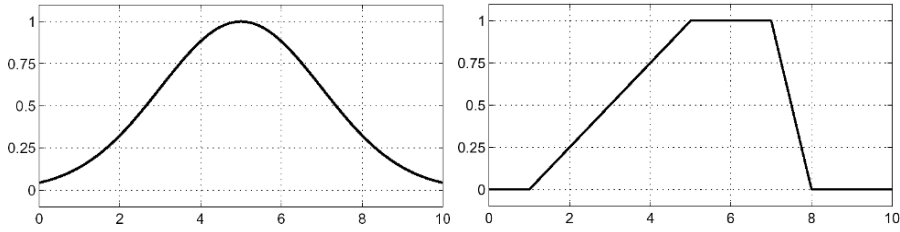
Fig. 4. Shapes of two commonly used MF's

$$A = \{(x, \mu_A(x)) \mid x \in X\}, \tag{4}$$

$\mu_A(x)$ is called the membership function of linguistic variable x in A, which maps X to the membership space M, $M = [0, 1]$, where M contains only two points 0 and 1, A is crisp and $\mu_A(x)$ is identical to the characteristic function of a crisp set. Triangular and trapezoidal membership functions are the simplest membership functions formed using straight lines. Some of the other shapes are Gaussian, generalized bell, sigmoidal and polynomial based curves.

Figure 4 illustrates the shapes of two commonly used MF's. The most important thing to realize about fuzzy logical reasoning is the fact that it is a superset of standard Boolean logic.

Fuzzy Logic Operators

It is interesting to note about the correspondence between two-valued and multi-valued logic operations for AND, OR, and NOT. It is possible to resolve the statement A AND B, where A and B are limited to the range (0,1), by using the operator minimum (A, B). Using the same reasoning, we can replace the OR operation with the maximum operator, so that A OR B becomes equivalent to maximum (A, B). Finally, the operation NOT A becomes equivalent to the operation $1 - A$. In fuzzy logic terms these are popularly known as fuzzy intersection or conjunction (AND), fuzzy union or disjunction (OR), and fuzzy complement (NOT). The intersection of two fuzzy sets A and B is specified in general by a binary mapping T, which aggregates two membership functions as follows.

$$\mu_{A \cap B}(x) = T(\mu_A(x), \mu_B(x)) \tag{5}$$

The fuzzy intersection operator is usually referred to as T-norm (Triangular norm) operator. The fuzzy union operator is specified in general by a binary mapping S.

$$\mu_{A \cup B}(x) = S(\mu_A(x), \mu_B(x)) \tag{6}$$

This class of fuzzy union operators are often referred to as T-conorm (or S-norm) operators 5.

If–then Rules and Fuzzy Inference Systems

The fuzzy rule base is characterized in the form of *if–then* rules in which preconditions and consequents involve linguistic variables. The collection of these fuzzy rules forms the rule base for the fuzzy logic system. Due to their concise form, fuzzy *if–then* rules are often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. A single fuzzy *if–then* rule assumes the form:

$$if\ x\ is\ A\ then\ y\ is\ B,$$

where A and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. The *if*-part of the rule $'x$ is A' is called the antecedent (pre-condition) or premise, while the *then*-part of the rule $'y$ is B' is called the consequent or conclusion. Interpreting an *if–then* rule involves evaluating the antecedent (fuzzification of the input and applying any necessary fuzzy operators) and then applying that result to the consequent (known as implication). For rules with multiple antecedents, all parts of the antecedent are calculated simultaneously and resolved to a single value using the logical operators. Similarly all the consequents (rules with multiple consequents) are affected equally by the result of the antecedent. The consequent specifies a fuzzy set be assigned to the output. The implication function then modifies that fuzzy set to the degree specified by the antecedent. For multiple rules, the output of each rule is a fuzzy set. The output fuzzy sets for each rule are then aggregated into a single output fuzzy set. Finally the resulting set is defuzzified, or resolved to a single number. The defuzzification interface is a mapping from a space of fuzzy actions defined over an output universe of discourse into a space of non-fuzzy actions, because the output from the inference engine is usually a fuzzy set while for most practical applications crisp values are often required. The three commonly applied defuzzification techniques are, max-criterion, center-of-gravity and the mean- of- maxima. The max-criterion is the simplest of these three to implement. It produces the point at which the possibility distribution of the action reaches a maximum value. Reader may please refer to [7] for more information related to fuzzy systems. It is typically advantageous if the fuzzy rule base is adaptive to a certain application. The fuzzy rule base is usually constructed manually or by automatic adaptation by some learning techniques using evolutionary algorithms and/or neural network learning methods [6].

Fuzzy Image Processing

The adoption of the fuzzy paradigm is desirable in image processing because of the uncertainty and imprecision present in images, due to noise, image sampling, lightning variations and so on. Fuzzy theory provides a mathematical

tool to deal with the imprecision and ambiguity in an elegant and efficient way. Fuzzy techniques can be applied to different phases of the segmentation process; additionally, fuzzy logic allows to represent the knowledge about the given problem in terms of linguistic rules with meaningful variables, which is the most natural way to express and interpret information. Fuzzy image processing [10, 68, 73, 102, 112] is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. An image I of size $M \times N$ and L gray levels can be considered as an array of fuzzy singletons, each having a value of membership denoting its degree of brightness relative to some brightness levels. For an image I, we can write in the notation of fuzzy sets:

$$I = \bigcup_{MN} \frac{\mu(g_{mn})}{g_{mn}} \ ,$$

(7)

where g_{mn} is the intensity of (m, n) the pixel and μ_{mn} its membership value.

The membership function characterizes a suitable property of image (e.g., edginess, darkness, textural property) and can be defined globally for the whole image or locally for its segments. In recent years, some researchers have applied the concept of fuzziness to develop new algorithms for image processing tasks, for example image enhancement, segmentation, etc. Fuzzy image processing system is a rule-based system that uses fuzzy logic to reason about image data. Its basic structure consists of four main components, as depicted in Fig. 5.
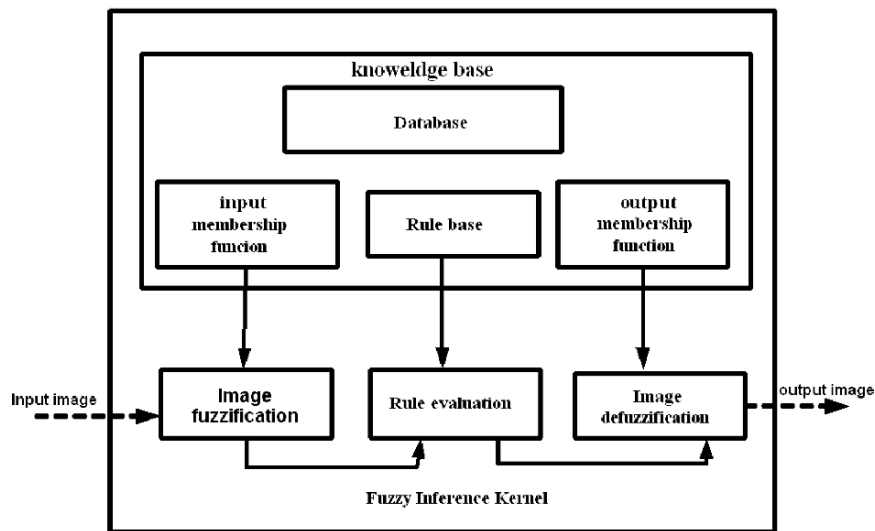

Fig. 5. Fuzzy image processing system [10]

- The coding of image data (fuzzifier), which translates gray-level plane to the membership plane
- An inference engine that applies a fuzzy reasoning mechanism to obtain a fuzzy output
- Decoding the result of fuzzification (defuzzifier), which translates this latter output into a gray-level plane; and
- Knowledge base, which contains both an ensemble of fuzzy rules, known as the rule base, and an ensemble of membership functions known as the database

The decision-making process is performed by the inference engine using the rules contained in the rule base. These fuzzy rules define the connection between input and output fuzzy variables. The inference engine evaluates all the rules in the rule base and combines the weighted consequents of all relevant rules into a single output fuzzy set.

2.5 Evolutionary Algorithms

Evolutionary algorithms (EA) are adaptive methods, which may be used to solve search and optimization problems, based on the genetic processes of biological organisms. Over many generations, natural populations evolve according to the principles of natural selection and 'survival of the fittest', first clearly stated by Charles Darwin in *The Origin of Species*. By mimicking this process, evolutionary algorithms are able to 'evolve' solutions to real world problems, if they have been suitably encoded [12]. Usually grouped under the term evolutionary algorithms or evolutionary computation, we find the domains of genetic algorithms [15, 16], evolution strategies [21], evolutionary programming [11], genetic programming [18] and learning classifier systems. They all share a common conceptual base of simulating the evolution of individual structures via processes of selection, mutation, and reproduction. The processes depend on the perceived performance of the individual structures as defined by the environment (problem).

EA's deal with parameters of finite length, which are coded using a finite alphabet, rather than directly manipulating the parameters themselves. This means that the search is unconstrained neither by the continuity of the function under investigation, nor the existence of a derivative function.

Figure 6 depicts the functional block diagram of a Genetic Algorithm (GA) and the various aspects are discussed below. It is assumed that a potential solution to a problem may be represented as a set of parameters. These parameters (known as genes) are joined together to form a string of values (known as a chromosome). A gene (also referred to a feature, character or detector) refers to a specific attribute that is encoded in the chromosome. The particular values the genes can take are called its alleles. The position of the gene in the chromosome is its locus. Encoding issues deal with representing a solution in a chromosome and unfortunately, no one technique works best for all problems.
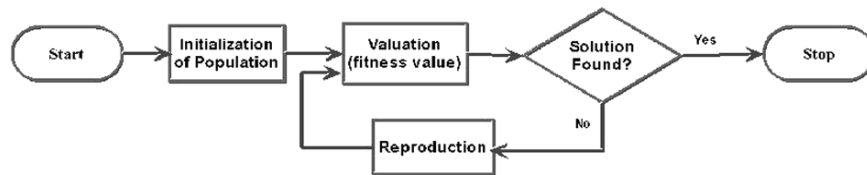
Fig. 6. The functional block diagram of a genetic algorithm

A fitness function must be devised for each problem to be solved. Given a particular chromosome, the fitness function returns a single numerical fitness or figure of merit, which will determine the ability of the individual, which that chromosome represents. Reproduction is the second critical attribute of GA's where two individuals selected from the population are allowed to mate to produce offspring, which will comprise the next generation. Having selected two parents, their chromosomes are recombined, typically using the mechanisms of crossover and mutation.

There are many ways in which crossover can be implemented. In a single point crossover two chromosome strings are cut at some randomly chosen position, to produce two 'head' segments, and two 'tail' segments. The tail segments are then swapped over to produce two new full-length chromosomes. Crossover is not usually applied to all pairs of individuals selected for mating. Another genetic operation is mutation, which is an asexual operation that only operates on one individual. It randomly alters each gene with a small probability. Traditional view is that crossover is the more important of the two techniques for rapidly exploring a search space. Mutation provides a small amount of random search, and helps ensure that no point in the search space has a zero probability of being examined. If the GA has been correctly implemented, the population will evolve over successive generations so that the fitness of the best and the average individual in each generation increases towards the global optimum. Selection is the survival of the fittest within GA's. It determines which individuals are to survive to the next generation. The selection phase consists of three parts. The first part involves determination of the individual's fitness by the fitness function. A fitness function must be devised for each problem; given a particular chromosome, the fitness function returns a single numerical fitness value, which is proportional to the ability, or utility, of the individual represented by that chromosome. For many problems, deciding upon the fitness function is very straightforward, for example, for a function optimization search; the fitness is simply the value of the function. Ideally, the fitness function should be smooth and regular so that chromosomes with reasonable fitness are close in the search space, to chromosomes with slightly better fitness. However, it is not always possible to construct such ideal fitness functions. The second part involves converting the fitness function into an expected value followed by the last part where the expected value is then converted to a discrete number of offspring. Some

of the commonly used selection techniques are roulette wheel and stochastic universal sampling. Genetic programming applies the GA concept to the generation of computer programs. Evolution programming uses mutations to evolve populations. Evolution strategies incorporate many features of the GA but use real-valued parameters in place of binary-valued parameters. Learning classifier systems use GAs in machine learning to evolve populations of condition/action rules.

2.6 Intelligent Paradigms: Probabilistic Computing and Swarm Intelligence

Probabilistic models are viewed as similar to that of a game, actions are based on expected outcomes. The center of interest moves from the deterministic to probabilistic models using statistical estimations and predictions. In the probabilistic modeling process, risk means uncertainty for which the probability distribution is known. Therefore risk assessment means a study to determine the outcomes of decisions along with their probabilities. Decision-makers often face a severe lack of information. Probability assessment quantifies the information gap between what is known, and what needs to be known for an optimal decision. The probabilistic models are used for protection against adverse uncertainty, and exploitation of propitious uncertainty.

Swarm intelligence is aimed at collective behavior of intelligent agents in decentralized systems. Although there is typically no centralized control dictating the behavior of the agents, local interactions among the agents often cause a global pattern to emerge. Most of the basic ideas are derived from the real swarms in the nature, which includes ant colonies, bird flocking, honeybees, bacteria and microorganisms, etc. Ant Colony Optimization (ACO), have already been applied successfully to solve several engineering optimization problems. Swarm models are population-based and the population is initialised with a population of potential solutions. These individuals are then manipulated (optimised) over many several iterations using several heuristics inspired from the social behavior of insects in an effort to find the optimal solution. Ant colony algorithms are inspired by the behavior of natural ant colonies, in the sense that they solve their problems by multi agent cooperation using indirect communication through modifications in the environment. Ants release a certain amount of pheromone (hormone) while walking, and each ant prefers (probabilistically) to follow a direction, which is rich of pheromone. This simple behavior explains why ants are able to adjust to changes in the environment, such as optimizing shortest path to a food source or a nest. In ACO, ants use information collected during past simulations to direct their search and this information is available and modified through the environment. Recently ACO algorithms have also been used for clustering data sets.

3 Computational Intelligence on Speech, Audio and Image Processing

Computational intelligence techniques are being used for processing speech, audio and image for several years [59, 64, 98]. Some of the applications in speech processing where computational intelligences are extensively used include speech recognition, speaker recognition, speech enhancement, speech coding and speech synthesis; in audio processing, computational intelligence are used for speech/music classification, audio classification and audio indexing and retrieval; while in image processing include image enhancement, segmentation, classification, registration, motion detection, etc. For example, Vladimir et al. [17] proposed a fuzzy logic recursive scheme for motion detection and spatiotemporal filtering that can deal with the Gaussian noise and unsteady illumination conditions in both the temporal and spatial directions. Our focus is on applications concerning tracking and de-noising of image sequences. An input noisy sequence is processed with fuzzy logic motion detection to determine the degree of motion confidence. The proposed motion detector combines the membership of the temporal intensity changes, appropriately using fuzzy rules, where the membership degree of motion for each pixel in a 2D sliding window is determined by a proposed membership function. Both the fuzzy membership function and the fuzzy rules are defined in such a way that the performance of the motion detector is optimized in terms of its robustness to noise and unsteady lighting conditions. Tracking and recursive adaptive temporal filtering are simultaneously performed, where the amount of filtering is inversely proportional to the confidence in the existence of motion. Finally, temporally filtered frames are further processed by a proposed spatial filter to obtain a de-noised image sequence. The proposed motion detection algorithm have been evaluated using two criteria: (1) robustness to noise and to changing illumination conditions and (2) motion blur in temporal recursive de-noising.

Speech and Audio Processing

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the intersection of digital signal processing and natural language processing.It can be divided in the following categories: (1) Speech recognition, which deals with analysis of the linguistic content of a speech signal; (2) Speaker recognition, where the aim is to recognize the identity of the speaker; (3) Enhancement of speech signals, e.g., audio noise reduction, Speech coding, a specialized form of data compression, is important in the telecommunication area; (4) Voice analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords; (5) Speech synthesis (i.e., the artificial synthesis of speech), which usually means computer generated speech; and (6) Speech enhancement, which deals

with enhancing the perceptual quality of speech signal by removing the destructive effects of noise, limited capacity recording equipment, impairments, etc. Reader may refer to [64] for an extensive overview of the advances on pattern recognition for speech and audio processing.

The feasibility of converting text into speech using an inexpensive computer with minimal memory is of great interest. Speech synthesizers have been developed for many popular languages (e.g., English, Chinese, Spanish, French, etc.), but designing a speech synthesizer for a language is largely dependant on the language structure. Text-to-speech conversion has traditionally been performed either by concatenating short samples of speech or by using rule-based systems to convert a phonetic representation of speech into an acoustic representation, which is then converted into speech. Karaali et al. [56] described a system that uses a Time-Delay Neural Network (TDNN) to perform this phonetic-to-acoustic mapping, with another neural network to control the timing of the generated speech. The neural network system requires less memory than a concatenation system, and performed well in tests comparing it to commercial systems using other technologies. It is reported that the neural network approach to speech synthesis offers the benefits of language portability, natural sounding speech, and low storage requirements as well as provide better voice quality than traditional approaches.

Hendessi et al. [55] developed a Persian synthesizer that includes an innovative text analyzer module. In the synthesizer, the text is segmented into words and after preprocessing, a neural network is passed over each word. In addition to preprocessing, a new model (SEHMM) is used as a post-processor to compensate for errors generated by the neural network. The performance of the proposed model is verified and the intelligibility of the synthetic speech is assessed via listening tests.

The use of neural networks to synthesize speech from a phonetic representation and to generate a frame of input to a vocoder. This requires the neural network to compute one output for each frame of speech from the vocoder, this can be computationally expensive. Corrigan et al. [57] introduced an alternative implementation to model the speech as a series of gestures, and let the neural network generate parameters describing the transitions of the vocoder parameters during these gestures. Their experiments have shown that acceptable speech quality is produced when each gesture is half of a phonetic segment and the transition model is a set of cubic polynomials describing the variation of each vocoder parameter during the gesture. Empirical results reveal a significant reduction in the computational cost.

Frankel et al. [60] described a speech recognition system which uses articulatory parameters as basic features and phone-dependent linear dynamic models. The system first estimates articulatory trajectories from the speech signal. Estimations of x and y coordinates of seven actual articulator positions in the midsagittal plane are produced every 2 ms by a recurrent neural network, trained on real articulatory data. The output of this network is then passed to a set of linear dynamic models, which perform phone recognition.

In recent years, the features derived from posteriors of a Multilayer Perceptron (MLP), known as tandem features, have proven to be very effective for automatic speech recognition. Most tandem features to date have relied on MLPs trained for phone classification. Cetin et al. [105] illustrated on a relatively small data set that MLPs trained for articulatory feature classification can be equally effective. They provided a similar comparison using MLPs trained on a much larger data set – 2,000 h of English conversational telephone speech. Also, authors explored how portable phone- and articulatory feature- based tandem features are in an entirely different language – Mandarin – without any retraining. It is reported that while phone-based features perform slightly better in the matched-language condition, they perform significantly better in the cross-language condition. Yet, in the cross-language condition, neither approach is as effective as the tandem features extracted from an MLP trained on a relatively small amount of in-domain data. Beyond feature concatenation, Cerin et al. explored novel observation modeling schemes that allow for greater flexibility in combining the tandem and standard features at hidden Markov model (HMM) outputs.

Halavati et al. [42] presents a novel approach to speech recognition using fuzzy modeling. The task begins with conversion of speech spectrogram into a linguistic description based on arbitrary colors and lengths. While phonemes are also described using these fuzzy measures, and recognition is done by normal fuzzy reasoning, a genetic algorithm optimizes phoneme definitions so that to classify samples into correct phonemes. The method is tested over a standard speech data base and the results are presented.

One of the factors complicating activity with speech signals is its large degree of acoustic variability. To decrease influence of acoustic variability of speech signals, it is offered to use genetic algorithms in speech processing systems. Bovbel and Tsishkoual [43] constructed a model which implements the technology of speech recognition using genetic algorithms. They made experiments on their model with a database of separated Belarussian words and achieved optimal results.

Ding [49] presented a fuzzy control mechanism for conventional Maximum Likelihood Linear Regression (MLLR) speaker adaptation, called FLC-MLLR, by which the effect of MLLR adaptation is regulated according to the availability of adaptation data in such a way that the advantage of MLLR adaptation could be fully exploited when the training data are sufficient, or the consequence of poor MLLR adaptation would be restrained otherwise. The robustness of MLLR adaptation against data scarcity is thus ensured. It is reported that the proposed mechanism is conceptually simple and computationally inexpensive and effective; the experiments in recognition rate show that FLC-MLLR outperforms standard MLLR especially when encountering data insufficiency and performs better than MAPLR at much less computing cost.

Kostek and Andrzej [47] discussed some limitations of the hearing-aid fitting process. In the fitting process, an audiologist performs tests on the wearer

of the hearing aid, which is then adjusted based on the results of the test, with the goal of making the device work as best as it can for that individual. Traditional fitting procedures employ specialized testing devices which use artificial test signals. Ideally, however, the fitting of hearing aids should also simulate real-world conditions, such as listening to speech in the presence of background noise. Therefore, more satisfying and reliable fitting tests may be achieved through the use of multimedia computers equipped with a properly calibrated sound system. Kostek and Andrzej developed a new automatic system for fitting hearing aids. It employed fuzzy logic and a computer makes choices for adjusting the hearing aid's settings by analyzing the patient's responses and answering questions with replies that can lie somewhere between a simple *yes* or *no*.

With the increase in access to multimedia computers, speech training can be made available to patients with no continuous assistance required from speech therapists. Another function the system can easily perform is screening testing of speech fluency providing directed information to patients who have various speech disorders and problems with understanding speech. Andrzej et al. [51] programmed speech therapy training algorithm consisting of diagnostic tools and rehabilitation devices connected with it. The first function the system has to perform is data acquisition where information about the patient's medical history is collected. This is done through electronic questionnaires. The next function is analysis of the speech signal articulated by the patient when prompted by the computer followed by some multimedia tests carried out in order to assess the subject's ability to understand speech. Next, the results of the electronic questionnaire, the patient's voice and patient's reactions are automatically analyzed. Based on that the system automatically diagnoses possible speech disorders and how strong they are. A large number of school children were tested and reported.

The process of counting stuttering events could be carried out more objectively through the automatic detection of stop-gaps, syllable repetitions and vowel prolongations. The alternative would be based on the subjective evaluations of speech fluency and may be dependent on a subjective evaluation method. Meanwhile, the automatic detection of intervocalic intervals, stop-gaps, voice onset time and vowel durations may depend on the speaker and the rules derived for a single speaker might be unreliable when trying to consider them as universal ones. This implies that learning algorithms having strong generalization capabilities could be applied to solve the problem. Nevertheless, such a system requires vectors of parameters, which characterize the distinctive features in a subject's speech patterns. In addition, an appropriate selection of the parameters and feature vectors while learning may augment the performance of an automatic detection system. Andrzej et al. [52] reported an automatic recognition of stuttered speech in normal and frequency altered feedback speech. It presents several methods of analyzing stuttered speech and describes attempts to establish those parameters that represent

stuttering event. It also reports results of some experiments on automatic detection of speech disorder events that were based on both rough sets and artificial neural networks.

Andrzej and Marek [54] presented a method for pitch estimation enhancement. Pitch estimation methods are widely used for extracting musical data from digital signal. A brief review of these methods is included in the paper. However, since processed signal may contain noise and distortions, the estimation results can be erroneous. The proposed method was developed in order to override disadvantages of standard pitch estimation algorithms. The introduced approach is based on both pitch estimation in terms of signal processing and pitch prediction based on musical knowledge modeling. First, signal is partitioned into segments roughly analogous to consecutive notes. Thereafter, for each segment an autocorrelation function is calculated. Autocorrelation function values are then altered using pitch predictor output. A music predictor based on artificial neural networks was introduced for this task. The description of the proposed pitch estimation enhancement method is included and some details concerning music prediction are discussed.

Liu et al. [48] proposed an improved hybrid support vector machine and duration distribution based hidden Markov (SVM/DDBHMM) decision fusion model for robust continuous digital speech recognition. The probability outputs combination of Support Vector Machine and Gaussian mixture model in pattern recognition (called FSVM), and embedding the fusion probability as similarity into the phone state level decision space of the Duration Distribution Based Hidden Markov Model (DDBHMM) speech recognition system (named FSVM/DDBHMM) were investigated. The performances of FSVM and FSVM/DDBHMM are demonstrated in Iris database and continuous mandarin digital speech corpus in four noise environments (white, volvo, babble and destroyer-engine) from NOISEX-92. The experimental results show the effectiveness of FSVM in Iris data, and the improvement of average word error rate reduction of FSVM/DDBHMM from 6% to 20% compared with the DDBHMM baseline at various signal noise ratios (SNRs) from $-5\,\mathrm{dB}$ to $30\,\mathrm{dB}$ by step of $5\,\mathrm{dB}$.

Andrzej [50] investigated methods for the identification of direction of the incoming acoustical signal in the presence of noise and reverberation. Since the problem is a non-deterministic one, thus applications of two learning algorithms, namely neural networks and rough sets were developed to solve it. Consequently, two sets of parameters were formulated in order to discern target source from unwanted sound source position and then processed by learning algorithms. The applied feature extraction methods are discussed, training processes are described and obtained sound source localizing results are demonstrated and compared.

Kostek et al. [53] presented an automatic singing voice recognition using neural network and rough sets. For this purpose a database containing singers' sample recordings has been constructed and parameters are extracted from recorded voices of trained and untrained singers of various voice types.

Parameters, which are especially designed for the analysis of the singing voice are described and their physical interpretation is given. Decision systems based on artificial neutral networks and rough sets are used for automatic voice type/voice quality classification.

Limiting the decrease in performance due to acoustic environment changes remains a major challenge for continuous speech recognition (CSR) systems. Selouani and Shaughnessy [25] proposed a hybrid enhancement noise reduction approach in the cepstral domain in order to get less-variant parameters. It is based on the Karhunen–Loeve Transform (KLT) in the mel-frequency domain with a Genetic Algorithm (GA). The enhanced parameters increased the recognition rate for highly interfering noise environments. The proposed hybrid technique, when included in the front-end of an HTK-based CSR system, outperformed the conventional recognition process in severe interfering car noise environments for a wide range of signal-to-noise ratios (SNRs) varying from $16\,$dB to $-4\,$dB. They also showed the effectiveness of the KLT-GA method in recognizing speech subject to telephone channel degradations.

CI in Speech Emotion Recognition

Speech emotion recognition is becoming more and more important in such computer application fields as health care, children education, etc. Only few works have been done on speech emotion recognition using such methods as ANN, SVM, etc., in the last years. Feature sets are broadly discussed within speech emotion recognition by acoustic analysis. While popular filter and wrapper based search help to retrieve relevant ones, we feel that automatic generation of such allows for more flexibility throughout search. The basis is formed by dynamic Low-Level Descriptors considering intonation, intensity, formants, spectral information and others. Next, systematic derivation of prosodic, articulatory, and voice quality high level functionals is performed by descriptive statistical analysis. From here on feature alterations are automatically fulfilled, to find an optimal representation within feature space in view of a target classifier. In addition, traditional feature selection method used in speech emotion recognition is computationally too expensive to determine an optimum or suboptimum feature subset. Focusing on these problems, many successful works have been addressed and discussed. For example, Zhou et al. [40] presented a novel approach based on rough set theory and SVM for speech emotion recognition. The experiment results illustrated that the introduced approach can reduce the calculation cost while keeping high recognition rate. Also, Schuller et al. [61] suggested the use of evolutionary programming to avoid NP-hard exhaustive search.

Fellenz et al. [44] proposed a framework for the processing of face image sequences and speech, using different dynamic techniques to extract appropriate features for emotion recognition. The features were used by a hybrid

classification procedure, employing neural network techniques and fuzzy logic, to accumulate the evidence for the presence of an emotional expression of the face and the speaker's voice.

Buscicchio et al. [19] proposed a biologically plausible methodology for the problem of emotion recognition, based on the extraction of vowel information from an input speech signal and on the classification of extracted information by a spiking neural network. Initially, a speech signal is segmented into vowel parts which are represented with a set of salient features, related to the Mel-frequency cesptrum. Different emotion classes are then recognized by a spiking neural network and classified into five different emotion classes.

Audio–Visual Speech Recognition

Audio–Visual Speech Recognition (AVSR) [63] is a technique that uses image processing capabilities in lip reading to aid speech recognition systems in recognizing undeterministic phones or giving preponderance among near probability decisions. A great interest in the research of AVSR systems is driven by the increase in the number of multimedia applications that require robust speech recognition systems. The use of visual features in AVSR is justified by both the audio and visual modality of the speech generation and the need for features that are invariant to acoustic noise perturbation. The performance of the AVSR system relies on a robust set of visual features obtained from the accurate detection and tracking of the mouth region. Therefore the mouth tracking plays a major role in AVSR systems. Moreover, A human listener can use visual cues, such as lip and tongue movements, to enhance the level of speech understanding, especially in a noisy environment. The process of combining the audio modality and the visual modality is referred to as speech reading, or lip reading. There are many applications in which it is desired to recognize speech under extremely adverse acoustic environments. Detecting a person's speech from a distance or through a glass window, understanding a person speaking among a very noisy crowd of people, and monitoring a speech over TV broadcast when the audio link is weak or corrupted, are some examples. Computational intelligence techniques plays an important role in this research direction. A number of CI-based AVSR methods have been proposed in the literature. For example, Lim et al. [39] presented an improvement version of mouth tracking technique using radial basis function neural network (RBF NN) with its applications to AVSR systems. A modified extended Kalman filter (EKF) was used to adjust the parameters of the RBF NN. Simulation results have revealed good performance of the proposed method.

Automatic Speech Recognition (ASR) performs well under restricted conditions, but performance degrades in noisy environments. AVSR combats this by incorporating a visual signal into the recognition. Lewis and Powers [62] discussed how to improve the performance of a standard speech recognition systems by using information from the traditional, auditory signals as well

as a visual signals. Using a knowledge from psycholinguistics, a late integration network was developed that fused the automatic and visual sources. An important first step in AVSR is that of feature extraction from the mouth region and a technique developed by the authors is briefly presented. Authors examined how useful this extraction technique in combination with several integration architectures is at the given task, demonstrates that vision does in fact assist speech recognition when used in a linguistically guided fashion, and gives insight remaining issues.

Alessandro et al. [38] focused the attention on the problem of audio classification in speech and music for multimedia applications. In particular, they presented a comparison between two different techniques for speech/music discrimination. The first method is based on zero crossing rate and Bayesian classification. It is very simple from a computational point of view, and gives good results in case of pure music or speech. The simulation results show that some performance degradation arises when the music segment contains also some speech superimposed on music, or strong rhythmic components. To overcome these problems, they proposed a second method, that uses more features, and is based on neural networks (specifically a multi-layer Perceptron). It is reported that the introduced algorithm is obtain better performance, at the expense of a limited growth in the computational complexity. In practice, the proposed neural network is simple to be implemented if a suitable polynomial is used as the activation function, and a real-time implementation is possible even if low-cost embedded systems are used.

Speech recognition techniques have been developed dramatically in recent years. Nevertheless, errors caused by environmental noise are still a serious problem in recognition. Employing algorithms to detect and follow the motion of lips have been widely used to improve the performance of speech recognition algorithms. Vahideh and Yaghmaie [65] presented a simple and efficient method for extraction of visual features of lip to recognize vowels based on the neural networks. The accuracy is verified by using it to recognize six main Farsi vowels.

Faraj and Bigun [41] described a new identity authentication technique by a synergetic use of lip-motion and speech. The lip-motion is defined as the distribution of apparent velocities in the movement of brightness patterns in an image and is estimated by computing the velocity components of the structure tensor by 1D processing, in 2D manifolds. Since the velocities are computed without extracting the speaker's lip-contours, more robust visual features can be obtained in comparison to motion features extracted from lip-contours. The motion estimations are performed in a rectangular lip-region, which affords increased computational efficiency. A person authentication implementation based on lip-movements and speech is presented along with experiments exhibiting a recognition rate of 98%. Besides its value in authentication, the technique can be used naturally to evaluate the liveness of someone speaking as it can be used in text-prompted dialogue. The XM2VTS database was used

for performance quantification as it is currently the largest publicly available database (300 persons) containing both lip-motion and speech. Comparisons with other techniques are presented.

Shan Meng and Youwei Zhang [58] described a method of visual speech feature area localization First, they propose a simplified human skin color model to segment input images and estimate the location of human face. Authors proposed a new localization method that is a combination of SVM and Distance of Likelihood in Feature Space (DLFS) derived from Kernel Principal Component Analysis (KPCA). Results show that the introduced method outperformed traditional linear ones. All experiments were based on Chinese Audio–Visual Speech Database(CAVSD).

4 Computational Intelligence in Video Processing

Edge extraction, texture classification, face recognition, character recognition, finger print identification, image/video enhancement, image/video segmentation and clustering, and image/video coding are some of the applications of computational intelligence in image processing. Here we demonstrated some reported examples of using the CI techniques in multimedia processing and, in particulars in image/video processing. As a result, there has been much recent research interest in this area. Many successful work towered this issue has been addressed and discussed. Here, we review some successful work to illustrate how CI could be applied to resolve video segmentation problem.

Computational Intelligence in Video Segmentation

Successful video segmentation is necessary for most multimedia applications. In order to analyze a video sequence, it is necessary to break it down into meaningful units that are of smaller length and have some semantic coherence. Video segmentation is the process of dividing a sequence of frames into smaller meaningful units that represent information at the scene level. This process serves as a fundamental step towards any further analysis on video frames for content analysis. In the past, several statistical methods that compare frame differences have been published in literature and a range of similarity measures between frames based on gray-scale intensity, color and texture have been proposed. Here we demonstrate a succuss works in using the CI techniques in video segmentation.

The organization of video information in video databases requires automatic temporal segmentation with minimal user interaction. As neural networks are capable of learning the characteristics of various video segments and clustering them accordingly. Cao and Suganthan [27] developed a neural network based technique to segment the video sequence into shots automatically and with a minimum number of user-defined parameters. They propose to employ Growing Neural Gas (GNG) networks and integrate multiple

frame difference features to efficiently detect shot boundaries in the video. Experimental results were presented to illustrate the good performance of the proposed scheme on real video sequences.

Lo and Wang [26] proposed a video segmentation method using a Histogram-Based Fuzzy C-Means (HBFCM) clustering algorithm. This algorithm is a hybrid of two approaches and is composed of three phases: the feature extraction phase, the clustering phase, and the key-frame selection phase. In the first phase, differences between color histogram are extracted as features. In the second phase, the Fuzzy C-Means (FCM) is used to group features into three clusters: the shot change (SC) cluster, the Suspected Shot Change (SSC) cluster, and the No Shot Change (NSC) cluster. In the last phase, shot change frames are identified from the SC and the SSC, and then used to segment video sequences into shots. Finally, key frames are selected from each shot. Authors simulation results indicate that the HBFCM clustering algorithm is robust and applicable to various types of video sequences.

Ford [20] presented a fuzzy logic system for the detection and classification of shot boundaries in uncompressed video sequences. It integrates multiple sources of information and knowledge of editing procedures to detect shot boundaries. Furthermore, the system classifies the editing process employed to create the shot boundary into one of the following categories: abrupt cut, fade-in, fade-out, or dissolve. This system was tested on a database containing a wide variety of video classes. It achieved combined recall and precision rates that significantly exceed those of existing threshold-based techniques, and it correctly classified a high percentage of the detected boundaries.

Video temporal segmentation is normally the first and important step for content-based video applications. Many features including the pixel difference, color histogram, motion, and edge information, etc., have been widely used and reported in the literature to detect shot cuts inside videos. Although existing research on shot cut detection is active and extensive, it still remains a challenge to achieve accurate detection of all types of shot boundaries with one single algorithm. Hui Fang et al. [24] proposed a fuzzy logic approach to integrate hybrid features for detecting shot boundaries inside general videos. The fuzzy logic approach contains two processing modes, where one is dedicated to detection of abrupt shot cuts including those short dissolved shots, and the other for detection of gradual shot cuts. These two modes are unified by a mode-selector to decide which mode the scheme should work on in order to achieve the best possible detection performances. By using the publicly available test data set from Carleton University, extensive experiments were carried out and the test results illustrate that the proposed algorithm outperforms the representative existing algorithms in terms of the precision and recall rates.

Mitra [71] proposed an evolutionary rough c-means clustering algorithm. Genetic algorithms are employed to tune the threshold, and relative importance of upper and lower approximations of the rough sets modeling the clusters. The Davies–Bouldin clustering validity index is used as the fitness

function, that is minimized while arriving at an optimal partitioning. A comparative study of its performance is made with related partitive algorithms. The effectiveness of the algorithm is demonstrated on real and synthetic data sets, including microarray gene expression data from Bioinformatics. In the same study, the author noted that the parameter threshold measures the relative distance of an object X_k from a pair of clusters having centroids cen_i and cen_j. The smaller the value of threshold, the more likely is X_k to lie within the rough boundary (between upper and lower approximations) of a cluster. This implies that only those points which definitely belong to a cluster (lie close to the centroid) occur within the lower approximation. A large value of threshold implies a relaxation of this criterion, such that more patterns are allowed to belong to any of the lower approximations. The parameter w_{low} controls the importance of the objects lying within the lower approximation of a cluster in determining its centroid. A lower w_{low} implies a higher w_{up}, and hence an increased importance of patterns located in the rough boundary of a cluster towards the positioning of its centroid.

Das et al. [103] presented a framework to hybridize the rough set theory with particle swarm optimization algorithm. The hybrid rough-PSO technique has been used for grouping the pixels of an image in its intensity space. Medical images become corrupted with noise very often. Fast and efficient segmentation of such noisy images (which is essential for their further interpretation in many cases) has remained a challenging problem for years. In there work, they treat image segmentation as a clustering problem. Each cluster is modeled with a rough set. PSO is employed to tune the threshold and relative importance of upper and lower approximations of the rough sets. Davies-Bouldin clustering validity index is used as the fitness function, which is minimized while arriving at an optimal partitioning.

Raducanu et al. [106] proposed a Morphological Neural Networks (MNN) algorithm as associative (with its two cases: autoassociative and heteroassociative) memories. It propose their utilization as a preprocessing step for human shape detection, in a vision-based navigation problem for mobile robots. It is reported that the MNN can be trained in a single computing step, they possess unlimited storing capacity, and they have perfect recall of the patterns. Recall is also very fast, because the MNN recall does not involve the search for an energy minimum.

Adaptation of C-Means to Rough Set Theory

C-means clustering is an iterative technique that is used to partition an image into C-clusters. Fuzzy C-Means (FCM) is one of the most commonly used fuzzy clustering techniques for different degree estimation problems, especially in medical image processing [104, 107, 116]. Lingras [70] described modifications of clustering based on Genetic Algorithms, K-means algorithm, and Kohonen Self-Organizing Maps (SOM). These modifications make it possible to represent clusters as rough sets [97]. In their work, Lingras established

a rough k-means framework and extended the concept of c-means by viewing each cluster as an interval or rough set [69]. Here is a brief summary of his pioneer clustering work.

K-means clustering is one of the most popular statistical clustering techniques used in segmentation of medical images [66, 72, 94, 108–110]. The name K-means originates from the means of the k clusters that are created from n objects. Let us assume that the objects are represented by m-dimensional vectors. The objective is to assign these n objects to k clusters. Each of the clusters is also represented by an m-dimensional vector, which is the centroid or mean vector for that cluster. The process begins by randomly choosing k objects as the centroids of the k clusters. The objects are assigned to one of the k clusters based on the minimum value of the distance $d(v, x)$ between the object vector $v = (v_1, \ldots, v_j, \ldots, v_m)$ and the cluster vector $x = (x_1, \ldots, x_j, \ldots, x_m)$. After the assignment of all the objects to various clusters, the new centroid vectors of the clusters are calculated as

$$x_j = \frac{\sum_{v \in x} v_j}{SOC}, where\ 1 \leq j \leq m \ , \tag{8}$$

where SOC is the size of cluster x.

Lingras [70] mentioned that incorporation of rough sets into K-means clustering requires the addition of the concept of lower and upper bounds. Calculation of the centroids of clusters from conventional K-Means needs to be modified to include the effects of lower as well as upper bounds. The modified centroid calculations for rough sets are then given by:

$$cen_j = W_{low} \times \frac{\sum_{v \in \underline{R}(x)}}{|\underline{R}(x)|} + w_{up} \times \frac{\sum_{v \in (\underline{BN_R(x)})}}{|\underline{BN_R(x)}|} \ , \tag{9}$$

where $1 \leq j \leq m$. The parameters w_{lower} and $w_{(upper)}$ correspond to the relative importance of lower and upper bounds, and $w_{low} + w_{up} = 1$. If the upper bound of each cluster were equal to its lower bound, the clusters would be conventional clusters. Therefore, the boundary region $BN_R(x)$ will be empty, and the second term in the equation will be ignored. Thus, the above equation will reduce to conventional centroid calculations. The next step in the modification of the K-means algorithms for rough sets is to design criteria to determine whether an object belongs to the upper or lower bound of a cluster, for more details refer to. The main steps of the algorithm are provided in Algorithm 1.

5 Computational Intelligence in Multimedia Watermarking

Multimedia watermarking technology has evolved very quickly during the last few years. A digital watermark is information that is imperceptibly and robustly embedded in the host data such that it cannot be removed.

Algorithm 1 Rough C-mean Algorithm

1: Set x_i as an initial means for the c clusters.
2: Initialize the population of particles encoding parameters threshold and w_{low}
3: Initialize each data object x_k to the lower approximation or upper approximation of clusters c_i by computing the difference in its distance by:

$$diff = d(x_k, cen_i) - d(x_k, cen_j), \tag{10}$$

where cen_i and cen_j are the cluster centroid pairs.
4: **if** diff $< \delta$ **then**
5:    $x_k \in$ the upper approximation of the cen_i and cen_j clusters and can not be in any lower approximation.
6:    Else
7:    $x_k \in$ lower approximation of the cluster c_i such that distance $d(x_k, cen_i)$ is minimum over the c clusters.
8: **end if**
9: Compute a new mean using equation 15
10: **repeat**
    statement 3-9
11: **until** convergence, i.e., there is no more new assignments

A watermark typically contains information about the origin, status, or recipient of the host data. The digital watermarking system essentially consists of a watermark encoder and a watermark decoder. The watermark encoder inserts a watermark onto the host signal and the watermark decoder detects the presence of watermark signal. Note that an entity called watermark key is used during the process of embedding and detecting watermarks. The watermark key has a one-to-one correspondence with watermark signal (i.e., a unique watermark key exists for every watermark signal). The watermark key is private and known to only authorized parties and it ensures that only authorized parties can detect the watermark. Further, note that the communication channel can be noisy and hostile (i.e., prone to security attacks) and hence the digital watermarking techniques should be resilient to both noise and security attacks. Figure 7 illustrates the digital watermark methodology in general.

The development of watermarking methods involves several design trade-offs: (1) *Robustness* which deals with the ability of the watermark to resist attempts by an attacker to destroy it by modifying the size, rotation, quality, or other visual aspects of the video; (2) *Security* which deals with the ability of the watermark to resist attempts by a sophisticated attacker to remove it or destroy it via cryptanalysis, without modifying the media itself; and (3) *Perceptual fidelity* the perceived visual quality of the marked media compared to the original, unmarked video. A copyright protection is the most prominent application of watermarking techniques, others exist, including data authentication by means of fragile watermarks which are impaired or destroyed by manipulations, embedded transmission of value added services within multimedia data, and embedded data labeling for other purposes
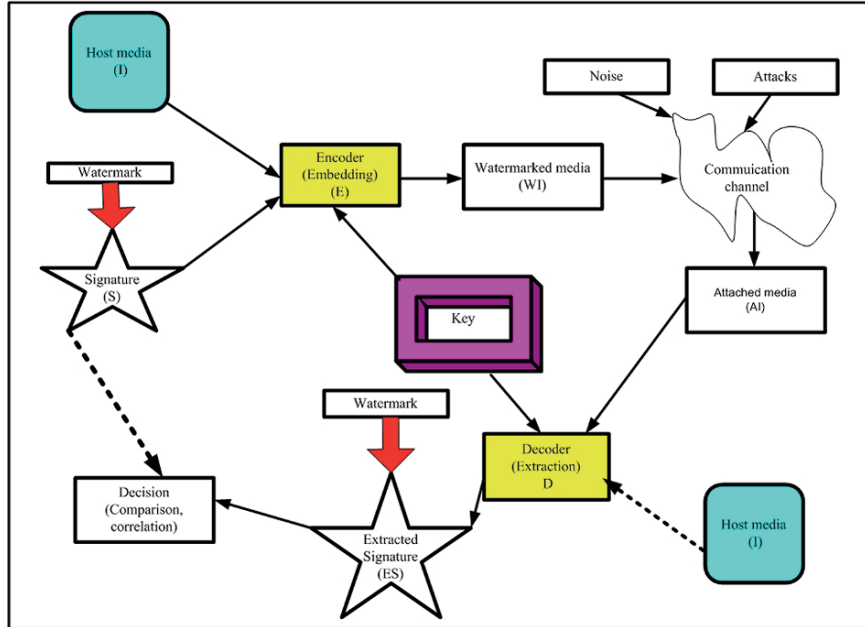
Fig. 7. General digital watermarking architecture [9]

than copyright protection, such as data monitoring and tracking. An example for a data-monitoring system is the automatic registration and monitoring of broadcasted radio programs such that royalties are automatically paid to the IPR owners of the broadcast data. Focusing on these problems, many successful works have been addressed and discussed. For example, Lou et al. [32] proposed a copyright protection scheme based on chaos and secret sharing techniques. Instead of modifying the original image to embed a watermark in it, the proposed scheme extracts a feature from the image first. Then, the extracted feature and the watermark are scrambled by a chaos technique. Finally, the secret sharing technique is used to construct a shadow image. The watermark can be retrieved by performing an XOR operation between the shadow images. It is reported that the introduced scheme compared with other works is secure and robust in resisting various attacks.

Cao et al. [37] proposed a novel audio watermarking algorithm based on neural networks. By transforming original audio sequence into 1D wavelet domain and selecting proper positions, several watermark bits were embed. Before transmitting, it effectively utilizes neural networks to learn the relation characteristics between original audio and watermarked audio. Due to the learning and adaptive capabilities of neural networks possessing, the trained neural networks almost exactly extract the watermark from the watermarked audio against audio processing attacks. Extensive experimental results showed that the proposed method significantly possesses robustness. It is immune against such attacks as low pass filtering, addition of noise, resampling and medium filtering.

Wei Lu et al. [33] presented a robust digital image watermarking scheme by using neural network detector. First, the original image is divided into four subimages by using subsampling. Then, a random binary watermark sequence is embedded into DCT domain of these subimages. A fixed binary sequence is added to the head of the payload watermark as the samples to train the neural network detector. Because of the good adaptive and learning abilities, the neural network detector can nearly exactly extract the payload watermark. Experimental results illustrated good performance of the proposed scheme on resisting common signal processing attacks.

Lou and Yin [30] proposed adaptive digital watermarking approach based upon human visual system model and fuzzy clustering technique. The human visual system model is utilized to guarantee that the watermarked image is imperceptible, while the fuzzy clustering approach has been employed to obtain the different strength of watermark by the local characters of image. In their experiments, the scheme provides a more robust and transparent watermark.

Cheng-Ri Piao et al. [34] proposed a new watermarking scheme in which a logo watermark is embedded into the Discrete Wavelet Transform (DWT) domain of the color image using Back-Propagation Neural networks (BPN). In order to strengthen the imperceptibility and robustness, the original image is transformed from RGB color space to brightness and chroma space (YCrCb). After transformation, the watermark is embedded into DWT coefficient of chroma component, CrCb. A secret key determines the locations in the image where the watermark is embedded. This process prevents possible pirates from removing the watermark easily. BPN learns the characteristics of the color image, and then watermark is embedded and extracted by using the trained neural network. Experimental results showed that the proposed method has good imperceptibility and high robustness to common image processing attacks.

Zheng Liu et al. [35] introduced a sensor-based authentication watermarking with the concept of authentication on demand, in which user requirements are adopted as parameters for authentication. In addition, fuzzy identification of multiple authentication sensors outputs has introduced the ability of finely tuning on authentication type and degree. With this approach, authentication sensitivity to malicious attacks is enhanced. It is reported that the introduced approach is more robust against allowed modifications. In addition, author's algorithm provide a new function, to detect the attack method.

Maher et al. [31] proposed a novel digital video watermarking scheme based on multi resolution motion estimation and artificial neural network. A multi resolution motion estimation algorithm is adopted to preferentially allocate the watermark to coefficients containing motion. In addition, embedding and extraction of the watermark are based on the relationship between a wavelet coefficient and its neighbor's. A neural network was given to memorize the relationships between coefficients in a 3×3 block of the image. Experimental results illustrated that embedding watermark where picture content is moving

is less perceptible. Further, empirical results demonstrated that the proposed scheme is robust against common video processing attacks.

Several discrete wavelet transform based techniques are used for watermarking digital images. Although these techniques are robust to some attacks, none of them is robust when a different set of parameters is used or some other attacks (such as low pass filtering) are applied. In order to make the watermark stronger and less susceptible to different types of attacks, it is essential to find the maximum amount of watermark before the watermark becomes visible. Davis and Najarian [111] used the neural networks to implement an automated system of creating maximum-strength watermarks.

Diego and Manuel [29] proposed an evolutionary algorithm for the enhancement of digital semi-fragile watermarking based on the manipulation of the image Discrete Cosine Transform (DCT). The algorithm searches for the optimal localization of the DCT of an image to place the mark image DCT coefficients. The problem is stated as a multi-objective optimization problem (MOP), that involves the simultaneous minimization of distortion and robustness criteria.

Chang et al. [28] developed a novel transform domain digital watermarking scheme that uses visually meaningful binary image as watermark. The scheme embeds the watermark information adaptively with localized embedding strength according to the noise sensitivity level of the host image. Fuzzy adaptive resonance theory (Fuzzy-ART) classification is used to identify appropriate locations for watermark insertion and its control parameters add agility to the clustering results to thwart counterfeiting attacks. The scalability of visually recognizable watermark is exploited to devise a robust weighted recovery method with composite watermark. The proposed watermarking schemes can also be employed for oblivious detection. Unlike most oblivious watermarking schemes, our methods allow the use of visually meaningful image as watermark. For automation friendly verification, a normalized correlation metric that suits well with the statistical property of their methods is used. The experimental results demonstrated that the proposed techniques can survive several kinds of image processing attacks and the JPEG lossy compression.

Tsai et al. [36] proposed a new intelligent audio watermarking method based on the characteristics of the HAS and the techniques of neural networks in the DCT domain. The method makes the watermark imperceptible by using the audio masking characteristics of the HAS. Moreover, the method exploits a neural network for memorizing the relationships between the original audio signals and the watermarked audio signals. Therefore, the method is capable of extracting watermarks without original audio signals. Finally, the experimental results are also included to illustrate that the method significantly possesses robustness to be immune against common attacks for the copyright protection of digital audio.

6 Computational Intelligence in Content-Based Multimedia Indexing and Retrieval

There are a growing number of applications, which extensively use the visual media. A key requirement in those applications is efficient access to the stored multimedia information for the purposes of indexing, fast retrieval, and scene analysis. The amounts of multimedia content available to the public and to researchers has been growing rapidly in the last decades and is expected to increase exponentially in the years to come. This development puts a great emphasis on automated content-based retrieval methods, which retrieve and index multimedia based on its content. Such methods, however, suffer from a serious problem: the semantic gap, i.e., the wide gulf between the low-level features used by computer systems and the high-level concepts understood by human beings.

Mats et al. [46] proposed a method of content-based multimedia retrieval of objects with visual, aural and textual properties. In their method, training examples of objects belonging to a specific semantic class are associated with their low-level visual descriptors (such as MPEG-7) and textual features such as frequencies of significant keywords. A fuzzy mapping of a semantic class in the training set to a class of similar objects in the test set is created by using Self-Organizing Maps (SOMs) trained from automatically extracted low-level descriptors. Authors performed several experiments with different textual features to evaluate the potential of their approach in bridging the gap from visual features to semantic concepts by the use textual presentations. Their initial results show a promising increase in retrieval performance. PicSOM [45] content-based information retrieval (CBIR) system was used with video data and semantic classes from the NIST TRECVID 20051 evaluation set. The TRECVID set contains TV broadcasts in different languages and textual data acquired by using automatic speech recognition software and machine translation where appropriate. Both the training and evaluation sets are were accompanied with verified semantic ground truth sets such as videos depicting explosions or fire.

In the PicSOM system the videos and the parts extracted from these were arranged as hierarchical trees as shown in Fig. 8, with the main video as the parent object and the different extracted media types as child objects. In this way, the relevance assessments can be transferred between related objects in the PicSOM algorithm. From each media type different features were extracted, and Self-Organizing Maps were trained from these as is shown with some examples in the Fig. 8.

Ming Li and Tong Wang [22] presented a new image retrieval technique based on concept lattices, named Concept Lattices- Based Image Retrieval, lattice browsing allows one to reach a group of images via one path. Because it can produce a lot of redundancies attributes when constructing the concept lattices by using a general method. In addition, authors proposed a method of attribute reduction of concept lattices based on discernibility matrix and
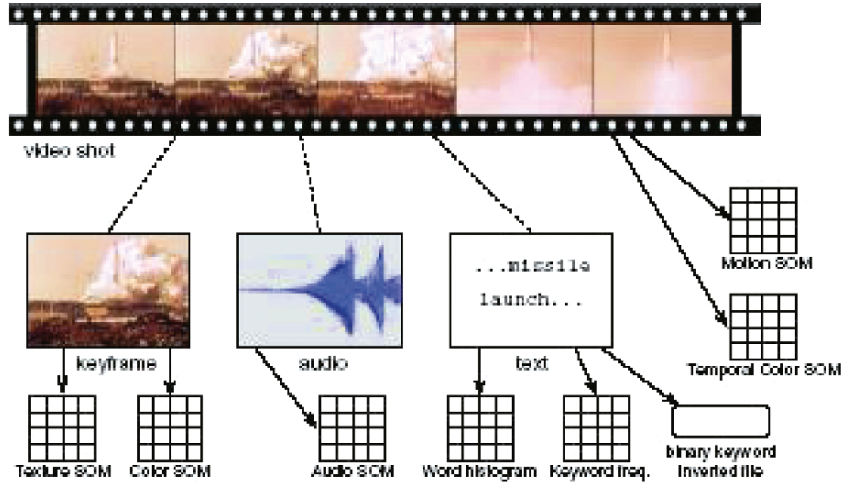
Fig. 8. The hierarchy of videos and examples of multi-modal SOMs [46]

boolean calculation to reduce the context of concept lattices. The scale of the problem is reduced by using this method. At the same time, the efficiency of image retrieval is improved, which is reflected in the experiments.

Fuzzy set methods have been already applied to the representation of flexible queries and to the modeling of uncertain pieces of information in databases systems, as well as in information retrieval. This methodology seems to be even more promising in multimedia databases which have a complex structure and from which documents have to be retrieved and selected not only from their contents, but also from the idea the user has of their appearance, through queries specified in terms of user's criteria. Dubois et al. [14] provided a preliminary investigation of the potential applications of fuzzy logic in multimedia databases. The problem of comparing semi-structured documents is first discussed. Querying issues are then more particularly emphasized. They distinguish two types of request, namely, those which can be handled within some extended version of an SQL-like language and those for which one has to elicit user's preference through examples.

Hassanien and Jafar [8] presented an application of rough sets to feature reduction, classification and retrieval for image databases in the framework of content-based image retrieval systems. The presented description of rough sets theory emphasizes the role of reducts in statistical feature selection, data reduction and rule generation in image databases. A key feature of the introduced approach is that segmentation and detailed object representation are not required. In order to obtain better retrieval results, the image texture features can be combined with the color features to form a powerful discriminating feature vector for each image. Texture features from the co-occurrence matrix are extracted, represented and, normalized in attribute vector then

the rough set dependency rules are generated directly from the real value attribute vector. Then the rough set reduction technique is applied to find all reducts of the data which contains the minimal subset of attributes that are associated with a class label for classification. A new similarity distance measure based on rough sets was presented. The classification and retrieval performance are measured using recall-precision measure, as is standard in all content based image retrieval systems. Figure 9 illustrates the image classification and retrieval scheme based on the rough set theory framework. (See also [114])

Chen and Wang [113] proposed a fuzzy logic approach UFM (Unified Feature Matching), for region-based image retrieval. In their retrieval system, an image is represented by a set of segmented regions, each of which is
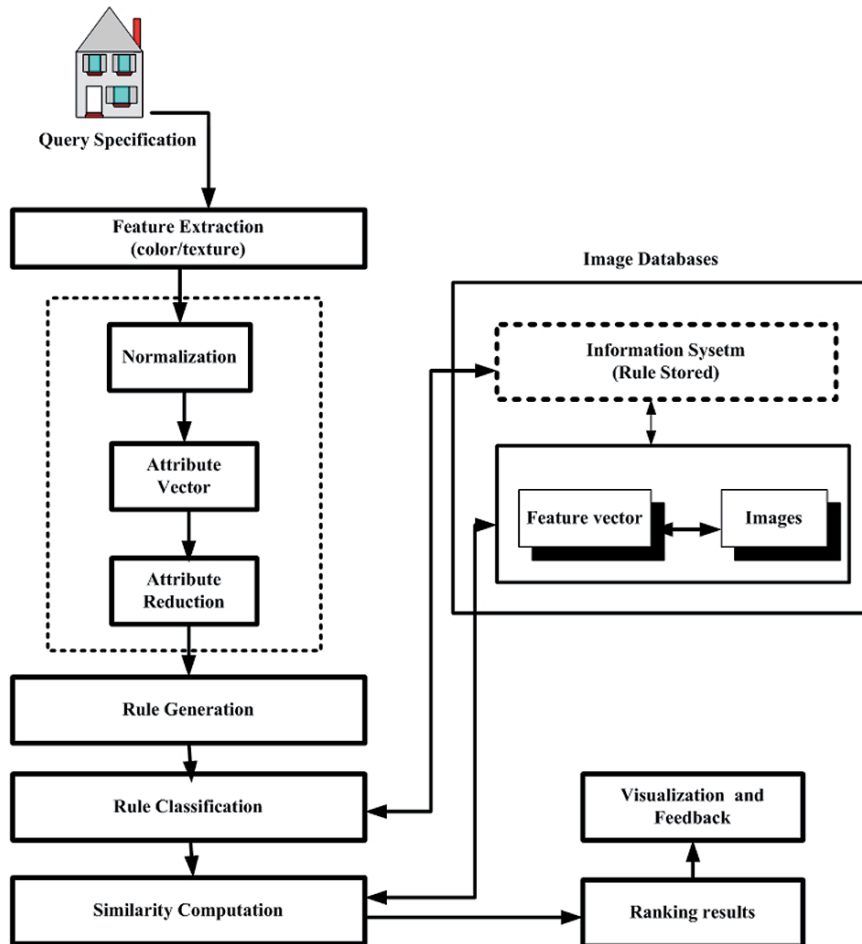

Fig. 9. CBIR in rough sets frameworks [8]

characterized by a fuzzy feature (fuzzy set) reflecting color, texture, and shape properties. As a result, an image is associated with a family of fuzzy features corresponding to regions. Fuzzy features naturally characterize the gradual transition between regions (blurry boundaries) within an image and incorporate the segmentation-related uncertainties into the retrieval algorithm. The resemblance of two images is then defined as the overall similarity between two families of fuzzy features and quantified by a similarity measure, UFM measure, which integrates properties of all the regions in the images. Compared with similarity measures based on individual regions and on all regions with crisp-valued feature representations, the UFM measure greatly reduces the influence of inaccurate segmentation and provides a very intuitive quantification. The UFM has been implemented as a part of author's experimental image retrieval system. The performance of the system was illustrated using examples from an image database of about 60,000 general-purpose images.

As digital video databases become more and more pervasive, finding video in large databases becomes a major problem. Because of the nature of video (streamed objects), accessing the content of such databases is inherently a time-consuming operation. Kulkarni [23] proposed a neural-fuzzy based approach for retrieving a specific video clip from a video database. Fuzzy logic was used for expressing queries in terms of natural language and a neural network is designed to learn the meaning of these queries. The queries were designed based on features such as color and texture of shots, scenes and objects in video clips. An error backpropagation algorithm was proposed to learn the meaning of queries in fuzzy terms such as very similar, similar and some-what similar. Preliminary experiments were conducted on a small video database and different combinations of queries using color and texture features along with a visual video clip; very promising results were achieved.

7 Conclusions, Challenges and Future Directions

During the last decades, multimedia processing has emerged as an important technology to generate content based on images, video, audio, graphics, and text. Furthermore, the recent new development represented by High Definition Multimedia content and Interactive television will generate a huge volume of data and important computing problems connected with the creation, processing and management of Multimedia content. Multimedia processing is a challenging domain for several reasons: it requires both high computation rates and memory bandwidth; it is a multirate computing problem; and requires low-cost implementations for high-volume markets. The past years have witnessed a large number of interesting applications of various computational intelligence techniques, such as neural networks; fuzzy logic; evolutionary computing; swarm intelligence; reinforcement Learning and evolutionary computation, rough sets, and a generalization of rough sets called near sets, to intelligent multimedia processing. Therefore, multimedia computing and

communication is another challenge and fruitful area for CI to play crucial roles in resolving problems and providing solutions to multimedia image/audio/video processing that understand, represent and process the media, their segments, indexing and retrieval.

Another challenge is to develop near sets-based methods, which offer a generalization of traditional rough set theory and a approach to classifying perceptual objects by means of features could be lead to new and will be useful in solving object recognition, particularly in solving multimedia problems such as classification and segmentation as well as to an application of the near set approach in 2D and 3D interactive gaming with a vision system that learns and serves as the backbone for an adaptive telerehabilitation system for patients with finger, hand, arm and balance disabilities. Each remote node in the telerehabilitation system includes a vision system that learns to track the behavior of a patient. Images deemed to be 'interesting' (e.g., images representing erratic behavior) are stored as well as forwarded to a rehabilitation center for follow up. In such a system, there is a need to identify images that are in some sense near images representing some standard or norm. This research has led to a study of methods of automating image segmentation as a first step in near set-based image processing.

In recent years, there has been a rapidly increasing demand for the development of advanced interactive multimedia applications, such as video telephony, video games and TV broadcasting have resulted in spectacular strides in the progress of wireless communication systems. However, these applications are always stringently constrained by current wireless system architectures because the request of high data rate for video transmission. To better serve this need, 4G broadband mobile systems are being developed and are expected to increase the mobile data transmission rates and bring higher spectral efficiency, lower cost per transmitted bit, and increased flexibility of mobile terminals and networks. The new technology strives to eliminate the distinction between video over wireless and video over wireline networks. In the meantime, great opportunities are provided for proposing novel wireless video protocols and applications, and developing advanced video coding and communications systems and algorithms for the next-generation video applications that can take maximum advantage of the 4G wireless systems. New video applications over 4G wireless systems is a challenge for the CI researchers.

The current third generation (3G) wireless systems and the next generation (4G) wireless systems in the development stages support higher bit rates. However, the high error rates and stringent delay constraints in wireless systems are still significant obstacles for these applications and services. On the other hand, the development of more advanced wireless systems provides opportunities for proposing novel wireless multimedia protocols and new applications and services that can take the maximum advantage of the systems.

In mobile ad hoc networks, specific intrusion detection systems are needed to safeguard them since traditional intrusion prevention techniques are not

sufficient in the protection of mobile ad hoc networks [1]. Therefore, intrusion detection system is another challenge and fruitful area for CI to play crucial roles in resolving problems and providing solutions to intrusion detection systems and authenticate the maps produced by the application of the intelligent techniques using watermarking, biometric and cryptology technologies.

A combination of kinds of computational intelligence techniques in application area of multimedia processing has become one of the most important ways of research of intelligent information processing. Neural network shows us its strong ability to solve complex problems for many multimedia processing. From the perspective of the specific rough sets approaches that need to be applied, explorations into possible applications of hybridize rough sets with other intelligent systems like neural networks [96], genetic algorithms, fuzzy approaches, etc., to multimedia processing and pattern recognition, in particulars in multimedia computing problems could lead to new and interesting avenues of research and is always a challenge for the CI researchers.

In conclusion, many successful algorithms applied in multimedia processing have been reported in the literature and the applications of rough sets in multimedia processing have to be analyzed individually. Rough set is a new challenge to deal with the issues that can not be addressed by traditional image processing algorithms or by other classification techniques. By introducing rough set, algorithms developed for multimedia processing and pattern recognition often become more intelligent and robust that provides a human-interpretable, low cost, exact enough solution, as compared to other intelligence techniques.

Finally, the main purpose of this article is to present to the CI and multimedia research communities the state of the art in CI applications to multimedia computing, and to inspire further research and development on new applications and new concepts in new trend-setting directions and in exploiting computational intelligence.

References

1. Abraham A., Jain R., Thomas J., and Han S.Y. (2007) D-SCIDS: Distributed soft computing intrusion detection systems. Journal of Network and Computer Applications, vol. 30, no. 1, pp. 81–98.
2. Bishop C.M. (1995) Neural Networks for Pattern Recognition. Oxford University Press, Oxford.
3. Kohonen T. (1988) Self-Organization and Associative Memory. Springer, Berlin Heidelberg New York.
4. Carpenter G. and Grossberg S. (1995) Adaptive Resonance Theory (ART). In: Arbib M.A. (ed.), The Handbook of Brain Theory and Neural Networks. MIT, Cambridge, pp. 79–82.
5. Grossberg S. (1976) Adaptive pattern classification and universal recoding: Parallel development and coding of neural feature detectors. Biological Cybernetics, vol. 23, pp. 121–134.

6. Abraham A. (2001) Neuro-fuzzy systems: State-of-the-art modeling techniques, connectionist models of neurons, learning processes, and artificial intelligence. In: Jose Mira and Alberto Prieto (eds.), Lecture Notes in Computer Science, vol. 2084, Springer, Berlin Heidelberg New York, pp. 269–276.

7. Nguyen H.T. and Walker E.A. (1999) A First Course in Fuzzy Logic. CRC, Boca Raton.

8. Hassanien A.E. and Jafar Ali (2003) Image classification and retrieval algorithm based on rough set theory. South African Computer Journal (SACJ), vol. 30, pp. 9–16.

9. Hassanien A.E. (2006) Hiding iris data for authentication of digital images using wavelet theory. International journal of Pattern Recognition and Image Analysis, vol. 16, no. 4, pp. 637–643.

10. Hassanien A.E., Ali J.M., and Hajime N. (2004) Detection of spiculated masses in Mammograms based on fuzzy image processing. In: 7th International Conference on Artificial Intelligence and Soft Computing, ICAISC2004, Zakopane, Poland, 7–11 June. Lecture Notes in Artificial Intelligence, vol. 3070. Springer, Berlin Heidelberg New York, pp. 1002–1007.

11. Fogel L.J., Owens A.J., and Walsh M.J. (1967) Artificial Intelligence Through Simulated Evolution. Wiley, New York.

12. Fogel D.B. (1999) Evolutionary Computation: Toward a New Philosophy of Machine Intelligence, 2nd edition. IEEE, Piscataway, NJ.

13. Pearl J. (1997) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco.

14. Dubois D., Prade H., and Sèdes F. (2001) Fuzzy logic techniques in multimedia database querying: A preliminary investigation of the potentials. IEEE Transactions on Knowledge and Data Engineering, vol. 13 , no. 3, pp. 383–392.

15. Holland J. (1975) Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Harbor.

16. Goldberg D.E. (1989) Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading.

17. Zlokolica V., Piurica A., Philips W., Schulte S., and Kerre E. (2006) Fuzzy logic recursive motion detection and denoising of video sequences. Journal of Electronic Imaging, vol. 15, no. 2.

18. Koza J.R. (1992) Genetic Programming. MIT, Cambridge, MA.

19. Buscicchio C.A., Grecki P., and Caponetti L. (2006) Speech emotion recognition using spiking neural networks. In: Esposito F., Ras Z.W., Malerba D., and Semeraro G. (eds.), Foundations of Intelligent Systems, Lecture Notes in Computer Science, vol. 4203, Springer, Berlin Heidelberg New York, pp. 38–46.

20. Ford R.M. (2005) Fuzzy logic methods for video shot boundary detection and classification. In: Tan Y.-P., Yap K.H., and Wang L. (eds.) Intelligent Multimedia Processing with Soft Computing, Studies in Fuzziness and Soft Computing, vol. 168, Springer, Berlin Heidelberg New York, pp. 151–169.

21. Back T. (1996) Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press, New York.

22. Ming Li and Tong Wang (2005) An approach to image retrieval based on concept lattices and rough set theory. Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies, 5–8 Dec., pp. 845–849.

23. Kulkarni S. (2004) Neural-fuzzy approach for content-based retrieval of digital video. Canadian Conference on Electrical and Computer Engineering, vol. 4, 2–5 May, pp. 2235–2238.

24. Hui Fang, Jianmin Jiang, and Yue Feng (2006) A fuzzy logic approach for detection of video shot boundaries. Pattern Recognition, vol. 39, no. 11, pp. 2092–2100.

25. Selouani S.-A. and O'Shaughnessy D. (2003) On the use of evolutionary algorithms to improve the robustness of continuous speech recognition systems in adverse conditions. EURASIP Journal on Applied Signal Processing, vol. 8, pp. 814–823

26. Lo C.-C. and Wang S.-J. (2001) Video segmentation using a histogram-based fuzzy C-means clustering algorithm. The 10th IEEE International Conference on Fuzzy Systems, vol. 2, 2–5 Dec., pp. 920–923.

27. Cao X. and Suganthan P.N. (2002) Neural network based temporal video segmentation. International Journal of Neural Systems, vol. 12, no. 3–4, pp. 263–629.

28. Chang C.-H., Ye Z., and Zhang M. (2005) Fuzzy-ART based adaptive digital watermarking scheme. IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 1, pp. 65–81.

29. Diego Sal Diaz and Manuel Grana Romay (2005) Introducing a watermarking with a multi-objective genetic algorithm. Proceedings of the 2005 conference on Genetic and evolutionary computation, Washington DC, USA, pp. 2219–2220.

30. Lou D.-C. and Yin T.-L. (2001) Digital watermarking using fuzzy clustering technique. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (Japan), vol. E84-A, no. 8, pp. 2052–2060.

31. Maher El-arbi, Ben Amar, and C. Nicolas, H. (2006) Video watermarking based on neural networks. IEEE International Conference on Multimedia and Expo, Toronto, Canada, pp. 1577–1580.

32. Der-Chyuan Lou, Jieh-Ming Shieh, and Hao-Kuan Tso (2005) Copyright protection scheme based on chaos and secret sharing techniques. Optical Engineering, vol. 44, no. 11, pp. 117004–117010.

33. Wei Lu, Hongtao Lu, and FuLai Chung (2005) Subsampling-based robust watermarking using neural network detector. Advances in Neural Networks, ISNN 2005, Lecture Notes in Computer Science, vol. 3497, pp. 801–806.

34. Cheng-Ri Piao, Sehyeong Cho, and Seung-Soo Han (2006) Color image watermarking algorithm using BPN neural networks. Neural Information Processing, Lecture Notes in Computer Science, vol. 4234, pp. 234–242

35. Zheng Liu, Xue Li, and Dong Z. (2004) Multimedia authentication with sensor-based watermarking. Proc. of the international workshop on Multimedia and security, Magdeburg, Germany, pp. 155–159

36. Hung-Hsu Tsai, Ji-Shiung Cheng, and Pao-Ta Yu (2003) Audio watermarking based on HAS and neural networks in DCT domain. EURASIP Journal on Applied Signal Processing, vol. 2003, no. 3, pp. 252–263

37. Cao L., Wang X., Wang Z., and Bai S. (2005) Neural network based audio watermarking algorithm. In: ICMIT 2005: Information Systems and Signal Processing, Wei Y., Chong K.T., Takahashi T. (eds.), Proceedings of the SPIE, vol. 6041, pp. 175–179

38. Alessandro Bugatti, Alessandra Flammini, and Pierangelo Migliorati (2002) Audio classification in speech and music: A comparison between a statisti-

cal and a neural approach. EURASIP Journal on Applied Signal Processing, vol. 2002, no. 4, pp. 372–378.

39. Lim Ee Hui, Seng K.P., and Tse K.M. (2004) RBF Neural network mouth tracking for audio–visual speech recognition system. IEEE Region 10 Conference TENCON2004, 21–24 Nov., pp. 84–87.

40. Jian Zhou, Guoyin Wang, Yong Yang, and Peijun Chen (2006) Speech emotion recognition based on rough set and SVM. 5th IEEE International Conference on Cognitive Informatics ICCI 2006, 17–19 July, vol. 1, pp. 53–61.

41. Faraj M.-I. and Bigun J. (2007) Audio–visual person authentication using lip-motion from orientation maps. Pattern Recognition Letters, vol. 28, no. 11, pp. 1368–1382.

42. Halavati R., Shouraki S.B., Eshraghi M., Alemzadeh M., and Ziaie P. (2004) A novel fuzzy approach to speech recognition. Hybrid Intelligent Systems. Fourth International Conference on Hybrid Intelligent Systems, 5–8 Dec., pp. 340–345.

43. Eugene I. Bovbel and Dzmitry V. Tsishkou (2000) Belarussian speech recognition using genetic algorithms. Third International Workshop on Text, Speech and Dialogue, Brno, Czech Republic, pp. 185–204.

44. Fellenz W.A., Taylor J.G., Cowie R., Douglas-Cowie E., Piat F., Kollias S., Orovas C., and Apolloni B. (2000) On emotion recognition of faces and of speech using neural networks, fuzzy logic and the ASSESS system. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, vol. 2, IJCNN 2000, pp. 93–98.

45. Laaksonen J., Koskela M., and Oja E. (2002) PicSOM-Self-organizing image retrieval with MPEG-7 content descriptions. IEEE Transactions on Neural Networks, Special Issue on Intelligent Multimedia Processing vol. 13, no. 4, pp. 841–853.

46. Mats S., Jorma L., Matti P., and Timo H. (2006) Retrieval of multimedia objects by combining semantic information from visual and textual descriptors. Proceedings of 16th International Conference on Artificial Neural Networks (ICANN 2006), pp. 75–83, Athens, Greece, September 2006.

47. Kostek B. and Andrzej C. (2001) Employing fuzzy logic and noisy speech for automatic fitting of hearing aid. 142 Meeting of the Acoustical Society of America, No. 5, vol. 110, pp. 2680, Fort Lauderdale, USA.

48. Liu J., Wang Z., and Xiao X. (2007) A hybrid SVM/DDBHMM decision fusion modeling for robust continuous digital speech recognition. Pattern Recognition Letter, vol. 28, No. 8, pp. 912–920.

49. Ing-Jr Ding (2007) Incremental MLLR speaker adaptation by fuzzy logic control. Pattern Recognition, vol. 40 , no. 11, pp. 3110–3119

50. Andrzej C. (2003) Automatic identification of sound source position employing neural networks and rough sets. Pattern Recognition Letters, vol. 24, pp. 921–933.

51. Andrzej C., Kostek B., and Henryk S. (2002) Diagnostic system for speech articulation and speech understanding. 144th Meeting of the Acoustical Society of America (First Pan-American/Iberian Meeting on Acoustics), Journal of the Acoustical Society of America, vol. 112, no. 5, Cancun, Mexico.

52. Andrzej C., Andrzej K., and Kostek B. (2003) Intelligent processing of stuttered speech. Journal of Intelligent Information Systems, vol. 21, no. 2, pp. 143–171.

53. Pawel Zwan, Piotr Szczuko, Bozena Kostek, and Andrzej Czyzewski (2007) Automatic singing voice recognition employing neural networks and rough sets. RSEISP 2007, pp. 793–802.
54. Andrzej C. and Marek S. (2002) Pitch estimation enhancement employing neural network-based music prediction. Proc. IASTED Intern. Conference, Artificial Intelligence and Soft Computing, pp. 413–418, Banff, Canada.
55. Hendessi F., Ghayoori A., and Gulliver T.A. (2005) A speech synthesizer for Persian text using a neural network with a smooth ergodic HMM. ACM Transactions on Asian Language Information Processing (TALIP), vol. 4, no. 1, pp. 38–52.
56. Orhan Karaali, Gerald Corrigan, and Ira Gerson (1996) Speech synthesis with neural networks. World Congress on Neural Networks, San Diego, Sept. 1996, pp. 45–50.
57. Corrigan G., Massey N., and Schnurr O. (2000) Transition-based speech synthesis using neural networks. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 2, pp. 945–948.
58. Shan Meng and Youwei Zhang (2003) A method of visual speech feature area localization. Proceedings of the International Conference on Neural Networks and Signal Processing, 2003, vol. 2, 14–17 Dec., pp. 1173–1176.
59. Sun-Yuan Kung and Jenq-Neng Hwang (1998) Neural networks for intelligent multimedia processing. Proceedings of the IEEE Workshop on Neural Networksm, vol. 86, no. 6, pp. 1244–1272.
60. Frankel J., Richmond K., King S., and Taylor P. (2000) An automatic speech recognition system using neural networks and linear dynamic models to recover and model articulatory traces. Proc. ICSLP, 2000.
61. Schuller B., Reiter S., and Rigoll G. (2006) Evolutionary feature generation in speech emotion. IEEE International Conference on Recognition Multimedia, pp. 5–8.
62. Lewis T.W. and Powers D.M.W., Audio–visual speech recognition using red exclusion and neural networks. Proceedings of the twenty-fifth Australasian conference on Computer science, vol. 4, Melbourne, Victoria, Australia, pp. 149–156.
63. Nakamura S. (2002) Statistical multimodal integration for audio–visual speech processing. IEEE Transactions on Neural Networks, vol. 13, no. 4, pp. 854–866.
64. Guido R.C., Pereira J.C., and Slaets J.F.W. (2007) Advances on pattern recognition for speech and audio processing. Pattern Recognition Letters, vol. 28, no. 11, pp. 1283–1284.
65. Vahideh Sadat Sadeghi and Khashayar Yaghmaie (2006) Vowel recognition using neural networks. International Journal of Computer Science and Network Security (IJCSNS), vol. 6, no. 12, pp. 154–158.
66. Hartigan J.A. and Wong M.A. (1979) Algorithm AS136: A K-means clustering algorithm. Applied Statistics, vol. 28, pp. 100–108.
67. Henry C. and Peters J.F. (2007) Image pattern recognition using approximation spaces and near sets. In: Proceedings of Eleventh International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2007), Joint Rough Set Symposium (JRS 2007), Lecture Notes in Artificial Intelligence, vol. 4482, pp. 475–482.
68. Kerre E. and Nachtegael M. (2000) Fuzzy techniques in image processing: Techniques and applications. Studies in Fuzziness and Soft Computing, vol. 52, Physica, Heidelberg.

69. Lingras P. and West C. (2004) Interval set clustering of web users with rough K-means. Journal of Intelligent Information Systems, vol. 23, no. 1, pp. 5–16.
70. Lingras P. (2007) Applications of rough set based K-means, Kohonen, GA Clustering. Transactions on Rough Sets, VII, pp. 120–139.
71. Mitra Sushmita (2004) An evolutionary rough partitive clustering. Pattern Recognition Letters, vol. 25, pp. 1439–1449.
72. Ng H.P., Ong S.H., Foong K.W.C., Goh P.S., and Nowinski, W.L. (2006) Medical image segmentation using K-means clustering and improved watershed algorithm. IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 61–65.
73. Nachtegael M., Van-Der-Weken M., Van-De-Ville D., Kerre D., Philips W., and Lemahieu I. (2001) An overview of classical and fuzzy-classical filters for noise reduction. 10th International IEEE Conference on Fuzzy Systems FUZZ-IEEE 2001, Melbourne, Australia, pp. 3–6.
74. Ning S., Ziarko W., Hamilton J., and Cercone N. (1995) Using rough sets as tools for knowledge discovery. In: Fayyad U.M. and Uthurusamy R. (eds.), First International Conference on Knowledge Discovery and Data Mining KDD'95, Montreal, Canada, AAAI, pp. 263–268.
75. Pawlak Z. (1991) Rough sets – Theoretical aspects of reasoning about data. Kluwer, Dordrecht.
76. Pawlak Z., Grzymala-Busse J., Slowinski R., and Ziarko W. (1995) Rough sets. Communications of the ACM, vol. 38, no. 11, pp. 88–95.
77. Polkowski L. (2003) Rough Sets: Mathematical Foundations. Physica, Heidelberg.
78. Peters J.F. (2007) Near sets: Special theory about nearness of objects. Fundamenta Informaticae, vol. 75, no. 1–4, pp. 407–433.
79. Peters J.F. (2007) Near sets. General theory about nearness of objects. Applied Mathematical Sciences, vol. 1, no. 53, pp. 2609–2029.
80. Peters J.F., Skowron A., and Stepaniuk J. (2007) Nearness of objects: Extension of approximation space model. Fundamenta Informaticae, vol. 79, pp. 1–16.
81. Peters J.F. (2007) Near sets. Toward approximation space-based object recognition, In: Yao Y., Lingras P., Wu W.-Z, Szczuka M., Cercone N., Ślęzak D. (eds.), Proc. of the Second Int. Conf. on Rough Sets and Knowledge Technology (RSKT07), Joint Rough Set Symposium (JRS07), Lecture Notes in Artificial Intelligence, vol. 4481, Springer, Berlin Heidelberg New York, pp. 22–33.
82. Peters J.F. and Ramanna S. (2007) Feature selection: Near set approach. In: Ras Z.W., Tsumoto S., and Zighed D.A. (eds.) 3rd Int. Workshop on Mining Complex Data (MCD'07), ECML/PKDD-2007, Lecture Notes in Artificial Intelligence, Springer, Berlin Heidelberg New York, *in press*.
83. Peters J.F., Skowron A., and Stepaniuk J. (2006) Nearness in approximation spaces. In: Lindemann G., Schlilngloff H. et al. (eds.), Proc. Concurrency, Specification & Programming (CS&P'2006), Informatik-Berichte Nr. 206, Humboldt-Universität zu Berlin, pp. 434–445.
84. Orłowska E. (1982) Semantics of vague concepts. Applications of rough sets. Institute for Computer Science, Polish Academy of Sciences, Report 469, 1982. See, also, Orłowska E., Semantics of vague concepts, In: Dorn G. and Weingartner P. (eds.), Foundations of Logic and Linguistics. Problems and Solutions, Plenum, London, 1985, pp. 465–482.

85. Orłowska E. (1990) Verisimilitude based on concept analysis. Studia Logica, vol. 49, no. 3, pp. 307–320.
86. Pawlak Z. (1981) Classification of objects by means of attributes. Institute for Computer Science, Polish Academy of Sciences, Report 429, 1981.
87. Pawlak Z. (1982) Rough sets. International Journal of Computing and Information Sciences, vol. 11, pp. 341–356.
88. Pawlak Z. and Skowron A. (2007) Rudiments of rough sets. Information Sciences, vol. 177, pp. 3–27.
89. Peters J.F. (2008) Classification of perceptual objects by means of features. International Journal of Information Technology and Intelligent Computing, vol. 3, no. 2, pp. 1–35.
90. Lockery D. and Peters J.F. (2007) Robotic target tracking with approximation space-based feedback during reinforcement learning. In: Proceedings of Eleventh International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2007), Joint Rough Set Symposium (JRS 2007), Lecture Notes in Artificial Intelligence, vol. 4482, pp. 483–490.
91. Peters J.F., Borkowski M., Henry C., and Lockery D. (2006) Monocular vision system that learns with approximation spaces. In: Ella A., Lingras P., Slezak D., and Suraj Z. (eds.), *Rough Set Computing: Toward Perception Based Computing*, Idea Group Publishing, Hershey, PA, pp. 1–22.
92. Peters J.F., Borkowski M., Henry C., Lockery D., Gunderson D., and Ramanna S. (2006) Line-crawling bots that inspect electric power transmission line equipment. Proc. 3rd Int. Conf. on Autonomous Robots and Agents 2006 (ICARA 2006), Palmerston North, NZ, 2006, pp. 39–44.
93. Peters J.F. (2008) Approximation and perception in ethology-based reinforcement learning. In: Pedrycz W., Skowron A., and Kreinovich V. (eds.), Handbook on Granular Computing, Wiley, New York, Ch. 30, pp. 1–41.
94. Peters J.F. and Borkowski M. (2004) K-means indiscernibility relation over pixels. Proc. 4th Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC 2004), Uppsala, Sweden, 1–5 June, pp. 580–585.
95. Peters J.F. and Pedrycz W. (2007) Computational intelligence. In: EEE Encyclopedia. Wiley, New York, *in press.*
96. Peters J.F., Liting H., and Ramanna S. (2001) Rough neural computing in signal analysis. Computational Intelligence, vol. 17, no. 3, pp. 493–513.
97. Peters J.F., Skowron A., Suraj Z., Rzasa W., Borkowski M. (2002) Clustering: A rough set approach to constructing information granules. Soft Computing and Distributed Processing. Proceedings of 6th International Conference, SCDP 2002, pp. 57–61.
98. Petrosino A. and Salvi G. (2006) Rough fuzzy set based scale space transforms and their use in image analysis. International Journal of Approximate Reasoning, vol. 41, no. 2, pp. 212–228.
99. Shankar B.U. (2007) Novel classification and segmentation techniques with application to remotely sensed images. Transactions on Rough Sets, vol. VII, LNCS 4400, pp. 295–380.
100. Otto C.W. (2007) Motivating rehabilitation exercise using instrumented objects to play video games via a configurable universal translation peripheral, M.Sc. Thesis, Supervisors: Peters J.F. and Szturm T., Department of Electrical and Computer Engineering, University of Manitoba, 2007.

101. Szturm T., Peters J.F., Otto C., Kapadia N., and Desai A. (2008) Task-specific rehabilitation of finger-hand function using interactive computer gaming, Archives for Physical Medicine and Rehabilitation, *submitted*.

102. Sandeep Chandana and Rene V. Mayorga (2006) RANFIS: Rough adaptive neuro-fuzzy inference system. International Journal of Computational Intelligence, vol. 3, no. 4, pp. 289–295.

103. Swagatam Das, Ajith Abraham, and Subir Kumar Sarkar (2006) A hybrid rough set – Particle swarm algorithm for image pixel classification. Proceedings of the Sixth International Conference on Hybrid Intelligent Systems, 13–15 Dec., pp. 26–32.

104. Bezdek J.C., Ehrlich R., and Full W. (1984) FCM: The fuzzy C-means clustering algorithm. Computers and Geosciences, vol. 10, pp. 191–203.

105. Cetin O., Kantor A., King S., Bartels C., Magimai-Doss M., Frankel J., and Livescu K. (2007) An articulatory feature-based tandem approach and factored observation modeling. IEEE International Conference on Acoustics, Speech and Signal, ICASSP2007, Honolulu, HI, vol. 4, pp. IV-645–IV-648.

106. Raducanu B., Grana M., and Sussner P. (2001) Morphological neural networks for vision based self-localization. IEEE International Conference on Robotics and Automation, ICRA2001, vol. 2, pp. 2059–2064.

107. Ahmed M.N., Yamany S.M., Nevin M., and Farag A.A. (2003) A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Transactions on Medical Imaging, vol. 21, no. 3, pp. 193–199.

108. Yan M.X.H. and Karp J.S. (1994) Segmentation of 3D brain MR using an adaptive K-means clustering algorithm. IEEE Conference on Nuclear Science Symposium and Medical Imaging, vol. 4, pp. 1529–1533.

109. Voges K.E., Pope N.K.L.I., and Brown M.R. (2002) Cluster analysis of marketing data: A comparison of K-means, rough set, and rough genetic approaches. In: Abbas H.A., Sarker R.A., and Newton C.S. (eds.), Heuristics and Optimization for Knowledge Discovery, Idea Group Publishing, pp. 208–216.

110. Chen C.W., Luo J.B., and Parker K.J. (1998) Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications. IEEE Transactions on Image Processing, vol. 7, no. 12, pp. 1673–1683.

111. Davis K.J. and Najarian K. (2001) Maximizing strength of digital watermarks using neural networks. International Joint Conference on Neural Networks, IJCNN 2001, vol. 4, pp. 2893–2898.

112. Sankar K. Pal (2001) Fuzzy image processing and recognition: Uncertainties handling and applications. International Journal of Image and Graphics, vol. 1, no. 2, pp. 169–195.

113. Yixin Chen and James Z. Wang (2002) A region-based fuzzy feature matching approach to content-based image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pp. 1252–1267.

114. Yu Wanga Mingyue Dingb, Chengping Zhoub, and Ying Hub (2006) Interactive relevance feedback mechanism for image retrieval using rough set. Knowledge-Based Systems, vol. 19, no. 8, pp. 696–703.

115. Zadeh L.A. (1965) Fuzzy sets. Information and Control, vol. 8, pp. 338–353.

116. Zbigniew W. (1987) Rough approximation of shapes in pattern recognition. Computer Vision, Graphics, and Image Processing, vol. 40, no. 2, pp. 228–249.